# PROTOCOLS FOR WIDE BAND SATELLITE SYSTEMS WITH A LARGE NUMBER OF SMALL VOICE AND DATA USERS

by

XU TAN

B.Sc., Shanghai Jiao Tong University, China, 1982

# A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE

in

### THE FACULTY OF GRADUATE STUDIES Department of Electrical Engineering

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA September, 1987

©Xu Tan, 1987

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ELECTRICAL ENGINEERING

The University of British Columbia 1956 Main Mall Vancouver, Canada V6T 1Y3

Date Oct. 14, 1987

#### ABSTRACT

Multiaccess protocols for integrated voice and data transmissions over satellite channels are studied, based on the SENET(slotted envelope network) structure. The satellite system is characterized as a wide band system with a large number of geographically distributed small voice and data users. Performance evaluations of commonly used protocols, i.e., ALOHA and reservation protocols, are first conducted, both analytically and by simulations. The effect of different voice backgrounds on data access protocols are shown explicitly.

Based on these results, a control algorithm is proposed. Analyses and simulations show that ALOHA protocol incorporated with such a control mechanism is globally stable under the integrated environment; moreover, the performance deterioration due to voice backgrounds is dramatically reduced. The actual implementation aspects of the control algorithm are considered.

An extension of the above results leads to a combined random/reservation protocol. Simulation and analysis results show that the combined protocol exhibits desired low delay and high throughput performance characteristics, with satisfactory voice blocking probability, under the worst user population assumption, i.e., all the voice sources and data sources are independent of each other — reflecting the nature of small earth station environment. The absence of the need for mini-slot structure lends ease and simplicity to the implementation of the combined protocol.

### TABLE OF CONTENTS

ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	ix
ACKNOWLEDGEMENTS	х

1	Introduction	1
	1.1 Basic Structure of Integrated Wide Band Systems	3
	1.1.1 The WB SATNET Structure	4
	1.1.2 The SENET Structure	5
	1.2 Scope of the Thesis	8

2	Voice Packet Transmission	10
	2.1 Virtual Connections Protocol for Voice Packets	10
	2.2 Voice Pattern Statistics	14
	2.2.1 Statistics of Voice Utilization	15
	2.2.2 State Correlation of Voice Occupied Slots	22

2.3 Simulations of Voice Pattern Structure	26
2.3.1 Choice of Parameters	26
2.3.2 Results of Simulations	28
Data Packet Transmission	38
3.1 Multiaccess Protocols in the IVDS Systems	39
3.2 ALOHA Protocol in the IVDS Systems	43
3.2.1 Acquisition of $ ilde{V}(n)$	45
3.2.2 Estimation of $\tilde{g}(t)$	45
3.2.3 Performance Analysis of ALOHA Protocol	52
3.3 Reservation Protocol in the IVDS Systems	56
3.3.1 Performance Analysis of the Reservation Protocol	59
3.4 Simulations of the Two Protocols	66
3.4.1 Simulations of the ALOHA Protocol	66
3.4.2 Simulations of the Reservation Protocol	75
3.5 ALOHA Protocol (The Second Version)	80
3.5.1 Performance Analysis of ALOHA-II Protocol	81
3.5.2 Effect of Estimation Errors	96
Combined Random/Reservation Protocol	104
	2.3 Simulations of Voice Pattern Structure         2.3.1 Choice of Parameters         2.3.2 Results of Simulations         3.1 Multiaccess Protocols in the IVDS Systems         3.2 ALOHA Protocol in the IVDS Systems         3.2.1 Acquisition of $\tilde{V}(n)$ 3.2.2 Estimation of $\tilde{g}(t)$ 3.2.3 Performance Analysis of ALOHA Protocol         3.3 Reservation Protocol in the IVDS Systems         3.3.1 Performance Analysis of the Reservation Protocol         3.4 Simulations of the Two Protocols         3.4.1 Simulations of the ALOHA Protocol         3.4.2 Simulations of the Reservation Protocol         3.5.1 Performance Analysis of ALOHA-II Protocol         3.5.2 Effect of Estimation Errors         Combined Random/Reservation Protocol

4.1 Mode Switching Information	105
4.2 Reservation Request Transmissions	106
4.3 The Combined Random/Reservation Protocol	107

	4.3.1 Protocol Description	108
	4.3.2 System Parameters for CP	114
	4.3.3 Asymptotic Analysis of CP	115
	4.4 Simulations of CP	<b>12</b> 1
	4.5 Comparison of CP with ALOHA and Reservation Protocols	129
5	Conclusions and Further Studies	131
	Appendix	134
	References	141

**
v
•

### LIST OF FIGURES

1-1	Basic Structure of WB SATNET	4
1-2	Illustration of Stream Subframe	6
1-3	Illustration of Data Transmission Procedures	6
1-4	The SENET Structure	7
2-1	(a) Voice Assignment Protocol I	12
2-1	(b) Voice Assignment Protocol II	12
2-2	(a) Blocking Probability $B(s,a)$ versus $S_f$	19
2-2	(b) Blocking Probability $B(s, a)$ versus Offered Load $a'$	19
2-3	Tail Probabilities, $P_{sp} = 0.005$	21
2-4	Illustration of Voice Correlation and Data Transmissions	23
2-5	(a) Correlation $\hat{\gamma}(k)$ for $ ho_v=0.5,$ VAP-I	31
2-5	(b) Slot Distribution $\hat{P}_r(l)$ for $\rho_v = 0.5$ , VAP-I	31
2-6	(a) Correlation $\hat{\gamma}(k)$ for $ ho_v=0.5,$ VAP-II	32
2-6	(b) Slot Distribution $\hat{P}_r(l)$ for $\rho_v = 0.5$ , VAP-II	32
2-7	(a) Correlation $\hat{\gamma}(k)$ for $ ho_v = 1/3,$ VAP-I	33
2-7	(b) Slot Distribution $\hat{P}_r(l)$ for $\rho_v = 1/3$ , VAP-I	33
2-8	(a) Correlation $\hat{\gamma}(k)$ for $ ho_v=2/3,$ VAP-I	34
2-8	(b) Slot Distribution $\hat{P}_r(l)$ for $\rho_v = 2/3$ , VAP-I	34

2-9	Correlation $\hat{\gamma}(k)$ and $S_f$	35
2-10	Slot Distribution $\hat{P}_r(l)$ and $S_f$	36
2-11	Tail Probability $P_j$ and $S_f$	37
3-1	Illustration of Delayed Rule	44
3-2	Bias and Variance of Maximum Likelihood Estimator $g_m$	49
3-3	Bias and Variance of Modified Arithmetic Estimator $\hat{M}$	53
3-4	Illustration of the Reservation Protocol	57
3-5	(a) Global Queueing Delay $D_q$ for VAP-I	60
3-5	(b) Global Queueing Delay $D_q$ for VAP-II	60
3-6	Excess Queue Size $\Delta q$ versus $\rho_d$	65
3-7	ALOHA Delay-Throughput Characteristics	68
3-8	(a)(b) Effect of Blocking Conditions	72
3-9	ALOHA Delay versus $K_{max}$	74
3-10	ALOHA Delay versus $S_f$	74
3-11	Reservation Delay-Throughput Characteristics	76
3-12	Output Rate $S_{out}$ versus $g$	86
3-13	$S_{in}$ / $S_{out}$ and EP's	88
3-14	Illustration of $D_2$ for VAP-II	94
3-15	Comparison of ALOHA-I and -II	95
3-16	Output Rate $S_{out}$ and Estimation Error $\delta(n)$	98
3-17	(a) Backlog Delay $D_q$ and $ ho_d$ , for $\mu^{-1}=500$	100
3-17	(b) $E_1$ and $E_2$ versus $\rho_d$ , for $\mu^{-1} = 500$	100
3-18	(a) Backlog Delay $D_q$ and $\rho_d$ , for $\mu^{-1} = 50$	101

3-18	(b) $E_1$ and $E_2$ versus $\rho_d$ , for $\mu^{-1} = 50$	101
4-1	Data Access Protocol Flowchart	110
4-2	Illustration of Continuous Reservation	119
4-3	CP Delay-Throughput Characteristics	124
4-4	CP Performance and User Number $N_u$	125
4-5	Relationship among $N_{um}$ , $S_f$ and $N_u$	127
4-6	Trial Number Distribution	128

ι

## LIST OF TABLES

3-1	ALOHA Delay-Throughput with VAP-I	69
3-2	ALOHA Delay-Throughput with VAP-II	70
3-3	Reservation Delay-Throughput with VAP-I	77

### ACKNOWLEDGEMENTS

I with to thank my supervisor, Dr. H. W. Lee, for his encouragement and suggestions throughout the research.

The financial support from Shanghai Jiao Tong University and from the research grant of Dr. H. W. Lee are gratefully acknowledged.

#### Chapter 1

# Introduction

The integrated transmission of heterogeneous types of information has been proposed as an important solution to the ever growing demand upon communication systems. In particular, extensive studies and experiments have been conducted in integrated voice/data transmissions. The integration of digitized voice with data in a common packet-switched system offers potential cost savings through sharing of switching and transmission resources, as well as enhanced services for users who require access to both voice and data communications[1-4].

Of vital importance here is the design of an efficient link level protocol for integrated voice/data transmissions. This is complicated, however, by the significant differences between voice traffic and data traffic. Statistically, the former is characterized by smoothly continuous, or so called "streamlike", flow in time; while the latter is characterized as "bursty", that is, short spurts of activities interspersed with relatively long idle periods. The streamlike traffic can be accommodated by circuit-switched networks; while bursty traffic can only be handled efficiently by packet-switched networks. In order for these two different types of traffic to be compatible in the same network, the transmission of streamlike traffic is *packetized*, i.e., the continuous stream of traffic is divided into a sequence of fixed sized packets before they are transmitted. In general, an integrated network is packet-switched in which the streamlike traffic is handled

#### Chapter 1: Introduction

by virtual connection protocols.

Experiments with packetized voice transmission show that satisfactory reproduction of speech at the receiver requires short packet end-to-end delay and small interpacket delay disparity. It is estimated that the end-to-end delay should be less than 150-350 ms and interpacket delay disparity should be less than 5-10 ms [5].

Because of stringent delay requirements placed by voice users on the network, integrated packet voice/data protocols have therefore tended to give precedence to voice packets and to insert data packets into any remaining network capacity after the voice requirements are fulfilled [4].

Current areas of research in integrated voice/data transmissions fall into the following categories:

- 1. Utilization of voice transmission silence periods for data packet transmissions. It is estimated that over 45% of time of a speech is silence period.
- 2. Data transmission via remaining channel capacity not used by voice transmissions.
- 3. Combinations of 1 and 2.

The present thesis is concerned with type 2 integration of voice and data transmissions. The environment under study is a global beam satellite system. There are a "large" number of geographically distributed, independent "small" data and voice users wishing to communicate with each other over the satellite channel. By a "large" number is meant anywhere from several tens to a few thousands, and by "small" user is meant each user would generate no more than one telephone call and/or interactive data connection.<sup>†</sup> The satellite channel is the sole communication medium among all the users.

A satellite channel provides excellent one hop connectivity among a large number of geographically distributed users. However, its one way propagation delay of 270 ms has been found a major difficulty in devising an efficient protocol for integrated voice/data transmissions.

#### 1.1 Basic Structure of Integrated Wide Band Systems

In order to support many voice calls simultaneously with reasonably small call blocking probability, an integrated voice/data(IVD\*) system is implemented in a wide band satellite system (details are given in Chapter 2).

Numerous protocols have been proposed for wide band IVDS systems, with experiments reported in [4, 6, 7, 8, 9]. Two representative protocols are of interest here and are described briefly as follows.

<sup>†</sup> In the consideration of implementations, such users usually have small antennas and are supported by multibeam satellite systems, which is not the topic of study of this thesis. The global beam system here can be viewed as one coverage zone in a multibeam system. See chapter 5 for further notes and references.

<sup>\* &</sup>quot;IVD" will stand for "integrated voice/data" in the following of the thesis. Furthermore, "IVDS" will stand for "integrated voice/data satellite" (systems).



Figure 1-1 Basic Structure of WB SATNET

#### 1.1.1 WB SATNET Structure

This is the protocol used in the first experimental wide band integrated satellite network (WB SATNET, [6]). In the WB SATNET, time is partitioned into fixed size frames of approximately 21.2 ms corresponding to  $2^{16}$  bits at a channel data rate of 3.088 Mbps. Each frame is further partitioned into several subframes. In the steady state, the frame structure is shown in figure 1-1.

The stream subframe in figure 1-1 supports the streamtype traffic, i.e. voice traffic. Stream transmission is characterized by periodic frame-to-frame recurrence of channel allocation, i.e., a dedicated portion of channel time in stream subframe is assigned to an ongoing voice user (station), as shown in

#### Chapter 1: Introduction

figure 1-2, for the length of the call; this dedicated assignment is referred to as virtual circuit or virtual connection.

The setup of virtual connections is via the reservation transmitted during control subframe. The stream subframe may be null if there is no voice traffic present.

The data subframe and control subframe support data transmission by a reservation protocol. Control subframe is divided into N fixed-size slots, each slot can be used to transmit a reservation for data packet or voice connection. The assignment of the N control slots to network stations can be either fixed: each station is assigned a dedicated slot within the control subframe; or contention based: the N slots are shared on a slotted ALOHA contention basis by all stations.

Data are transmitted in data subframe. The basic procedures are illustrated in figure 1-3. Delay for a data packet is 2 round trip propagation delay of approximately 0.54s plus the global queueing delay.

#### 1.1.2 The SENET Structure

Under the SENET structure (Slotted Envelop Network, [7]), the channel time is partitioned into fixed size frames; each frame is further partitioned into a number of fixed size slots; the size of one slot is equal to the transmission time of one voice or data packet. Virtual connections are also used for voice calls. Data packets are transmitted in the remaining slots in a frame. In the original SENET, data access protocol is TDMA, as illustrated in figure 1-4.



Figure 1-2 Illustration of Stream Subframe



Figure 1-3 Illustration of Data Transmission Procedures



C – On going voice call

D – Data transmission, assuming 5 data users

Figure 1-4 The SENET structure

#### Chapter 1: Introduction

In the SENET structure, there is no boundary between voice portion and data portion, thus making it distinctive from WB SATNET structure. In general, a higher degree of integration can be achieved by using SENET than WB SATNET structure. The performance analysis, however, tends to be much more complicated for the former than the latter, due to the entanglement between voice and data transmissions.

#### 1.2 Scope of the Thesis

In this thesis, IVD protocols are studied under SENET structure. In the original SENET structure, data multiple-access protocol TDMA was used. It is well-known, however, that TDMA does not give satisfactory performance under bursty and large user population environment.

There exist many multiple-access protocols with superior performance to TDMA. The large propagation delay of satellite channels, however, precludes use of schemes such as CSMA, polling, etc, since they only perform well when propagation delay is a small fraction of packet transmission time.

In a long propagation delay environment, simple random access scheme has found to be effective. Reservation scheme is another alternative to cope with the long propagation delay. Combined random/reservation schemes have also been proposed which transfer from one mode (random access) to another (reservation) according to the channel load.

Under IVD environment, these protocols, namely random access and reser-

#### Chapter 1: Introduction

vation, are studied; performance is evaluated analytically and compared to simulation results. A combined random/reservation scheme is proposed, which differs from most known combined random/reservation schemes in that it does not have to partition the channel time into mini-slot structures.

Chapter two begins with investigation of statistical properties of SENET structure.

In Chapter three, ALOHA and reservation protocols are analyzed and simulated. In order to derive the analytic results, an independence assumption is made, which is shown to be valid for low and medium range of voice traffic intensity and for certain voice assignment protocols.

A control algorithm is introduced in Chapter three. It is shown that ALOHA protocol incorporated with such a control mechanism is globally stable in the IVD environment, even for infinite number of users.

Based on the results of chapter three, the combined random/reservation protocol is proposed and simulated in chapter four. Analytic and simulation results show a preference on the combined protocol to non-combined ones.

Chapter five concludes the thesis with suggestions for further studies.

#### Chapter 2

# Voice Packet Transmission

Voice packet transmission protocol will be discussed in this chapter. Emphasis is placed on the study of statistical properties of voice pattern structure as a basis for the forthcoming chapters.

#### 2.1 Virtual Connection Protocols for Voice Packets

Because of the streamlike property and the delay requirement of voice packets, a virtual connection or circuit for an ongoing voice call is both reasonable and desirable. The basic procedure for establishing a virtual connection is described as follows:

The channel time is divided into frames, and each frame is further divided into fixed number of slots (SENET structure, see figure 1-4). The length of each frame in time will be denoted as  $T_f$ , and number of slots in a frame, called frame size, will be denoted as  $S_f$  in the sequel.

Reservation protocol is used to exchange setup information of a virtual call connection. Whenever a station wants to call, it sends a request to the controller or, in a distributed control network, to all the other stations. (The transmission of request will be discussed in chapter four.) After the request being heard by the controller or all the stations in the network, including the sender, on first come first served basis, the following actions will be taken:

- 1. One empty slot in the current frame is assigned to the call; this slot is assigned to the same call for the succesive frames until the completion of the call. (virtual connection).
- 2. If no empty slot is found (among all slots in the current frame), the call is rejected. (blocked)

In the case where there is more than one slot available, the slot assignment to the new call can be done according to certain predetermined rules. For example:

Immediate Assignment: The empty slot immediately following the slot in which the request arrives is assigned to the new call; (figure 2-1(a))

The begin-of-frame Assignment: Slot assignments take place only at the beginning of each frame; empty slots will be assigned to new calls in the order of left (earlier in time) to right (figure 2-1(b)).

There is little difference between two assignments as far as the voice performance is concerned. However, because of the order in which they make use of empty slots, the resultant voice pattern<sup>\*</sup> will differ, which have significant effects on data performance, as will be discussed in chapter 3.

<sup>\*</sup> will be discussed in section 2.2



Figure 2-1(a) Voice Assignment Protocol I



Figure 2-1(b) Voice Assignment Protocol II

In the subsequent discussions throughout the thesis, for the convenience of notation, let the immediate assignment protocol be called "voice assignment protocol I", abbreviated as VAP-I, and the begin-of-frame assignment protocol be called "voice assignment protocol II", abbreviated as VAP-II.

The two important SENET structure parameters  $T_f$  (frame length) and  $S_f$  (frame size) are determined as follows.

Since an ongoing voice call transmits one voice packet per frame, the frame length is equal to voice packet size in time. In order to minimize both the packetization delay at the transmitter and the perceptual effect of lost packet anomalies at the receiver, packets should be as short as possible. Experience with lost packet anomalies indicates that individual packets should contain no more than 50 ms of speech[10]; ideally, packets should be even shorter to minimize packetization delay. On the other hand, in order to maintain high channel utilization, it would be desirable to keep the number of speech bits per packet as large as possible relative to the overhead which must accompany each packet.

In WB SATNET(see section 1.1.1.), frame length was chosen as 21.2 ms. Typically, the frame length  $T_f$  is between 10 to 30 ms in a wide band network when considering the tradeoffs mentioned above.

Frame size  $S_f$  is numerically equal to the maximal number of voice calls supported by the network simultaneously. It depends on the voice coding rate and channel capacity. If coding rate is  $K_c$  and total channel capacity is C, then

$$S_f = \lfloor C/K_c \rfloor \tag{2-1}$$

where the symbol  $\lfloor \rfloor$  denotes the truncation function which gives the integer part of a real number. Equation (2-1) assumes that the portion of total channel capacity C used for controlling and synchronization, etc, is negligible.

