SPACE DIVISION MULTIPLE ACCESS PROTOCOLS FOR A SATELLITE SWITCHED COMMUNICATION NETWORK

A thesis submitted in fulfilment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

in the

Department of Electrical and Electronic Engineering,
The University of Adelaide

by

Ko King-Tim, B.E.(Hons)

December 1982

謹以此論文獻給
先嚴與家慈
子高敬添敬獻

TO MY MOTHER, AND

IN MEMORY OF MY FATHER

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

This study indicates that the use of multiple spotbeam antennas in a satellite switched communication network is a viable method of alleviating the problems of overcrowded frequency spectrum and orbital space. The spotbeam coverage allows high energy concentration, frequency re-use and jam proof operation, and the satellite switching provides the appropriate beam connections. These techniques are particularly appropriate for future domestic satellite communication, especially for low traffic density, mobile, maritime or direct to user applications.

Three possible data multiple access protocols suitable for a space division and satellite switched communication network, known collectively as the Space Division Multiple Access (SDMA) protocols, have been investigated. They are Fixed Assignment (FA) Protocol, the hybrid Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol and the fully Demand Assignment (DA) Protocol. Taking the network characteristics into consideration, DA-FA Protocol is most advantageous when operating with distributed reservation control, while DA Protocol is best with central reservation control with the controller located on the satellite. The reservation channel is accessed with a Time Division Multiple Access Protocol. Their extensions to integrated data and voice traffic are also indicated, where the data traffic utilises message switching and the voice traffic circuit switching.

The problems of optimal location of stations and spotbeam coverages are modelled by facility location and set covering solutions respectively. A detailed delay analysis is presented for an arbitrary compound arrival process. Two new analytical solutions are developed for calculation of the scheduling delay and the reservation delay with various reservation burst sizes. In order to avoid the calculation of a large number of complex roots from a transcendental equation, an Integration Method which is essentially independent to the capacity, is derived for the scheduling delay calculation. A solution has also been developed for the calculation of reservation delay under an environment of synchronous batch output. The two results have been verified by simulations. The multilplexing and demultiplexing buffer requirements for a system with and without acknowledgement have been solved using a constant and nonconstant output model, and their performances are well correlated with the simulation results. Comparisons of the delay and the buffer performances are made between the three SDMA protocols and other global beam data packet multiple access protocols. The results indicate that the DA Protocol has the best delay and buffer performances amongst the three SDMA protocols, and the delay performance approaches that of the global beam multiple access protocols in the region of high throughput. Relationships and possible tradeoffs between the performances, the terrestrial network and protocol parameters are also discussed in detail.

Statement

I declare that this thesis contains no material which has been accepted for the award of any other degree at this or any other University, and that, to the best of my knowledge, it contains no material previously published or written by any other person except where otherwise acknowledged in the text.

Ko King-Tim

Acknowledgements

I gratefully acknowledge the guidance and support of my supervisor, Dr. B. R. Davis, over the period that the work for this thesis was undertaken, and in particular I thank him for his assistance in the analysis of Chapter 4. In addition I would like to express my gratitude to Mr. G. Beilby for his enthusiastic support in programming, Dr. B. R. Davis and Dr. D.A.C. Evans for proof reading this thesis.

The assistance of Telecom Australia in the preparation of this thesis is acknowledged. The accurate typing of Mrs. Denise George is deeply appreciated. Finally I wish to thank my colleagues and the staff of the Department of Electrical and Electronic Engineering of The University of Adelaide, who helped to make my period in the department an enjoyable one.

Publications

Ko K.T., "Space division multiple access protocol for a satellite communication network", Proc. IREECON International, 1981.

Ko K.T. and Davis B.R., "Design of a demand assignment satellite switched space division multiple access communication network", Journal of Australia Telecommunication Research, Vol.16, No.2, pp.25-38, Nov. 82, (Appendix 3).

Ko K.T. and Davis B.R., "A space division multiple access protocol for spotbeam antenna and satellite switched communication network", to be published in IEEE Selected Areas in Commun., Jan. 83.

List of Symbols

$\tilde{A}_n,\tilde{A}(z),$ $A[M(z)]$	$\tilde{A}_n$ is the total number of network slot arrivals in the n-th frame, with z-transform of its probability distribution $\tilde{A}(z)$ or $A[M(z)]$.
$A_n,A(z)$	A_n is the total number of network message arrivals in the n-th frame, with z-transform $A(z)$.
$A_{ni},A_i(z)$	A_{ni} is the total number of capacity requests of the n-th frame into station i, with z-transform $A_i(z)$.
$B(x)$	probability of x slots in the buffer at the beginning of the frame.
$B_i(x)$	probability of $B(x)$ given i slots available.
$B_n,B(z)$	B_n is the queue size after the arrivals of the n-th reservation group, with z-transform $B(z)$.
B_d	total data blocking probability.
B_{dd}	data demultiplexing buffer blocking probability.
B_{dm}	data multiplexing buffer blocking probability.
B_{dr}	data request buffer blocking probability.
$B_{d\ ret}$	data retransmission buffer blocking probability.
B_v	total voice blocking probability.
B_{vi}	voice multiplexing buffer blocking probability.
B_{vj}	voice demultiplexing buffer blocking probability.
B_{vr}	voice channel request buffer blocking probability.
D_d	data message delay (s).
D_R	reservation delay (s).
D_S	data scheduling delay (s).
D_v	voice connection delay (s).
D_{vS}	voice scheduling delay (s).
DA	Demand Assignment uplink and downlink.
DA-FA	Demand Assignment uplink and Fixed Assignment downlink.
E_j	location of the j-th earth station.
$E_q(A)$	Erlang Loss Formula.
F	channel capacity (slots/s).
F_d,F_d'	distance and permitted distance function of the spotbeam coverage repectively.

FA	Fixed Assignment uplink and downlink.
FCFS	First Come First Served.
G_q	location of the q-th spotbeam coverage.
K	number of spotbeam coverage in the network.
L	number of voice circuit request buffer on-board the traffic controller.
$\tilde{L}_n$,$\tilde{L}(z)$	$\tilde{L}_n$ is the waiting time of the n-th message leader (slots), with z-transform $\tilde{L}(z)$.
N	number of data slots per frame (slots).
N_{ij}	number of data slots allocating for traffic originating from station i and destinated for zone j.
N_{kj}	number of data slots allocating for traffic originating from zone k and destinated for zone j.
Q	number of data request buffer.
S	channel utilisation.
SDMA	Space Division Multiple Access.
T_i	location of the i-th terminal.
V_i	number of duplex voice circuits in station i.
$\tilde{W}_n$,$\tilde{W}(z)$	$\tilde{W}_n$ is the waiting time of the message beyond its group leader (slots), with z-transform $\tilde{W}(z)$.
Y	number of message requests leftover from previous frames.
$\tilde{Z}_i$	amount of traffic destinated for station i (slots).
$\tilde{X}_{i,d}$	amount of traffic received by station i given d slots of demultiplexing buffer (slots).
$\tilde{a}(k)$,$\tilde{a}_I(k)$ $a_{II}(k)$	probability of k slot arrivals in a frame into a station, Group I and Group II respectively.
c	number of different terminal types.
c_d	demultiplexing channel rate (slots/s)
c_{ij}	cost of connecting T_i to E_j.
d	number of data demultiplexing buffer in Group I.
d_j	cost of establishing a station at site j.
f	framelength (s).
$f_d(i)$	destination function for data traffic destinated for station i.
f_v	voice framelength (s).

g	number of voice circuits in a frame.
l	number of data retransmission buffers.
$m(k), M(z)$ $\Phi_M(s)$	probability of the message length equal to k slots, with z-transform $M(z)$ and characteristic function $\Phi_M(s)$.
q	number of data multiplexing buffer in Group I.
rtpd	round trip propagation delay.
s	number of stations in the network.
t	number of terminals in the network.
u	number of control information carried by each control burst.
v, v_i	number of voice terminals in the network or connected to station i respectively.
y	number of message requests arrived in the current frame prior to the arrival under consideration.
z_r	roots of the transcendental equation.
α	expected output from Group I (slots).
$\alpha(i)$	expected output from Group I given i available slots (slots).
β	offered traffic of Group I (slots).
δ	information bandwidth assignment.
Δ, Δ_i	ratio of correctly transmitted message w.r.t. the total transmitted message into the controller and station i respectively.
ε_i	ratio of terminals connected to station i w.r.t. the total number of terminals in the network.
$1/\theta$	average message length (slots/message), and $\chi = 1-\theta$.
λ, λ_i	average message arrivals from the network and station i respectively.
$1/\mu$	average call holding time (s).
ρ	offered traffic per voice circuit.
ω, ω'	number of calls generated or received from the network by a terminal respectively.
$\Phi_I(s), \Phi_{II}(s)$	characteristic function of $\tilde{a}_I(k)$ and $\tilde{a}_{II}(k)$ respectively.

"Light Arose in Me in Things
Not Heard Before"
Dhammacakka Sutta

1. INTRODUCTION

The **first** commercial satellite communication service, INTELSAT I, became operational in April, 1965 and satellite communication is considered to have reached a mature phase by 1971 with the operation of INTELSAT IV. The latter spacecraft weighs 730 Kg in orbit and provides two 4° diameter spotbeams as well as global coverages. With the proliferation of data communication, the demand for a highly reliable and distance independent link such as the satellite channel, is increasing. Furthermore the trend towards a cheap direct to user satellite link has invited more innovation in the following two areas. First, improved or new technology is needed to provide low cost earth stations, and to overcome the acute problems of overcrowded orbital space and frequency spectrum. Second, suitable multiple access protocols are required in order to utilize the satellite channel as effective as possible. However the latter task is complicated by the diversity of user traffic characteristics and service requirements (see Section 1.1).

The use of a satellite switched multiple spotbeam network has the potential to overcome the problems of overcrowding in orbital space and frequency spectrum. It permits frequency re-use in each spotbeam coverage, since interference is avoided through physical separation of the coverages (space division). Hence the recommended protocols operating in such an environment are known as Space Division Multiple Access (SDMA) protocols. The required number of spacecraft can be reduced if the following hardware is available :-

1. fully scannable spotbeams;

2. switchable transponders;

3. intersatellite link.

Such facilities shift the complexity from the earth to the space segment. The financial gain of a space division network has been studied by Fordyce and Stamminger [FOR 79] and summarized in Table 1.0.1. They have shown a greater than 50% reduction in cost over conventional space technology (see Section 1.2 for hardware considerations).

The added advantage of a multiple spotbeam antenna is its higher effective isotropic radiated power (EIRP) due to the increase in satellite transmitter antenna gain as the resolution becomes finer. Such an increase in EIRP can be utilized either by a reduction in earth station antenna size or an increase in channel capacity. In either case such a benefit, and the single frequency operating feasibility, popularize direct to user satellite communication by reducing its cost. In addition the difficulty of unauthorized reception and jamming are side benefits of a multiple spotbeam network.

However in such a network, stations can only be connected by synchronizing their transmission with the beam switching on-board the satellite. This increases the complexity of the multiple access protocols and necessitates the maintenance of precise network synchronization. It is to the problems of protocol architecture and performance analysis that this thesis is addressed.

Section 1.3 gives a brief description of a feasible synchronization technique for a space division network. Section 1.4 reviews the available packet communication multiple access protocols for data traffic using a global coverage. Some of the desired features are adapted by the space division multiple access protocols. Finally, Section 1.5 examines the perspectives and reviews the work in this dissertation.

conventional ?

	Multiple spotbeam system	Satellite
Space segment	0.46	0.55
Earth segment	1.04	2.69
Total	1.50	3.24

(in thousands of 1979 dollars per circuit year)

Source : [FOR 79]

Table 1.01 Cost comparison of multiple spotbeam system and conventional satellite for the 10 year period of 1987-1996

1.1 TRAFFIC MODELS AND SERVICE REQUIREMENTS

With the diversity of users the space division network is proposed to serve, the modelling of all the traffic types is a difficult task. However attention is concentrated on the two prevalent traffic types, namely data and voice traffic.

For voice traffic operating in an environment of a large number of users, it is recognised that the Poisson and negative exponential distribution processes provide sufficiently accurate models for the voice call generation and circuit holding time respectively [SYS 58]. Voice traffic is extremely sensitive to the variation of packet interarrival time but reasonably tolerant to error and call request blocking. For these reasons, it is proposed that circuit switching be used for voice traffic in a space division network.

Data traffic on the other hand embraces a large range of requirements, ranging from bursty interactive traffic to continuous file transfer. Fuch and Jackson examined such systems and found their interarrival time can be modelled by a gamma or a lognormal distribution [FUC 70]. However an article by Pawlita casted doubts on the accepted simple models, especially on the behaviour of a medium speed terminal such as a video display unit. Pawlita showed that the traffic patterns are system and application dependent and can not be modelled by any particular distribution [PAW 81]. Hence the analytical results in this thesis are derived for an arbitrary arrival process. Pawlita also indicated that the traffic is generally extremely bursty. Hence message switching is proposed to be used for data traffic in a space division network.

A fixed boundary interleaving frame structure is proposed as a means of serving a network with integrated data and voice traffic (see Chapter 3). A recent analysis by Ross and Mowafi on such a frame structure indicates its performance viability [ROS 82].

1.2 HARDWARE CONSIDERATIONS

As the satellite switched multiple spotbeam network shifts the complexity of satellite communication from the earth to the space segment, the most vital piece of hardware in a space division network is the scanning spotbeam. Recent developments have indicated promising results with the phased array antenna [REU 80]. It is particularly favoured due to its excellent sidelobe suppression and fast switching time.

The phased array antenna combines the output of an array of radiating elements to form the desired pattern. A feed system distributes energy to the arrays and provides amplitude weighting. Digital phase shifters which are manipulated by a high speed control unit, alter the phases of the feed elements in order to shape and steer the spotbeams. The produced pattern is amplified and projected onto a large aperture which focusses on a particular coverage.

With ever increasing logic speed, switching time should not be an issue in the near future. The current capability for a RF satellite switch over 500 MHz bandwidth is 50 nanoseconds with 50 db input/output isolation. If RF is converted to IF with on-board processing, a switching unit with less space, weight and power can be utilized. Processing time of the control unit may be reduced if parallel processing and hardwire logic modules are used [STE 80].

Another important consideration is the technology involved in the production, steering and tracking of a spotbeam coverage. With existing technology, a 600 Km diameter spotbeam coverage can be obtained (the boundary of such coverage is defined by the -3db power level). It has been predicted that dramatic improvements can be achieved within the next twenty years. With an operating frequency of 10 GHz and a power level suitable for communication purposes (10^{-8}W/cm^2), a 50 Km spotbeam coverage can be obtained with a 30 m aperture in a geostationary orbit using 100 KW power input or even a 15 Km spotbeam coverage with a 100 m aperture and 10 KW power input [NOL 78].

The role of this section is to show that although the required technology for a multiple spotbeam space division network is complex, it could be achieved within the immediate future.

1.3 SYNCHRONIZATION

One of the characteristics of a space division network is its point-to-point nature, i.e. a transmitting station can only be connected with a destination station via an uplink and a downlink beams. Hence a transmission is made without the benefits of a feedback from the downlink beam. Nevertheless a precise network time has to be maintained in order to synchronize the station transmission and the beam switchings on-board the satellite. The network time is established by the stations using a particular synchronization procedure suitable for a space division network. In this section, one of these procedures is briefly described [NUS 77], [SMI 78].

A coarse network time is broadcast to each spotbeam coverage by a master earth station or the satellite. Further fine tuning of the network time can be made by the Window Method [SMI 78]. A coded burst is transmitting into a pre-assigned time slot and the transmission is looped-back. The location of the pre-assigned time slot is estimated with the knowledge of the propagation delay and the broadcast network time.

The coded burst has a period of the order of the window length. The station listens to its looped-back transmission. If it passes through the window in its entirety and is received by the transmitting station, the fine tuning is completed. However if the coded burst is truncated, the station adjusts its timing accordingly and repeats the transmission until a satisfactory accuracy is achieved. Such fine tuning can also be used for the calculation of the variation in network time due to satellite motion and oscillator frequency drift. Hence an accurate prediction of the network time can be obtained in the case of minor outage.