Theoretically, the larger  $S_f$  is, the smaller the call blocking probability will be.\* In practice, the maximum  $S_f$  is bounded by the total channel capacity (bandwidth) C. For a wideband voice packet network,  $K_c$  is from 16kbps (CVSD coding) to 64 kbps (PCM coding), and C from 3 Mbps to 6 Mbps (commercial C band), thus  $S_f$  is from 50 to 400.†

In a satellite channel, the single hop propagation delay of 270 ms, denoted as R, corresponds to roughly 12 times the frame length  $T_f$ .

#### 2.2 Voice Pattern Statistics

Slots occupied by voice virtual connections form a pattern (see figure 2-1 for example), which will be referred to as *voice pattern* in this thesis. As each slot can be in either one of two states, i.e., occupied or unoccupied, if the occupied state is assigned value 1 and unoccupied state value 0, the pattern structure can be modelled as a 2-valued random process in slot time n, denoted as  $\tilde{V}(n)$ .

Since the voice pattern serves as the background for data transmissions, some of its statistics, which have significant effects on data performance, are discussed here.

<sup>\*</sup> will be discussed in section 2.2.

<sup>&</sup>lt;sup>†</sup> Experiments on  $K_u$  band with C = 500 Mbps has been reported in [11]. A much higher value of  $S_f$  is thus possible.

#### 2.2.1 Statistics of Voice Utilization

Define  $\rho_v(n)$  to be the first moment of  $\tilde{V}(n)$ , i.e.,

$$ho_v(n) = \mathbb{E}\left[ ilde{V}(n)
ight] = P_r\left\{ ilde{V}(n) = 1
ight\}$$

Assume that  $\tilde{V}(n)$  is ergodic and stationary, then,

$$\rho_{v}(n) = \rho_{v} = P_{r} \left\{ \tilde{V}(n) = 1 \right\} = \lim_{T \to \infty} \frac{1}{2T} \sum_{n=-T}^{T} \tilde{V}(n)$$
(2-2)

If V(n) is a sample of  $\tilde{V}(n)$ ,  $n = 0, 1, \dots, T$ , then  $\rho_v$  can be estimated by

$$\hat{\rho}_{v} = \frac{1}{T} \sum_{n=0}^{T} V(n)$$
(2-3)

By definition,  $\rho_v$  is the voice utilization, which is the portion of total channel capacity occupied by voice calls. The portion of remaining capacity available to data transmissions is  $1 - \rho_v$ . To estimate the steady state value of  $\rho_v$  by (2-3), T should be chosen long enough in order to average out the fluctuations. On the other hand, a small value of T gives short time average  $\rho_v$ ; a sequence of short time average  $\rho_v$ 's can serve as a measure of changing rate and extent of fluctuation of voice utilization, especially when T is chosen as  $S_f$ .

The variance of  $\tilde{V}(n)$ , denoted as  $\sigma_{v}^{2}$ , is

$$\sigma_{v}^{2} = E\left[\tilde{V}(n)^{2}\right] - \{\rho_{v}\}^{2} = \rho_{v}(1 - \rho_{v})$$
 (2-4)

The steady state value of  $\rho_v$  can be derived from the following voice source and system models.

#### The voice source model

Calls are generated from an ensemble of N individual voice sources, where  $N \rightarrow \infty$ , with the following characteristics:

1. The aggregated number of calls,  $N_c$ , generated during the time interval T, obeys a Poisson distribution, i.e.,

$$P_r(N_c,T) = rac{(\lambda T)^{N_c} e^{-\lambda T}}{N_c!}$$
  $N_c = 1, 2, \cdots$ 

where  $\lambda$  is call arrival rate. (Typical value of  $\lambda$  for a large satellite network would be in the order of a few calls per minute).

2. The calls are packetized. The length of each call is geometrically distributed, i.e. if l denotes number of packets in one call, then

$$P_r(l) = \mu (1-\mu)^{l-1}$$
  $l = 0, 1, \cdots$ 

where  $\mu$  is the *call completion rate*, and its reciprocal  $\mu^{-1}$  is the average number of packets in one call. For packet size of 20 to 30 ms,  $\mu^{-1}$  is of the order of several ten thousand (packets).

#### The system model

The analysis of virtual connection of voice transmissions is equivalent to that of circuit switched telephone systems, which are modelled as an M/M/s blocked customer removed  $(M/M/s \ b.c.r)$  queueing system. In the sense of providing virtual connections, slot size  $S_f$  is equivalent to the server number s.

If state j of the M/M/s b.c.r system is defined as the number of 1's in a frame (equal to the number of voice occupied slots), then state transition rate  $\lambda_j$  from state j to j + 1 is

$$\lambda_j = \begin{cases} \lambda, & j = 0, 1, \cdots, s - 1 \\ 0, & j = s. \end{cases}$$
(2-5)

from state j to j-1 is

$$\mu_j = j\mu$$
  $j = 1, 2, \cdots, s.$  (2-6)

where  $\lambda$  and  $\mu$  are the call arrival rate and the call completion rate, respectively; s is equal to  $S_f$ . The meaning of  $\lambda_s = 0$  is that additional voice arrivals after all  $s(=S_f)$  slots are occupied are blocked and removed. The state probability  $P_j$ for an M/M/s b.c.r system is given by [12],

$$P_{j} = \begin{cases} \frac{(\lambda/\mu)^{j}/j!}{\sum\limits_{k=0}^{s} (\lambda/\mu)^{k}/k!}, & j = 0, 1, \cdots, s\\ 0, & j > s. \end{cases}$$
(2-7)

Note  $\lambda$  and  $\mu$  appear in (2-7) only through the ratio  $\lambda/\mu$ , which is the offered load in the terminology of queueing theory, denoted as a. (2-7) is also known as truncated Poisson distribution.

The average number of busy servers  $\overline{n}$  can be calculated as

$$\overline{n} = \sum_{j=1}^{s} j P_j \tag{2-8}$$

 $\overline{n}$  is also the average number of voice occupied slots in a frame, thus,  $\rho_v$  can be obtained as

$$\rho_v = \overline{n}/s \tag{2-9}$$

substituting (2-7), (2-8) in (2-9) yields,

$$\rho_{v} = \frac{a}{s} \left( 1 - \frac{a^{s}/s!}{\sum_{k=0}^{s} a^{k}/k!} \right)$$
(2-10)

where  $s = S_f$ ,  $a = \lambda/\mu$ .

The blocking probability of voice calls is recognized as  $P_j$  for  $j = s = S_f$ , which is the probability that a customer finds all the s servers are busy, or equivalently, a call arrival finds all the  $S_f$  slots in a frame are occupied, denoted as B(s, a),

$$B(s,a) = \frac{\frac{a^{s}/s!}{\sum\limits_{k=0}^{s} \frac{a^{k}}{k!}}$$
(2-11)

(2-11) is well known as the Erlang loss formula [12]. Combining (2-9), (2-10) and (2-11) gives,

$$\overline{n} = a \left( 1 - B(s, a) \right) \tag{2-12}$$

which suggests that  $\overline{n}$  can also be seen as the carried load of the system. For convenience, define the normalized offered load  $a' = a/S_f$ ; the normalized carried load  $\overline{n}/S_f$  is just  $\rho_v$ . Figure 2-2(a) plots B(s,a) versus  $S_f$ , for a' = 0.8; figure 2-2(b) plots B(s,a) versus a' for  $S_f = 25$ , 50, 100 and 150, respectively.

As seen from the figures, the larger  $S_f$  is, the smaller the blocking probability will be. This can be easily understood by the statistic averaging effect of a large number of cooperating servers. Formula (2-11) also justifies the use of a wide band system which can provide a large number of virtual connections simultaneously. Calculations of (2-11) suggests that for  $S_f \geq 50$ , the blocking probability B(s, a) will be less than 0.02 for normalized offered load  $a' \leq 0.8$ . In





Figure 2-2(b) Blocking Probability B(s, a) versus Offered Load a'

such case where B(s, a) is very small, the offered load is approximately equal to the carried load, as can be seen from (2-12).

Thus, in a wide band system  $(S_f \ge 50)$  for  $\rho_v \le 0.8$ ,  $\rho_v$  can be approximately calculated by

$$\rho_v = \frac{\lambda}{\mu S_f} \tag{2-13}$$

Generally speaking, if blocking probability is negligible, in steady state, the flow balance condition gives,

flow in rate = flow out rate  

$$\lambda = \overline{n}\mu, \qquad \overline{n} = \frac{\lambda}{\mu}$$
  
 $\rho_v = \frac{\overline{n}}{S_f} = \frac{\lambda}{\mu S_f}$ 
(2-14)

that is

which is the same as (2-13). (2-14) is true regardless of distributions of arrival and call length processes.

It is important to point out that, in a wide band system, although the blocking probability can be neglected for *voice calls*, it plays an important role in data performance, this is to be discussed in chapter 3.

The variance of slot occupancy j, denoted as  $\sigma_j^2$ , is given by

$$\sigma_j^2 = \sum_{j=0}^s j^2 P_j - \overline{n}^2 = \overline{n} \left[ 1 - \frac{B(s,a)(s-\overline{n})}{1 - B(s,a)} \right]$$
(2-15)

where  $s = S_f$ .

Of more importance is the relative variation rate,  $\Delta P_j$ , defined as

$$\Delta P_j = (j_{max} - j_{min})/S_f$$

such that when  $0 \leq j \leq j_{min}$  and  $j_{max} \leq j \leq S_f$ , the (tail) probabilities  $P_j \leq P_{sp}$ , where  $P_{sp}$  is a pre-specified value. To data transmissions, these tail probabilities indicate directly the availability of slots. Figure 2-3 plots  $P_j$   $(j = 0, 1, \dots, S_f)$ for  $S_f = 50,100,200$  and  $\rho_v = 1/3, 1/2, 2/3$ ,  $P_{sp} = 0.005$ . It can be noticed that  $\Delta P_j$  decreases with the increase of  $S_f$ .



Figure 2-3 Tail Probabilities,  $P_{sp} = 0.005$ 

Changing rate of  $\rho_v$  in *time* is the same as that of *j*. From (2-5) and (2-6), the arrival rate is  $\lambda$ , and the completion rate is  $\mu j$ . The total changing rate of

 $\rho_v$  for state *j*, denoted as  $\Delta \rho_v$ , is

$$\Delta \rho_v = \lambda + \mu j \tag{2-16}$$

maximum  $\Delta \rho_v$  and average  $\Delta \rho_v$  are thus

$$\Delta \rho_{v_{max}} = \lambda + \mu \cdot S_f$$

and

$$\Delta 
ho_{v_{ave}} = \lambda + \mu \cdot \overline{n} pprox 2\lambda$$

A typical numerical example of  $\Delta \rho_v$  follows: take  $T_f = 22$ ms,  $\lambda = 1.0$ call/sec = 0.022 call/frame,  $\mu^{-1} = 1.1$  minutes = 3000 frames/call,  $S_f = 100$ , then

 $\overline{n} \approx a = 66$ 

 $\Delta 
ho_{vave} = 0.022 + 66/3000 = 0.044$  change/frame = 1 change/23 frames

that means  $\rho_v$  will not change for 22 frames on the average. Compared to data arrival rate, which is on the order of 1 arrival in a few slots,  $\rho_v$  can be said a very slowly time varying process.

#### 2.2.2 State Correlation of Voice Occupied Slots

One of the complexities in evaluating the effect of voice pattern background on data performance arises from the fact that, even for the backgrounds of the same voice utilization, different distributions of voice occupied slots produce significant difference in data packet performance. This can be illustrated using two extreme cases(figure 2-4).



Figure 2-4 Illustration of Voice Correlations and Data Transmissions

In general, states of slots (state 1, occupied by a voice virtual connection; state 0, not occupied) are not completely randomly distributed, indicating that there exists correlations among states of slots.

Define  $\gamma(k)$  to be the correlation coefficient for states of slots k slots apart,

then

$$\gamma(k) = \frac{E\left[\tilde{V}(n)\tilde{V}(n+k)\right] - \left\{E\left[\tilde{V}(n)\right]\right\}^{2}}{\sigma_{v}^{2}}$$

$$= \frac{P_{r}\{\tilde{V}(n) = 1\}P_{r}\{\tilde{V}(n+k) = 1|\tilde{V}(n) = 1\} - \rho_{v}^{2}}{\rho_{v}(1-\rho_{v})}$$

$$= \frac{P_{r}\{\tilde{V}(n+k) = 1|\tilde{V}(n) = 1\} - \rho_{v}}{1-\rho_{v}}$$
(2-17)

There are two important special cases:

1) 
$$k = S_f$$
,  
 $P_r\{\tilde{V}(n+k) = 1 | \tilde{V}(n) = 1\} = 1 - \mu$  (2-18)

and

$$\gamma(k) = 1 - rac{\mu}{1-
ho_v}$$

Take  $\mu = 0.0005$ ,  $\rho_v = 0.5$ , then

$$\gamma(k)|_{k=S_f}=0.999\approx 1.0$$

That means if slot j in frame i is in state 1 (or 0), then slots j in frames (i-1) and (i+1) are also very likely to be in state 1 (or 0), due to the virtual connection protocol.

From (2-17) and (2-18),

$$1.0 \approx \gamma(k)|_{k=S_f} = \frac{\mathbf{E}\left[\tilde{V}(n)\tilde{V}(n+S_f)\right] - \rho_v^2}{\mathbf{E}\left[\tilde{V}(n)\tilde{V}(n)\right] - \rho_v^2} \tag{2-19}$$
$$\iff \tilde{V}(n) \approx \tilde{V}(n+S_f) \quad \forall n$$

The above is interpreted as:  $\tilde{V}(n)$  is (quasi) periodic in the sense of probability, with period  $S_f$ ; the (quasi) periodicity of  $\tilde{V}(n)$  determines that  $\gamma(k)$ is also (quasi) periodic with the same period  $S_f$ .

2) Define random variable L to be the number of consecutively occupied slots plus 1, as shown in Figure 2-4, and  $P_r(l)$  be its probability density function; then there exists an important relationship between  $P_r(l)$  and  $\gamma(k)$ , i.e.,

$$\gamma(k) = 0 \iff P_r\{\tilde{V}(n+k) = 1 | \tilde{V}(n) = 1\} = \rho_v \quad \forall n, k \neq 0$$
$$\iff P_r(l) = \rho_v^{l-1}(1-\rho_v)$$
(2-20)

The forward equivalence comes directly from the derivation of geometric distribution; the backward equivalence states the fact that geometric distribution is memoryless.

For an arbitrary k, however, the correlation among slots is much harder to predict. It is related to many factors such as frame size  $S_f$ , voice utilization  $\rho_v$ , and in particular, voice assignment protocols (VAP-I and VAP-II).

In the next chapter where analysis of data performance is carried out, the independence of states of the slots is assumed, i.e.,  $\gamma(k) = 0$ ,  $\forall k \neq 0$ . The assumption is necessary for the performance evaluations to be mathematically tractable. The question is how strongly the slots are correlated. If the correlations are weak,<sup>\*</sup> it would be reasonable to assume the independence. Unfortunately, the question is very difficult to answer analytically. As the correlations are important factors on data performance, they will be obtained by the simulations

<sup>\*</sup> Except at  $k = nS_f$ , where n is an integer. With  $S_f$  in the wide band range, these "spikes" are sparsely spaced.
in the next section.

Since  $\tilde{V}(n)$  is assumed to be stationary and ergodic,

$$\gamma(k) = \mathbb{E}[\tilde{V}(n)\tilde{V}(n+k)] = \lim_{T \to \infty} \frac{1}{2T} \sum_{n=-T}^{T} \tilde{V}(n)\tilde{V}(n+k)$$
(2-21)

which can be estimated as

$$\hat{\gamma}(k) = \frac{1}{M-k} \sum_{n=1}^{M-k} V(n) V(n+k)$$
(2-22)

where  $V(n), n = 1, 2, \cdots, M$ , denotes a sample of voice pattern process  $\tilde{V}(n)$ .

# 2.3 Simulations of Voice Pattern Structure

Voice arrival, virtual connection and voice completion processes were simulated. Samples of simulated  $\tilde{V}(n)$  are used to obtain the estimates of  $\gamma(k)$  and  $P_{\tau}(l)$  in this section.

# 2.3.1 Choice of Parameters

Basic parameters relevant to the discussion here include frame length  $T_f$ , frame size  $S_f$ , channel capacity C, voice coding rate  $K_c$ , voice arrival rate  $\lambda$ , and voice completion rate  $\mu$ . The voice assignment protocols VAP-I and VAP-II are also treated as one parameter in the system.

The ranges of parameters used in the simulations are as follows:

- 1. frame length  $T_f$ : 22ms (typically 10-30 ms);
- 2. frame size  $S_f$ : from 50 to 400 slots per frame. The length of a slot in time unit is thus 0.44(=22/50) ms to 0.055(=22/400) ms. Because a slot is the entity with the smallest time unit in the system, the length of a slot is selected as the time unit of the simulation. Since frame length  $T_f$  is fixed as 22 ms, the increase or decrease of  $S_f$  means a corresponding increase or decrease in total channel capacity C (see (2-1));
- 3. Voice arrival rate λ and completion rate μ: their ratio normalized to S<sub>f</sub> determines the voice utilization ρ<sub>v</sub> (see (2-14)). In order to simulate Δρ<sub>v</sub>, average call length μ<sup>-1</sup> should be chosen as close as possible to the real situations, which is typically about a few minutes. However, an actual calculation suggests that the real time parameter is an enormously large number in the simulation time units of frames and slots. Take μ<sup>-1</sup> = 3 minutes for example, the call length in frames is 3mi × 60 × 1000/22ms = 8181 frames, if S<sub>f</sub> is 100, the call length in slots is 818100 slots, a simulation of very long length has to be carried out in order to get any useful information. In considering the tradeoffs between the reliability of simulation and cost of simulation, μ<sup>-1</sup> is chosen as 44 seconds in real time, which corresponds to 2000 frames in simulation time. (μ = 0.0005 completion/frame).

In the following simulations whenever  $\rho_v$  needs to be changed, it is done by varying  $\lambda$  while fixing  $\mu = 0.0005$  completion/frame.

### 2.3.2 Results of Simulations

The simulation results are presented in 3 groups to demonstrate the effects of  $S_f$ ,  $\rho_v$  and VAP on voice pattern statistics, respectively. 3 runs of different seeds for random number generators were carried out for each result. Results presented here are the means of 3 runs. The standard deviation of the 3 runs is within 5-10% of the mean, unless otherwise noted.

# 1) VAP group

Figures 2-5(a) and 2-5(b) show the estimates of  $\gamma(k)$  and  $P_r(l)$  for  $\rho_v = 0.5$ ,  $S_f = 100$ , VAP-I. In figure 2-5(b), the geometric distribution is shown as dotted curve as comparison with simulated curve(solid). The following can be noticed in the figures:

- In figure 2-5(a), slots of 100 slots(=S<sub>f</sub>) apart are strongly correlated with <sup>ˆ</sup>γ(k) ≈ 1.0 (the calculated value is 0.999, see (2-18)). The (quasi) period-icity of V
   <sup>˜</sup>(n) is reflected by that of γ(k), with period equal to S<sub>f</sub>.
- 2)  $\hat{\gamma}(k)$  decreases with the increase of k, for  $0 < k \leq 5$ , and fluctuates around 0.1, with maximum  $\hat{\gamma}(k) \approx 0.2$ , for  $5 < k \leq 50$ .\*

As comparisons, figures 2-6(a) and 2-6(b) show  $\hat{\gamma}(k)$  and  $\hat{P}_r(l)$  for the same  $\rho_v$  and  $S_f$ , except that VAP-I is changed to VAP-II. As seen from figure 2-6(a),

<sup>\*</sup>  $\hat{\gamma}(k)$  is shown to be symmetric to  $k = 50 (= S_f/2)$ , this is because autocorrelation functions are always even functions plus in this case  $\gamma(k)$  is quasi periodic.

average  $\hat{\gamma}(k)$  increases significantly compared to figure 2-5(a); difference between  $\hat{P}_r(l)$  and the geometric distribution is also much larger in figure 2-6(b) than in figure 2-5(b).