1.4 MULTIPLE ACCESS PROTOCOLS FOR A SATELLITE PACKET COMMUNICATION NETWORK USING A GLOBAL BEAM

Whenever an expensive resource is required to be shared amongst a number of independent users, a multiple access protocol or a commonly agreed scheme is used to distribute the resource. In a packet satellite communication network, the resource is the communication channel and the multiple access protocol is the tool for allocating the resource.

In the case of a global beam satellite communication network, the satellite transmits and receives signal to and from any of the stations within its beam coverage; therefore connectivity between stations within the beam coverage is easily obtained. For a geostationary communication satellite, the beam coverage is approximately one third of the earth´s surface. Another important characteristic of satellite communication is its relatively long round trip propagation delay (rtpd) of 0.27 second, where the round trip is defined as the distance from the earth to the satellite and back to earth. This delay influences the selection of an appropriate multiple access protocol.

In order to accommodate bursty data traffic, attention is focussed on packet switching. Several review papers discuss multiple access protocols for packet communication in detail [TOB 80], [RET 80], [LAM 79]. From these references, four main categories of satellite packet communication multiple access protocols were identified.

1. fixed assignment scheme - the channel bandwidth is allocated to each station in a static manner.

2. random access scheme - each station accesses the channel in a random fashion; the protocol provides a mechanism to reduce and/or resolve conflicts due to the collision of packets.

3. distributed control demand assignment scheme - implicit or explicit control information is to be exchanged amongst all the stations. The stations execute the same scheduling algorithm in order to obtain a co-ordinated response.

4. centrally controlled demand assignment scheme - the transmission schedule is allocated by a central controller. The schedule is then relayed to each station.

Each of the above schemes is discussed in detail in the following sections.

1.4.1 Fixed Assignment Scheme

The fixed assignment scheme is the simplest of all the available schemes. It partitions the communication channel into either frequency or time domains. A Frequency Division Multiple Access (FDMA) Protocol is obtained by dividing the frequency bandwidth into separate non-overlapping channels. Each station has access to a dedicated channel. The advantage is its ease of implementation. The disadvantages include

i. poor channel utilisation for bursty data traffic;

ii. inflexibility in bandwidth re-allocation - requires alteration in hardware;

iii. in order to avoid intermodulation interference, the travelling wave tube (TWT) operates in a backoff linear region and is therefore less efficient.

In the case of the Time Division Multiple Access (TDMA) Protocol, each station is scheduled to transmit in a short non-overlapping time interval on the same channel. Hence, the TDMA Protocol requires a frame structure and the stations need a global timing mechanism. Timing is established either explicitly by a reference station, or implicitly by measuring the propagation delay from the station to the satellite [SCHM 74]. The main advantages are :-

i. the capacity assignment can be tailored to the station's requirement;

ii. the TWT can be operated in a region of maximum efficiency, since only one station is transmitting at any one time;

iii. the connectivity between stations is an intrinsic property of the TDMA Protocol, since every station listens and transmits on the same channel.

The disadvantages include

i. poor channel utilisation for bursty data traffic;

ii. a more complex implementation in comparison with the FDMA Protocol.

The TDMA Protocol also exhibits a superior delay performance than a comparable system using a FDMA Protocol [RUB 79]. The average message delay for the TDMA Protocol is given in Appendix 2 for the case of a Poisson message input and geometrical message length distribution.

Another quasi-static access protocol is the Code Division Multiple Access (CDMA) Protocol. The transmission for each station is combined with a pseudo-random code so as to cause the transmission to occupy the entire bandwidth of the transponder. The destination station extracts the message using a correlation detector. The CDMA Protocol provides cryptograhic security and jam proof operation. However the multiple orthogonal codes are obtained at the expense of increased bandwidth and are therefore highly inefficient.

The main disadvantage of any one of the above protocols, based on the fixed assignment scheme, is its inability to accommodate the burstiness of the data traffic. Thus they cannot provide an instantaneous response to the traffic requirement of a user which is the very quality that a small user requires. For this reason the random access and the demand assignment schemes have been examined.

1.4.2 Random Access Scheme

In the random access scheme, the entire channel is accessed by all stations in a random manner. The two well known examples of the random access scheme are the Aloha and Slotted Aloha Protocols. Both protocols have been extensively studied for satellite applications by [ABR 77] and [KLI 75II]. The delay performance for multiple packet messages with Poisson input assumption is given in [BEL 80].

This scheme requires some mechanism to resolve conflicts due to the collision of packets. The Aloha Protocol permits the stations to transmit as soon as the messages arrive. Messages from different stations may overlap and destroy part or entire messages. With a global beam network, the transmission station detects the collisions and the messages are retransmitted after a randomised delay to avoid a repeated collision. In the Slotted Aloha Protocol, the channel is slotted into time segments and the stations are synchronised to transmit at the beginning of a slot. This eliminates the packet partial overlaps.

Both protocols are relatively simple to implement and adaptive to the traffic requirement of each individual user. However, they are inherently unstable. Retransmissions could lead to more retransmissions by introducing a larger number of inputs and therefore further increase the probability of collision. In order to maintain stability, the maximum channel throughput has to keep below 18% and 37% of the channel capacity for the Aloha and Slotted Aloha Protocols respectively. The average multiple packet message delay for the superior of the two protocols, Slotted Aloha Protocol, is given in Appendix 2 for a Poisson channel input process.

1.4.3 Distributed Control Demand Assignment Scheme

The characteristic long round trip propagation delay (rtpd) of satellite communication influences the selection of demand assigned multiple access protocols for packetized data communication. All distributed control demand assignment protocols require the dissemination of capacity requirements to all stations. With this information, the stations execute the same scheduling algorithm, resulting in a co-ordination in their transmissions. Hence, it has a minimum message delay of two rtpd's. The requirement of all stations receiving the same control information can be met by a global beam satellite channel, because of its broadcast nature. In the following two distributed control demand assignment protocols, Implicit Reservation and Reservation Aloha Protocol, are discussed.

The Implicit Reservation Protocol [CRO 73] utilises the frame concept in a Slotted Aloha channel. The frame has a framelength longer than one rtpd. The protocol uses the reservation-by-use principle, whenever a station successfully transmits in a slot, that slot is subsequently assigned to the successful station in the next frame. With each station maintaining a history of usage of each slot for the duration of one frame, the idle and the contended slots of the previous frame are accessed by all stations in the present frame in a Slotted Aloha contention mode. The Implicit Reservation Protocol is effective for continuous stream type traffic or messages with long message length. It is inefficient for short messages, since at least one slot is left vacant after a successful transmission. It also has the disadvantage of being unable to prevent a station from capturing most or all of the slots for an indefinite period.

The Reservation Aloha Protocol [ROB 73] utilises a separate reservation channel for capacity reservations. The frame is divided into data and reservation sections. The reservation slots are accessed by the Slotted Aloha Protocol, and used by all stations to send capacity reservations and acknowledgements. The data slots are used on a reservation basis. Whenever a message enters a station, the station temporarily stores the message and randomly chooses one of the reservation slots for the transmission of the capacity request. All the successful reservations are received by the stations within the global beam coverage, and they update their reservation count. To maintain co-ordination in transmission, each station sends the status of its queue with its data message. This information is used by the entering and out of co-ordination stations.

The amount of bandwidth assigned to the reservation channel can be made adaptive to the traffic load. The smaller the load, the larger the reservation channel. The robustness of Reservation Aloha Protocol can be improved by a proper encoding of the reservation packets to increase the probability of their correct reception by all stations.

Robert [ROB 73] uses M+1 large slots with the last slot converted into V small reservation slots. In Appendix 2, the average delay is calculated for an arbitrary size reservation channel. The result is obtained for a Poisson message input with geometrical message length distribution.

1.4.4 Centrally Controlled Demand Assignment Scheme

With a centrally controlled demand assignment scheme and using a master earth station as the central controller, a minimum message delay of three rtpd's is required for a capacity reservation scheme. This could be reduced to two rtpd's if an on-board central controller is used. The advantage of this scheme is that a simple earth station can be used. Two centrally controlled demand assignment protocols suitable for satellite communication are illustrated below.

A circuit oriented system divides the bandwidth into smaller circuits in the time or frequency domain, and the circuits are assigned on demand. One such method known as the SPADE system, which is used for satellite telephony traffic, dynamically assigns a pool of FDMA circuits to each active station. It also has a reservation channel and each station is assigned with one reservation slot using a TDMA technique. The requests and the releases of FDMA circuits are handled via the reservation slots.

The Split-channel Reservation Multiple Access (SRMA) Protocol utilises a centrally controlled capacity reservation scheme [TOB 76]. The channel bandwidth is divided into control and data channels in the frequency domain, the control channel is further divided into reservation and scheduling subchannels. The stations send capacity requests in the reservation subchannel whenever required, and the central controller manages a queue of requests and informs the stations of their allocated capacity via the scheduling subchannel. The request subchannel uses the random access Slotted Aloha contention method. In Appendix 2, the average message delay is derived for a Poisson message input with geometrical message delay for the SRMA Protocol.

1.5 REVIEW

In the following chapters, the features and the characteristics of protocols operating in a satellite switched multiple spotbeam antennas communication network are examined. Attention is focussed on the point-to-point data communication, while extensions to an integrated data and voice traffic system are also discussed. The space division system has a single transponder and the spotbeam coverages are not necessarily contiguous in order to obtain the simplest possible space segment. The earth segment has a single transmission rate. The transmission error in the control channel is assumed to be negligible, i.e. the reservation request and scheduling information are nearly error free. This could be achieved by robust redundant coding. These features are chosen in order to highlight the architecture and performance studies of the space division multiple access protocols.

With simplified traffic arrival processes, Stern et al [STE 80] derived analytical results for the delay performance of a transponder switched space division network, where the number of transponders is equal to the number of spotbeam coverages. For the case of a network where the number of spotbeam coverages is greater than the number of transponders, only simulation results are obtained. The emphasis of this dissertation is on the detailed study on the protocol architecture, delay and buffer performances under an arbitrary traffic arrival process, using a beam switched single transponder system. It also considers the topology of the terrestrial network, why? the protocol parameters and the hardware configurations in the calculation of the delay and buffer performances. It is hoped that this work will shed some light on the beam switched space division network and consequently the more complex combined beam and transponder switched system as recommended in [SCHM 74].

Chapter 2 models the connection of the terminals to earth stations at minimum cost by a facility location problem. A set covering solution is used to assign the earth stations so as to achieve the minimum number of spotbeam coverages for a non-contiguous network.

Connection between the stations in a beam switched space division network is made by an uplink and a downlink beam. Hence, the multiple access protocols can be regarded as procedures of allocating the two beams in a time division manner. All the feasible architecture of the space division multiple access protocols are examined in Chapter 3.

In Chapter 4, the average data message delay and the average voice connection delay are derived for a First Come First Served discipline. It identifies various components of the message delay and derives new analytical solutions where appropriate.

The calculations of various buffer performances for voice, unacknowledged and acknowledgement data traffic systems in the space division multiple access protocols are presented in Chapter 5. Their performances are obtained using a constant and a non-constant output model.

Comparisons of the delay and buffer performances are made in Chapter 6 between the space division multiple access protocols and other global beam data packet multiple access protocols. Relationships and possible tradeoffs between the performances and the terrestrial configurations and protocol parameters are also discussed in detail.

The simulations in Chapter 7 validate the analytical results of delay and buffer performances. The simulation model can also be used to obtain results which are too difficult to obtain analytically.

2. TERRESTRIAL NETWORK DESIGN

For a satellite network utilising a Space Division Multiple Access protocol there exists a close relation between the protocol and the terrestrial network. Each of the terminals in the network is connected to a station. In our context, a terminal could be either a data source and sink, or a voice source and sink, or a combination of both. The number of stations, and the number of spotbeam coverages, directly affects the delay and the buffer performances of the protocols (see Chapter 4, 5 and 6). This chapter describes the formulation, and reviews some available solution methods for the problem of connecting the terminals to the earth stations at minimum cost, and the problem of assigning the earth stations such as to achieve a minimum number of spotbeam coverages for a non-contiguous network. The review is by no means complete as an enormous number of works can be consulted, e.g.[BOO 77] and [GAR 72]. However it indicates some of the salient features of the solutions of the suggested formulations.

2.1 TERMINALS TO STATIONS OPTIMIZATION

The incentive to optimize the number and the locations of the earth stations is one of economics. The problem of associating a terminal with an earth station can be reduced to three subproblems:-

1. determination of the number and the locations of stations,

2. assignment of the terminals to the stations,

3. determination of a layout for the terminals.

In order to focus onto the first two subproblems, a star connection is assumed for the terminal layout (Fig.2.1.1). Further extensions of the layout are studied in the next section.

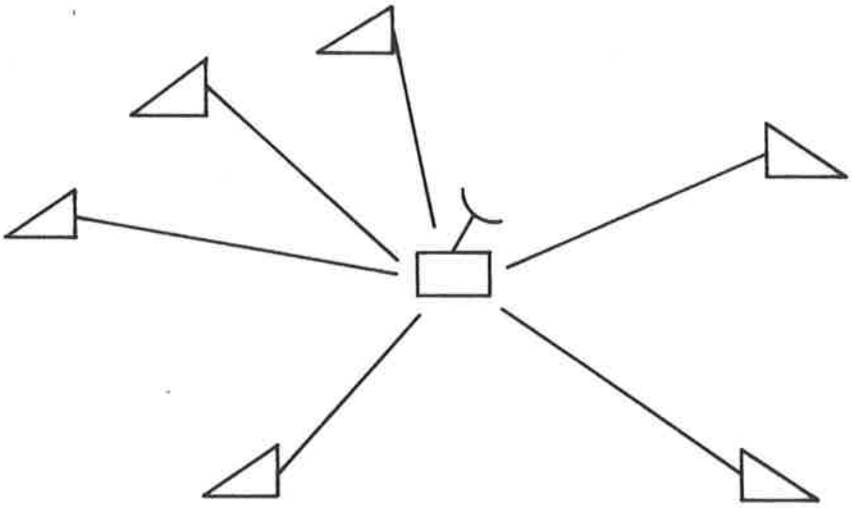


Fig.2.1.1 A star connection

2.1.1 Definition of the Terminals to Stations Optimization

The following parameters of the network are assumed known:-

1. the locations of the terminals, T_i, $i=1,2,\ldots,t$
2. the cost c_{ij} of connecting a terminal T_i to an earth station E_j, $j=1,2,\ldots,s$
3. the cost d_j of establishing a station at site j.

The cost of connecting a terminal to a station is normally a function of capacity and distance between the end points; it also includes the costs of modems and connections at the station. The cost of establishing a station at site j includes site acquisition, cost and maintenance of the station. Due to the limitations of logical channel units, line capacity and buffer capacities, the station capacity is expressed in the number of terminals (k) that a station can accommodate. k is the number of voice or data terminals or a number taking the integrated nature of the local network into account. The problem is to choose the number of stations, their sites and terminal assignments such that the cost Z is minimised :-

$$Z = \sum_{i=1}^{t} \sum_{j=1}^{s} c_{ij} x_{ij} + \sum_{j=1}^{s} d_j y_j \qquad (2.1.1)$$

where $x_{ij} = 1$ if T_i is connected to E_j

$= 0$ otherwise

$y_j = 1$ if a station is located at site j

$= 0$ otherwise

Since each terminal must connect to at least one station

$$\sum_{j=1}^{s} x_{ij} \geqslant 1 \qquad (2.1.2)$$

and each station accommodates at most k terminals

$$\sum_{i=1}^{t} x_{ij} \leq ky_j \qquad (2.1.3)$$

The problem is now to choose zero-one integers for x_{ij} and y_j such that Z is minimised subject to constraints (2.1.2) and (2.1.3).

The formulation as stated above is similar to the plant, warehouse and facility location problem found in operations research literature. The exact solution employs either exhaustive search or a branch-and-bound procedure both of which require long computation times. However, a large number of heuristic (non-optimum) yet computable algorithms are available [BOO 77]. The selection of the heuristic Centre of Mass (COM) Algorithm is discussed in the next section.

The terminal layout routine is often used as an ingredient for the heuristic facility location algorithm. The problem is to find a tree connecting all terminals to the station with the minimum cost. Since the multidrop line constraint is incorporated into the COM Algorithm during the terminal asignment, the terminal layout can be regarded as an unconstrained Minimum Spanning Tree problem for which two efficient optimium algorithms exist. They are the heap sorted Kruscal´s Algorithm [KER 72] and the Prim´s Algorithm [PRI 57], both of which are described in the next section.

2.1.2 Discussion On Optimization

A. STATION LOCATION PROBLEM

A review article by Boorstyn and Frank [BOO 77] evaluates some of the available heuristic algorithms for the station location problem. The evaluating criteria are

1. the cost performance with the optimal solution or other heuristic solutions,
2. the required computation time.

The Centre of Mass (COM) Algorithm [MCG 77] is recommended due to its improvement in performance and computation time. In addition, the following extensions are possible within COM Algorithm to tailor for the requirements of a satellite network :-

1. provision for different types of terminals,
2. provision for different types of stations,
3. accommodate a list of preferred station sites in order to allow for the political or strategical importance of the candidate sites or to take advantage of the established terrestrial network,
4. incorporate a terminal layout algorithm.

The COM Algorithm is divided into four stages :- a. clustering, b. partitioning, c. local optimizing, d. line layout. The clustering stage replaces a group of weighted terminals with a single node at the Centre-Of-Mass site (or COM site) of the group. The multidrop line constraint, which is the maximium number of stations that a line can support, limits the cluster size. The partitioning stage investigates each of the COM sites as a station location. It uses a heuristic terminal assignment algorithm to evaluate the cost saving of establishing a station at each of the COM sites. This stage is terminated when no saving is made for an additional station. Since the COM site may not be on a terminal site, the local optimizing stage selects the most suitable terminal near the COM site as the actual station site. A line layout algorithm is invoked to connect all the terminals in the associated clusters to the chosen station site. An example of the COM Algorithm is given in [MCG 77].

B. TERMINAL LAYOUT PROBLEM

1. Kruscal's Algorithm orders all the costs and connects the cheapest pair of terminals first. It continues this process until all the terminals are connected in a tree configuration. It essentially orders the branches and checks to prevent a loop being formed. The ordering requires $n^2\log_2 n$ comparisons, where n is the number of terminals. The more efficient heap sorted Kruscal's Algorithm [KOR 72] requires n^2 comparisons. Instead of ordering all possible branches, it only orders the required number.

2. Prim's Algorithm [PRI 57] connects the closest terminal to the station and forms a group of two. It continues to find the cheapest connection between the remaining terminals and the group until all terminals are connected, and it requires n^2 comparisons.

Prim's Algorithm is chosen here for its efficiency and simplicity in machine coding.

2.2 SPOTBEAM COVERAGE OPTIMIZATION

In a contiguous space division network, the spotbeam coverages fill the entire terrestrial network and every station is assigned to a coverage. However, certain advantages can be gained by a non-contiguous network, e.g. reduced number of feedhorns and protocol overhead; and for some applictions, a non-contiguous network is more suitable, e.g. mobile or maritime traffic. In this section, with a given set of station locations, the problem of associating the stations with the minimun number of coverages for a non-contiguous space division network under certain coverage area constraints, is modelled by a set-covering problem; and a strategy is suggested for its solution. It is believed that such modelling had not been revealed in any published work.

2.2.1 Definition of the Spotbeam Coverage Optimization

Once the station sites are known, the sites are assigned to a set of spotbeam coverages such that the required number of coverages is minimised. This problem is similar to an emergency service facility location problem - a specialised set-covering type problem. The similarity between these two problems is the objective of the optimization taking into account the maximum separation distance between the population (stations) and the emergency service facility (centre of spotbeam coverage).

The objective of this optimization is to assign the j-th station (E_j) to the q-th spotbeam coverage, which is represented by the location of the "centre" of the spotbeam G_q; such that

1. the total number of the spotbeam coverages (z) is minimized,

2. all stations are covered by at least one spotbeam,

3. the distance $F_d(j,q)$ between E_j and G_q is less than the permitted distance function F_d'; in the case of a circular spotbeam coverage, the distance between any station and the centre of coverage must be less than the radius of the circle.

The above requirement leads to the formulation of the q-th feasible set - N_q :-

$$N_q = \{j \mid F_d(j,q) \leq F_d'\} \qquad (2.2.1)$$

where $q \in Q$, $Q=1,2,\ldots,s'$

There exists s′ candidate coverage centres, as the centre of coverage may differ from the station sites. Requirement 1 and 2 can be stated as a set-covering problem and its solution is given by an integer linear program [GAR 72] :-

$$\min \; z = \sum_{q=1}^{s'} x_q \qquad (2.2.2)$$

such that

$$\sum_{q=1}^{s'} a_{jq} x_q \geq 1 \qquad j=1,2,\ldots,s'$$

where $x_q = 1$ if coverage centre is on site q

$= 0$ otherwise

$a_{jq} = 1$ if $j \in N_q$

$= 0$ otherwise

x_q can also be specified at the beginning of the optimization to take into account decisions based on past experience.

2.2.2 Discussion On Optimizaton

There exists a number of exact solutions for the set-covering problem. However, for a large network (s or $s' \geq 100$), such methods become impractical. The following strategy is suggested by Toregas et al [TOR 73]. Roth's Reduction Method [ROT 69] is used to produce an optimal solution in most cases; otherwise, a cyclic matrix may be produced. In such a case, a partial solution set (S_p) is obtained with a new matrix $[a_{jq}]$ of reduced dimension $s_1 \times s_1'$, where a_{jq} is the element of the still active cyclic matrix. The original problem is now reduced to

$$\min \quad z_1 = \sum_{q \varepsilon (Q-S_p)} x_q \tag{2.2.3}$$

such that

$$\sum_{q \varepsilon (Q-S_p)} a_{jq} \, x_q \geq 1 \qquad\qquad j=1,2,\ldots,s_1$$

The optimization (2.2.3) can now be solved as linear program, i.e. $x_q \geq 0$ instead of an integer linear program where $x_q \ \varepsilon \ (0,1)$. The results of the linear program give three classes of solutions :-

1. z_1 and all x_q are integers, i.e. the optimal slution,

2. z_1 is not an integer,

3. z_1 is an integer but some x_q are not.

To overcome case 2, an additional constraint is added to the formulation of (2.2.3) until an integer solution is obtained

$$\min \quad z_2 = \sum_{q \varepsilon (Q-S_p)} x_q \tag{2.2.4}$$

such that

$$\sum_{q\varepsilon(Q-S_p)} a_{jq}\ x_q \geqslant 1 \qquad\qquad j=1,2,\ldots,s_1$$

$$\sum_{q\varepsilon(Q-S_p)} x_q \geqslant [z_1] + 1$$

where $[z_1]$ is a integer portion of the solution obtained in (2.2.3).

For case 3, in which z_1 is an integer but some x_q are not, an exhaustive search among the alternate optimums or branch-and-bound techniques guarantees an optimal solution. However, the required computuation time could be formidable. The following "Reduction Method" [TOR 73] may not always derive an optimal solution, but it is efficient. In the case where the Reduction Method fails to arrive at an optimum, the exhaustive search or the branch-and-bound algorithm (see [TOR 73] for descriptions) may be used.

The Reduction Method can only apply if some number (b) of x_q have unity value at optimum. Then the constraint matrix $[a_{jq}]$ can be reduced by two stages :-

1. Since b of the x_q are fixed with the value of one, some of the constraints have been satisfied. The b columns and rows can now be eliminated at those columns which have unity x_q - Toregas Reduction.

2. Roth´s Reduction Method can now be applied to the above matrix as it may no longer be cyclic.

If the matrix is not yet fully reduced, an optimal solution may be obtained from the newly reduced matrix of e nodes with additional linear programming. If b+e=z, the solution obtained by various methods is an optimal one. Otherwise, either an exhaustive search or a branch-and-bound algorithm is invoked to solve the outstanding case. Fig.2.2.1 summarizes the solution procedure in a flow chart. An example of the Reduction Method

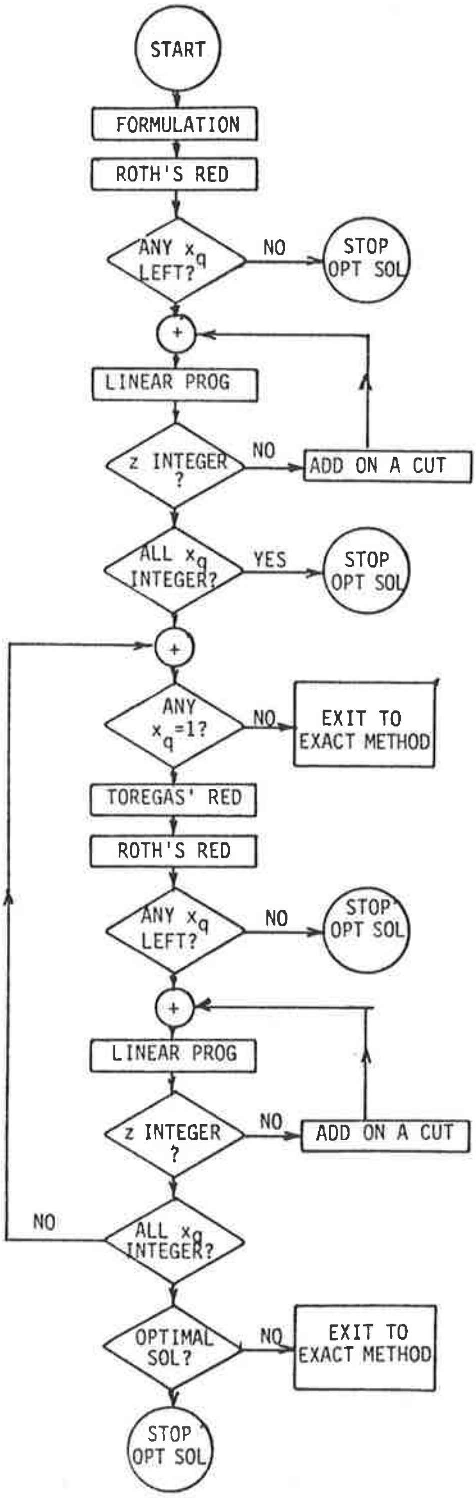


Fig.2.2.1 Solution procedure of the Reduction Method

is given in [TOR 73].

2.3 CONCLUSION

This chapter attempts to clearly model the two terrestrial network design problems, and reviews some available solution methods. The problem of connecting the terminals to the earth stations at minimum cost is modelled by a facility location problem. The heuristic Centre-Of-Mass Algorithm is used for the optimization of station locations and terminals assignment, because of its efficiency, performance and possible extensions. A simple unconstrained minimum spanning tree solution - Prim's Algorithm is used to optimize the terrestrial layout problem.

The problem of associating the stations with the minimum number of coverage zones for a non-contiguous space division network is modelled by a set-covering problem. This approach is believed to be new. The problem can be solved by either exhaustive search or a branch-or-bound algorithm. However, for a large network of greater than one hundred stations, such time consuming algorithms are impracticable. The use of a reductive strategy is appropriate, and none of the examples tested failed to arrive at an optimum.

3. SPACE DIVISION MULTIPLE ACCESS PROTOCOLS

With no on-board storage facility, the uplink and the downlink of a mutiple spotbeam satellite network can be regarded as a re-allocatable resource and assigned on a

1. fixed sequence, i.e. Fixed Assignment (FA) Protocol,
2. allocation on demand, i.e. Demand Assignment (DA) Protocol,
3. a combination of fixed and demand assignment, i.e. hybrid system of Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol.

Two forms of demand assignment protocols can be identified. The first is to allocate the uplink and the downlink for the duration of the message - Demand Assignment/Message Switched (DA/MS) (Fig.3.0.1a). In the second, a slot per frame is allocated to the call until it is terminated - Demand Assignment/Time Division Multiple Access (DA/TDMA) (Fig.3.0.1b).

The relative merits of the two forms of demand assignment protocols are as follows :- for the case of the DA/MS Protocol, a temporary storage of the message in the station is required before it is sent; therefore, a large message buffer is needed. As the entire message is sent in a burst, the average service time is given by

$$\text{average service time} = \text{average message length} / \text{bandwidth}$$

On the other hand, the DA/TDMA Protocol sends a fixed amount of information from the terminal to the station per frame. It reduces the required message buffer length, since only a segment of the message is sent in each frame. Hence, the DA/MS Protocol is more suitable for traffic with a short message length, while the DA/TDMA Protocol a long message length. The disadvantage of the DA/TDMA Protocol is its long service time.

$$\text{average service time} = \frac{\text{average message length x framelength}}{\text{number of assigned slots per frame}}$$

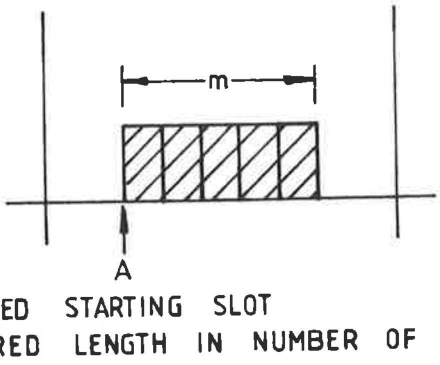


Fig.3.0.1a Demand assignment/message switched

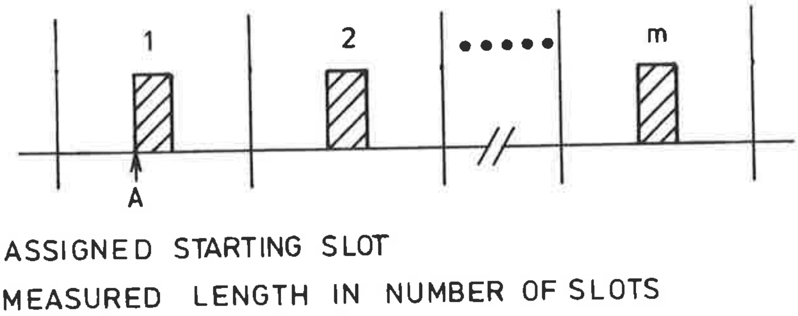


Fig.3.0.1b Demand assignment/time division multiple access

Both forms of the demand assignment algorithms will be used in the Space Division Multiple Access Protocols described in this thesis.

In this chapter, various Space Division Multiple Access (SDMA) protocols and their frame and delay strucutures are examined in detail. In designing the SDMA protocols, it is assumed that each station maintains precise frame and slot timing and knows its propagation delay from the satellite. Two independent random components of the delay strucuture can be identified (Fig.3.0.2):-

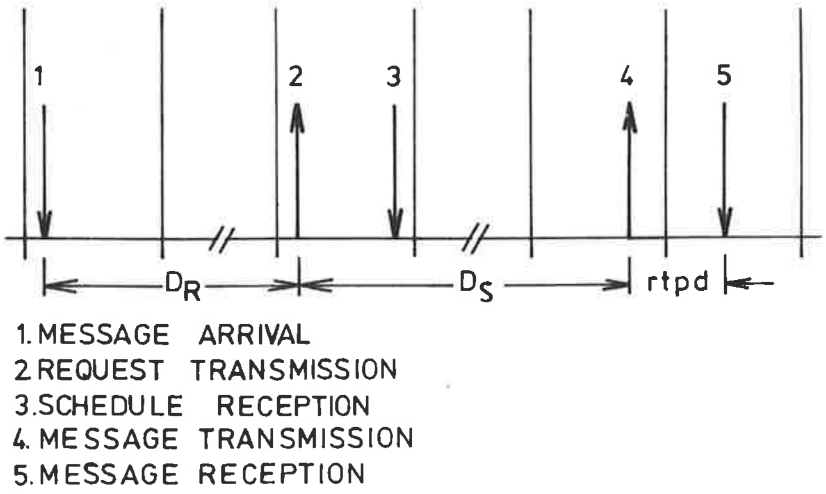


Fig.3.0.2 Signal-time diagram

1. Reservation Delay (D_R) - the duration between the arrival of a message or a notification of the impending message into a station and the submission of a capacity request for the message. A delay can occur since only a fixed number of capacity requests are carried by each reservation burst and the capacity requests are served by a defined discipline.