These results indicate that VAP has a strong influence on  $\gamma(k)$ . For VAP-I, it seems reasonable if  $\gamma(k)$  is assumed to be 0; whereas it would be nonrealistic to make the assumption for VAP-II.

Simulations were also run for  $\rho_v = 1/3$ , 2/3,  $S_f = 50$ , 200, and 400 cases with VAP-I and -II. Very similar results as above were obtained.

### 2) $\rho_v$ group

Figures 2-7(a) and 2-7(b) show  $\hat{\gamma}(k)$  and  $\hat{P}_r(l)$  for the same parameters as in figures 2-5 and 2-6, except that  $\rho_v$  is lowered to 1/3. No significant change is found in  $\hat{\gamma}(k)$ .  $\hat{P}_r(l)$  is seen to be slightly closer to the geometric distribution than in  $\rho_v = 0.5$  case.

 $\rho_{v}$  is then increased to 2/3,  $\hat{\gamma}(k)$  and  $\hat{P}_{r}(l)$  are shown in figures 2-8(a) and 2-8(b). Again no significant change in  $\hat{\gamma}(k)$  can be seen. The  $\hat{P}_{r}(l)$  curve shows slightly more deviation from that of the geometric distribution.

Similar results were obtained for VAP-II cases.

# 3) $S_f$ group

Simulation results are shown in figures 2-9, 2-10 and 2-11 for  $S_f = 10, 25, 50$  and 200.

Inspections of these figures suggest that  $S_f$  has effect on correlations, with smallest correlation for  $S_f = 200$  and largest for  $S_f = 10$  (figure 2-9). The differences are not significant, however. Similar situation can be seen for  $P_r(l)$ . For all the four cases,  $\hat{P}_r(l)$  has no significant changes(figure 2-10).

The most noticeable changes of voice pattern statistics are changes of  $P_j$ (figure 2-11). For the same  $\rho_v$ , relative variation rate of  $P_j$ ,  $\Delta P_j$ , increases as  $S_f$ decreases. As a consequence, the probability of blocking condition occurrence in small  $S_f$  case ( $S_f = 10$ , for example) is much more noticeable than  $S_f \geq 50$ , i.e., the wide band range. Compared to the effect of  $\rho_v$ , it can be seen that the major effect of both  $\rho_v$  and  $S_f$  is on the tail probabilities  $P_j$ . The smaller  $S_f$  is and the larger  $\rho_v$  is, the larger  $\Delta P_j$  will be. Besides these major effects, they also influence correlation slightly.

From the above simulation results, the following conclusions can be reached:

- 1. It is a reasonably good approximation to assume the independence of the slot states for VAP-I, while this is not true for VAP-II.
- 2. High  $\rho_v$  and/or small  $S_f$  are related to relatively high probabilities of blocking condition occurrence. It is seen in the next chapter that blocking condition has significant influence on data performance.



Figure 2-5(a) Correlation  $\hat{\gamma}(k)$  for  $\rho_v = 0.5, \text{ VAP-I}$ 



Figure 2-5(b) Slot Distribution  $\hat{P}_r(l)$  for  $\rho_v = 0.5$ , VAP-I



Figure 2-6(a) Correlation  $\hat{\gamma}(k)$  for  $\rho_v = 0.5$ , VAP-II



Figure 2-6(b) Slot Distribution  $\hat{P}_r(l)$  for  $\rho_v = 0.5$ , VAP-II



Figure 2-7(a) Correlation  $\hat{\gamma}(k)$  for  $\rho_v = 1/3$ , VAP-I



Figure 2-7(b) Slot Distribution  $\hat{P}_r(l)$  for  $\rho_v = 1/3$ , VAP-I



Figure 2-8(b) Slot Distribution  $\hat{P}_r(l)$  for  $\rho_v = 2/3$ , VAP-I



Figure 2-9  $\hat{\gamma}(k)$  and  $S_f$ 

35



Figure 2-10  $\hat{P}_r(l)$  and  $S_f$ 

Chapter 2: Voice Packet Transmission

 $g_{\mathcal{B}}$ 



Figure 2-11  $P_j$  and  $S_f$ 

37

### Chapter 3

# **Data Packet Transmission**

Extensive work has been done in recent years in devising multiaccess protocols suitable to a variety of communication environments. A good summary can be found in numerous literature, such as [13,14].

A satellite channel is a multiaccess broadcast environment characterized by long propagation delay of approximately 270 ms for a single hop. In such an environment, random access protocols, i.e., ALOHA protocols are found to be suitable because of their insensitivities to long propagation delay. Reservation protocols are also extensively used in such environments. Combined random/reservation protocols extract the merits from both protocols, at the expense of increased protocol complexities.

In this chapter, ALOHA protocol and reservation protocol are modified to the IVD environment. Approximate analyses are made and compared to the simulation results. With voice being integrated into the system, the effect of slowly time varying availability of channel capacity on the performance of these data multiaccess protocols is shown by these results. A combined random/reservation protocol will be studied in chapter 4.

### 3.1 Multiaccess Protocols in the IVDS Systems

A distributed multiaccess environment is defined as:

- a number of geographically distributed stations(users) wish to communicate by transmitting messages via a single communication channel shared by all stations;
- 2. the channel is the sole means of communication among the stations;
- 3. only a single message can be successfully transmitted over the channel at any given time. If two or more messages are simultaneously transmitted over the channel, these messages interfere (collide) with each other and none of them will be correctly received by the stations for which they were destined.

The multiaccess environment is also known as a *broadcast channel* since a message sent by one station can be heard by all the other stations.

A satellite communication channel with distributed earth stations is a multiaccess environment.

In the multiaccess environment, certain coordination is necessary among the stations in order for them to share the channel. Several issues complicate the problem of channel sharing. First, since the stations are distributed, they must either explicitly or implicitly communicate to each other if they wish to coordinate channel sharing. However, there is only a single communication channel, coordination among the users about channel sharing necessarily requires use of the channel itself. Second, since the stations are distributed, they can never instantaneously know the present status of other stations in the environment; information about other stations is always at least as old as the amount of time required for a message to propagate from one station to another. This problem is especially important for a satellite multiaccess environment with a 270 ms long propagation delay.

A distributed algorithm by which the stations are coordinated in the sharing of the channel is known as a *multiaccess protocol*.

Multiaccess protocols can be classified by the extent of coordination they exercise [14], ranging from weak coordination (such as ALOHA) to strong coordination (such as TDMA, Polling).

The following types of information are commonly used in multiaccess protocols:

- Predetermined rules known to each station when it begins operation; an example is the transmission rules specified in the multiaccess protocols. These predetermined rules will not be changed by individual users during the operation.
- 2. Global information obtained by each station by listening to the channel; An important example of global information is the instantaneous total channel traffic factor, denoted as  $\tilde{g}(t)$ , which is the short time average of the ratio

of total transmissions to the channel capacity, i.e.,

$$\tilde{g}(t) = rac{\text{Total transmissions in time interval}(t, t + \Delta t)}{\Delta t}$$
(3-1)

Only delayed version  $\tilde{g}(t-R)$  of  $\tilde{g}(t)$  can be obtained by individual users because of the propagation delay R.

In an IVD system, the voice pattern structure is another example of global information.

3. Local information known only to a single station; One example of local information is the queue length of individual users. Local information can be transformed into global information by transmission activities.

Based on these information, each station estimates the channel status and schedules its transmission activities accordingly. The designed purpose of a multiaccess protocol is such that the *overall* probability of successful transmission is maximized, so that the desired low-delay high-throughput characteristics is realized.

The propagation delay R plays a very important role in the design of multiaccess protocols. In the environments where propagation delay is insignificant compared to the transmission time of a message, such as in a local area network, many protocols are proposed (e.g. CSMA, polling) which give performance close to that of the so called *perfect scheduling*, — the performance upper bound of multiaccess protocol which is only achievable in a non-distributed environment, i.e., R = 0 environment [13]. In the environment where R is comparable or even larger than the transmission time of a message, such as in a satellite channel, the increase of the delay of information exchange impairs the coordination ability of these protocols. The consequence is that these protocols, though sophisticatedly designed, give no better performance than the simple and little coordinated protocols, e.g., the ALOHA protocol [14]. Because of this reason, ALOHA protocol has been one of the major protocols used in satellite channel. However, the almost complete lack of coordination of stations in ALOHA protocols results in a maximum of only 36% of the channel capacity usable. For the applications where higher throughput is concerned, reservation protocol is used which provides up to 90% of channel utilization. The increased throughput, however, is achieved at the expense of increased delay, since in a reservation protocol, delay is at least 2 times of propagation delay R no matter how small the throughput may be.

Many combined random/reservation protocols have been proposed. They are intended to have the advantages of both random protocol (low delay) and reservation protocol (high utilization). The problem associated with combined protocols are usually the implementation complexities and reliabilities versus the advantage of the protocols. Up to date, for the long propagation delay environment, there has been no single protocol that gives overall satisfactory performance.

On top of the long propagation delay in a satellite channel, the integration of voice into the system introduces fluctuations of channel capacity available to data users. The slowly changing voice background makes it more difficult to transmit data efficiently than if only the reduced channel capacity were available to data users with no voice presence (no fluctuations).

In the following sections, ALOHA protocol and reservation protocol are modified to the IVD environment, so that no conflict between voice transmission and data transmission will occur; and their performance are studied.

### **3.2 ALOHA Protocol in the IVDS System**

Since available slots are interleaved with voice connections, data arrivals during voice occupied slots have to delay their transmission. A natural extension of slotted ALOHA protocol is defined as follows:

- 1. Each user remembers the voice pattern  $\tilde{V}(n)$  in the previous frame; By listening to the channel, it also estimates the instantaneous total channel traffic factor  $\tilde{g}(t)$ , defined in (3-1) with  $\Delta t = S_f$ ; the estimate is denoted as  $\hat{g}$ .\*
- 2. A transmission delay factor  $\alpha$  is defined as

$$\alpha = \min\left(\frac{1}{\hat{g}l}, 1\right) \tag{3-2}$$

where l is the length of consecutively voice occupied slots plus 1, e.g. in figure 3-1, l = 4.

3. Upon a new arrival of data packet, with probability  $\alpha$ , transmit in the next available slot (e.g, 7th slot in figure 3-1); with probability  $1 - \alpha$ , delay the transmission K slots, where K is a random variable uniformly distributed between  $[1, K_{max}]$ .† The delayed transmission repeats the same rule until

<sup>\*</sup> The problems involved with acquisition of  $\tilde{V}(n)$  and  $\hat{g}$  will be discussed in the following subsections.

<sup>&</sup>lt;sup>†</sup>  $K_{max}$  is about one to two times of frame size  $S_f$ . see footnote on page 73.



Figure 3-1 Illustration of Delayed Rule

it is transmitted.

4. If there is a collision, delay the collided transmission K slots where K is the same as above. Note a collision can only be known R time later, where R is usually more than 10 times of frame size (see Chapter two). During this time interval R, newly arrived packets repeat the same rules as above, i.e, selective repeat ARQ is used.<sup>‡</sup>

The following are the discussions involved with  $\tilde{V}(n)$  and  $\hat{g}$  used in the protocol.

<sup>&</sup>lt;sup>‡</sup> Those packets whose acknowledgement has not yet received are buffered. Potentially a large buffer is required, this is assumed to be always satisfied.

# **3.2.1** Acquisition of $\tilde{V}(n)$

Because of the propagation delay R, the voice pattern structure on the ground is different from that in the sky. In the above description, present voice pattern structure was supposed to be known, this actually means the ground voice pattern recorded by every station. Whenever a new voice virtual connection is established, each station marks the corresponding slot occupied, thus, there will be no conflict between data transmissions and voice virtual connections. Due to R, however, the termination of a virtual connection can not be known instantaneously. Suppose in the header of each voice packet, there is an indication whether it is the last packet. This information will not propagate to the other stations until R later. During this time period R, all the other stations assume the virtual connection is still there and no one attempts to transmit in that particular slot, thus a portion of capacity corresponding to the time R is wasted. Since the average call length is much longer than the propagation delay R, however, the actual percentage of the wasted capacity is relatively small. As an example, take average call length to be 60 second, the percentage of unused capacity then is equal to 0.27 s/60 s and is less than 0.5%. Thus it is considered negligible here. Of course, this portion can be reused, as well as the silence period in speech; the topics will be briefly discussed in the last chapter.

### **3.2.2** Estimation of $\tilde{g}(t)$

Two problems are involved with acquisition of  $\tilde{g}(t)$ . First, the earliest available information about channel status is at time t - R; consequently, the available estimation of  $\tilde{g}(t)$  is the delayed version of  $\tilde{g}(t - R)$ . The validity of the use

of such a delayed quantity in coordinating transmission activities relies on the following 2 facts: 1) The load placed on the channel will not change significantly during time R; 2) The voice pattern is very slowly time-varying and does not change much during R.\*

The second problem arises from the fact that due to collision, the exact number of transmissions in  $(t, t + S_f)$  can not be obtained. For a time interval T, the following 3 types of information are available to each user: 1) the number of slots in which no transmissions were attempted,  $n_0$ ; 2) the number of slots in which exactly 1 transmission was attempted,  $n_1$ ; and 3) the number of slots in which two or more transmissions were attempted,  $n_2$ . If there are total of n slots in the time interval T, then  $n_0 + n_1 + n_2 = n$ .

Based on these outcomes of transmission activities during T, an estimator for  $\tilde{g}(t)$  can be constructed. The following are the discussions of 2 estimators of  $\tilde{g}(t)$ , with different suitable situations.

### 1) The maximum likelihood estimator $g_m$ :

Under quite general conditions, the maximum likelihood estimator can be shown to be most efficient (minimum variance) in an asymptotic sense (that is, as the sample size approaches infinity). However, the estimator may be biased for small sample sizes.

<sup>\*</sup> Later on in section 3.5, another estimation algorithm will be studied; the algorithm is designed particularly to minimize the effect of propagation delay R.

Assume that data arrivals obey Poisson distribution and that the average arrival rate in a slot is g, which is the parameter under estimation. If the observations are  $n_1$  and  $n_2$ , ( $n_0 = n - n_1 - n_2$  is a dependent observation), then the probability distribution is an extension of Binomial distribution, i.e,

$$p(n_1, n_2) = \frac{n!}{n_1! n_2! (n - n_1 - n_2)!} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}$$
(3-3)

where

$$p_1 = ge^{-g}, \qquad p_2 = 1 - e^{-g}(1+g)$$

 $p_1$  is the probability of one arrival in a slot;  $p_2$  is the probability of two or more arrivals in a slot. The estimator  $\hat{g}$  is the value of g for which (3-3) is maximized with observations  $n_1$  and  $n_2$ .

With  $n_1$  and  $n_2$  given, the maximization of (3-3) involves only the exponentials

$$p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}$$
(3-4)

Take the logarithm of (3-4) and differentiate with respect to g, and set the result to zero, the following maximization equation is obtained

$$\frac{n_1}{g} + \frac{n_2 g e^{-g}}{1 - e^{-g}(1 + g)} - n + n_2 = 0$$
(3-5)

There is no general analytical solution to (3-5), except for some special cases. The following considers two such cases.

i)  $n_2 = 0$ 

(3-5) becomes

$$\frac{n_1}{g} - n = 0 \qquad g = \frac{n_1}{n}$$

which is the same as (3-1).

ii) 
$$n_1 = 0, n_2 = n$$

$$\frac{nge^{-g}}{1-e^{-g}(1+g)}=0$$

The solution that maximizes (3-3) is  $g \to \infty$ . This actually indicates that the maximum likelihood estimator is disabled. A regularity condition must be imposed so that  $\hat{g}$  will converge. This regularity condition is usually the *a priori* knowledge of g. For instance, in an actual network, g will never exceed certain maximum value, say 5 (i.e,  $g \leq 5$ ). This maximum value, denoted as  $g_x$ , should be used as the value of the estimation  $g_m$  for this case.

The bias and variance of the estimator is

$$b(g_m) = \sum_{n_1, n_2} g_m p(n_1, n_2) - g = \overline{g_m} - g$$

$$Var(g_m) = \sum_{n_1, n_2} g_m^2 p(n_1, n_2) - \overline{g_m}^2$$
(3-6)

Figure 3-2 shows  $b(g_m)$  and  $Var(g_m)$  in (3-6), for  $g = \{0.5, 1, 2.0\}$ , and  $n = \{1 \rightarrow 50\}$ . It is seen that both bias and variance decrease with the increase of n. At  $n \geq 30$ , the estimation errors for  $g \leq 1.0$  are virtually negligible.

### 2) The arithmetic estimator $\hat{M}$

It is often desirable to estimate the number of transmissions, M, during time interval T (in which there are n slots). The slowly time-varying channel load can be estimated by the estimator  $g_m$  introduced in the last subsection. However,





49

Chapter 3: Data Packet Transmission

the change of M is much faster and is in a much wider range than channel load g. (Theoretically M can change from 0 to infinity).

By using the similar method, a maximum likelihood estimator for M can be constructed. However, in terms of actual implementation complexity, a simpler arithmetic estimator  $\hat{M}$  is proposed, i.e.

$$\hat{M} = n_1 + 2n_2 \tag{3-7}$$

It is based on the idea that if g is not very large ( $\leq 1.0$ ), the probability of 3 or more transmissions in a slot is small. (3-7) assumes that in any collision slot there is only 2 transmissions involved, thus  $\hat{M}$  is an under-biased estimator for M. Since M can vary in a wide range, the estimator error is defined in terms of a relative error factor  $\epsilon$ 

$$\epsilon = rac{\hat{M}-M}{M}$$

and its bias and variance are defined as the mean and variance of  $\epsilon$ , respectively,

$$b(\hat{M}) = E[\epsilon] = E\left[\frac{\hat{M} - M}{M}\right]$$

$$Var(\hat{M}) = Var[\epsilon] = Var\left[\frac{\hat{M} - M}{M}\right]$$
(3-8)

To evaluate (3-8), it is assumed that data arrivals obey Poisson distribution, with arrival rate g per slot. In the estimation of M, g is the *a priori* knowledge. The distribution of  $n_1$ ,  $n_2$  with M given, is not the same as (3-3), since slot occupancies are related. If each of the M transmissions is equally likely in any one of the n slots, then the distribution function is given by [Appendix],

$$p(n_1, n_2, M, g) = \sum_{l=l_0}^{n_2} \sum_{m=0}^{l} \frac{(-1)^m m^{(l+M-n_1-n_2)} n! e^{-gn} g^M}{n_1! (n_2-l)! (n-n_1-n_2)! (l+n-n_1-n_2)! (l-m)! m!}$$
(3-9)

where  $l_0 = \max(n_1 + n_2 - n, 0)$ .

The calculations of (3-8) suggest that, although the variance decreases, the bias increases with the increase of n (equivalently, the length of observation period). It is therefore necessary to modify  $\hat{M}$  as follows

$$\hat{M} = n_1 + 2n_2 + d(g, n) \tag{3-10}$$

where d(g, n) is chosen to offset the increase of bias with n.

Denote the bias  $m_{\epsilon}$ , which is given in (3-8), then the variable  $\epsilon - m_{\epsilon}$  has zero mean and the same variance as  $\epsilon$ , i.e.

$$E\left[rac{\hat{M}-M}{M}-m_{\epsilon}
ight]=0$$

which implies that the estimator  $\hat{M}$  should be modified according to

$$rac{\hat{M}-M}{M}-m_\epsilon=rac{(\hat{M}-M\cdot m_\epsilon)-M}{M}$$

The correction term is  $M \cdot m_{\epsilon}$ . Since M is unknown, it is substituted by  $\hat{M} \cdot m_{\epsilon}$ , thus, the modified estimator is

$$\hat{M} = (n_1 + 2n_2)(1 - m_\epsilon)$$
 (3-11)

where  $m_{\epsilon}$  is given in (3-8).