2. Scheduling Delay (D_S) - the duration between the sending of the capacity request and the beginning of the message transmission. The capacity allocation strategy and the traffic volume determines the value of the scheduling delay. This includes the time required to relay the scheduling information form the controller to the requesting station.

In our analysis, the scheduling delay calculation for the two Demand Assignment formats of Message Switched (DA/MS) and Time Division Multiple Access (DA/TDMA) are similar, since the scheduling delay is defined as the duration between the sending of the request and the beginning of the message transmission.

3.1 Fixed Assignment Uplink and Downlink (FA) Protocol

The simplest multiple access protocol for a multiple spotbeam network is to have the uplink and the downlink switching through a fixed sequence. The advantage of such protocol is its simplicity in implementaton on both the earth and space segments. Also, its regularity is ideal for continuous traffic such as voice traffic. The disadvantages include low channel utilisation for bursty data traffic and difficulty in reallocating capacity if there is a variation in the traffic pattern.

3.1.1 Frame Structure of the FA Protocol

In the case of the FA Protocol, the uplink and the downlink are switched through a fixed sequence. The frame has a fixed structure. For full connectivity, each station must possess K segments of equal or different duration within the frame, one for each spotbeam coverage and K is the total number of spotbeam coverages or zones in the terrestrial network. For stability, each segment must contain a capacity larger than the expected carried traffic. With s stations in the network, the frame contains K x s segments, i.e. the frame is dependent on the topological formation of the earth segment. The structure may be arranged to allow the daily or seasonal variation in the traffic pattern, i.e. a quasi-static assignment. The frame structure of a FA Protocol for data communication is shown in Fig.3.1.1. The synchronization is maintained by the looped-back synchronistion burst which is sent by the station into its own zone. In addition, a master earth station periodically sends the network time to all stations in the network.

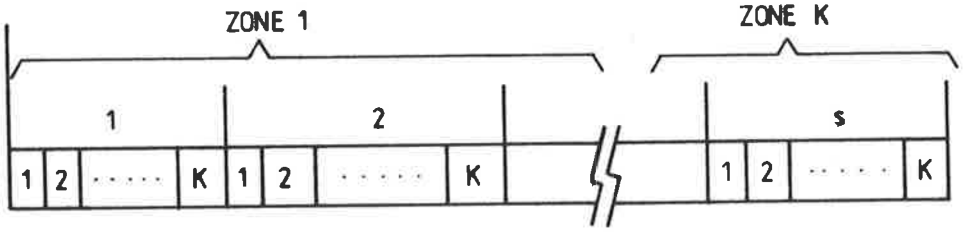


Fig.3.1.1 Frame structure of a FA Data Protocol

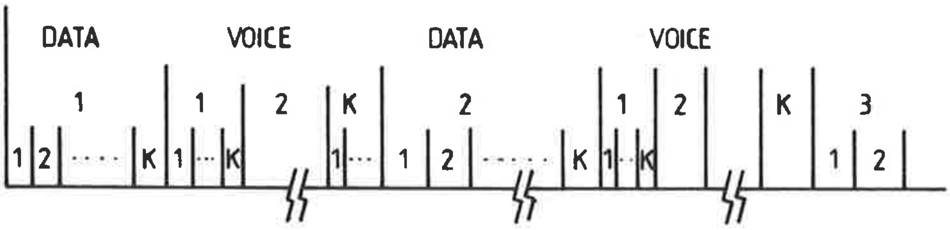


Fig.3.1.2 Frame structure of a FA Integrated Traffic Protocol

For an integrated data-voice protocol with the voice traffic utilising circuit switching, the voice subframes have a smaller length (f_v) and regularly intersperse with the data within the frame (Fig.3.1.2). The higher effective frame rate of the voice subframes permits the utilisation of a smaller buffer and a framelength (f) of greater than a round trip propagation delay (rtpd) without an exorbitant overhead.

3.1.2 Delay Structure of the FA Protocol

The delay structure of the Fixed Assignment (FA) Protocol is similar to the delay strucuture of the Time Division Multiple Access (TDMA) Protocol, since in both protocols a fixed amount of capacity is assigned to a station per frame. In the FA Protocol, each station must contain K queues of buffers, one for each zone. When a message or a notification of impending message enters a station, its destinaton is interrogated and the message or the capacity requirement is stored in the appropriate queue by a defined discipline - e.g. First Come First Served (FCFS). The station then waits for its turn to transmit, a pre-assigned number of slots are sent from each queue to the corresponding zone at the appropriate time. The discipline on queue positioning, and the traffic volume determine the waiting time of each message, i.e. the scheduling delay (D_S). The delay strucuture is shown in Fig.3.1.3, the average delay E(D) per messsage and its variance are given by

$$E(D) = E(D_S) + rtpd \qquad (3.1.1)$$

$$var(D) = var(D_S) \qquad (3.1.2)$$

where rtpd=round trip propagation delay

3.2 Demand Assignment Uplink and Fixed Assignment Downlink (DA-FA) Protocol

A hybrid system of demand assignment uplink and fixed assigned downlink can be used to improve the efficiency. The demand assignment uplink can be operated in either a message switched or a TDMA mode. Since the traffic is expected to be bursty in nature, message switching is used in the proposed DA-FA Protocol. The inefficiency of the channel utilisation is caused by the difference between the assigned capacity and the actual traffic. With the DA-FA Protocol, the capacity is allocated on

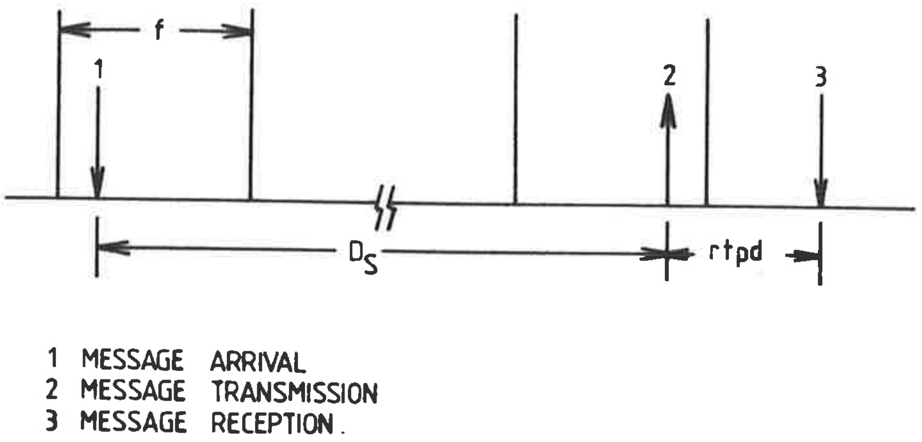


Fig.3.1.3 Signal-time diagram for a FA Protocol

a zone-to-zone basis rather than a station-to-station basis as in the FA Protocol. The channel utilisation is improved due to the statistical multiplexing of all the stations within a zone. From the space segment point of view, both the uplink and the downlink are still switching through a fixed sequence. Hence the frame structure is again dependent on the traffic pattern. The average delay is given by

$$E(D) = \underline{E(D_R)} + E(D_S) + rtpd \qquad \text{for very long messages?} \qquad (3.2.1)$$

with the delay structure shown in Fig.3.0.2. All the stations within a zone are now competing for the allocated zone capacity. Hence a control mechanism is required for the uplink transmission. There exist four possible alternatives :- A. centrally controlled with satellite link, B. centrally controlled with on-board controller, C. centrally controlled with terrestrial link and D. distributed control. Each of the possibilities will now be examined in detail.

A. Centrally controlled with satellite link

The simplest in implementation of all the DA-FA Protocols is a centrally controlled protocol via a satellite link. The frame is divided into two sections :- the control and the information. Let δ be the bandwidth assignment of the information section, let N_{kj} be the number of data slots allocated for traffic originating form zone k and destined for zone j; while the whole control section is further divided into the reservation and the scheduling subframes. A master terrestrial station is assigned for each zone. Once a message enters the station, a reservation burst consisting of a certain number of capacity requests is submitted from the station to the master station via the reservation subframe after a reservation delay (D_R). The master station receives all the requests from the zone after a round trip propagation delay. The resultant schedule is sent off via the scheduling subframe. The scheduling delay (D_S) includes two round trip propagation delays, as this is the minimum length of time in which a request can be sent to the master station and a schedule received by the requesting station. Synchronization is maintained through the looped-back reservation burst. The advantage of such a protocol is its simplicity in implementation. The disadvantages include the required overhead on the satellite link and the lengthy relay of the reservation request and transmission schedule.

B. Centrally controlled with on-board controller

To reduce the transmission time of the reservation and the scheduling information, one could place the central controller on-board the satellite. Hence a more complex space segment is required. Exactly the same operational procedure as in the case of a master terrestrial station with satellite link can be used, but the scheduling delay contains one round trip propagation delay only, since the relay for both the reservation request and the transmission schedule are now reduced by half.

C. Centrally controlled with terrestrial link

A terrestrial control link can be used to exchange control information such as capacity requirements, transmission schedules and network synchronization. A master terrestrial station which has a terrestrial link with all the stations within the zone is assigned to each zone. Such a link is feasible, since a sufficiently small spotbeam coverage is envisaged. It has the advantages of reducing the required propagation time of the control information and imposing no overhead on the satellite channel. The disadvantage is the additional cost incurred in the terrestrial control network. The delay is given by

$$E(D) = E(R) + E(D_S) + rtpd$$

where E(R) is the average reservation delay required by the adopted terrestrial control network and the scheduling delay $E(D_S)$ does not contain any component of round trip propagation delay (rtpd) between the station and the satellite.

D. Distributed control

In this protocol, each station sends its reservation burst and receives the capacity requests from all the stations within the zone through the looped-back control section. The transmission schedule is then determined by each station individually. By executing the same scheduling program, a uniform transmission schedule can be obtained. Synchronization is maintained by the looped-back reservation burst. The disadvantages of this protocol are the requirement of a more complex earth segment and a control overhead. The advantages are that a less complex space segment can be used and the scheduling delay contains only one round trip propagation delay, as only the capacity requests are relayed to the station via the satellite channel.

The criteria which govern the selection of the control mechanisms include :-

1. the available technology and the network cost,

2. the delay requirement of the traffic.

With the first criterion in an ever changing state, the present comparison is based on the delay performance. The centrally controlled protocol by a master terrestrial station via the satellite link has an unduly large delay because of the additional relay of reservation requests to the master terrestrial station and transmission schedule to the requesting stations. The reservation delay of the centrally controlled protocol by a master terrestrial station via the terrestrial link is network dependent, and no absolute comparison of the delay performance is possible. However, it has the potential of being the best protocol if one could afford the additional cost of the control network. The distributed control protocol has an equal or better delay performance than the centrally controlled protocol with on-board controller, since the former enjoys a larger information section bandwidth assignment. The distributed control protocol requires a more complex earth segment while the protocol with an on-board controller requires a more complex space segment.

With such views in mind, the distributed control protocol is selected to illustrate the delay performance of the hybrid protocol of a demand assignment uplink and a fixed assignment downlink. However, the methods of calculating the reservation and the scheduling delays are common to all the described DA-FA protocols.

3.2.1 Frame Structure of the Distributed Control DA-FA Protocol

The frame structure of the distributed control DA-FA Protocol consists of a control and an information section. Fig.3.2.1 shows the frame structure of a data network. A regular assessment of the capacity requirement of each station is needed. Hence, a fixed assignment TDMA Protocol is used for the reservation channel. The s reservation slots are looped-back to the sending zone in order to maintain the network synchronization (Fig.3.2.2). A master earth station periodically sends a coarse network time to all stations in the network. Each reservation burst consists of certain number (u) of capacity requests and a synchronization word. For the stability of the reservation channel, u must be greater than the average number of requests that are expected to enter the station.

For an integrated data-voice traffic protocol, with the voice traffic utilising circuit switching, the voice subframes are regularly interspersed with the data within the frame (Fig.3.2.3). Each of the voice subframes consists of a certain number of voice slots. This configurations permits the utilisation of a small buffer and a long framelength.

3.2.2 Delay Structure of the Distributed Control DA-FA Protocol

When a message enters the station, it is temporarily stored in a set of buffers. After a reservation delay, a capacity request is sent with a reservation burst which contains at most u capacity requests. The reservation subframe is looped-back to the originating zone. After the reception of capacity requests from all the stations within the zone, each station independently executes the same scheduling algorithm, resulting in co-ordination in transmission schedule. The message is sent from the buffers to the corresponding zone after a scheduling delay which is

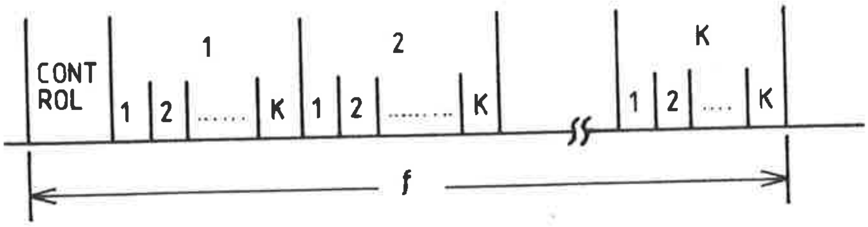


Fig.3.2.1 Frame structure of a Distributed Control DA-FA Data Protocol

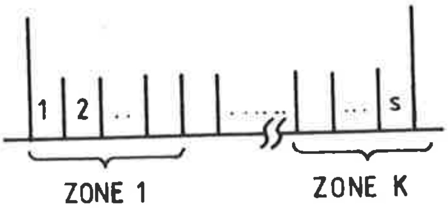


Fig.3.2.2 Reservation channel

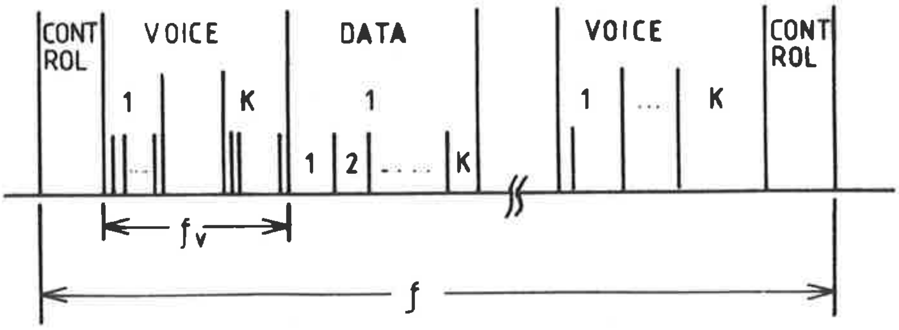


Fig.3.2.3 Frame structure of a Distributed Control DA-FA Integrated Traffic Protocol

determined by the traffic volume and the scheduling algorithm. The scheduling delay includes one round trip propagation delay as this is the minimum length of time in which a request can be relayed to every station in the zone. The delay structure is similar to Fig.3.0.2. The average delay per message E(D) and the variance var(D) are given by the following equations.

$$E(D) = E(D_R) + E(D_S) + rtpd \qquad (3.2.1)$$

$$var(D) = var(D_R) + var(D_S) \qquad (3.2.2)$$

3.3 DEMAND ASSIGNMENT UPLINK AND DOWNLINK (DA) PROTOCOL

The next logical extension of the multiple access protocol is to allow both the uplink and the downlink to be assigned on demand. The channel utilisation is increased since the entire capacity is statistically multiplexed amongst all the stations in the network. The penalties one pays for such an improvement are an increase in both the control overhead and the complexity of the control mechanism.

Due to the nature of the multiple spotbeam coverages, a station can only receive what is directed to it whereas distributed control requires a comprehensive knowledge of the instantaneous network capacity requirement. Hence, the approach of a central controller via a satellite control link is adopted. The controller assesses all requests and transmits the allocated transmission schedule to the requesting stations. This schedule is also used to co-ordinate the switching of the uplink and the downlink beams. The satellite is an ideal location for such a traffic controller, since it is the only node in the network which has a comprehensive knowledge of the traffic requirements. The alternative is to appoint a master terrestrial station and suffer an additional propagation delay.