Figure 3-3 shows the bias and variance of  $\hat{M}$  (modified). In the range  $0 \le g \le 1.0$ , the estimator has small bias and variance. The bias still increases with n, but at a very slow rate after  $n \ge 20$ .

Estimation is an important and independent topic. Instead of going into detailed discussion of it, the above 2 algorithms serve as the examples which show the possibility of acquiring  $\tilde{g}(t)$  by estimation to a satisfactory accuracy. In section 3.5, another algorithm will be proposed which shows the possibility of minimizing the effect of propagation delay R. Based on these examples, it is assumed in the following simulations and analysis that  $\tilde{g}(t)$  can be obtained free of error. The effect of estimation error of  $\tilde{g}(t)$  is studied in section 3.5.

### **3.2.3 Performance Analysis of ALOHA Protocol**

There are basically 2 analytic approaches to ALOHA protocol, i.e., the steady state approach by Abramson [16] and dynamic Markovian process approach by Kleinrock et al [17]. In this section, the former approach is used.

The following assumptions are made:

 Infinite population model is used for the analysis. The data packet arrival process is Poisson. It has been proven that for number of users larger than 10, the performance of infinite population model is very close to that of a corresponding finite population model.



Figure 3-3 Bias and Variance of Modified Arithmetic Estimator  $\hat{M}$ 

2. The Independence Assumption:

$$\gamma(k)=0 \qquad orall k
eq 0$$

(chapter 2). Note this implies that

$$P_{r}(l) = \rho_{v}^{l-1}(1 - \rho_{v}) \tag{3-12}$$

where l is the length of consecutively occupied slots plus 1.

The results in chapter two (the last section) suggest that the independence assumption is reasonable for low and medium voice utilization ( $\rho_v = 0.0$  to approximately 0.7) and VAP-I.

- 3. Two general assumptions for all analysis and simulations in this thesis:
  - a) The size of packets is of the fixed length of 1 (slot);
  - b) Transmission/reception errors are caused only by packet collisions; those caused by channel noise, etc, are negligible.

### Define:

- random variable L to be number of consecutively voice occupied slots plus 1 (figure 3-1);
- random variable M to be number of data arrivals during the L interval, obeys Poisson distribution;
- variable  $\alpha_l$  to be delayed transmission factor as defined in (3-2) with the difference that  $\tilde{g}(t)$  is replaced by steady state average total channel traffic

factor G, defined as

$$G = \overline{\tilde{g}(t)} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \tilde{g}(t) dt$$
 (3-13)

Let

$$S(l,m) = P_r[1 \text{ successful transmission } | \text{ given } L = l, M = m]$$
  
=  $m(1 - \alpha_l)^{m-1} \alpha_l$  (3-14)

Let S be the average of S(l,m) over all values of l and m, then

$$S = \sum_{l,m} S(l,m) P_r(m|l) P_r(l)$$
 (3-15)

where

$$P_r(m|l) = \frac{(Gl)^m}{m!} e^{-Gl}$$

Thus

$$S = \sum_{l,m} m(1 - \alpha_l)^{m-1} \alpha_l P_r(l) \frac{(Gl)^m}{m!} e^{-Gl}$$
  
=  $\sum_l \alpha_l e^{-Gl} (1 - \alpha_l)^{-1} P_r(l) \sum_{m=1}^{\infty} \frac{[(Gl)(1 - \alpha_l)]^m}{(m-1)!}$  (3-16)  
=  $\sum_l \alpha_l Gl e^{-\alpha_l Gl} P_r(l)$ 

Define the normalized data utilization,  $\rho_d$ , as

 $\rho_d = \frac{\text{number of slots used by data transmissions}}{\text{number of slots available to data transmissions}}$ 

then since in every L = l slots, (l-1) slots are occupied by voice and one slot is available to data, S is numerically equal to the normalized data utilization  $\rho_d$ . The probability of successful transmission  $P_s$  is given by

$$P_s = S(1 - \rho_v)/G$$
  
=  $\sum_l \alpha_l l(1 - \rho_v) e^{-\alpha_l G l} P_r(l)$  (3-17)

The average retransmission number E is given by

$$E = 1/P_s - 1 \tag{3-18}$$

The total delay is given by

$$D_A = E(R + K_{max}/2) + 1.5 + R \tag{3-19}$$

As a verification, if  $P_r(l) = \delta(l-1)$ , that is no voice case, from (3-17) and (3-18),

$$E=e^G-1$$

which is the familiar formula for no voice case.

If  $P_r(l) = (1 - \rho_v)\rho_v^{l-1}$  as the result of independence assumption, then from (3-17),

$$P_{s} = \sum_{l=1}^{\infty} \alpha_{l} l e^{-\alpha_{l} G l} (1 - \rho_{v})^{2} \rho_{v}^{l-1}$$
(3-20)

### **3.3 Reservation Protocol in the IVDS Systems**

Under voice pattern structure, reservation protocols can be straightforwardly applied, with the adaption that reserved transmissions are only scheduled in the slots not being occupied by voice connections.



Figure 3-4 Illustration of the Reservation Protocol

As in the pure data environment, there are many variations of reservation protocols depending on how the request packets are transmitted. In the ideal reservation protocol, it is assumed that a hypothetical dedicated subchannel for request transmissions exists and the subchannel is error free and collision free. This reservation protocol is assumed here. The problem of request packet transmission will be discussed in chapter 4.

The following is the protocol description:

- Whenever a new packet arrives, send a request via the request subchannel. When the echo of the request is heard, enter the request into the global queue;
- 2. Wait until the beginning of the next frame. Suppose this is the *i*th frame, and suppose the request is in the *k*th position in the global queue at the beginning of the *i*th frame, and the total number of slots not being occupied by voice in the *i*th frame is SSN, where  $0 \leq SSN \leq S_f$ .

If  $k \leq SSN$ , transmit the packet corresponding to the request in the kth unoccupied slot.

If k > SSN, delay to the beginning of the (i+1)th frame; repeat the procedure 2. The updated position of the request at the beginning of the (i+1)th frame is (k - SSN)th, since SSN requests ahead of it have been transmitted during the *i*th frame.

An illustration is given in figure 3-4.

In the above protocol, transmissions take place only at the beginning of each frame, instead of transmissions for each request at the time of its arrival at the head of the global queue. The difference between the two is that the total delay of the former is  $S_f/2$  slots larger than that of latter, because in the former, the time from a request enters the queue to the beginning of the following frame is on the average  $S_f/2$ . However, since the basic overhead of reservation protocol 2R is more than 20 times larger than  $S_f/2$ , the difference will be considered negligible in the IVD environment under discussion.

### 3.3.1 Performance Analysis of the Reservation Protocol

The total delay of reservation protocol  $D_R$  consists of the  $2R + S_f/2$  overhead plus the queueing time  $D_q$ , which is the time a request resides in the global queue until the corresponding reserved packet is transmitted, i.e.,

$$D_R = 2R + S_f / 2 + D_q (3-21)$$

 $D_q$  is different for VAP-I and VAP-II. It is noted however, that the total number of reservations in the global queue at the beginning of the *i*th frame, denoted as  $q_i$ , is the same for VAP-I and VAP-II. If  $Q_d$  denotes the expectation of  $q_i$ , then  $D_q$  for VAP-I is given by (figure 3-5(a))

$$D_q = ar{l} \cdot rac{Q_d+1}{2}$$

Under independence assumption,

$$\bar{l}=\frac{1}{1-\rho_v}$$

thus

$$D_q = \frac{Q_d + 1}{2(1 - \rho_v)} \tag{3-22}$$

For VAP-II,  $D_q$  can be calculated by (see figure 3-5(b)),

$$D_q = \rho_v S_f + \frac{Q_d + 1}{2} + \left\lfloor \frac{Q_d}{(1 - \rho_v)S_f} \right\rfloor \cdot S_f$$
(3-23)

An exact analysis for  $Q_d$  is given in [18]. The outline is as follows.



Figure 3-5(a) Global Queueing Delay  $D_q$  for VAP-I



Figure 3-5(b) Global Queueing Delay  $D_q$  for VAP-II

## Define

 $q_i$  = number of data packets in the global queue at the beginning of *i*th frame;

- $v_i$  = number of slots occupied by voice in the *i*th frame;
- $d_{i+1}$  = number of data packet arrivals during the *i*th frame.

Then the distribution of  $q_i$  in the steady state is derived using the moment generating function (mgf) approach. For the data queue size  $q_i$  the following relationship holds between adjacent frames,

$$q_{i+1} = [q_i - (s - v_i)]^+ + d_{i+1}$$
(3-24)

where  $[k]^+ \triangleq \max(0, k)$ ,  $s = S_f$ . The sequence  $\{q_i, v_i\}$  is a two-dimensional Markov chain. For the steady state joint probabilities  $P\{q_i = l, v_i = n\}$ , define the mgf (moment generating function)

$$X_n(z) \stackrel{\Delta}{=} \sum_{l=0}^{\infty} P\{q_i = l, v_i = n\} z^l$$
(3-25)

Then L(z), the mgf for  $q_i$ , can be expressed as

$$L(z) = \sum_{n=0}^{s} X_{n}(z)$$
 (3-26)

where  $s = S_f$ .

It is shown that the following matrix equation holds for the vector  $\mathbf{X}(z) = [X_0(z), X_1(z), \dots, X_s(z)]^t$ , where  $[\ ]^t$  denotes the transpose operation:

$$[I - G(z)\tilde{\mathbf{A}}(z)]\mathbf{X}(z) = G(z)\mathbf{B}(z)$$
(3-27)
where G(z) is the mgf of  $d_i$ ; with Poisson data arrival process,

$$G(z) = \exp[-gs(1-z)]$$

where g is the average arrival rate,  $s = S_f$ . The (s+1)(s+1) elements of  $\tilde{\mathbf{A}}(z)$  are given by

$$A_{mn}=\phi_{nm}z^{-s+n}; \qquad 0\leq m\leq s, \ 0\leq m\leq s$$

where  $\phi_{mn} = P\{v_{i+1} = n | v_i = m\}$  and is given in (2-5) and (2-6). The (s+1) elements of the vector  $\mathbf{B}(z)$  are given by

$$\mathbf{B}(z) = ilde{\phi}^t \mathbf{D}(z)$$

where matrix  $ilde{\phi}^t = [\phi_{mn}], \, \mathbf{D}(z)$  is an (s+1) vector, given by

$$D_n(z) = \sum_{k=0}^{s-n-1} [1-z^{-(s-n)+k}]\pi_{kn}, \quad 0 \le n \le s$$

where  $\pi_{kn} = P\{q_i = k, v_i = n\}$ , which are (s+1)(s/2) unknown boundary probabilities. It can be shown that the equation

$$\det[I - G(z)\tilde{\mathbf{A}}(z)] = 0$$

has exactly (s+1)(s/2) roots on and inside the unit circle [18]. Once the roots are found, a set of (s+1)(s/2) linear simultaneous equations ensue from (3-27) which are solved to determine  $\pi_{kn}$ 's. Then (3-27) can be solved for  $\mathbf{X}(z)$ , and  $Q_d$  is determined by using the relationship

$$Q_d = rac{dL(z)}{dz}|_{z=1}$$

where L(z) is given in (3-26).

Since the number of unknown boundary probabilities  $\{\pi_{kn}\}$  is proportional

to  $s^2$  ( $s = S_f$  throughout the discussion), the above approach may become computationally infeasible when  $S_f$  is large. In [19], an approximation is proposed. The average data queue length  $Q_d$  is given as

$$Q_{d} = \frac{\rho_{d} \sigma_{v}^{2} h(\rho_{d})}{(1 - \rho_{d})(1 - \overline{v})(1 - \omega)}$$
(3-28)

where  $\rho_d$  is the normalized data utilization,  $\overline{v}$  and  $\sigma_v^2$  are the mean and variance of  $v_i$ , given in (2-12) and (2-15), respectively.  $\omega = \gamma(k)|_{k=S_f}$  and is given in (2-18). The most crucial factor here is  $h(\rho_d)$ , which is a convex cup function and ranges from 0 to 1. Numerical (simulation) results indicate that  $h(\rho_d)$  is heavily dependent on the values of s and N, where N is the number of voice sources. In the Poisson arrival case,  $N \to \infty$ . The choice of  $h(\rho_d)$  is highly empirical. For the special case s = N, [19] recommends  $h(\rho_d)$  to be the reciprocal of the normalized queue improvement for M/M/s queueing model, which is given by

$$h(\rho_d)^{-1} = 1 + \frac{(s-1)!(1-\rho_d)}{(s\rho_d)^{s-1}} \sum_{k=0}^{s-2} \frac{(s\rho_d)^k}{k!}$$
(3-29)

When applying the approximation to the reservation protocol under study here, it is found that the validity of the approximation is heavily dependent on the voice statistics, especially the voice blocking probability. For  $S_f$  in the wide band range, blocking probability B(s, a) was shown to be negligible for voice arrivals in chapter 2. But this is not true for the evaluation of  $Q_d$ . It was found that with blocking probability  $10^{-20} \leq B(s, a) \leq 10^{-10}$ , (3-28) gives good approximation to  $Q_d$ ; with  $10^{-10} \leq B(s, a) \leq 10^{-5}$ ,  $h(\rho_d)$  should be modified by using s' in place of s in (3-29), where  $1 \leq s' < s$ . In these B(s, a) range the approximation (3-28) has the accuracy of  $\pm 20\%$ . With  $B(s, a) \geq 10^{-5}$ , the approximation does not apply. The special case s = N considered in [19] corresponds to B(s, a) = 0. Although the exact solution according to (3-27) may be computationally infeasible for  $S_f$  in the wide band range, it is noticed from (3-24) that the condition

$$[q_i - (S_f - v_i)]^+ = 0$$

can be easily satisfied when  $S_f$  is large. If the condition is satisfied at the beginning of every frame, then the queue size is simply the same as the average data arrivals in a frame, i.e.

$$Q_d = \overline{d}$$

Figure 3-6 shows the excess queue size  $\Delta q = (Q_d - \overline{d})/\overline{d}$  for  $\rho_v = 0.5$ ,  $S_f = 100$  and 200. It is seen that  $\Delta q$  is virtually equal to zero in  $0 \le \rho_d \le 0.5$  for  $S_f = 100$  and in  $0 \le \rho_d \le 0.6$  for  $S_f = 200$ . Also shown is  $\Delta q$  for shorter voice length  $\mu^{-1} = 200$ , the range for which  $\Delta q$  is very close to zero is  $0 \le \rho_d \le 0.6$ .

Based on these observations, the following approximation is taken to evaluate the queueing delay  $D_q$ : express  $D_q$  as

$$D_q = D_1 + D_2$$

where  $D_1$  is given in (3-22) for VAP-I and in (3-23) for VAP-II, with  $Q_d$  replaced by  $\overline{d}$ .  $D_2$  is used to represent the missing part  $\Delta Q = Q_d - \overline{d}$ . Choice of  $D_2$ is empirical. Here it is chosen as the delay of M/G/1 queue with geometric service time distribution, this corresponds to the distribution of the length of voice consecutively occupied slots under the independence assumption. Hence  $D_2$  is given by

$$D_2 = rac{\lambda au^2}{2(1-\lambda au)} \left(1+rac{\sigma^2}{ au^2}
ight)$$

The right hand side is the well-known expression for M/G/1 queueing delay,



Figure 3-6 Excess Queue Size  $\Delta q$  versus  $\rho_d$ 

where  $\tau$  and  $\sigma^2$  are the mean and variance of the service time distribution, respectively;  $\lambda = \rho_d$ . With the service time distribution given in (3-12), the above becomes

$$D_2 = rac{
ho_d (1+
ho_v)}{2(1-
ho_v)(1-
ho_d)}$$

Simulation results indicate that the above approximation gives the same accuracy as (3-28). Both approximations are reasonable for low and medium  $\rho_d$  and  $\rho_v$ . Discrepancies become significant when  $\rho_v$  is large (> 2/3).

The total delay of reservation protocol  $D_R$  can be expressed as

$$D_R = 2R + S_f / 2 + D_q (3-30)$$

It can be seen that the queueing delay  $D_q$  contributes only an insignificant amount in the total delay, except for very high  $\rho_v$  and/or  $\rho_d$ . As an example, take  $S_f = 100$ ,  $\rho_v = 0.5$  and  $\rho_d = 0.5$ ,  $D_q = 28.2$  slots from the simulation and is 27.5 by the above approximation, which accounts for 1.2% of total delay  $D_R$ , which is 2478 slots, for  $R = 12S_f$ .

### **3.4 Simulations of the two protocols**

A difficulty in the simulation of IVD system is due to the disparity in the time scales for voice arrival rate and data arrival rate. Voice call interarrival time is in the order of minutes or seconds while data interarrival time is in the order of slots. To run a simulation that is statistically significant in terms of the number of voice calls processed, the number of data packets that are processed will be in the order of a hundred million. This results in very long simulation time and expensive computing cost. Because of the limitation, most results in this section represent means of 9 runs of different random generator seeds. The standard deviations of the means are within 5% of the means unless otherwise specified.

## **3.4.1** Simulations for ALOHA protocol

Simulation results are shown in Table 3-1 for  $\rho_v = 0.0$  (no voice case), 1/3(low

voice case), 1/2 (medium voice), 2/3 (high voice) and 4/5 (very high voice case), with VAP-I. Figure 3-7 plots the curves of mean packet delay with analytic results as comparison.

In general, it can be seen that the analysis gives reasonable accuracies of the real (simulated) delay-throughput characteristics, especially for  $\rho_v$  in the low( $\leq 1/3$ ) and medium (from 1/2 to 2/3) ranges. Discrepancies between analytic and simulated increase with the increase of  $\rho_v$ .

The following factors contribute to the discrepancy between analytic and simulated results:

## 1) Independence of voice pattern:

As the analysis is done under independence assumption, it is expected to have certain discrepancy between the analytic and simulated results. As a check, simulation was run for the simulated independent voice pattern. In Table 3-1, rows corresponding to  $\rho_v = 0.5*$  are the results for simulated independent voice pattern. Compared with  $\rho_v = 0.5$  row which corresponds to VAP-I, the results of simulated independent voice pattern coincides with that of analysis very well.

For VAP-II voice pattern background, simulation results are shown in Table 3-2. The differences between analytic and simulated results are significant even for small  $\rho_v$  range, indicating that the independence assumption is not valid for VAP-II voice pattern. Table 3-1 and 3-2 also suggest that ALOHA performance deteriorates significantly on aggregatively distributed voice pattern compared to the more evenly distributed pattern with exactly the same  $\rho_v$  statistics. Cor-



Figure 3-7 ALOHA Delay-Throughput Characteristics

## Table 3-1 ALOHA delay-throughput with VAP-I

 $R = 1200, S_f = 100, K_{max} = 120$ 

ρυ	Ра	Simulation	Analysis	%Difference
0.0	0.05	1281.6	1269.7	0.94
	0.10	1345.8	1350.7	-0.36
	0.15	1451.5	1448.6	0.20
	0.20	1572.7	1574.4	-0.11
	0.25	1763.3	1742.9	1.17
	0.05	1315.4	1295.0	1.58
	0.10	1422.7	1407.0	1.12
1/3	0.15	1571.3	1550.1	1.36
	0.20	1765.4	1743.7	1.24
	0.25	2027.4	2020.3	0.35
1/2	0.05	1332.2	1307.6	1.88
	0.10	1480.7	1439.8	2.84
	0.15	1675.4	1609.6	4.09
	0.20	1880.9	1847.8	1.79
	0.25	2182.1	2207.3	-1.14
1/2*	0.05	1310.8	1307.6	0.25
	0.10	1431.9	1439.8	-0.54
	0.15	1621.3	1609.6	0.73
	0.20	1819.8	1847.8	-1.52
	0.25	2214.5	2207.3	0.33
2/3	0.05	1407.8	1323.5	6.37
	0.10	1539.6	1473.1	4.51
	0.15	1771.6	1677.4	5.61
	0.20	2029.4	1974.2	2.80
	0.25	2483.5	2442.8	1.67
4/5	0.05	1425.3	1337.2	6.59
	0.10	1629.6	1504.3	8.33
	0.15	1896.1	1739.7	8.99
	0.20	2201.2	2093.2	5.16
	0.25	2853.8	2685.1	6.28
4/5†	0.05	1424.1	1337.2	6.50
	0.10	1627.8	1504.3	8.21
	0.15	1913.9	1739.7	10.01
	0.20	2354.3	2093.2	12.47
	0.25	3111.5	2685.1	15.88

\* Independent voice pattern. †Without input voice control.