3.3.1 Frame Structure of the DA Protocol

In a multiple spotbeam centrally controlled Demand Assignment (DA) Protocol, part of the capacity of the satellite link is allocated to the control mechanism. The frame structure of such a protocol consists of control and information sections (Fig.3.3.1). The control section (Fig.3.3.2) consists of

1. reservation subframe - the controller regularly assesses the capacity requirement of each station via the reservation channel. A TDMA Protocol is used by the subframe because of its simplicity in implementation and its regularity,
2. scheduling subframe - this is used by the controller to inform the stations of their allocated transmission schedule.

The network synchronization is maintained by the looped-back reservation burst and the timing dissemination by the controller. Each burst consists of a certain number (u) of capacity requests and a synchronization word. Again for reservation channel stability, u must be greater than the expected incoming requests.

The information section for a data network consists of N data slots. For an integrated data-voice traffic protocol with the voice traffic using circuit switching, the voice subframes are regularly interspersed with the data within the frame (Fig.3.3.3). The higher frame rate of voice subframes permits the utilisation of a small buffer and a long overall framelength.

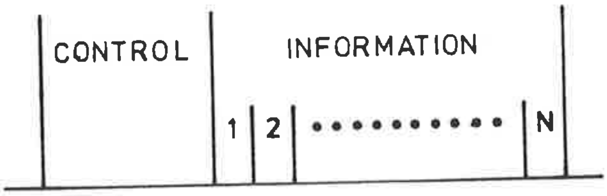


Fig.3.3.1 Frame structure of a Centrally Controlled DA Data Protocol

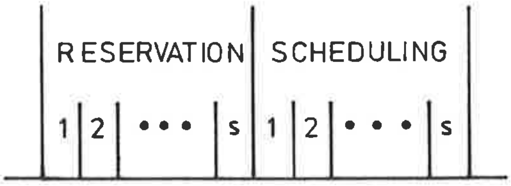


Fig.3.3.2 Control subframe

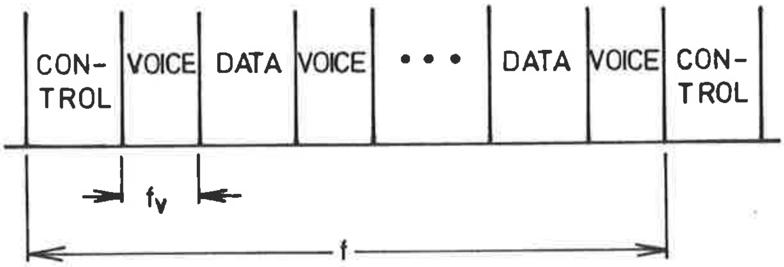


Fig.3.3.3 Frame structure of a Centrally Controlled DA Integrated Traffic Protocol

3.3.2 Delay Structure of the DA Protocol

When a message enters the station, it is temporarily stored into a queue of buffers. For a priority queue, the ordering of the messages may be shuffled. A capacity request is submitted for the message after a reservation delay $E(D_R)$. The controller receives the capacity requests from all the stations and serves them using a defined discipline (e.g. FCFS). The transmission schedule is sent via the scheduling subframe. The capacity allocation strategy and the channel congestion give rise to the scheduling delay. The scheduling delay includes one round trip propagation delay as it is the minimum length of time that a request and a schedule can be relayed between the station and the satellite. A connection is made when the message arrives at the satellite in synchronization with the on-board switching. The delay strucure is again similar to Fig.3.0.2, since D_R and D_S are statistically independent variables. The average delay per message E(D) and the variance var(D) are given by the following relations. one message per slot?

$$E(D) = E(D_R) + E(D_S) + rtpd \tag{3.3.1}$$

$$var(D) = var(D_R) + var(D_S) \tag{3.3.2}$$

4. DELAY ANALYSIS

With the Space Division Multiple Access protocols, two outstanding features are observed:-

1. the multiple reservation carrying facility, and

2. the large data capacity.

In order to accommodate both features, new methods have been developed for the delay calculations. All the required delay calculations for the Space Division Multiple Access protocols are presented in this chapter.

4.1 DATA DELAY ANALYSIS

For Demand Assignment (DA) and the hybrid Fixed Assignment uplink-Demand Assignment downlink (DA-FA) Protocol described, the structure of the message delay consists of two statistically independent variables :- the reservation delay and the scheduling delay. For the case of the Fixed Assignment (FA) Protocol, the delay consists of the scheduling delay only. The calculation of the average message delay for various Space Division Mulitple Access (SDMA) protocols requires different sets of parameters, but the calculation methods are common to all protocols. In this chapter, the average values of the two delay components are derived.

When a message enters a station, it is temporarily stored. A capacity request is submitted for each message after a reservation delay (D_R) which is defined as the duration between the message arrival and the instant that the request is sent. A delay can occur since only a fixed number (u) of reservation requests is carried by each reservation burst and the requests are served on a First Come First Serve (FCFS) discipline. The traffic controller receives the capacity reservation from the stations and serves them using a defined discipline. The allocation strategy and the channel congestion gives rise to the scheduling delay (D_S), which is defined as the duration between the sending of the capacity request and the beginning of the message transmission. In the case of the DA Protocol, the scheduling delay effectively includes one round trip propagation delay (rtpd), as this is the minimum length of time in which a request and a schedule can be relayed to and from the satellite. In the case of the DA-FA Protocol, one rtpd is the minimum length of time in which a request can be relayed to every station in the zone. The average data message delay $E(D_d)$ is given by

$$E(D_d) = E(D_R) + E(D_S) + \text{rtpd} \qquad (4.1.1)$$

The data arrival process and the duration of the scheduling delay have been discussed by [RUB 79]. A solution for the scheduling delay is derived by finding a set of complex roots for a transcendental equation. This technique is illustrated in Section 4.2.2. However, the complexity of the solution increases as the capacity increases. A simpler method, which is essentially independent of capacity, derived by Ko and Davis [KO 82], [KO 83] is presented in Section 4.2.3.

A new calculation method for the reservation delay, for the case of multiple request per reservation burst, is derived in Section 4.3. The facility of multiple requests is required for a long frame to ensure the stability of the reservation channel, i.e. the request transmission must be greater or equal to the request arrival. In Section 4.4, the result for a single terminal type is expanded to a multiple terminal type network.

In Section 4.1.1, a generalised message arrival process, which is used for the derivation of scheduling and reservation delay, is described. Simplifications of the delay calculation can be made if a compound Poisson arrival process with geometrical message length distribuation is assumed.

4.1.1 Data Arrival Process

Let A_n be the number of messages generated by the network in the n-th frame, and $\{A_n, n \geq 1\}$ is a sequence of independent identical distributed (i.i.d.) random variables governed by the relations :-

$$a_k = \Pr(A_n = k), \qquad k \geq 0$$

and $$\sum_{k=0}^{\infty} a_k = 1$$

The z-transform or the generating function of $\{a_k, k \geq 0\}$ is :

$$A(z) = \sum_{k=0}^{\infty} a_k z^k \qquad |z| \leq 1$$

Let $m(k) = \Pr(\text{i-th message length} = k \text{ slots})$,

and the z-transform of $\{m(k), k \geq 1\}$ is

$$M(z) = \sum_{k=1}^{\infty} m(k) z^k \qquad |z| \leq 1$$

Furthermore, let m_j be the j-th moment of the message length and $\tilde{A}_n$ the traffic (in number of slots) generated by the network in the n-th frame, where a slot consists of a fixed number of bits. $\{\tilde{A}_n, n \geq 1\}$ is an i.i.d. sequence governed by

$$\tilde{a}_k = \Pr(\tilde{A}_n = k)$$

$$= \sum_{i=1}^{k} m^{i*}(k)\, a_i \qquad k \geq 1$$

and $\tilde{a}_0 = a_0$

where $m^{i*}(k)$ = i-th fold self convolution of $m(k)$.

For the n-th frame

$$\sum_{k=0}^{\infty} \tilde{a}_k = 1$$

The z-transform of $\{\tilde{a}_k, k \geq 0\}$

$$\tilde{A}(z) = \sum_{k=0}^{\infty} \tilde{a}_k z^k$$

With the above definition of $\tilde{a}_k$, it can be shown that

$$\tilde{A}(z) = A[M(z)] \tag{4.1.2}$$

For the particular case of a Poisson message arrival process, $A(z)$ becomes

$$A(z) = \exp[-\lambda(1-z)] \tag{4.1.3}$$

where λ = average number of message arrivals for the network per frame.

If the message has a geometrical message length distribution, $M(z)$ becomes

$$M(z) = \frac{\theta z}{1 - \chi z} \tag{4.1.4}$$

where $1/\theta$ = average message length in number of slots

$$\chi = 1 - \theta \tag{4.1.5}$$

Under such circumstances, the packet arrival process is a compound Poisson input process with

$$\tilde{A}(z) = \exp\left[\frac{-\lambda(1-z)}{1 - \chi z}\right] \tag{4.1.6}$$

4.2 SCHEDULING DELAY

The data scheduling delay (D_S) is defined as the duration between the sending of the capacity request and the beginning of the message transmission. It is caused by the channel congestion and allocation strategy. Let the n-th reservation group (G_n) be the set of capacity reservations made within the n-th frame, the reservation which will serve first among a reservation group is designated as the group leader. The group size process $\{G_n, n \geq 1\}$ is a sequence of i.i.d. random variables with distribution $\{g_k\}$ given by

$$g_k = Pr(G_n = k) \qquad k \geq 1 \qquad n \geq 1$$

$$= \frac{a_k}{1 - a_o}$$

The number of slots contained in G_n is denoted by $\tilde{G}_n$. $\{\tilde{G}_n, n \geq 1\}$ is a sequence of i.i.d. random variables with the following distribution :-

$$\tilde{g}_k = Pr(\tilde{G}_n = k) \qquad k \geq 1 \qquad n \geq 1$$

$$= \frac{\tilde{a}_k}{1 - \tilde{a}_o}$$

The n-th message scheduling delay $\tilde{D}_n^S$, in number of slots, can be expressed as the sum of two random variables :-

$$\tilde{D}_n^S = \tilde{L}_n + \tilde{W}_n$$

where $\tilde{L}_n$ is the waiting time of the n-th group leader in number of slots; and $\tilde{W}_n$ is the waiting time of the message beyond its group leader in number of slots. $\tilde{W}_n$ is service discipline dependent, a random service discipline is assumed for the analysis.

The average scheduling delay - $E(\tilde{D}_S)$:-

$$E(\tilde{D}_S) = E(\tilde{L}) + E(\tilde{W}) \qquad (4.2.1)$$

The associated z-transforms are

$$\tilde{L}(z) = \sum_{n=0}^{\infty} \tilde{L}_n z^n$$

$$\tilde{W}(z) = \sum_{n=0}^{\infty} \tilde{W}_n z^n$$

$$\tilde{D}_S(z) = \sum_{n=0}^{\infty} \tilde{D}_n^S z^n$$

Since $\tilde{L}_n$ and $\tilde{W}_n$ are statistically independent variables, using the definitions of $\tilde{D}_n^S$ and $\tilde{D}_S(z)$,

$$\tilde{D}_S(z) = \tilde{L}(z)\,\tilde{W}(z) \qquad (4.2.2)$$

4.2.1 Derivation of $E(\tilde{W})$ and $\mathrm{var}(\tilde{W})$

Let us consider an i.i.d. sequence of group size $\{G_n, n\geq 1\}$ and the associated discrete time renewal point process $\{C_n, n\geq 1\}$,

$$C_n = \sum_{j=1}^{n} G_j \qquad n\geq 1$$

i.e. C_n represents the leader of the n-th group.

With random service within a group, the time elapsed between a randomly selected discrete time unit n and the preceding group leader (η_n) gives the number of messages preceding the n-th message in its own group. From renewal theory, the limiting distribution of η_n always exists ([COHJ 69]pp.114) and is identical with the corresponding stationary distribution of the renewal process η_n, $n\geq 1$.

$$\tilde{W}(z) = \eta[M(z)]$$

$$= \frac{1 - A[M(z)]}{E(A)[1 - M(z)]} \qquad |z| \leq 1 \qquad (4.2.3)$$

By repeatedly differentiating $\tilde{W}(z)$ and setting $z=1$, the mean and the variance can be obtained

$$E(\tilde{W}) = \frac{m_1}{2}\left[\frac{E(A^2)}{E(A)} - 1\right] \qquad (4.2.4)$$

if $E(A^3) < \infty$

$$\mathrm{var}(\tilde{W}) = \frac{1}{2}[m_2 - m_1^2]\left[\frac{E(A^2)}{E(A)} - 1\right] + m_1^2\left[\frac{E(A^3)}{3E(A)} - \frac{E(A^2)^2}{4E(A)^2} - \frac{1}{12}\right] \qquad (4.2.5)$$

where m is the j-th moment of the message length. As a particular case, the compound Poisson packet arrival process have a simplified mean and variance expression :-

$$E(\tilde{W}) = \frac{\lambda}{2\theta} \qquad (4.2.6)$$

$$\mathrm{var}(\tilde{W}) = \frac{\lambda}{2\theta^2}[2 - \theta + \lambda/6] \qquad (4.2.7)$$

4.2.2 Derivation of $E(\tilde{L})$ and $var(\tilde{L})$ by the Root Finding Method

The waiting time of the group leader $\tilde{L}_n$ is a Markov process governed by the recurrence relationship

$$\tilde{L}_{n+1} = [\tilde{L}_n + \tilde{W}_n - T_{n+1}]^+ \qquad n \geq 1 \qquad (4.2.8)$$

where $[x]^+ = \max(0,x)$

i.e. $\tilde{L}_{n+1}$ is the difference between $\tilde{L}_n + \tilde{W}_n$ and the group interarrival time $- T_{n+1}$. To evaluate the steady state distribution of $\tilde{L}_n$, let the queue size prior to the n-th frame be defined as X_n. $\{X_n, n \geq 1\}$ is governed by the recurrence relationship

$$X_{n+1} = [X_n + \tilde{A}_n - N]^+ \qquad n \geq 1 \qquad (4.2.9)$$

where N = number of slots in a frame

$\{\tilde{L}_n, n \geq 1\}$ corresponds to $\{X_n, n \geq 1\}$ at the beginning of the frame at which time some reservations are made. The distribution of the sampling process (t_n) is a geometrical distribution - a memoryless occurrence.

$$Pr(t_n = kN) = (1-a_o)a_o^{k-1} \qquad k \geq 1$$

Hence, $\tilde{L}_n$ and X_n have an identical steady state distributions from the Renewal Theorem ([COHJ 69]pp.102).

i.e. $\tilde{L}(z) = X(z)$ (4.2.10)

To solve (4.2.9), consider the sequence $\{B_n, n \geq 1\}$ - where B_n denotes the queue size after the arrival of the n-th reservation group.

i.e. $B_n = X_n + \tilde{A}_n$ (4.2.11)

With (4.2.9), B_n has a recurrence relationship

$$B_{n+1} = [B_n - N]^+ + \tilde{A}_{n+1} \qquad (4.2.12)$$

With (4.2.10),

$$\tilde{L}(z) = B(z)/\tilde{A}(z) \qquad (4.2.13)$$

To find $B(z)$, the following generating functions are defined. They are analytic in the region of $|z|<1$ and $|w|<1$.

$$B(z) = \sum_{k=0}^{\infty} Pr(B_n = k) z^k$$

$$B(z,w) = \sum_{n=0}^{\infty} B_n(z)w^n$$

Now (4.2.12) has a stationary transition equation

$$Pr(B_n=a|B_{n-1}=b) = Pr(\tilde{A}_n=a-[b-N]^+)$$

and $B_j(z)$ can be obtained by multiplying the above equation by $z^a Pr(B_{n-1}=b)$ and summing over the indices a and b. Again, B(z,w) can be obtained by multiplying $B_j(z)$ by w^j and sum over j.

$$B(z,w) = \frac{B_o(z)z^N + w\tilde{A}(z) \sum_{b=0}^{N-1} (z^N - z^b)c_b(w)}{z^N - w\tilde{A}(z)}$$

where $\quad c_b(w) = \sum_{j=0}^{\infty} Pr(B_j=b)w^j$

B(z,w) can be calculated [KON 72] as :

$$B(z,w) = \frac{B_o(z)z^N + w(z-1)\tilde{A}(z) \sum_{j=1}^{N} \frac{B_o[z_j(w)]}{1 - z_j(w)} \prod_{\substack{t=1 \\ t\neq j}}^{N} \frac{z - z_t(w)}{z_j(w) - z_t(w)}}{z^N - w\tilde{A}(z)}$$

With $\quad B(z) = \lim_{w\to 1} (1-w)B(z,w)$

$$B(z) = \frac{[N-E(\tilde{A})][1-z]\tilde{A}(z)}{\tilde{A}(z) - z^N} \prod_{r=1}^{N-1} \left[\frac{z-z_r}{1-z_r}\right] \tag{4.2.14}$$

With (4.2.13),

$$\tilde{L}(z) = \frac{[N-E(\tilde{A})][1-z]}{\tilde{A}(z) - z^N} \prod_{r=1}^{N-1} \left[\frac{z-z_r}{1-z_r}\right] \tag{4.2.15}$$

where z_r are roots of the equation

$$\tilde{A}(z) - z^N = 0$$

or

$$1-z^N \tilde{A}^{-1}(z) = 0 \tag{4.2.16}$$

The moments of $\tilde{L}(z)$ are easily obtained by differentiating (4.2.15),

$$E(\tilde{L}) = \frac{\tilde{A}^{(2)}(1) - N(N-1)}{2[N - E(\tilde{A})]} + \sum_{r=1}^{N-1} \frac{1}{1-z_r}$$

or
$$E(\tilde{L}) = \frac{\mathrm{var}(\tilde{A})}{2[N - E(\tilde{A})]} - \frac{E(\tilde{A})}{2} - \frac{N-1}{2} + \sum_{r=1}^{N-1} \frac{1}{1-z_r} \qquad (4.2.17)$$

$$\mathrm{var}(\tilde{L}) = \frac{1}{12} + \frac{E(\tilde{A}^3) - N^3}{3[N - E(\tilde{A})]} + \left\{\frac{E(\tilde{A}^2) - N^2}{2[N - E(\tilde{A})]}\right\}^2 - \sum_{r=1}^{N-1} \frac{z_r}{(1-z_r)^2} \qquad (4.2.18)$$

One can obtain the average scheduling delay if the complex roots for (4.2.16) are known. They can be calculated by the Newton-Raphson Method, but the number of calculations increases as the number of data slots in a frame (N) increases. A new and simpler method which is essentially independent to the value of N is presented in the next section.