٠

F				·····
ρυ	Ρd	Simulation	Analysis	%Difference
0.0	0.05	1279.5	1269.7	0.77
	0.10	1351.2	1350.7	0.04
	0.15	1460.4	1448.6	0.82
	0.20	1565.4	1574.4	-0.57
	0.25	1760.9	1742.9	1.04
1/3	0.05	1471.4	1295.0	13.62
	0.10	1538.3	1407.0	9.33
	0.15	1627.4	1550.1	4.99
	0.20	1731.9	1743.7	-0.67
	0.25	2085.7	2020.3	3.24
1/2	0.05	1501.2	1307.6	14.80
	0.10	1641.6	1439.8	14.02
	0.15	1810.7	1609.6	12.49
	0.20	1991.3	1847.8	7.77
	0.25	2148.1	2207.3	-2.69
	0.05	1780.3	1323.5	34.51
	0.10	1829.8	1473.1	24.21
2/3	0.15	1941.2	1677.4	15.73
	0.20	2109.9	1974.2	6.88
	0.25	2524.7	2442.8	3.35
4/5	0.05	1772.3	1337.2	32.54
	0.10	2063.1	1504.3	37.14
	0.15	2225.6	1739.7	27.93
	0.20	2457.9	2093.2	17.42
	0.25	3166.9	2685.1	17.95

Table 3-2 ALOHA delay-throughput with VAP-II

 $R = 1200, S_f = 100, K_{max} = 120$ 

.

relation of voice pattern is an important factor in the performance of ALOHA protocols in an IVD environment.

## 2) Effect of blocking condition:

Blocking condition occurs in high  $\rho_v$  range. During the blocked period, no data packet can be transmitted. As new packets are continuously arriving, the total channel traffic grows rapidly during the period and the system is very likely to be driven into unstable region [17]. This will result in large increase in the average packet delay. Figure 3-8(a) shows the fluctuation of available slots to data transmission in each frame for a randomly sampled period of time, for  $\rho_v = 0.85$ , together with the corresponding short time average packet delay. Effect of blocking condition can be seen very clearly in the picture.

In order to avoid the occurrence of blocking condition, voice input control mechanism should be implemented, i.e., the number of voice connections in any given frame should be limited to a number  $S_d$ ,  $S_d < S_f$ , so that at least  $S_f - S_d$  slots are always available to data transmissions. Figure 3-8(b) shows the same system of Figure 3-8(a) with voice input control. The improvement on data performance is dramatic, while it has little adverse effect on voice performance. (see Chapter 2 equation (2-11) and figure 2-2).

In Table 3-1, data shown are those with the input voice control introduced as above. As comparison, data in  $\rho_v = 0.8$ <sup>†</sup> rows show the same system without input voice control.<sup>‡</sup>

<sup>&</sup>lt;sup>‡</sup> Since the variance of the delays for high  $\rho_v$  range without input voice control is large, these data can vary significantly.





3) The total channel traffic factor:

Average G is used in the steady state performance analysis while in the simulation instantaneous  $\tilde{g}(t-R)$  is used. This can also contribute to the discrepancy of results between simulation and analysis.

It turned out from the simulations that the instantaneous control mechanism of  $\tilde{g}(t-R)$  is indispensable. The system is likely to go into unstable state because of combined fluctuation of voice background and data arrival, if  $\tilde{g}(t-R)$ is replaced by average G. The stability of ALOHA protocol in the IVD environment is an important measure of performance. In section 3.5, some interesting results about stability will be derived.

- 4) Effects of other system parameters:
  - a)  $K_{max}$ : figure 3-9 shows simulated data delay versus  $K_{max}$  for  $\rho_v = 1/2$ ,  $\rho_d = 0.1$  and 0.2. Dashed curves show corresponding analysis. As in the pure data case,  $K_{max}$  has little effect on ALOHA performance. \*
  - b)  $S_f$ : As can be seen from chapter 2, effect of small  $S_f$  should be similar to that of high  $\rho_v$ , as both are related to the probability of blocking conditions. Figure 3-10 plots simulated data delay versus  $S_f$  for  $\rho_v = 0.5$ ,  $\rho_d = 0.1$  and 0.2. As long as  $S_f$  is in the wide band

<sup>\*</sup>  $K_{max}$  is related to stability. It is shown that for  $K_{max} \ge K_T$ , where  $K_T$  is related to system parameters such as channel load and user population, the further increase in  $K_{max}$  has insignificant effect on ALOHA performance, this is especially so if additional transmission/retransmission control mechanisms are used.



Figure 3–9 ALOHA Delay versus  $K_{max}$ 



Figure 3-10 ALOHA Delay versus S<sub>f</sub>

range (> 50), there is little difference on data performance. For very small  $S_f$ , delay increases as blocking condition is more likely to occur for the same  $\rho_v$  than if  $S_f$  is large.

## **3.4.2** Simulations for reservation protocol

Table 3-3 lists the results for global queueing delay (Total delay  $D_R - 2R$ ) for  $\rho_v = \{0.0, 1/3, 1/2, 2/3, 4/5\}$  with VAP-I.

Mean curves are plotted in figure 3-11. These results indicate the range of  $\rho_v$  for which the independence assumption and the approximate analysis for reservation protocol are valid, together with the accuracies of the approximations.

As in ALOHA protocol, effect of voice pattern structure is shown in Table 3-3 (rows of  $\rho_v = 0.5*$ ) and figure 3-11 for the protocol performance on VAP-II voice background. The voice blocking condition occurs for  $\rho_v \ge 0.8$  which causes the significant increase of the variance of simulation results. The results presented for  $\rho_v = 0.8$  is with the voice input control described in the last subsection.

### Some Observations

1) Performance analysis of multiaccess protocols in the IVD systems:

From the simulation results of ALOHA and reservation protocol, it is seen that Independence Assumption is valid for voice patterns of strong random na-



Figure 3-11 Reservation Delay-Throughput

'n

ρυ	ρd	Simulation	Analysis	%Difference
0.0	0.10	56.4	55.6	1.52
	0.30	66.7	65.7	1.48
	0.50	77.9	76.0	2.44
	0.70	89.0	86.7	2.64
	0.90	102.8	100.0	2.82
	0.10	57.0	55.9	1.94
	0.30	66.6	66.2	0.65
1/3	0.50	75.8	76.7	-1.28
	0.70	86.7	88.1	-1.58
	0.90	102.6	104.7	-2.06
	0.10	58.3	56.2	3.74
	0.30	69.0	66.6	3.49
1/2	0.50	78.2	77.5	0.88
	0.70	88.1	89.5	-1.56
	0.90	106.2	109.5	-2.98
1/2*	0.10	93.2	103.2	-9.73
	0.30	104.7	113.6	-7.87
	0.50	115.9	124.5	-7.39
	0.70	142.4	136.5	4.45
	0.90	173.6	156.5	10.97
2/3	0.10	60.0	57.0	5.33
	0.30	68.0	67.6	0.69
	0.50	80.4	79.0	1.29
	0.70	99.3	92.3	7.67
	0.90	130.1	119.0	9.37
4/5	0.10	64.4	58.8	10.05
	0.30	74.7	69.5	7.54
	0.50	92.7	82.0	13.31
	0.70	115.9	98.0	17.37
	0.90	170.2	138.0	23.21

Table 3-3 Reservation Delay-Throughput with VAP-I

 $R = 1200, S_f = 100$ 

\* with VAP-II

ture such as generated by VAP-I. However, the assumption does not hold with VAP-II. It was seen in Chapter 2 that VAP-II voice pattern tends to have strong correlations. It can be intuitively generalized that for other voice assignment protocols of strong predetermined nature (for example, a protocol which assigns empty slots to new voice calls from the center of the current frame), the Independence Assumption will not hold.

It was shown in Chapter 2 that voice pattern background is a very slowly time varying process compared to the time scale of data activities. Based on this, a *Fixed Pattern Assumption* can be applied to the analysis of voice patterns of strong predetermined nature. The basic idea is to assume that the voice pattern is fixed, and delay for each fixed pattern is calculated; the total delay will then be a weighted sum of those of all possible patterns with weighting factors being probabilities of occurrence of each possible pattern. Two major difficulties are associated with this approach: the first, it is easily seen that the number of all possible patterns increases at a *factorial* rate, with the increase of  $S_f$ . Thus, for  $S_f$  in the wide band range, the complexity of computation could be expected; the second problem is that there are a number of patterns for which average data input rate is larger than the output rate, thus no steady state solution exits for these patterns. When dealing with these situations, generally speaking the method of fluid approximation is used.\*

### 2) Performance degradation of multiaccess protocols in the IVD systems:

It is axiomatic in queueing theory that the performance of a system worsens

<sup>\*</sup> Such an approach was reported in [20], yielding a reasonable result for high  $\rho_d$  range, but is rather inaccurate for low  $\rho_d$  range. The analysis was done on reservation protocols.

as the system becomes more "irregular". Thus the performance in an integrated system can be no better, and in general will be worse, than that of a system in which it is allocated the same fixed capacity. This is shown by the analytic and simulation results of ALOHA and reservation protocols in the last section and this section.

In the distributed IVD environment, the difference of reservation protocol on different voice pattern background is small, since the 2R plus half  $S_f$  overhead contributes the most to the total delay. Consequently, the total delay is not affected significantly by voice utilization. The effect of blocking condition on reservation performance is also smaller than that on ALOHA performance. It does not seem important to distinguish among different voice pattern backgrounds.

On the other hand, ALOHA protocol is affected significantly by difference of voice pattern background (different  $\rho'_v s$ , VAP's and blocking, etc).

Comparison of ALOHA performance with reservation performance leads to an important observation: with the introduction of voice pattern background into IVD system, data access protocols of little coordination deteriorates significantly while data access protocols of strong coordination are not as much affected.

Since the natural extension of ALOHA protocol for pure data environment shows undesired performance deterioration in an integrated environment, especially for VAP-II voice pattern background, another version of ALOHA protocol is proposed in the following section. The basic idea is, by incorporating certain extent of coordination into random nature of ALOHA protocol, it could be less susceptible to the fluctuation of voice pattern background, so that an improved performance in the IVD environment is expected.

## 3.5 ALOHA protocol (The Second Version)

The protocol description is as follows:

- 1. Any new arrival during *i*th frame postpones the transmission until the beginning of the (i + 1)th frame.
- 2. At the beginning of the (i + 1)th frame, with probability p, delay the transmission further to the beginning of the (i+2)th frame; with probability 1-p, go to the next step.
- 3. Choose a number randomly with equal probability from among 1 to  $M_{i+1}$ , where  $M_{i+1}$  is the total number of unoccupied slots in the (i+1)th frame. Suppose the outcome is  $j, 1 \leq j \leq M_{i+1}$ , transmit in the *j*th unoccupied slot.
- 4. Collided transmissions repeat the same procedure as above, as well as delayed transmissions.

p is also called delay factor and is defined as

$$p = \min\left(M/(S_f \cdot g), 1\right) \tag{3-31}$$

where  $g = \tilde{g}(t - R), M = M_{i+1}$ .

Because the unoccupied slot in which data transmission is to take place is chosen randomly and *evenly* from among all unoccupied slots, the positions of these unoccupied slots should have no influence on the slot be chosen; as a consequence, the performance of the protocol should not be affected by the distribution of voice occupied (or unoccupied) slots (VAP-I or -II, for example); moreover, it is shown in the following subsection that with the control in step 2, the effect of fluctuation of voice connections can be greatly reduced.

For the convenience of notation, let the ALOHA protocol described in section 3.2 be called ALOHA-I, and in this section ALOHA-II.

## 3.5.1 Performance Analysis of ALOHA-II

Let  $M_i$  be the number of slots available to data transmissions, and  $n_i$  the number of packets awaiting transmission at the beginning of the *i*th frame.  $n_i$  is also called the backlog. Then

$$n_{i+1} = (n_i - k_i) + d_i + r_i \tag{3-32}$$

where

 $k_i$  - number of packets attempted transmissions (step 3 in the protocol);

- $d_i$  number of new packet arrivals during the *i*th frame;
- $r_i$  number of retransmissions from the previous collisions.

Since  $k_i$  is dependent on  $M_i$  and  $n_i$ , (3-32) describes a 3-dimensional Markov chain  $\{n_i, M_i, r_i\}$ . The transition probabilities for  $M_i$  is known as (see (2-5) and (2-6))

$$P\{M_{i+1} = M - 1 | M_i = M\} = M\mu$$
  
 $P\{M_{i+1} = M | M_i = M\} = 1 - \lambda - M\mu$   
 $P\{M_{i+1} = M + 1 | M_i = M\} = \lambda$ 

The dependency of  $k_i$  on  $n_i$  and  $M_i$  complicates the analysis, which is further complicated by the feedback introduced by the control in step 2. The  $r_i$ term also complicates the analysis.

In the following analysis, the equilibrium point analysis (EPA) and fluid approximation approach is used. To make the analysis tractable, it is assumed tentatively that the estimation of  $\tilde{g}$  is perfect (zero bias and variance for any sample size). The results thus obtained are modified later by taking the estimation error into account.

The basic principle of EPA is that at EP (equilibrium point), the input to the system,  $S_{in}$ , equals to the output of the system,  $S_{out}$ , i.e,

$$S_{in} = S_{out} = S$$

where S is the throughput of the system.

In the following analysis, the expressions of  $S_{out}$  and  $S_{in}$ , conditioned on  $M_i$  and  $n_i$ , are derived; then the above EPA principle is used to evaluate the protocol performance.

# 1) Sout:

Given  $M_i = M$ ,  $n_i = n$ , the probability of number of attempted transmissions, k, (all the subscripts are dropped from now on), is given by

$$p(k|M,n) = \binom{n}{k} p^k (1-p)^{n-k}$$
(3-33)

where

$$p = \min\left(\frac{M}{n}, 1\right) \tag{3-34}$$

the same as (3-31).

Given M, n and k, the probability of number of successful transmissions, m, is given by [Appendix]

$$p(m|k, M, n) = \frac{(-1)^m M! k!}{m! M^k} \sum_{j=0}^M (-1)^j \frac{(M-j)^{k-j}}{(M-j)! (j-m)! (k-j)!}$$
(3-35)

The average number of success,  $\overline{m}$ , given M and n, is

$$\overline{m} = E[m|M,n]$$

$$= \sum_{k} E[m|k,M,n]p(k|M,n) \qquad (3-36)$$

It can be shown that [Appendix]

$$E[m|k, M, n] = k(1 - \frac{1}{M})^{k-1}$$
(3-37)

which does not depend on n when k is given.

Combine (3-33), (3-36) and (3-37)

$$\overline{m} = \sum_{k=0}^{n} \binom{n}{k} p^{k} (1-p)^{n-k} k \left(1 - \frac{1}{M}\right)^{k-1}$$
(3-38)

After some algebraic manipulations,

.

$$\overline{m} = p \ n \left[ p \left( 1 - \frac{1}{M} \right) + 1 - p \right]^{n-1}$$
(3-39)

By (3-34),

$$p = egin{cases} 1, & n \leq M \ rac{M}{n}, & n > M \end{cases}$$

(3-39) is

$$\overline{m} = \begin{cases} n\left(1 - \frac{1}{M}\right)^{n-1}, & n \le M \\ \\ M\left(1 - \frac{1}{n}\right)^{n-1}, & n > M \end{cases}$$
(3-40)

The output of the system,  $S_{out}$ , is

$$S_{out} = \frac{\overline{m}}{\overline{M}} = \begin{cases} \frac{n}{\overline{M}} \left(1 - \frac{1}{\overline{M}}\right)^{n-1}, & n \le M \\\\ \left(1 - \frac{1}{n}\right)^{n-1}, & n > M \end{cases}$$
(3-41)

Define \*

$$g=rac{n}{M}$$

<sup>\*</sup> This definition is slightly different from that in (3-1). Here the denominator is the number of slots not occupied by voice, while in (3-1) it is the total number of slots (occupied as well as occupied by voice) during the observation period.

Then

$$S_{out} = \begin{cases} g \left( 1 - \frac{1}{M} \right)^{M(g - \frac{1}{M})}, & g \le 1 \\ \\ \left( 1 - \frac{1}{gM} \right)^{gM(1 - \frac{1}{gM})}, & g > 1 \end{cases}$$
(3-42)

Figure 3-12 shows  $S_{out}$  versus g, for M = 5, 10, 20 and 100. Two important features of  $S_{out}$  are noticed:

- i) For different M, the difference between  $S_{out}$  is small. In fact, after  $M \ge 10$ , the difference is negligible.
- ii) All curves approach  $e^{-1}$ , when g is large. In fact,  $e^{-1}$  is the asymptotic line for all the  $S_{out}$  curves. From (3-40), when  $n \to \infty$ ,

$$S_{out} = \left(1 - \frac{1}{n}\right)^{n-1} \to e^{-1} \tag{3-43}$$

i) For finite population model and stop-and-wait transmission rule,  $S_{in}$  is given by

$$S_{in} = \sigma(N - n) \tag{3-44}$$

where N is the size of population;  $\sigma$  is the probability that a user has a packet to transmit; and n is the backlogged number of users.

ii) For infinite population model,  $\sigma$  in (3-44) goes to infinitesimal and the



Figure 3-12 Output Rate  $S_{out}$  versus g

product  $\sigma N$  is constant and is the aggregated  $S_{in}$  rate, i.e,

$$S_{in} = \sigma N = \text{constant}$$

It is also the  $S_{in}$  for a finite population if the transmission rule is selective repeat, as is the case for both ALOHA-I and ALOHA-II.

### 3) Equilibrium point and stability:

Putting  $S_{out}$  and  $S_{in}$  together in figure 3-13 (M = 5); the following can be noticed:

- i) If  $S_{in} < e^{-1}$ , that is region I in figure 3-13, there is only one intersection between  $S_{out}$  and  $S_{in}$ , where  $S_{out} = S_{in}$ , that is the EP. Moreover, it is a globally stable EP.
- ii) In region II of figure 3-13, there are 3 intersections between  $S_{out}$  and  $S_{in}$ ; except point A, which is a *locally* stable EP, the rest are unstable EP's. Region II is unstable.

## 4) Throughput S and voice fluctuation:

÷

i) The maximum channel throughput normalized to  $S_f(1-\rho_v)$ ,  $S_{max}$ , is  $e^{-1}$ , which is independent of voice fluctuation. This is shown as the following:



Figure 3-13  $S_{in}$  /  $S_{out}$  and EP's

Chapter 3: Data Packet Transmission

Define

$$p(M) = P_r\{M_i = M\}$$

which is given in (2-7), then the maximum throughput in terms of number of packets is

$$\overline{m}_{max} = \sum_{M} \overline{m}_{max}(M) p(M) \tag{3-45}$$

where  $\overline{m}_{max}(M)$  is

$$\overline{m}_{max}(M) = \lim_{n \to \infty} \overline{m}(M)$$

by (3-40)

$$\overline{m}_{max}(M) = M \mathrm{e}^{-1}$$

thus (3-45) becomes

$$\overline{m}_{max} = M_0 \mathrm{e}^{-1} \tag{3-46}$$

where  $M_0$  is the average number of available slots to data transmissions, from (2-8) and (2-9)

$$M_0 = S_f(1-\rho_v)$$

The normalized throughput is thus  $e^{-1}$ .

ii) For  $S_{in} < e^{-1}$ , the throughput is closely approximated by

$$S = S_{in} = S_{out}$$
  
=  $g_e \left(1 - \frac{1}{M_o}\right)^{-M_o(g_e - 1/M_o)}$  (3-47)

where  $M_o$  is the average number of slots available to data transmissions;  $g_e = n_e/M_o$ , where  $n_e$  represents the EP value of n for  $M = M_0$ . This is shown as the following: The number n - k represents the delayed transmissions per frame; and k - m the number of collided packets after attempting transmissions. Given M and n, the averages are  $n - \overline{k}$  and  $\overline{k} - \overline{m}$ , where

$$\overline{k}=E[k|n], \qquad \overline{m}=E[m|n]$$

respectively. From (3-33),

$$n-\overline{k}=n(1-p)$$

$$\overline{k}-\overline{m}=np\{1-[p(1-rac{1}{M})+1-p]^{n-1}\}$$

From (3-34)

$$n - \overline{k} = \begin{cases} 0, & n \leq M \\ n - M, & n > M \end{cases}$$
(3-48)

and

$$\overline{k} - \overline{m} = \begin{cases} n[1 - (1 - 1/M)^{n-1}], & n \le M \\ M[1 - (1 - 1/n)^{n-1}], & n > M \end{cases}$$
(3-49)

If the input rate per frame is  $n_{in}$ , then when  $M < n_{in}$ , the backlog  $n - \overline{k}$  increases *linearly*, while the backlog  $\overline{k} - \overline{m}$  is *held* at its maximum value, which is

$$M(1 - e^{-1})$$

when available slots returns to  $M \ge n_{in}$ , the excess number of packets in the backlog  $n - \overline{k}$  is cleared; as long as

$$n_{in} \leq M_0 \mathrm{e}^{-1}$$

the steady state condition can be reached, i.e,

$$n_{out} = n_{in}$$

or equivalently,

$$S_{out} = S_{in}$$

the same as (3-47).

Another point of view, when considering the effect of the fluctuation of Mis the following: The variation of M can be equivalent to a perturbation in  $S_{in}$ ,  $\Delta S_{in}$ . As long as  $S_{in}$  is in region I, no matter how large  $\Delta S_{in}$  may be, after a finite period of time after  $\Delta S_{in}$  disappears, EP is obtained with certainty, and the condition  $S_{in} = S_{out} = S$  prevails. The average throughput is thus well approximated by (3-47). However, the larger  $\Delta S_{in}$  is, the longer it will take the system return to the EP. It is thus expected that the fluctuation of M will cause the increase of the variance of throughput S.

The exact value of S is

$$S = \sum_{n,M} S_{out} p(M, n)$$
(3-50)

where

$$p(M,n) = P_r\{M_i = M, n_i = n\}$$

(3-47) approximates (3-50) by letting  $p(M,n) = \delta(M - M_o, n - n_e)$ . From previous literatures[21,22], it has been concluded that for a globally stable EP, such an approximation is accurate.

Simulations in the next subsection well confirm the approximation of (3-47).

## 5) Delay $D_q$ :

 $D_q$  is defined as the delay incurred by a packet due to the delayed transmission and collision retransmission.<sup>\*</sup> Conditioned on n and M,  $D_q(M, n)$  is

$$D_q(M,n) = \left(\frac{n}{\overline{k}} - 1\right) S_f + \left(\frac{\overline{k}}{\overline{m}} - 1\right) (R+K)$$
(3-51)

where R is the propagation delay; K is the retransmission delay;  $R + K \gg S_f$ ;  $\overline{k} = E[k|n], \overline{m} = E[m|n].$ 

By (3-33), (3-39), (3-48) and (3-49),  

$$D_q(M,n) = \begin{cases} \left[ \left(1 - \frac{1}{M}\right)^{1-n} - 1 \right] (R+K), & n \le M \\ \\ (n/M-1) S_f + \left[ \left(1 - \frac{1}{n}\right)^{1-n} - 1 \right] (R+K), & n > M \end{cases}$$
(3-52)

and total  $D_q$  is

$$D_q = \sum_{M,n} D_q(M,n) P_r\{M_i = M, n_i = n\}$$
(3-53)

By the same EP approach about throughput,

$$D_q \approx D_q(M_o, n_e) \\ = \left[ (1 - \frac{1}{M_o})^{1 - n_e} - 1 \right] (R + K), \quad (n_e \le M_o)$$
(3-54)

Since  $n_e < M_o$  for EP, the approximation ignores the term  $(n/M-1)S_f$ .

<sup>\*</sup> The symbol  $D_q$  is used here in the same context as in section 3.3, yet the quantity it represents is different from that in section 3.3.

However, it is true that because of voice fluctuation, the condition n > M is satisfied easily when  $M < M_0$ . The justification of the approximation of (3-53) by (3-54) is due to the fact that  $S_f \ll (R + K)$ . For  $S_f \approx (R + K)$ , such an approximation may cause significant error.

Using the notation  $g_e = n_e/M_o$ , and note  $M_o \gg 1$ , (3-54) is well approximated as

$$D_q = (e^{g_e} - 1)(R + K)$$
(3-55)

### Some Observations

- 1) From (3-46), the normalized maximum throughput is little affected by voice fluctuations.
- 2) From (3-47), the normalized throughput is little affected by voice fluctuations, but the variance of that increases as the result of voice fluctuations.
- 3) From (3-54), very similar conclusion as in item 2 above is reached for delay  $D_q$ , provided that  $S_f \ll (R+K)$ .

## 6) The Total Delay $D_T$ :

The total delay of ALOHA-II,  $D_T$ , consists of the backlogged delay  $D_q$ , as discussed above, plus a header delay  $D_h$  and propagation delay R, i.e.

$$D_T = D_q + D_h + R$$

where the header delay  $D_h$  is the time difference between a packet arrival and its first transmission activity, which can be either a transmission trial or a delayed transmission.  $D_h$  can be further decomposed into

$$D_h = D_1 + D_2$$

where  $D_1$  is the delay from the time of arrival of a packet to the beginning of the next frame, which equals to  $S_f/2$  on the average;  $D_2$  is the time between the beginning of the frame and the actual transmission of the packet, which equals to  $S_f/2$  on the average for VAP-I. For VAP-II it is given by

$$D_2=rac{1}{2}S_f(1+
ho_v)$$

as illustrated in Figure 3-14. Since  $S_f$  is one twelfth of R, the contribution of  $D_h$  to total delay  $D_T$  is insignificant.



Figure 3-14 Illustration of  $D_2$  for VAP-II



Figure 3-15(a) and figure 3-15(b) show simulation results of ALOHA-II, assuming perfect estimation. The effect of voice background is explicitly shown using  $\rho_v$  as x axis variable. The analytic results for ALOHA-II from (3-55) are the points on y axis ( $\rho_v=0$ ). Also shown are the curves of ALOHA-I for comparisons. It can be seen that ALOHA-II does not change with voice pattern background as much as ALOHA-I. Its advantage over ALOHA-I is most clearly seen in the medium and high  $\rho_v$ ,  $\rho_d$  and VAP-II voice pattern background. The values of  $\rho_v$ at the intersections of ALOHA-I and ALOHA-II indicate the boundaries beyond which ALOHA-II is preferred, and vice versa.

## **3.5.2 Effect of Estimation Error**

Let  $n_r$  be the estimated value and n be the real value of backlogged number at the beginning of an arbitrary frame. The estimation error  $\Delta n$  is

$$\Delta n = n_r - n$$

The delay factor p (3-34) now becomes

$$p = \min\left(rac{M}{n+\Delta n},1
ight)$$

From (3-39), expected number of success,  $\overline{m}$ , becomes,

$$\overline{m} = \begin{cases} n\left(1 - \frac{1}{M}\right)^{n-1}, & n + \Delta n \le M \\ \frac{n}{n + \Delta n} \left(1 - \frac{1}{n + \Delta n}\right)^{n-1}, & n + \Delta n > M \end{cases}$$
(3-56)

Let  $\delta(n)$  be the relative estimation error, i.e,

$$\delta(n) = \frac{\Delta n}{n}$$

then the case  $n + \Delta n > M$  in (3-56) becomes

$$\overline{m} = \frac{1}{1+\delta(n)} \left[ 1 - \frac{1}{n[1+\delta(n)]} \right]^{n-1}, n+\Delta n > M$$
(3-57)

Let  $l = n[1 + \delta(n)]$ , then (3-57) becomes

$$\overline{m} = \frac{1}{1+\delta(n)} \left\{ \left(1-\frac{1}{l}\right)^l \right\}^{\frac{1}{1+\delta(n)}-\frac{1}{l}}, n+\Delta n > M$$
(3-58)

Figure 3-16 illustrates (3-56) for M = 5,  $\delta(n) = \pm 0.3$ ,  $\pm 0.5$  and  $\pm 0.7$ . The asymptotic value of  $\overline{m}$ , when  $n \to \infty$ , is

$$\lim_{n \to \infty} \overline{m} = \frac{1}{1 + \delta(n)} e^{-\frac{1}{1 + \delta(n)}}$$
(3-59)

where  $\delta(n) = \pm 0.3$ ,  $\pm 0.5$  and  $\pm 0.7$ . Maximum value of (3-59) is obtained at  $\delta(n) = 0.0$  and is  $e^{-1}$ .

In general, however,  $\delta(n)$  is a random variable. Given n, its probability distribution function is defined as

$$p(e|n) = P_r\{\delta(n) = e|n\}$$
(3-60)

assuming that the probability distribution exists.

Finding function p(e|n) is difficult. A few approximations can be used. For instance, one can assume, if  $\delta(n)$  is not too large, that  $\delta(n)$  is uniformly distributed between  $(\delta_1, \delta_2), -1 \leq \delta_1 \leq \delta_2$ , where both  $\delta_1$  and  $\delta_2$  may be dependent on n.


Figure 3-16 Output Rate  $S_{out}$  and Estimation Error  $\delta(n)$ 

On the other hand, (3-59) gives a general picture of how the performance can be affected by estimation error  $\delta(n)$ . In particular, it can be seen from figure 3-16 that a positive estimation error, which corresponds to a positively biased estimator, does not shrink the boundary of the globally stable region I (figure 3-13) as much as a negative error, which corresponds to a negatively biased estimator. (3-59) becomes 0 if  $\delta(n) = -1$ , which corresponds to the case where no control mechanism is used.

Simulation results of estimation error are shown in figures 3-17 and 3-18.<sup>\*</sup> The estimator error  $\delta(n)$  is assumed to be uniformly distributed in [a, b] and independent of n, where a and b are given as the first and the second number in the figure legends. Figure 3-17(a) shows mean backlogged delay  $D_q$  in frames; For various estimation errors, the differences are small. It is noticed that unbiased estimator of error range [-0.3,0.3] (curve marked 'o') gives almost the same results as an underbiased estimator of halved error range [-0.3,0.0] (curves marked ' $\Delta$ '), and an overbiased estimator of halved error range [0.0, 0.3] (curves marked ' $\diamond$ '). The curves marked ' $\oplus$ ' correspond to ALOHA delay-throughput curve for pure data environment ( $\rho_v = 0$ ).

Figure 3-17(b) gives some insight of how the control mechanism works, where  $E_1$  is the expected number of delayed transmissions, and  $E_2$  the expected number of collisions. They are defined as

$$E_1=\overline{n}/\overline{k}-1, \qquad E_2=\overline{k}/\overline{m}-1$$

respectively. From figure 3-17(b), it is seen that  $E_2$  is close to that of pure ALOHA curve( $\oplus$ ) due to the control mechanism, while  $E_1$  increases as  $\rho_d$  and

<sup>\*</sup> In these simulations, the average voice length  $\mu^{-1}$  is 500 frames.



Figure 3-17 (a) Backlog Delay  $D_q$  and  $\rho_d$ , for  $\mu^{-1} = 500$ 

Figure 3-17 (b)  $E_1$  and  $E_2$  versus  $ho_d$ , for  $\mu^{-1}=500$ 

100



Figure 3-18 (a) Backlog Delay  $D_q$  and  $\rho_d$ ,  $\mu^{-1} = 50$ 

Figure 3-18 (b)  $E_1$  and  $E_2$  versus  $\rho_d$ ,  $\mu^{-1} = 50$ 

Chapter 3: Data Packet Transmission

101

 $\rho_v$  increase. The pattern that  $E_1$  changes with  $\rho_d$  is very similar to the excess queue size in the reservation protocol discussed in section 3.3, in that both of them remain near zero when  $\rho_d$  is below a certain value, and abruptly increases after that value (which is about 0.25 in figure 3-17(b)). Figure 3-18(a) and (b) show the same system with  $\mu^{-1} = 50$  frames. It is seen that  $E_2$  remains basically unchanged while  $E_1$  has been significantly reduced. In this sense,  $E_1$  makes a good analogue to the excess queue size in the reservation protocol (see figure 3-6).

The actual probability distribution of  $\delta(n)$  depends on the estimation algorithm. The above simulations, on the other hand, indicate that ALOHA-II does not require a very accurate estimation of n. This can be explained by the S-G curve:  $s = ge^{-g}$ , which is rather flat in the region around g = 1.0.

In section 3.2, two examples show the possible implementation of estimators. In the following, an algorithm is proposed which minimizes the effect of propagation delay R.

First it is realized that a randomized retransmission for collided packets is unneccessary, since the delayed transmission control mechanism serves perfectly the purpose of controlling new transmissions as well as retransmissions (there is actually no distinction between the two). As a result, any collided packet can start retransmission, according to the protocol description, as soon as the collision is heard after one propagation delay R. The relationship among the backlogged number at the beginning of the (i+1)th frame,  $n_{i+1}$ , and that at the *i*th frame,  $n_i$ ; the total new data arrivals during the *i*th frame,  $d_i$ ; and the total retransmissions heard during the *i*th frame,  $r_i$ , is (a variation of (3-32))

$$n_{i+1} = (n_i - k_i) + d_i + r_i \tag{3-61}$$

(3-61) suggests an algorithm for the estimation of  $n_{i+1}$  by

- 1) Estimation of number of delayed transmissions  $(n_i k_i)_{est}$  from the previous frame;
- 2) Estimation of number of arrivals in the previous frame,  $d_{i,est}$ ;
- 3) Estimation of number of retransmissions heard in the previous frame,  $r_{i,est}$ .

Both  $d_{i,est}$  and  $r_{i,est}$  can be estimated using the similar methods as in section 3.2. For  $(n_i - k_i)_{est}$ , it can be estimated by

$$(n_i - k_i)_{est} = n_{i,est}(1 - p)$$
 (3-62)

where  $n_{i,est}$  is the estimate of  $n_i$  at the *i*th frame, and

$$p = \min\left(rac{M}{n_{i,est}},1
ight)$$

It is possible that  $n_{i,est}$  was an erroneous estimate of  $n_i$ . It can be shown, by exhaustive numerical values, that (3-62) always converges, i.e, error  $(n_{(i+1)est} - n_{i+1}) < (n_{i,est} - n_i)$ , for arbitrary initial value of  $n_i$ .

In the next chapter, a combined random/reservation scheme is studied. ALOHA-II will be assumed throughout the discussions.

#### Chapter 4

# **Combined Random/Reservation Protocol**

It has been shown in the previous chapter that, with the control algorithm introduced in section 3.5, ALOHA protocol can be used in an IVD environment, giving performance close to that in the pure data case. Delay of less than 2Rcan be achieved, with guaranteed stability. However, the maximum throughput of the protocol is still limited to  $e^{-1}$ . On the other hand, although the delay for reservation protocol is larger than 2R, its throughput can be extended to almost 1.0. It is generally true that there does not exit a single multiaccess protocol that performs better than all the others over the entire range of system throughput. Although some characteristics of a system are unlikely to change during the operation, it is certain that the load placed upon the system will be time varying. An adaptive strategy for choosing an access protocol suitable to the varying need is desired, so that optimality is maintained at all times.

In order to extract the merits of low delay of the ALOHA protocol and high throughput of reservation protocol, while avoiding the disadvantages of them, many combined random/reservation protocols have been proposed in the pure data environment. In this chapter, a combined random/reservation protocol is proposed and studied under the IVDS environment.

## 4.1 Mode Switching Information

The adaption ability is achieved by exchange of information among the distributed decision makers. The type and amount of information required by an adaptive strategy, as well as the implementation of the information acquisition mechanism, are among the most crucial factors in determining the performance and robustness of the strategy.

Local information and global information or a combination of both can be used in order to accomplish adaptivity of protocol switching according to the system status.

One example of local information is the length of user queue. If the queue length is smaller than some prespecified threshold, random access protocol is selected. With the increase of channel load, more collisions will occur which increases the length of user queue. Once the queue length is larger than the threshold, access mode is switched to reservation protocol.

An important example of global information, which was used repeatedly in the previous chapter, is the total channel traffic factor  $\tilde{g}(t)$ . If the channel is lightly loaded,  $\tilde{g}(t)$  will be small most of the time, which indicates that ALOHA access protocol can be most suitably used. On the other hand, in the heavily loaded condition,  $\tilde{g}(t)$  is large most of the time, indicating that reservation protocol is prefered.

As global information provides an overview of the system status, it is expected to be a better candidate than local information in achieving the adaptivity of access protocol selections. The disadvantage with global information is that acquisition of global information is much more difficult than that of local information.

Since the instantaneous total traffic factor  $\tilde{g}(t)$  is used in the control of ALOHA protocol discussed in the last chapter, it is natural to extend its use as the mode switching information as well.

#### 4.2 Reservation Request Transmissions

In chapter 2, it was assumed that requests for setup of voice virtual connections were transmitted via a dedicated subchannel. In chapter 3, the same assumption was made on the transmission of reservation requests. For the multiaccess environment where there is only one channel available, the implementation of request transmissions requires that a certain portion of channel capacity be reserved as the subchannel, e.g., a subframe can be reserved for exchange of control information in each frame, as in the WB SATNET system. Due to this requirement, the actual achievable channel utilization for reservation protocol is smaller than what was seen in chapter 3 where a free hypothetical subchannel was available. As the portion of channel capacity reserved for request transmission is kept as small as possible, the maximum rate at which requests can be sent is limited. In the case of fixed request subchannel assignment, this limitation implies that the maximal number of users a reservation protocol can actually support is also limited. In the case of random request subchannel assignment, the maximum rate of request transmission, with any specified probability of success, is limited by the capacity reserved to the request subchannel.

The request subchannel in SENET structure can be implemented by reserving a number of slots in each frame and partition the slots further into the mini-slots, each mini-slot is of the length of a request packet plus the guard times which can no longer be neglected since guard time is comparable with the size of request packets. The mini-slots can be assigned to each individual user dedicatedly if the number of users is not large, or on ALOHA contention basis if the number of users is large, since the minimal size of mini-slots is limited by the guard times and implementation difficulties of synchronizations. (The size of request packet is usually very small).

Combined random/reservation protocols provide a natural way of request transmission, i.e., attaching the request packets to the data packets transmitted via random access mode, this method is also called piggybacking. Since no extra guard times and synchronization mechanism are necessary, piggybacking causes very little increase of channel load and implementation costs. Moreover, since no partition of mini-slots is necessary, the access protocol can be much simpler than the case where mini-slots are used.

### 4.3 The Combined Random/Reservation Protocol

The basic idea is to make use of the instantaneous channel traffic factor  $\tilde{g}(t)$  on deciding one of the 2 access modes: random access and reservation access, both have been discussed in chapter 3.