4.2.3 Derivation of $E(\tilde{L})$ and $\mathrm{var}(\tilde{L})$ by the Integration Method

Equation (4.2.14) can be rewritten as

$$B(z) = \frac{[N - E(\tilde{A})][1-z]}{1-z^N\tilde{A}^{-1}(z)} \prod_{r=1}^{N-1} \left[\frac{z-z_r}{1-z_r}\right] \qquad (4.2.19)$$

Since B(z) is regular and bounded within the unit circle, $|z|<1$. (4.2.19) can be expressed as

$$B(z) = \frac{\sum_{j=0}^{N-1} b_j\,(z^j - z^N)}{1-z^N\tilde{A}^{-1}(z)}$$

$$= \frac{-z^{-N}\tilde{A}(z) \sum_{j=0}^{N-1} b_j\,(z^j - z^N)}{1-z^{-N}\tilde{A}(z)}$$

or
$$\frac{B(z)}{\tilde{A}(z)} = \frac{-z^{-N} \sum\limits_{j=0}^{N-1} b_j \, (z^j - z^N)}{1-z^{-N}\tilde{A}(z)} \triangleq \frac{\gamma(z)}{\phi(z)} \tag{4.2.20}$$

From Appendix 1,

$$\ln B(z) = \ln \tilde{A}(z) + \ln \sum_{j=0}^{N-1} b_j - \frac{1}{j2\pi} \int_{\Gamma} \frac{\ln \phi(x)}{x-z} \, dx \tag{4.2.21}$$

where Γ is a circle of radius $1+\varepsilon$ and $|z| \leq 1$.

Differentiate (4.2.21) and set z=1, (see Appendix 1)

$$E(B) = B^{(1)}(1) = \tilde{A}^{(1)}(1) - \frac{1}{j2\pi} \int_{\Gamma} \frac{\ln \phi(x)}{(x-1)^2} \, dx$$

knowing $E(\tilde{L}) = E(B) - E(\tilde{A})$

$$E(\tilde{L}) = -\frac{\psi^{(1)}(1)}{\psi(1)} - \frac{1}{\pi} \int_0^{\pi} G(\rho) \, d\rho \tag{4.2.22}$$

where
$$G(\rho) = \frac{-1}{2(1-\cos\rho)} \ln\left\{\frac{1}{2\psi(1)}\left[\frac{2(m^2+n^2)}{1-\cos\rho}\right]^{\frac{1}{2}}\right\} - \frac{\psi^{(1)}(1)}{2\psi(1)} \qquad \rho \neq 0$$

$$= \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{2\psi(1)^2} \qquad \rho = 0 \tag{4.2.23}$$

$$m+jn = \phi(e^{j\rho}) = 1 - e^{-jN\rho}\tilde{A}(e^{j\rho}) \tag{4.2.24}$$

The values of ψ are listed in Appendix 1, by repeatedly differentiating (4.2.21), and using

$$\text{var}(\tilde{L}) = \text{var}(B) - \text{var}(\tilde{A})$$

$$\text{var}(\tilde{L}) = E(\tilde{L}) - \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{2\psi(1)^2} - \frac{2}{\pi} \int_0^{\pi} H(\rho) \, d\rho \tag{4.2.25}$$

where $$H(\rho) = \mathrm{Re}\left\{\left[\frac{1 - \cos\rho + j\sin\rho}{4(1 - \cos\rho)^2}\right]\left[\ln\left\{\frac{m + jn}{2\psi(1)}\left[1 - \frac{j\sin\rho}{1-\cos\rho}\right]\right\}\right]\right\}$$

$$+ \frac{\psi^{(1)}(1)}{2\psi(1)(1-\cos\rho)} - \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{4\psi(1)^2} \qquad \rho\neq 0$$

(4.2.26)

$$= [\psi^{(3)}(1)\psi(1)^2 - 3\psi^{(2)}(1)\psi^{(1)}(1)\psi(1) + 2\psi^{(1)}(1)^3]/[6\psi(1)^3]$$

$$\rho=0$$

The value of $E(\tilde{L})$ and $\mathrm{var}(\tilde{L})$ can be calculated for an arbitrary compound arrival process $\tilde{A}(z)$. For the specific case of compound Poisson process with geometrical message length, the value of m^2+n^2 in (4.2.23) and $m+jn$ in (4.2.26) can be simplified as followed:-

$$m+jn = 1 - e^{X}\cos Y - je^{X}\sin Y \qquad (4.2.27)$$

$$m^2+n^2 = 1 - 2e^{X}\cos Y + e^{2X} \qquad (4.2.28)$$

where $$X = -\lambda + \frac{\lambda\theta(\cos\rho - \chi)}{1 + \chi^2 - 2\chi\cos\rho} \qquad (4.2.29a)$$

$$Y = -N\rho + \frac{\lambda\theta\sin\rho}{1 + \chi^2 - 2\chi\cos\rho} \qquad (4.2.29b)$$

where λ, θ and χ are as defined in Section 4.1.1. The integration in (4.2.25) can be done by one of the conventional numerical integration algorithms. Since the singular parts of the integral have been removed, this integration is straight forward.

4.2.4 Calculation of $E(D_S)$ using a M/G/1 Queueing Model

Further simlification for the calculation of the average scheduling delay $E(D_S)$ is possible, if $\tilde{A}(z)$ is a compound Poisson arrival process and the channel is assumed to be continuously available with a reduced capacity of N slots/frame. Using a M/G/1 model, the average scheduling delay is estimated from the Pollaczek-Khintchine Formula [SYS 58].

$$E(D_S) = \left[m_1 (1+c^2) \frac{S_d}{2(1-S_d)} \right] \frac{f}{N} \qquad (s)$$

where $c = \sigma/m_1$ coefficient of variation

$\sigma^2 = m_2 - m_1^2$ variance

$S_d = \lambda/(\theta N)$ data channel throughput

$$E(D_S) = \left[\frac{\lambda/\theta}{2(N - \lambda/\theta)} \frac{m_2}{m_1} \right] \frac{f}{N} \qquad (s) \qquad (4.2.30)$$

For a geometrical message length distribution, the moments become

$m_1 = 1/\theta$ (slots)

$m_2 = (2-\theta)/\theta^2$ (slots^2)

$$E(D_S) = \left\{ \frac{\lambda/\theta}{2(N - \lambda/\theta)} \left[\frac{2-\theta}{\theta} \right] \right\} \frac{f}{N} \qquad (s) \qquad (4.2.31)$$

The above simpler relation can only apply to a compound Poisson arrival process, while the Root Finding or Integration Methods find the average and the variance of the scheduling delay for any arbitrary arrival process. However, as shown in Chapter 7, simulation results indicate that Pollaczek-Khintchine Approximation is a poor estimte of the average scheduling delay due to the adopted assumptions.

4.2.5 Conclusion for the Scheduling Delay

The scheduling delay is composed of two terms :-

1. $\tilde{L}_n$ is the waiting time of the group leader, the leader is the request which will be served first amongst all requests from the same frame.

2. $\tilde{W}_n$ is the waiting time of the requests beyond its group leader.

Since the two terms are statistically independent, the average and the variance of the scheduling delay $E(D_S)$ are proportional to the sum of the two respective averages and variances. In the case of the Fixed Assignment (FA) Protocol, $E(D_S)$ is the sum of the two averages. In the case of the Demand Assignment (DA) Protocol, $E(D_S)$ includes one round trip propagation delay (rtpd), as it is the minimum length of time that a request and a schedule can be relayed to and from the satellite. In the case of the hybrid Demand Assignment uplink-Fixed Assignment Downlink (DA-FA) Protocol, one rtpd is the minimum length of time that a request can be relayed to every station in the zone. Hence, it is included in the $E(D_S)$. For all protocols, the variance of the scheduling delay is proportional to the sum of the two variances.

FA : $E(D_S) = \{E(\tilde{L}) + E(\tilde{W})\}f/N$ (s) (4.2.32)

DA,DA-FA : $E(D_S) = \{E(\tilde{L}) + E(\tilde{W})\}f/N + \text{rtpd}$ (s) (4.2.33)

DA,DA-FA,FA : $\text{var}(D_S) = \{\text{var}(\tilde{L}) + \text{var}(\tilde{W})\}f^2/N^2$ (s^2) (4.2.34)

The value of $E(\tilde{W})$, $\text{var}(\tilde{W})$, $E(\tilde{L})$ and $\text{var}(\tilde{L})$ for a generalised arrival process are given by (4.2.4), (4.2.5), (4.2.17) and (4.2.18) respectively with the latter two using the Root Finding Method. The complexity of this calculation is depending on the number of slots per frame (N). The Integration Method is essentially independent of N. (4.2.22) and (4.2.25) give $E(\tilde{L})$ and $\text{var}(\tilde{L})$ respectively for any arbitrary arrival process, while (4.2.27) to (4.2.29) give the simplification for a compound Poisson arrival process with geometrical message length distribution.

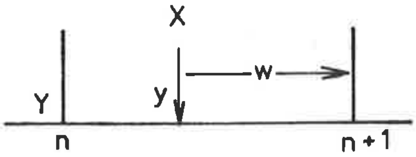
4.3 RESERVATION DELAY

The reservation delay (D_R) is the duration between the arrival of a message, or a notification of an impending voice traffic call, and the instant at which a reservation request is sent for that message or call. Since the request is sent on a First Come First Served (FCFS) basis, and a fixed number (u) of control packets carried by each control burst is transmitted by each station. A waiting time is imposed on each message before a request is sent.

The reservation delay is a function of the number of requests which have arrived before the arrival under consideration (point X in Fig.4.3.1) and the number of requests carried by each reservation burst (u). The total number of previously arrived messages consists of two components :-

1. Y - the number of message requests leftover from previous frames, i.e. the number of requests at the beginning of the n-th frame.

2. y - the number of message requests that have arrived during the current frame prior to the arrival under consideration, i.e. the number of requests that have arrived within the n-th frame and prior to point X.



X - ARRIVAL UNDER CONSIDERATION

Y - NUMBER OF MESSAGES LEFTOVER FROM PREVIOUS FRAMES

y - NUMBER OF ARRIVALS PRIOR TO X IN THE CURRENT FRAME

Fig.4.3.1 Reservation delay

4.3.1 Derivation of E(Y) and var(Y)

Consider the number of leftover requests at the beginning of the n+1 frame (Y_{n+1}), a recurrence relationship can be established :-

$$Y_{n+1} = [Y_n + A_{ni} - u]^+ \qquad (4.3.3)$$

where $[x]^+ = \max(0,x)$

A_{ni} = number of arrivals during the n-th frame into station i

u = number of control packets carried by each control slot

That is, the number of requests at the beginning of the frame consists of the leftover requests and arrivals from the previous frames minus the number of requests sent. The recurrence relationship of (4.3.3) is similar to the expression of (4.2.9). From Section 4.2, using suitable substitutions

$$E(z^Y) = Y(z) = \frac{[u-E(A_i)][1-z]}{A_i(z)-z^u} \prod_{r=1}^{u-1} \left[\frac{z-z_r}{1-z_r}\right] \qquad (4.3.4)$$

Hence, the solution can be derived from the Root Finding Method or Integration Method as discussed in Section 4.2. With the Root Finding Method :-

$$E(Y) = \frac{var(A_i)}{2[u-E(A_i)]} - \frac{E(A_i)}{2} - \frac{u-1}{2} + \sum_{r=1}^{u-1} \frac{1}{1-z_r} \qquad (4.3.5)$$

$$var(Y) = \frac{1}{12} + \frac{E(A_i^3) - u^3}{3[u-E(A_i)]} + \left\{\frac{E(A_i^2) - u^2}{2[u-E(A_i)]}\right\}^2 - \sum_{r=1}^{u-1} \frac{z_r}{(1-z_r)^2}$$

(4.3.6)

where z_r are the roots for the equation

$$1 - z^u A_i^{-1}(z) = 0 \qquad (4.3.7)$$

and $A_i(z)$ is the z-transform of the message arrival process into station i. For a Poisson message arrival process,

$$A_i(z) = \exp -[\lambda_i(1-z)]$$

where λ_i = average mesage arrival rate into station i

With the Root Finding Method, E(Y) and var(Y) becomes

$$E(Y) = \frac{\lambda_i}{2(u - \lambda_i)} - \frac{\lambda_i}{2} - \frac{u-1}{2} + \sum_{r=1}^{u-1} \frac{1}{1-z_r} \tag{4.3.8}$$

$$var(Y) = \frac{1}{12} + \frac{\lambda_i^3 + 3\lambda_i^2 + \lambda_i - u^3}{3(u - \lambda_i)} + \left[\frac{\lambda_i - (u^2-\lambda_i^2)}{2(u - \lambda_i)}\right]^2 -$$

$$\sum_{r=1}^{u-1} \frac{z_r}{(1 - z_r)^2} \tag{4.3.9}$$

where z_r are the roots for

$$1 - z^u \exp[\lambda_i(1-z)] = 0 \tag{4.3.10}$$

With the uniform traffic assumption, the average number of message arrivals into station i (λ_i) is inversely proportional to the number of stations in the network.

$$\lambda_i = \lambda/s \tag{4.3.11}$$

For the simplest case of u=1, i.e. one reservation request per frame, the mean and the variance of the leftover request become :-

$$E(Y) = \frac{\lambda_i^2}{2(1 - \lambda_i)} \tag{4.3.12}$$

$$var(Y) = \frac{- \lambda_i^4 - 2\lambda_i^3 + 6\lambda_i^2}{12(1 - \lambda_i)^2} \tag{4.3.13}$$

4.3.2 Derivation of $E(D_R)$ and $var(D_R)$

The reservation delay D_R is a function of :-

1. Y - the leftover requests from the previous frame,

2. y - requests arrived within the current frame,

3. u - maximum number of requests carried by each reservation burst.

The reservation delay can be expressed in two different components,

$$D_R = r + w$$

$$= \left[\frac{y+Y}{u}\right] + w$$

The probability density function $p(D_R|w,r)$:-

$$p(D_R|w,r) = \Delta(D_R - w - r)$$

or $$p(D_R|w) = \sum_{n=0}^{\infty} Pr(r=n)\Delta(D_R-w-n)$$

$$= \sum_{n=0}^{\infty} \int_0^{\infty} Pr(r=n)\Delta(D_R-w-r)dr$$

Taking the Laplace Transform of the above equation,

$$\hat{p}(D_R|w)(s) = \sum_{n=0}^{\infty} Pr(r=n)exp-(n+w)s$$

Knowing

$$\hat{p}(D_R|w)(s) = 1-E(D_R|w)s+E(D_R^2|w)s^2/2!-\ldots$$

$$E(D_R|w) = \sum_{n=0}^{\infty} (n+w)Pr(r=n)$$

$$=E(r|w) + w \qquad (4.3.14)$$

$$E(D_R^2|w) = \sum_{n=0}^{\infty} (n+w)^2 Pr(r=n)$$

$$= E(r^2|w) + 2E(r|w) + w^2 \qquad (4.3.15)$$

Hence $$E(D_R) = \int_0^1 E(D_R|w)dw \qquad (4.3.16)$$

$$E(D_R^2) = \int_0^1 E(D_R^2|w)dw \qquad (4.3.17)$$

$$var(D_R) = E(D_R^2) - E(D_R)^2 \qquad (4.3.18)$$

A. $u=1$

If each reservation burst can only carry one request, i.e.u=1.

$$r = [y+Y]$$

$$= y+Y$$

The value of r is only a summation of y and Y. For an arbitrary arrival process, the average and the variance of the reservation delay is the summation of the two respective components. For a simple Poisson message arrival process

$$E(y) = \lambda_i(1-w)$$

$$E(y^2) = \lambda_i(1-w) + \lambda_i^2(1-w)^2$$

From Section 4.2 :-

$$E(Y) = \frac{\lambda_i^2}{2(1-\lambda_i)}$$

$$E(Y^2) = \frac{\lambda_i^4 - \lambda_i^3 + 3\lambda_i}{6(1-\lambda_i)^2}$$

after summation and integration,

$$E(D_R) = \frac{1}{2(1-\lambda_i)} \qquad (4.3.19)$$

$$var(D_R) = \frac{1+\lambda_i}{12(1-\lambda_i)^2} \qquad (4.3.20)$$

B. $u>1$

If the number of requests carried by each reservation burst is greater than one, r is no longer a simple relation. In fact,

$$r = \left\lceil \frac{y+Y}{u} \right\rceil$$

$$= (y+Y-i)/u \qquad \text{for a value of } i$$

$$i=0,1,2,\ldots,u-1$$

With a fixed value of Y,

$$E(r|w) = \{E(y|W)+Y-E(i|w)\}/u$$

and $E(y|w) = \lambda_i(1-w)$

$$\begin{aligned} E(i|w) = {} & 0.\{Pr(-Y)+Pr(u-Y)+Pr(2u-Y)+\ldots\ldots\ldots\} \\ & +1.\{Pr(1-Y)+Pr(u+1-Y)+Pr(2u+1-Y)+\ldots\ldots\} \\ & \quad \cdot \\ & \quad \cdot \\ & +(u-1).\{Pr(u-1-Y)+Pr(2u-1-Y)+Pr(3u-1-Y)+\ldots\} \end{aligned}$$

$$= \sum_{i=0}^{u-1} i \sum_{m=0}^{u-1} Pr(i-Y-mu)u(i-Y-mu)$$

$$\text{where } u(x)=0 \qquad x\leq 0$$

$$=1 \qquad x>1$$

For a simple Poisson arrival process, $E(i|w)$ can be written in a finite series :-

Let $\alpha = \exp(j2\pi/u)$ (4.3.21)

Knowing

$$1+\alpha^K+\alpha^{2K}+\ldots+\alpha^{(u-1)K}=u \qquad \text{if } K=0,u,2u,\ldots$$

$$=0 \qquad \text{otherwise}$$

$$\exp(x) \sum_{m=0}^{\infty} Pr(i-Y-mu) = x^{i-Y}/(i-Y)! + x^{i-Y+u}/(i-Y+u)! + \ldots$$

Hence $$E(i|w) = \frac{1}{u}\exp(-x)\sum_{i=0}^{u-1} i \sum_{s=0}^{u-1}\exp(\alpha^s x)\alpha^{-s(i-Y)}$$

$$= \frac{1}{u}\exp(-x)\sum_{s=1}^{u-1}\frac{\exp(\alpha^s x)}{\alpha^{s(u-Y-1)}}\left\{\frac{\alpha^{us}-1}{(\alpha^s-1)^2} - \frac{u}{\alpha^s-1}\right\} + \frac{u-1}{2}$$

$$= \frac{u-1}{2} - \exp(-x)\sum_{s=1}^{u-1}\frac{\exp(\alpha^s x)\alpha^{s(Y+1)}}{\alpha^s-1}$$

where $x=\lambda_i(1-w)$

$$E(i) = \int_0^1 E(i|w)dw$$

$$= \frac{u-1}{2} - \frac{1}{\lambda_i}\sum_{s=1}^{u-1}\frac{[\exp\lambda_i(\alpha^s-1) - 1]\alpha^{s(Y+1)}}{(\alpha^s-1)^2}$$

$$E(D_R) = \int_0^1 w + E(r|W)\ dw$$

$$= \frac{1}{2} + \frac{1}{u}\left\{\frac{\lambda_i}{2} + Y - \frac{u-1}{2} + \frac{1}{\lambda_i}\sum_{s=1}^{u-1}\frac{[\exp\lambda_i(\alpha^s-1) - 1]\alpha^{s(Y+1)}}{(\alpha^s-1)^2}\right\}$$

Since E(r) is always positive, $E(D_R)>\frac{1}{2}$, or the average reservation delay has a minimum value of half a framelength.

Next, averaging $E(D_R)$ over the range of Y. From (4.3.4), with simple Poisson arrival process :-

$$E(z^Y) = Y(z) = \frac{(u-\lambda_i)(1-z)}{\exp[-\lambda_i(1-z)] - z^u}\prod_{r=1}^{u-1}\left[\frac{z - z_r}{1 - z_r}\right]$$

where z_r are the roots for

$$z^u - \exp{-\lambda_i(1-z)} = 0 \qquad (4.3.22)$$

inside the unit circle of z.

$$E(Y) = -\frac{u(u-1) - \lambda_i^2}{2(u-\lambda_i)} + \sum_{r=1}^{u-1}\frac{1}{1-z_r}$$

$$E(\alpha^{sY}) = Y(\alpha^s) = \frac{(u - \lambda_i)(\alpha^s - 1)}{1 - \exp\lambda_i(\alpha^s - 1)} \prod_{r=1}^{u-1} \frac{\alpha^s - z_r}{1 - z_r}$$

$$E(D_R) = \frac{1}{2} + \frac{1}{u}\left\{\frac{\lambda_i}{2} + \sum_{r=1}^{u-1} \frac{1}{1 - z_r} - \frac{u(u - 1) - \lambda_i^2}{2(u - \lambda_i)} - \frac{u-1}{2} - \right.$$

$$\left. \frac{u - \lambda_i}{\lambda_i} \sum_{s=1}^{u-1} \frac{\alpha^s}{\alpha^s - 1} \prod_{r=1}^{u-1} \frac{\alpha^s - z_r}{1 - z_r}\right\} \text{ (frame)} \qquad (4.3.23)$$

4.3.3 Conclusion for the Reservation Delay

The reservation delay D_R is a function of :-

1. Y - the number of requests leftover from the previous frame,
2. y - the number of requests arriving in the current frame,
3. u - the number of requests carried by each burst.

Except for the simplest case of u=1, D_R is not a linear relation of Y and y. For the case of u=1, the mean and variance of any arbitrary arrival process is simply a summation of the respective means and variances of the two components. However, for the case of u>1, and an arbitrary arrival process, the mean and the variance can only be expressed as an infinite series. For a simple Poisson message arrival process, the mean can be written as a finite series in terms of the complex roots of a transcendental equation.

4.4 EXPANSION TO A MULTIPLE TERMINAL TYPE NETWORK

Up until now, the delay analysis for a single terminal type network has been presented. For a multiple terminal type network, the total probability of k arrivals into the network :-

$$\tilde{a}_{k\ tot} = Pr(\tilde{A}_{n\ tot} = k)$$

$$= \left(\sum_{i=1}^{k} m_1^{i*}(k)a_{i1}\right)*\left(\sum_{i=1}^{k} m_2^{i*}(k)a_{i2}\right)*\ldots$$

$$\ldots*\left(\sum_{i=1}^{k} m_c^{i*}(k)a_{ic}\right) \qquad (4.4.1)$$

i.e. $\tilde{a}_{k\ tot}$ is a convolution of arrivals from each of the c terminal-types. The z-transform of $\tilde{a}_{k\ tot}$ is simply the multiplication of each of the c individual z-transforms.

$$\tilde{A}_{tot}(z) = \tilde{A}_1(z)\tilde{A}_2(z)\ldots\tilde{A}_c(z) \qquad (4.4.2)$$

Hence, all the moments of a multiple terminal type network arrival can be derived from (4.4.2).

In the particular case of the multiple compound Poisson input process, the arrival process of the j-th terminal type is

$$\tilde{A}_j(z) = \exp\{-\lambda_j[1-M_j(z)]\}$$

$$= f(\lambda_j;M_j(z))$$

From (4.4.2),

$$\tilde{A}_{tot}(z) = \exp\{-\lambda_1[1-M_1(z)]\}\exp\{-\lambda_2[1-M_2(z)]\}\ldots\exp\{-\lambda_c[1-M_c(z)]\}$$

$$= f\left\{\sum_j \lambda_j;\ \frac{\sum_j \lambda_j M_j(z)}{\sum_j \lambda_j}\right\} \qquad (4.4.3)$$

Therefore, the convolution of the multiple compound Poisson arrival process is another compound Poisson arrival process with modified rate and average message length. Hence, all the analysis for the single terminal type network is still valid for a multiple terminal type network with a compound Poisson arrival process.

4.5 VOICE CONNECTION DELAY

The voice connection delay (D_v) is the duration between the arrival of a voice circuit request into the station and the instant at which a connection is made with the calling station. It consists of three statistically independent components :-

1. The reservation delay (D_R) is the duration between the arrival of the voice circuit request into a station and the instant that the request is sent. Since voice and data requests are sent by the station in a First Come First Served (FCFS) basis, the average reservation delay $E(D_R)$ has been discussed in Section 4.3.

2. The scheduling delay (D_{vS}) is the duration between the transmisson of the voice circuit request and the beginning of the call. One of the g duplex voice circuits is allocated to the request if a vacancy exists. If all circuits are busy, the overflow requests are stored in one of the L request buffers and the channel is connected when a vacancy arises, i.e. a blocked-customer-delay model. The average scheduling delay $E(D_{vS})$ is given in the next section.

3. The first voice packet transmisson delay of one round trip propagation delay (rtpd).

The average voice connection delay $E(D_v)$ is given by

$$E(D_v) = E(D_R) + E(D_{vS}) + \text{rtpd}$$

4.5.1 Voice Scheduling Delay

The voice scheduling delay is defined as the duration between the sending of the voice circuit request and the beginning of the call. If g is the number of voice circuits available and L the number of voice circuit request buffers on-board the traffic controller, assume the call has a Poisson arrival process with an average network arrival of $v\omega$ calls per second and a negative exponential holding time distribution with an average of $1/\mu$ seconds. The average voice scheduling delay $E(D_{vS})$ in seconds is given by the delay component of the M/M/g/L delay-loss queueing model [TEL 78] :-

$$E(D_{vS}) = \frac{\rho[1 - (L+1)\rho^{L} + L\rho^{L+1}]/\omega}{(1-\rho)^{2}\ E_{g}^{-1}(A) + \rho(1-\rho)(1-\rho^{L})} \qquad \text{(s)}$$

where $E_g(A) = \dfrac{A^g/g!}{\sum_{j=0}^{g} A^j/j!}$ Erlang Loss Formula

$A = v\omega/\mu$ offered traffic

$\rho = A/g$ offered traffic per voice circuit

v=total number of voice terminals in the network

ω=average call generation rate per terminal

4.6 CONCLUDING REMARKS

In this chapter, the average data message delay and the average voice connection delay are derived for a First Come First Served discipline. The message delay consists of the sum of reservation, scheduling delay and one round trip propagation delay. In turn, the scheduling delay is proportional to the sum of the group leader waiting time and waiting time beyond its group leader. The latter quantity has a simple analytical solution. Two exact methods are presented for the solution of the group leader waiting time. The first is the Root Finding Method, the complexity of which is dependent on the number of data slots available per frame or data capacity, while the new Integration Method based on a single integration and essentially independent of the data capacity is illustrated.

A new approach to the reservation delay for a multiple request carrying burst is derived. It is a function of the leftover message requests, the message requests arriving during the current frame and the maximum number of requests carried by each reservation burst. The former quantity has a similar recurrence relation to the group leader waiting time of the scheduling delay. Therefore, its solution can be used. With an arbitrary message arrival process, the average and the variance of the reservation delay can only be expressed as an infinite series. However, if the message arrival process assumes a simple Poisson distribution, the average reservation delay can be written as a finite series.

To accommodate a multiple terminal type network, the probability of a particular arrival from the network is the convolution of the probabilities of arrival from each terminal type in the network. The transform of the total arrival process is simply the product of the transforms of the individual arrival processes. For a multiple compound Poisson input process, the total arrival process is another compound Poisson process with modified average arrival rate and message length generating function. Hence, the analysis for a single terminal type network is still valid for a multiple terminal type network.

Finally, for the voice traffic, the voice connection delay consists of the reservation delay and the voice circuit scheduling delay. The calculation method for the reservation delay is similar to the one outlined for the data traffic. In the case of integrated data and voice traffic, the average capacity request arrival rate is the sum of the average data and voice arrival rates. The average voice circuit scheduling delay can be obtained as the delay of a Markovian delay-loss queueing model.

5. BUFFER ANALYSIS

To calculate the buffer requirements of the Space Division Multiple Access (SDMA) protocols, some of its characteristics have to be examined, and specialised methods are used to overcome the associated peculiarities of the SDMA protocols.

In the case of voice traffic operating in the circuit switching mode, the well known solutions of blocking probabilities which use the assumptions of Markovian input process and negative exponential holding time are used. However, the non-constant output nature of the demand assignment protocols calls for the adoption of new techniques. In addition, a new model is developed for the calculation of buffer blocking probability for an acknowledgement system.