The requests of reservation are sent by attaching to the packet currently being transmitted, which can either be a random transmission or a pre-reserved transmission.

The protocol is assumed to operate under the simplest and worst user population case, i.e, each user generates one voice traffic and/or data traffic. Data traffic and voice traffic are assumed to be *independent* of each other even if they are from the same user. In a more general case where data traffic and voice traffic are not independent, it is possible to improve the performance of the proposed protocol by incorporating several known methods, e.g, by making use of voice silence period for data transmissions, or reservation request transmission by attaching to the voice packet in virtual connections, if there is one. On the other hand, the voice setup request transmissions can also be sent via data packets in transmission. Some of these topics will be discussed in Chapter 5.

Since voice traffic and data traffic are assumed to be independent of each other, the setup transmission of voice virtual connection is via random access mode. It will be shown later in section 4.3.3 that required blocking probability can be easily satisfied.

#### **4.3.1 Protocol Description**

The protocol is described as follows: first, the data access part; then the voice access part; and finally the interactions between them.

#### The data access protocol (figure 4-1)

Each user keeps a record of its present transmissions. The record includes: i) the number of total packets in the buffer of the user,  $N_t$ ; ii) the number of packets being transmitted via random access mode yet the acknowledgement has to wait to come,  $N_a$ ; iii) the number of packets on the process of reservation (i.e., request has been sent for them yet reservation grant has not been received) plus the number of packets for which reservation has been granted,  $N_r$ .

The number of packets which are neither being sent via random access nor on the process of reservation,  $N_w$ , is thus

$$N_w = N_t - N_a - N_r$$

- 1. Upon any new arrival, the user checks the total channel traffic factor  $g = \tilde{g}(t)$  against a prespecified threshold, denoted as  $G_{th}$  (node 1). If  $g \leq G_{th}$ , transmit the packet via ALOHA mode, thus  $N_a$  is incremented by 1 (node 4); if  $g > G_{th}$ , put the new packet in the buffer (add 1 to  $N_w$ ), then go to the next step (node 2).
- (node 2) If there is any packet scheduled to be sent via ALOHA mode (the schedule is due to retransmissions), wait until the transmission of the packet (node 6), then attach a piggybacking request to that packet (node 5). The request consists of two pieces of information: the identity of the user (station) and the number of packets awaiting for reservation, i.e., N<sub>w</sub>.
- 3. If there is no transmission scheduled via ALOHA, check if there is any



Figure 4-1 Data Access Protocol Flowchart

packet being reserved (node 3); if yes, wait (node 14) until the transmissions of these reserved packets and attach the request piggybacked to the reserved packets (node 11).

4. If there is neither transmission scheduled via random access nor granted reservation, i.e., if  $N_a = N_r = 0$ , then schedule a transmission via ALOHA at the time (back to node 4).

The node 4, "ALOHA transmission", means that transmissions follow the ALOHA protocol described in the last chapter. Similarly, node 10, "reservation transmission", means that transmissions execute the reservation protocol described in the last chapter.

5. Whenever a transmission is successful either via ALOHA or reservation (node 7), add the number of packets requesting for reservation,  $N_w$ , to  $N_r$  (node 9). All the  $N_r$  packets wait in global queue until their scheduled transmissions (node 9). If upon successful transmission, no request is attached, i.e.,  $N_w = 0$  (node 8), then the cycle is completed (STOP).

### Voice Access Protocol

- 1. Whenever a new call arrives, send a request via ALOHA mode.
- 2. If the transmission of the request is successful, the voice request is entered in a global voice request queue. The voice global queue is given higher priority to the data reservation queue and is served by assigning an empty

slot and a virtual connection in that slot to the requested call.

If no empty slot is found, the call is rejected.

3. If the request transmission is collided, add 1 to the total number of trials of request transmission  $N_{tl}$ . If  $N_{tl}$  greater than a prespecified threshold  $N_{tr}$ , the call is rejected; otherwise, schedule a retransmission as in the ALOHA access protocol.

## Interactions:

As is seen in the last chapter, ALOHA transmissions can be blocked by voice connections. Similarly, it can be blocked by reservation transmissions while reservation and ALOHA are in the same system. In order to avoid the blocking condition, a maximal number of voice connections plus reservations accommodated in each frame,  $S_{th}$ , is specified, where  $S_{th} < S_f$ . At the beginning of each frame, the following information are available to each user: i) the entries of the voice request queue, and the number in queue,  $N_v'$ ; ii) the entries of the global data reservation queue and the number in queue,  $N_r'$ ; iii) the total number of unoccupied slots,  $S_f - N_v$ , where  $N_v$  is the number of voice occupied slots before updating. According to these information, the following actions will be taken:

1. An empty slot is assigned to a voice request if there is any while keeping the total number of voice connections  $N_v$  (which is incremented by 1 every time a slot is assigned to a new call) in that frame no larger than  $S_{th}$ . Mark the correspondent slot as state 1. If  $N_v = S_{th}$ , the remaining voice requests in the voice global queue are rejected.

2. After all voice requests having been accommodated, transmit the reservations in the global queue while keeping the total number of slots being occupied by voice,  $N_v$ , plus the number of slots just being used by reservation transmissions,  $N_r^{t}$ , no larger than  $S_{th}$ , i.e.,  $N_v + N_r^{t} \leq S_{th}$ . If the sum exceeds  $S_{th}$ , the rest of the reservations are *delayed* to the beginning of the next frame.

In the case where  $N_v = S_{th}$ , all the reservations are delayed to the beginning of the next frame.

3. Finally, for each user attempting ALOHA transmissions, choose randomly and uniformly 1 slot among SSN remaining slots, where  $SSN = S_f - (N_v + N_r^t)$ . Because  $(N_v + N_r^t) \leq S_{th}$ ,  $SSN \geq S_f - S_{th} > 0$ .

It is necessary to mention that the introduction of quantities  $N_v$ ,  $N_r^t$  and  $S_{th}$  does not imply the partitioning of slot regions among voice occupied slots, reservation slots and slots for ALOHA, for any slot can be occupied by voice, reservation or used for ALOHA transmission. However, a record is kept for the number of each type of slots used, and the following relationship among them is maintained,

$$N_{v} + N_{r}^{t} \leq S_{th}, \qquad S_{th} < S_{f}, \qquad SSN = S_{f} - (N_{v} + N_{r}^{t})$$

In the remaining of the thesis, the abbreviation CP will stand for the

"combined protocol".

### 4.3.2 System Parameters for CP

The following list basic system parameters and their effects to the combined protocol(CP).

1. Voice utilization and voice pattern distribution.

Since both the ALOHA protocol and the reservation protocol are insensitive to the voice utilization and voice pattern distribution, the performance of CP is expected to be insensitive to the 2 system parameters.

For discussions following, these 2 system parameters will be chosen to represent typical situations, e.g.,  $\rho_v$  can be either 0.0 representing no voice case or 0.5 representing voice case.

2.  $S_{th}$ . It can influence the performance of CP, especially for high voice utilization and data utilization range. It does not have significant effect for voice utilization and data utilization of low and medium range. An intuitive choice of  $S_{th}$  is made by reserving at least 5% of the total capacity or at least 1 slot to the data transmissions, whichever is smaller, i.e,

$$S_{th} = S_f - \max(1, 0.05(1 - \rho_v)S_f)$$

3.  $G_{th}$ . Since it determines the switching point between ALOHA mode

and reservation mode, it is an important parameter in CP. The range of  $G_{th}$  can be estimated as follows: since g larger than 1 is not desired for ALOHA protocol, it should be between 0.0 and 1.0. The effect of  $G_{th}$  to CP will be shown in the simulations in the next section.

4.  $S_f$  and number of users in the system,  $N_u$ .

For low throughput range, the system operates basically in ALOHA mode, thus  $S_f$  and  $N_u$  will have no significant effect on the performance of CP. For throughput greater than  $e^{-1}$ , there is a value for the number of users,  $N_{um}$ , below which the delay of data packets is minimum. A certain relationship exits between  $N_{um}$  and  $S_f$ , as will be discussed in the next section.

5. The maximal number of trials for voice setup request transmissions,  $N_{tr}$ .

 $N_{tr} \cdot R$  is the maximum waiting time a caller would incur before he knows whether his call will be connected or blocked. it is discussed in the next section.

#### 4.3.3 Asymptotic Analysis of CP

Due to the introduction of mode switching parameter  $G_{th}$ , a detailed and accurate analysis for CP is difficult, especially at the throughput range around  $e^{-1}$ , where the system is frequently switching between the ALOHA mode and reservation mode. For throughput much less than  $e^{-1}$ , however, the system operates basically in ALOHA mode; while for throughput much higher than  $e^{-1}$ , the system operates most of the time in reservation mode. As both ALOHA and reservation modes have been studied in the previous chapter, an asymptotic analysis for CP, which gives a good indication of the performance of CP, can be readily obtained.

**1**)  $g \ll G_{th}$ :

For low throughput range where  $g \leq G_{th}$  is always satisfied, the protocol operates in the loop



in figure 4-1, which is the ALOHA mode; thus for very low throughput range, the performance of CP is very close to that of ALOHA.

# 2) $g \gg G_{th}$ :

For high throughput range where  $g > G_{th}$  is always satisfied, CP can be in



or

ii) the loop

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow 12 \longrightarrow 13$$

which is the random transmission and reservation mode;

or

iii) the entire loop of figure 4-1 except the ALOHA mode for  $g \leq G_{th}$ ,

determined by the total number of users  $N_u$  in the system. There exits a value for the number of users,  $N_{um}$ , below which the system will operate (most of the time) in loop i; As  $N_u$  increases, the probability that the system enters loop ii increases, for throughput greater than  $e^{-1}$ . An extreme case is that the number of users in the system is infinite. For finite aggregative total channel throughput, the throughput for any individual user (assuming load is offered evenly from all the users) is infinitesimally small. That in turn means that no user can have more than one packet at any time, thus no packet can be transmitted via reservation. As all packets are transmitted via ALOHA and total channel load is greater than  $e^{-1}$ , the chance of successful transmissions is approaching zero and delay approaching infinite, this is the second loop indicated above (for infinite population models). In the throughput range greater than  $e^{-1}$ , it is desired that most of the packets should be transmitted via reservation mode; i.e, the first loop indicated above.

The following derives the asymptotic value of  $N_{um}$  under symmetric load assumption,\* i.e., the aggregative load to the system, denoted as  $\tilde{D}$ , is generated *evenly* from among all users. If the individual data arrival rate is  $\tilde{d}$ , the total number of the users in the system is  $N_u$ , then

$$ilde{D} = ilde{d} \cdot N_{u}$$

In order to *keep* the system in the first loop, there should be at least one packet arriving during the period in which the previous packets are in the process of reservations (figure 4-2), i.e.,

$$\frac{\tilde{D}}{N_u} = \tilde{d} \ge 1, \quad \frac{\tilde{D}}{R \cdot S_f \cdot (1 - \rho_v)} \le 1$$
(4-1)

where the unit of R is  $S_f$ . The left hand side equation in (4-1) represents the average number of arrivals during time R for each user evenly;  $\tilde{d} \ge 1$  represents the continuous reservation condition. the right hand side equation assumes that

<sup>\*</sup> This assumption is realistic for a large number of small users.

the normalized maximum data throughput is 1.0. Re-arrange (4-1),

$$rac{R \cdot S_f \cdot (1-
ho_v)}{N_u} \geq rac{ ilde{D}}{N_u} \geq 1$$

or equivalently,

$$N_{um} \le R \cdot S_f \cdot (1 - \rho_v) \tag{4-2}$$

(4-2) gives an upper estimate of  $N_{um}$ .



Figure 4-2 Illustration of Continuous Reservation

A lower estimate of  $N_{um}$  is obtained as follows. If the normalized maximum data throughput is no greater than  $e^{-1}$ , then the system will operate in the third loop and will always (even if the continuous reservation condition, i.e.,  $\tilde{d} \ge 1$ , is not satisfied) be able to enter the first loop, thus

$$\frac{\tilde{D}}{N_u} = \tilde{d} \le 1, \quad \frac{\tilde{D}}{R \cdot S_f \cdot (1 - \rho_v)} \ge e^{-1}$$
(4-3)

the right hand side equation reverses the condition that the data throughput is no greater than  $e^{-1}$ , in order to obtain  $N_{um}$ . Simplifying (4-3) gives,

$$N_{um} \ge e^{-1} \cdot R \cdot S_f \cdot (1 - \rho_v) \tag{4-4}$$

Combining (4-2) and (4-4) gives,

$$e^{-1} \cdot R \cdot S_f \cdot (1 - \rho_v) \le N_{um} \le R \cdot S_f \cdot (1 - \rho_v)$$

$$(4-5)$$

Average delay for system operating in the first loop is that of reservation protocol discussed in the previous chapter, plus R/2, as illustrated in figure 4-2.

The effect of  $S_f$  on CP is also seen from (4-2) and (4-4).  $S_f$  has no significant effect on low throughput range where ALOHA mode dominates. It influences the system on high throughput range as (4-2) and (4-4) indicate, i.e.,  $S_f$  is proportional to  $N_{um}$ . Moreover, (4-2) and (4-4) suggest that  $N_{um}$  is proportional to  $(1 - \rho_v)$ .

#### 3) Trial numbers for voice request transmissions:

If trials are assumed to be independent with probability of success  $p_s$ , then the distribution of trial numbers  $l_T$  is given by

$$P_r[l_T = l] = p_s(1 - p_s)^{l-1}$$
(4-6)

As voice arrival rate is much smaller than data arrival rate, the contribution of voice request transmissions to the total traffic of random access is actually negligible. Moreover, since voice request transmissions follow the same ALOHA protocol as data packets do, probability  $p_s$  is the same as that for data packets transmitted via ALOHA. For  $g < e^{-1}$ , it is given in section 3.5. For  $g > e^{-1}$ , an accurate calculation for  $p_s$  is difficult; as an estimation, it can be noticed that if  $N_u \leq N_{um}$ , the average delay of data packets is about 2.5 times of propagation delay R. Thus,  $p_s$  can be estimated by

$$p_s = 1/2.5 = 0.4$$

For such  $p_s$ , calculation of (4-6) suggests that no more than 2% of trials exceed 10 times, and virtually no trial numbers is larger than 20 ( $20 \times R = 5.4$  seconds), which should be acceptable for most applications.

#### 4.4 Simulations for CP

Simulations were carried out to obtain the delay-throughput performance curves with different values of  $G_{th}$  as parameters.<sup>\*</sup> The relationship among  $\rho_v$ ,  $S_f$  and  $N_{um}$  as indicated in (4-5) was verified. The simulated trial number distribution  $P_r[l_T]$  was also obtained.

#### 1) Effect of $G_{th}$

Figure 4-3(a) shows delay versus throughput curves for  $\rho_v = 0.5$ ,  $S_f = 100$ ,  $N_u = 100$ ,  $G_{th} = \{0.0, 0.15, 0.35, 0.75, 1.0\}$ . As can be seen,  $G_{th}$  determines the point at which the system switches from ALOHA mode to reservation mode.

<sup>\*</sup> The estimation of  $\tilde{g}$  is assumed to be perfect.

For  $G_{th} \geq 0.35$ , change on  $G_{th}$  actually has no effect on low  $\rho_d$  range  $(0.0 \leq \rho_d \leq 0.32)$ , i.e., ALOHA throughput range). On the other hand, the delay in the  $0.32 \leq \rho_d < 1.0$  range increases with the increase of  $G_{th}$ .

For  $0.0 \leq G_{th} \leq 0.35$ , the major change of delay is in low  $\rho_d$  range. In  $G_{th} = 0.0$  case, the curve departs from that of ALOHA right at  $\rho_d = 0.0$ ; in  $G_{th} = 0.15$  case, departure occurs at  $\rho_d \approx 0.15$ . These curves have larger delay than that of ALOHA. However, for  $G_{th} = 0.35$ , the curve departs from that of ALOHA at  $\rho_d \approx 0.32$  and the delay is smaller than that of ALOHA, indicating that for  $\rho_d \geq 0.32$ , reservation mode should be *partially* used instead of pure ALOHA mode. The desirable range of  $G_{th}$  can also be seen from the figure, i.e., the optimal value of  $G_{th}$  is about 0.35.

Figure 4-3(b) shows delay versus throughput curves for  $\rho_v = 0.0$ ,  $S_f = 100$ ,  $N_u = 200$ ,  $G_{th} = \{0.0, 0.3, 0.7, 1.5\}$ . Similar observations can be made on  $G_{th}$ , except that the desirable value of  $G_{th}$  has been doubled to around 0.7. Also noticed is the similarity between figures 4-3(a) and 4-3(b): curves of  $\rho_v = 0.5$  (figure 4-3(a)) correspond to curves of  $\rho_v = 0.0$  (figure 4-3(b)) with  $G_{th}$  doubled. This similarity can be explained by the fact that as  $\rho_v$  changes from 0.5 to 0.0, the total channel capacity available to data users doubles. If the total channel traffic factor g were normalized to  $(1 - \rho_v)$ ,  $G_{th}$  would have the same values for  $\rho_v = 0.5$  and  $\rho_v = 0.0$ . The similarity between performance with and without voice presence indicates that CP is not susceptible to voice background. The system parameters  $S_f$ ,  $N_u$  and  $\rho_v$  for figures 4-3(a) and 4-3(b) are chosen such that the ratio

$$\frac{N_u}{S_f(1-\rho_v)}$$

is identical. The relationship among  $S_f$ ,  $N_u$  and  $\rho_v$  is investigated further in the

following section.

# 2) The relationship among $S_f$ , $N_u$ and $\rho_v$

The relationship among  $S_f$ ,  $N_u$  and  $\rho_v$  is given in (4-5). In this section, simulation results are shown to verify (4-5).

Figure 4-4(a) shows delay versus throughput for  $\rho_v = 0.5$ ,  $S_f = 100$ ,  $G_{th} = 0.25$ ,  $N_u = \{100, 200, 300, 400, 500\}$ . It is seen that low throughput range of  $\rho_d$  ( $0.0 \le \rho_d \le 0.36$ ) is little affected by  $N_u$ , since ALOHA performance is not affected by  $N_u$ . For  $\rho_d \ge 1/e$ , delay increases with the increase of  $N_u$ . However, there are significant differences among curves for  $N_u \le 200$  and for  $N_u > 200$ , this indicates that  $N_{um}$  of the system is between 200 and 300, a calculation of  $N_{um}$  according to (4-5) gives

$$220 \le N_{um} \le 600$$

thus the simulation results lie between the upper estimate and the lower estimate of  $N_{um}$  by (4-5).

For  $N_u$  below  $N_{um}$ , the change of  $N_u$  does not change the delay versus throughput performance significantly, e.g., curve for  $N_u = 100$  is very close to that for  $N_u = 200$ .

Similar results can be seen for  $\rho_v = 0.0$  case from the delay versus throughput curves shown in figure 4-4(b). The simulation shows that  $N_{um}$  is greater



Figure 4-3 CP Delay-Throughput Characteristics



Figure 4-4 CP Performance and User Number  $N_u$ 

than 400 and (much) below 800. According to (4-5),

$$440 \leq N_{um} \leq 1200$$

Also noticed is the similarity between the curve for  $N_u = 400(800)$ ,  $\rho_v = 0.0$ , shown in figure 4-4(b), and the curve for  $N_u = 200(400)$ ,  $\rho_v = 0.5$ , shown in figure 4-4(a). The ratio  $N_u/(1-\rho_v)$  is the same for both of them.

For  $N_u$  greater than  $N_{um}$ , delay increases rapidly with the increase of  $\rho_d$ . For  $\rho_d = 0.75$ , figure 4-5(a) and figure 4-5(b) normalized delay versus  $N_u$  for  $\rho_v = 0.0$  and 0.5, respectively. Effect of  $N_u$  can be seen more clearly from these figures. Calculated  $N_{um}$  according to (4-5) for each curve is also shown as a comparison.

Again the similarity for curves with the same  $N_u/(1 - \rho_v)$  ratio, i.e, the curves  $S_f = (100)200$ ,  $\rho_v = 0.5$  (in figure 4-5(a)) and  $S_f = 200(400)$ ,  $\rho_v = 0.0$  (in figure 4-5(b)) can be seen.

#### 3) Trial number distribution

Figure 4-6(a) shows trial number distributions (simulated) for  $\rho_v = 0.0$ ,  $\rho_d = \{0.1, 0.3, 0.5, 0.75\}$ . As  $\rho_d$  increases, the trial number increases. The maximal number of trials shown for  $\rho_d = 0.75$  is 6 with frequency  $10^{-5}$ .

Figure 4-6(b) shows trial number distribution for  $\rho_v = 0.5$ , The maximal number of trials shown for  $\rho_d = 0.75$  is 8 with frequency of  $10^{-5}$ . These results



Figure 4-5 Relationship among  $N_{um}$ ,  $S_f$  and  $N_u$ 



Figure 4-6 Trial Number Distribution

suggest that for practical time requirement of voice connections, random access protocol for setup request transmission is satisfactory.

## 4.5 Comparison of CP with ALOHA and Reservation

1. For low throughput range  $0 \le \rho_d \le e^{-1}$ , with proper choice of  $G_{th}$ , CP gives the same delay versus throughput performance as ALOHA.

2. In throughput range  $\rho_d \geq e^{-1}$ , for proper choice of  $G_{th}$  and  $N_u$  not exceeding  $N_{um}$ , delay of CP is 0.5 imes R larger than that of the reservation protocol in which a freely available subchannel for request transmission is assumed. However, actual implementable reservation protocols have larger delay and smaller utilization than such a hypothetical reservation protocol. The maximal number of users supported by an implementable reservation protocol is also limited. The comparisons between reservation protocols using subchannels for request transmission and CP can be made in terms of overhead delay, maximal number of users supported by the system (for delay not exceeding certain specified value), and complexity of implementation. As an example, if a minislotted reservation protocol is to operate under SENET structure for  $\rho_v = 0.5$ ,  $S_f = 100$ ,  $N_u = 200$ and maximum  $\rho_d$  no less than 0.85, at most 10% of average remaining capacity can be used for subchannel request transmissions, i.e.,  $0.1 \times 0.5 \times 100 = 5$  slots, or 1.1 ms in real time units; if minislots are to be assigned to each user dedicatedly, 200 minislots have to be used, the resulting length of each minislot will be less than  $1.1 \text{ms}/200 = 5.5 \mu \text{s}$ , including guard times. On the other hand, if minislots are assigned on a contention basis, the average delay for request transmissions may as well be larger than 1.5 times of R due to collisions, especially in high  $\rho_d$  range. Thus the delay and the throughput of CP is comparable to those of minislotted reservation protocols.

The advantage of CP in terms of complexity of implementation over subchannel oriented reservation protocols is obvious, since no additional partitions of channel time is necessary in CP.

### Chapter 5

# **Conclusions and Further Studies**

The major conclusions are summarized as follows:

- 1. The state correlation function, defined in (2-17), is heavily influenced by voice assignment protocols. Two voice assignment protocols, i.e., random assignment(VAP-I) and left to right assignment(VAP-II) are studied. The independence assumption(p.26) produces good approximations for performance analysis of ALOHA and reservation protocols studied in chapter 3, under the voice background generated by VAP-I.
- 2. The performance of both ALOHA and reservation protocols deteriorates in the IVD environment, due to the channel capacity fluctuations caused by voice connections. However, the deterioration of the former is much more severe than the latter. Under IVD environment, it is indicated, by simulation results in section 3.4 and analysis in section 3.5, that an adaptive control mechanism is indispensable for ALOHA type protocols. If they are to be used in an such environment, more coordinations are needed in the protocol definitions.
- 3. Based on the above observation, ALOHA-II is proposed and studied. Under the assumption that estimation of backlog number is perfect, the protocol sig-

nificantly reduces the effect of voice background fluctuation; as a consequence, the protocol stability is guaranteed for throughput under  $e^{-1}$ . Furthermore, if the estimation error is given, the upper boundary of the stable region can be estimated by (3-59). An estimation algorithm, which is aimed at reducing the effect of propagation delay R, is proposed in section 3.5.

4. A natural extension of the results from chapter 3 is the combined protocol in chapter 4. Asymptotic analysis gives a good estimation of the user number, under which minimum delay can be achieved when throughput is larger than  $e^{-1}$ .

The future work can be conducted in the following areas:

- 1. The performance evaluation of ALOHA-I protocol under VAP-II type voice background needs further studies.
- 2. The properties of the algorithm for estimating the number of backlog users, proposed in section 3.5, are worth further studies.
- 3. For a complete performance evaluation of ALOHA-II and the combined protocol, the 2-dimensional probability function in (3-50) needs to be derived.
- 4. Voice packet multiplexing and/or data packet transmission, by taking advantage of voice silence period, can provide potentially more than 30% increase in channel utilization, due to the fact that up to 50% of a voice connection is silence period. Furthermore, voice multiplexing by taking advantage of silence period reduces frame-to-frame correlations, since the durations of talkspurts

and silence periods are much shorter than that of voice connections. It is known that [4,18,19] data performance improves as the voice frame-to-frame correlation weakens. In satellite systems, a hybrid voice multiplexing and random access protocol, which is particularly suitable for small users, has been reported in [23]. This is an important topic of further studies.

5. Onboard processing and multibeam satellites are the future trends of satellite systems [24-26]. They are essential for supporting users with small earth station antennas, such as mobile users. IVD protocols under such environments require further studies.
In this appendix, the 1-dimensional and 2-dimensional distributions given in (3-35) and (3-9) are derived.<sup>\*</sup> Since the transmission of n packets in M slots can be equivalent to placing n balls randomly in M boxes, such model and terms are used. The expectation of 1-dimensional distribution given in (3-37) is also derived.

#### 1. The basic formula

Consider the model of placing n ordered balls randomly in M boxes. The total number of ways to do so is given by  $M^n$ . Let  $n_1$  denote the number of all the patterns<sup>†</sup> with no 1-ball boxes.<sup>‡</sup> Let  $N_k$  denotes the number of the patterns that have at least k 1-ball boxes, for example,  $N_1$  is the number of patterns that has at least one 1-ball box. Then  $n_1$  is given by

$$n_1 = M^n - N_1$$

<sup>\*</sup> The 1-dimensional distribution is also given in [15]. A more general approach is used in this appendix to derive 2-dimensional (and in principle multi-dimensional) distributions as well as 1-dimensional distributions.

<sup>&</sup>lt;sup> $\dagger$ </sup> A pattern has one to one correspondence to a way of placing *n* balls in *M* boxes.

 $<sup>\</sup>ddagger$  A *j*-ball box is a box containing *j* balls.

 $N_1$  is unknown, but it is easy to see that one way of getting at least one 1-ball box is to choose one ball from the *n* ordered balls randomly and place it in the *M* boxes randomly; there are \*

$$\binom{M}{1}A_n^1$$

ways to do so. In each of them, the remaining m-1 boxes and n-1 balls have  $(M-1)^{n-1}$  patterns. But the product

$$\binom{M}{1}A_n{}^1(M-1)^{n-1} \ge N_1$$

since in  $(M-1)^{n-1}$ , the ways of getting at least two 1-ball boxes are included. Thus

$$N_1 = \binom{M}{1} A_n^{-1} (M-1)^{n-1} - N_2$$

But  $N_2$  is also unknown, the number

$$\binom{M}{2}{A_n}^2(M-2)^{n-2} \ge N_2$$

and should be subtracted by  $N_3, \ldots$ , etc, thus

$$egin{aligned} n_1 &= M^n - \left\{ inom{M}{1} A_n{}^1 (M-1)^{n-1} - \left\{ inom{M}{2} A_n{}^2 (M-2)^{n-2} \ & - \left\{ inom{M}{3} A_n{}^3 (M-3)^{n-3} - & \cdots & \pm N_L 
ight\} 
ight\} \end{aligned}$$

the last number  $N_L$  is

$$N_L = \binom{M}{L} A_n{}^L (M - L)^{n-L}$$

\* The function  $A_n^i$  is defined as

$$A_n^i \stackrel{\Delta}{=} \binom{n}{i} i!$$

where  $L = \min(M, n)$ .  $N_L$  does not contain  $N_{L+1}$ , since  $N_j = 0, \forall i > L$ .

Thus

$$n_{1} = M^{n} - \binom{M}{1} A_{n}^{1} (M-1)^{n-1} + \binom{M}{2} A_{n}^{2} (M-2)^{n-2} - \binom{M}{3} A_{n}^{3} (M-3)^{n-3} + \dots \pm \binom{M}{L} A_{n}^{L} (M-L)^{n-L}$$
(A1)
$$= \sum_{k=0}^{L} (-1)^{k} \binom{M}{k} A_{n}^{k} (M-k)^{n-k}$$

(A1) can also be written as

$$n_1 = \sum_{k=0}^{L} (-1)^k G(M, n, k) B(M, n, k)$$
(A2)

where

$$G(M, n, k) = \binom{M}{k} A_n^k$$
$$B(M, n, k) = (M - k)^{n-k}$$

Name G(M, n, k) the weighting function and B(M, n, k) the sample function. The above results does not depend on particular forms of G(M, n, k) and B(M, n, k), thus (A2) is applicable to many similar problems. For example, if the order of the *n* balls does not matter, then

$$G(M, n, k) = \binom{M}{k}$$
$$B(M, n, k) = \binom{M + n - k}{n - k}$$

and corresponding  $n_1$  is \*

$$n_1 = \sum_{k=0}^{L} (-1)^k \binom{M}{k} \binom{M+n-k}{n-k}$$
(A3)

In order to obtain  $n_j$ , G(M, n, k) and B(M, n, k) are modified as

$$egin{aligned} G(M,n,k) = & igg( M \ k igg) rac{A_n^{kj}}{\left( j! 
ight)^k} \ B(M,n,k) = & (M-k)^{n-kj} \end{aligned}$$

and

$$n_{j} = \sum_{k=0}^{L} (-1)^{k} {\binom{M}{k}} \frac{A_{n}^{kj}}{(j!)^{k}} (M-k)^{n-kj}$$
(A4)

where  $L = \min(M, n/j)$ . An important special case is j = 0,

$$n_0 = \sum_{k=0}^{M} (-1)^k \binom{M}{k} (M-k)^n$$
 (A5)

### 2. The one-dimensional distribution

Choose i balls and place in i boxes, there are

$$\binom{M}{i}A_n{}^i$$

different ways of doing so. In each case, in the remaining M - i boxes and n - i balls, the total number of patterns that do not contain 1-ball box,  $n_1$ , is given in (A1) with M being substituted by M - i and n by n - i; thus the total number

<sup>\*</sup> This expression is not used in the thesis.

of patterns containing exactly i 1-ball boxes, denoted by  $c_1(i)$ , is

$$c_1(i) = \binom{M}{i} A_n^{i} \sum_{k=0}^{L} (-1)^k \binom{m}{k} A_l^{k} (m-k)^{l-k}$$
(A6)

where m = M - i, l = n - i,  $L = \min(m, l)$ .

In general,  $c_j(i)$ , the number of exactly i *j*-ball boxes, is given by

$$c_{j}(i) = \binom{M}{i} \frac{A_{n}^{ij}}{(j!)^{i}} \sum_{k=0}^{L} (-1)^{k} \binom{m}{k} \frac{A_{l}^{kj}}{(j!)^{k}} (m-k)^{l-kj}$$
(A7)

where m = M - i, l = n - ij,  $L = \min(m, l/j)$ .

If all the possible patterns are equally likely, the probability of exactly i j-ball boxes,  $p_1(i)$ , is given by (A6) divided by  $M^n$ , i.e.,

$$p_1(i) = \frac{c_1(i)}{M^n} \tag{A8}$$

(3-35) is obtained after algebraic simplification of (A8).

### 3. The two-dimensional distribution

Denote  $c(i_0, i_1)$  the number of patterns each has  $i_0$  empty boxes and  $i_1$  1-ball boxes. Then

$$c(i_0, i_1) = c(i_0)c(i_1|i_0)$$

where  $c(i_0) = {M \choose i_0}$ . To find  $c(i_1|i_0)$ , it is sufficient to apply (A2) with M substi-

tuted by  $M - i_0$ , and

$$G(M - i_0, n, k) = \binom{M - i_0}{k} A_n^k$$
  
 $B(M - i_0, n, k) = c_0(0) = n_0$ 

where  $n_0$  is given in (A5) with M substituted by  $M-i_0$ , which means the sample function in the remaining  $M-i_0$  boxes should not contain patterns with empty boxes. It follows that

$$c(i_{0}, i_{1}) = c(i_{0})c(i_{1}|i_{0})$$

$$= \binom{M}{i_{0}}\binom{M-i_{0}}{i_{1}}A_{n}^{i_{1}} \times$$

$$\sum_{k=0}^{L} (-1)^{k}\binom{m}{k}A_{l}^{k}\sum_{j=0}^{m-k} (-1)^{j}\binom{m-k}{j}(m-k-j)^{l-k}$$
(A9)

where  $m = M - i_0 - i_1$ ,  $l = n - i_1$ ,  $L = \min(m, l)$ . (3-9) is obtained by combining (A9) with Poisson distribution and simplification.

## 4. The expectation of (A8)

Let the expectation be

$$E(i) = \sum_{i=0}^{\min(M,n)} ip(i)$$

By (A6) and (A8),

$$E(i) = \frac{M!n!}{M^n} \sum_{i+j=0}^{M} (-1)^j \frac{i[M-(i+j)]^{n-(i+j)}}{i!j![M-(i+j)]![n-(i+j)]!}$$

Let k = i + j, then

$$E(i) = \frac{M!n!}{M^n} \sum_{k=0}^{M} \left\{ \sum_{j=0}^k \frac{k!(-1)^j(k-j)}{j!(k-j)!} \right\} \frac{(M-k)^{n-k}}{k!(M-k)!(n-k)!}$$
(A10)

It is easy to show that

$$\sum_{j=0}^{k} \frac{k!(-1)^{j}(k-j)}{j!(k-j)!} = k\delta(k-1)$$
(A11)

Substituting (A11) into (A10), one obtains (3-37).

# References

- T. Bially, A. McLaughlin, and C. Weinstein, "Voice communication in integrated digital voice and data networks", *IEEE Trans. Commun.*, Vol.COM-28, pp. 1478-1490, Sept. 1980.
- [2] I. Gitman and H. Frank, "Economic analysis of integrated voice and data networks: A case study", Proc. IEEE, Vol. 66, pp.1549-1570, Nov. 1978.
- [3] D. Johnson and C. Weinstein, "A user interface using recognition of LPC speech transmitted over the ARPANET", EASCON Rec., Sept. 1977.
- [4] C. Weinstein and J. Forgie, "Experience with Speech Communication in Packet Networks", *IEEE Journal on selected areas in commun.*, Vol.SAC-1, No.6, pp.963-980, Dec. 1983.
- [5] G. Coviello, O. Lake, and G. Redinbo, "System Design Implications of Packetized Voice", Proc. Int. Conf. Commun., Chicago, IL., pp.38.3-4 to 38.3-53, June 1977.
- [6] G. Falk, J. Groff, W. Milliken, M. Nodine, S. Blumenthal, W. Edmond, "Integration of Voice and Data in the Wideband Packet Satellite Network", *IEEE*

Journal on Selected Areas in Commun., Vol.SAC-1, No.6, pp.1076-1083, Dec. 1983.

- [7] G. Coviello and P. Vena, "Integration of Circuit/Packet Switching by SENET (Slotted Envelope Network) Concept", Proc. Nat. Telecommun. Conf., pp.42.12-42.17, Dec. 1975.
- [8] C. Weinstein and H. Heggestad, "Multiplexing of Packet Speech on an Experimental Wideband Satellite Network", Proc. AIAA 9th Commun. Satellite Syst. Conf., San Diego, CA., March 1982.
- [9] H. Heggestad and C. Weinstein, "Experiments in Voice and Data Communications on a Wideband Satellite/Terrestrial Internetwork System", Conf. Rec. Int. Conf. Commun., Boston, MA., June 1983.
- [10] J. Forgie, "Network Speech Implications of Packetized Speech", Annu. Rep. Defense Commun. Agency, MIT Lincoln Lab., Lexington, MA., DTIC AD-A45455, Sept. 30, 1976.
- [11] A. Rustako, G. Vannucci, C. Woodworth, "An Experimental Scanning Spot Beam Satellite System Implementing 600 Mbps TDMA", Bell Lab. report., 1982.
- [12] R. Syski, Introduction to Congestion Theory in Telephone Systems, North Holland, Amsterdam, 1986.

#### References

- [13] F. Tobagi, "Multiaccess Protocols in Packet Communication Systems", IEEE Trans. Commun., Vol.COM-28, No.4, April 1980.
- [14] J. Kurose, M. Schwartz, and Y. Yemini, "Multiaccess Protocols and Time-Constrained Communication", *Computing Surveys*, Vol.16, No.1, pp.43-70, March 1984.
- [15] W. Feller, An Introduction to Probability Theory and Its Applications, pp.112, Vol.I, John Wiley & Sons, Inc., 1968.
- [16] N. Abramson, "The ALOHA System Another Alternative for Computer Communications", 1970 Fall Joint Comput. Conf., AFIPS Conf. Proc., Vol.37, Montvale, NJ, AFIPS Press, 1970, pp.281-285.
- [17] L. Kleinrock and S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation", *IEEE Trans. Commun.*, Vol.COM-23, No.4, pp.410-423, April 1975.
- [18] K. Sriram, P. Varshney, J. Shanthikumar, "Discrete-Time Analysis of Integrated Voice/Data Multiplexers With and Without Speech Activity Detectors", *IEEE Journal on Selected Areas in Commun.*, Vol.SAC-1, No.6, pp.1124-1132, Dec. 1983.
- [19] S. Li and J. Mark, "Performance of Voice/Data Integration on a TDM System", *IEEE Trans. on Commun.*, Vol.COM-33, No.12, Dec. 1985.

### References

- [20] A. Leon-Garcia, R. Kwong and G. Williams, "Performance Evaluation Methods for an Integrated Voice/Data Link", *IEEE Trans. on Commun.*, Vol. COM-30, No.8, pp.1848-1857, August 1982.
- [21] A. Fukuda and S. Tasaka, "The Equilibrium Point Analysis A Unified Analytic Tool for Packet Broadcast Networks", *IEEE Proc.*, pp.33.4.1-33.4.8, 1983.
- [22] S. Tasaka, Performance Analysis of Multiple Access Protocols, The MIT Press, Cambridge, MA., 1986.
- [23] H. Lee and J. Xu, "A Hybrid Scheme of TASI and Random Access for Integrated Voice and Data Transmission over a Satellite Channel", *Pacific Rim* Conference Record, June 1987.
- [24] H. Lee, Hybrid Random/Reservation Access Protocols for Communication Satellites, CCNG Report No.E-117, Dept. of EE, Univ. of Waterloo, August 1983.
- [25] P. Jain, "Onboard Processing in Future Satellite Communication Systems", Journal of Telecommunication Networks, pp.139-151, Computer Science Press, Inc., 1983.
- [26] A. Frank and T. Stern, "Onboard Demand Scheduling of a SS/TDMA Multibeam Satellite with Integrated Circuit and Packet Switching", Proc. 6th Int. Conf. Digital Satellite Commun., pp.x-27 to x-34, 1983.