5.1 BUFFERING REQUIREMENTS AND BLOCKING PROBABILITIES

The functional block diagrams of the on-board controller and the earth station with and without message acknowledgement for the Fixed Assignment (FA), Demand Assignment (DA) and the hybrid Demand Assignment uplink-Fixed Assignment downlink (DA-FA) Protocols are shown in the following series of diagrams (Fig.5.1.1 to 5.1.10). The voice call request is serviced by one of the vacant duplex voice circuits, or stored in one of the request buffers until a vacancy arises. The blocking probability of the voice request buffers is discussed in Section 5.2.1 and given by the loss component of the M/M/g/L queueing model. The Engset Loss Formula is used for the calculation of the sending and the receiving station blocking probabilities for the voice traffic in Section 5.2.2. For a system

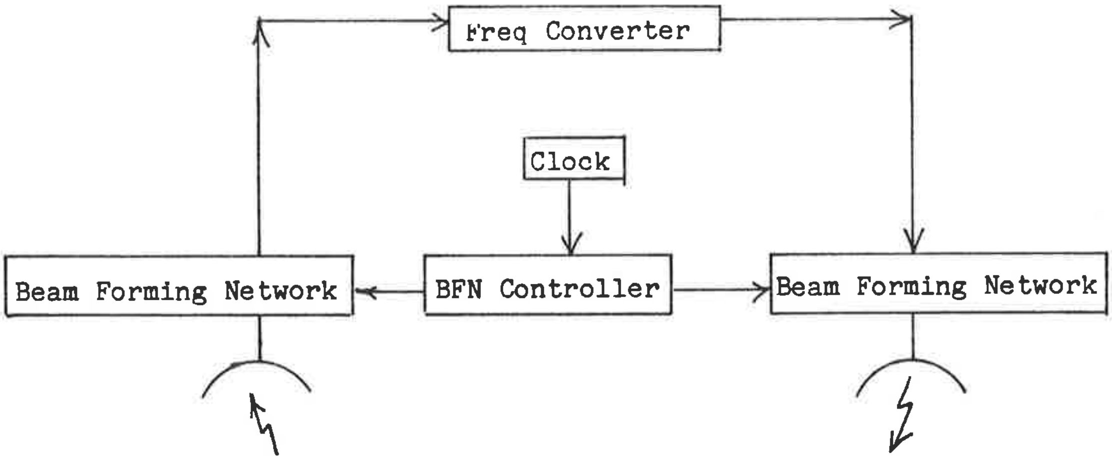


Fig.5.1.1 On-board controller functional block diagram for integrated traffic using the FA Protocol

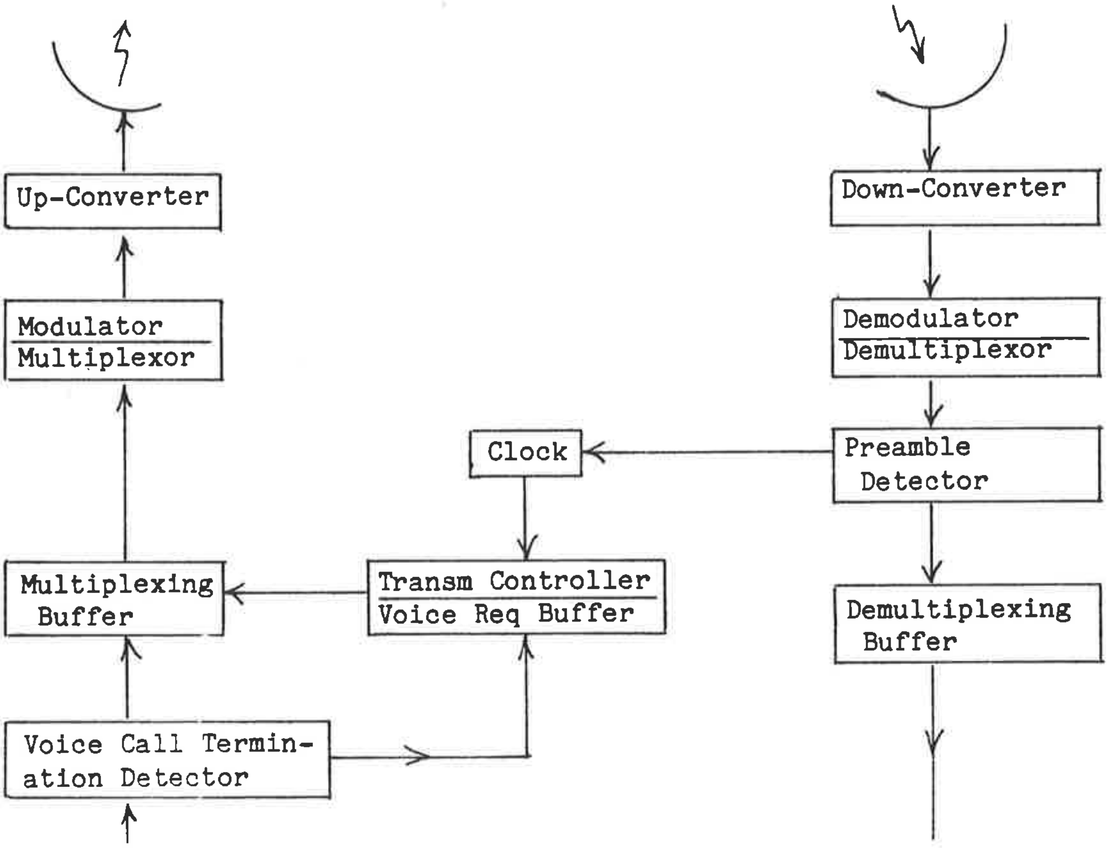


Fig.5.1.2 Earth station functional block diagram for integrated traffic using the FA Protocol

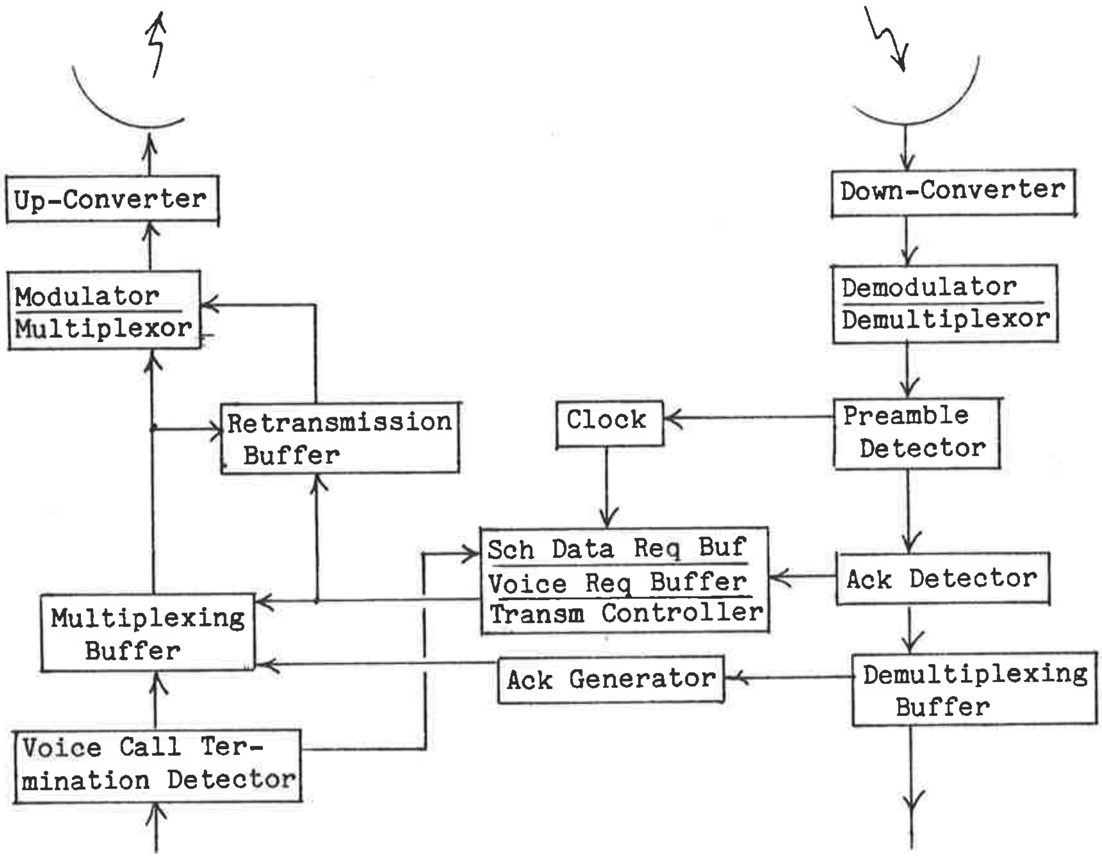


Fig.5.1.3 Earth station functional block diagram for integrated traffic using the Acknowledgement FA Protocol

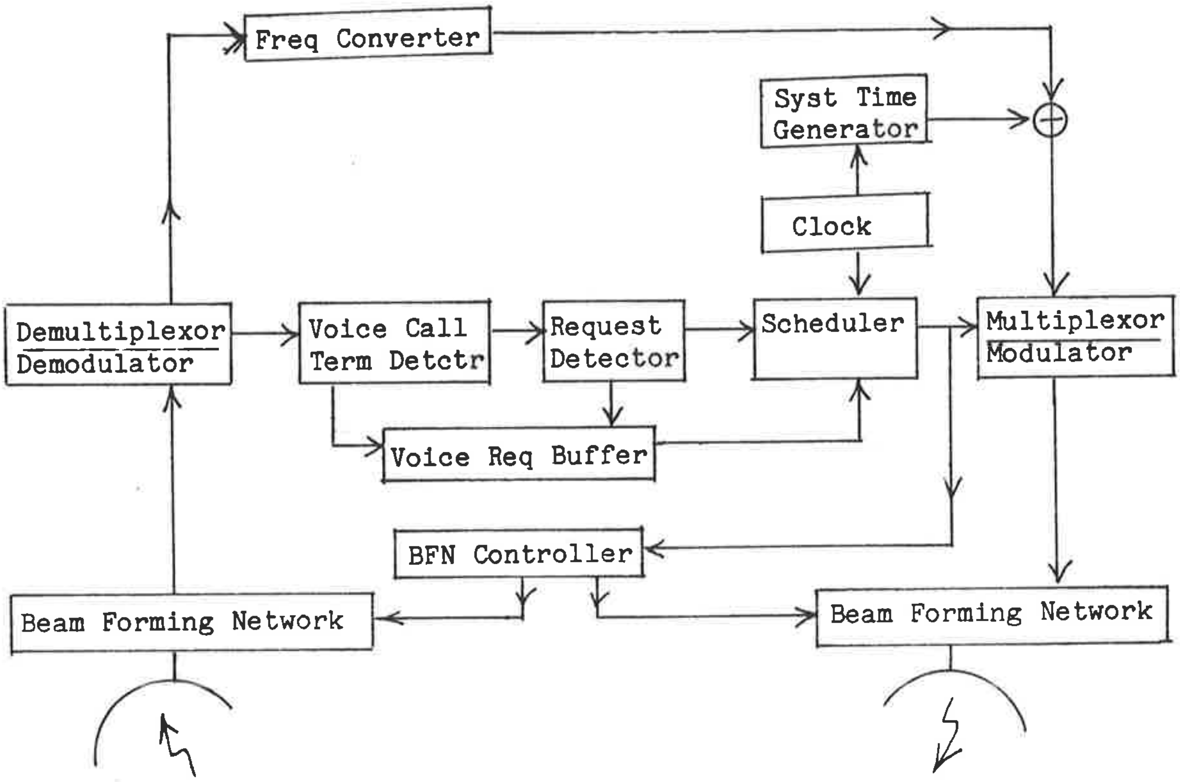


Fig.5.1.4 On-board controller functional block diagram for integrated traffic using the DA Protocol

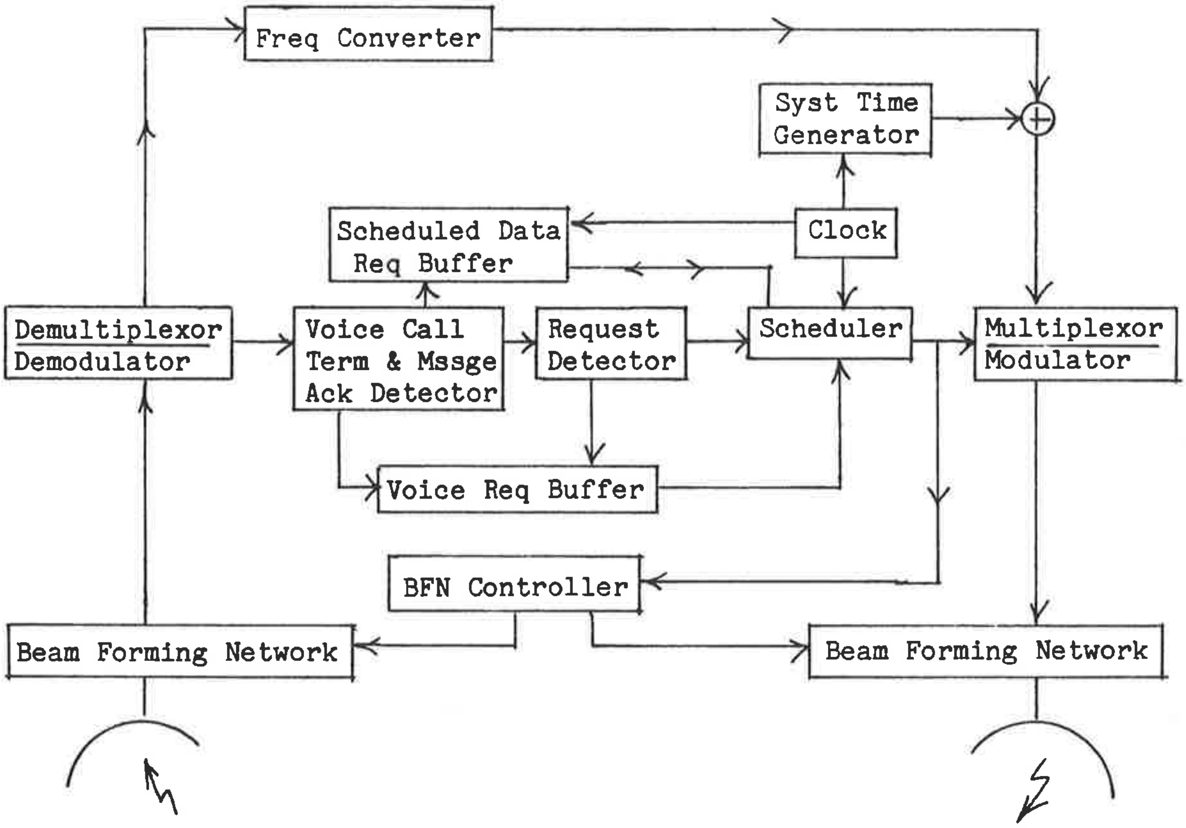


Fig.5.1.5 On-board controller functional block diagram for integrated traffic using the Acknowledgement DA Protocol

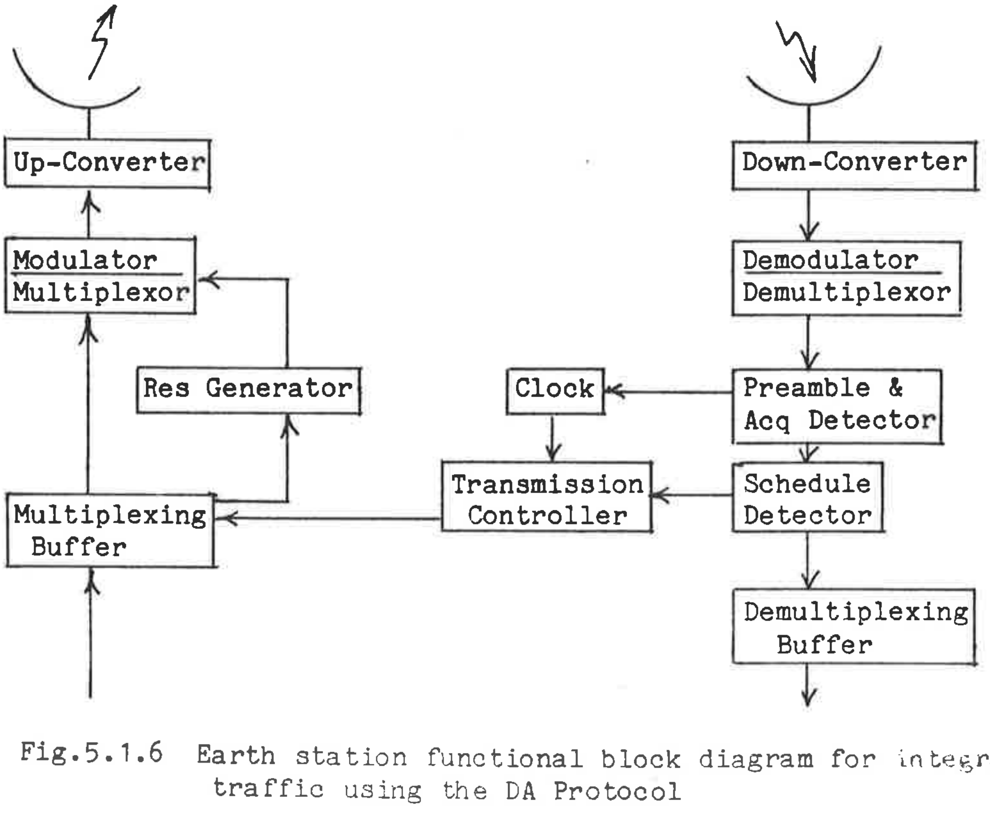


Fig.5.1.6 Earth station functional block diagram for integrated traffic using the DA Protocol

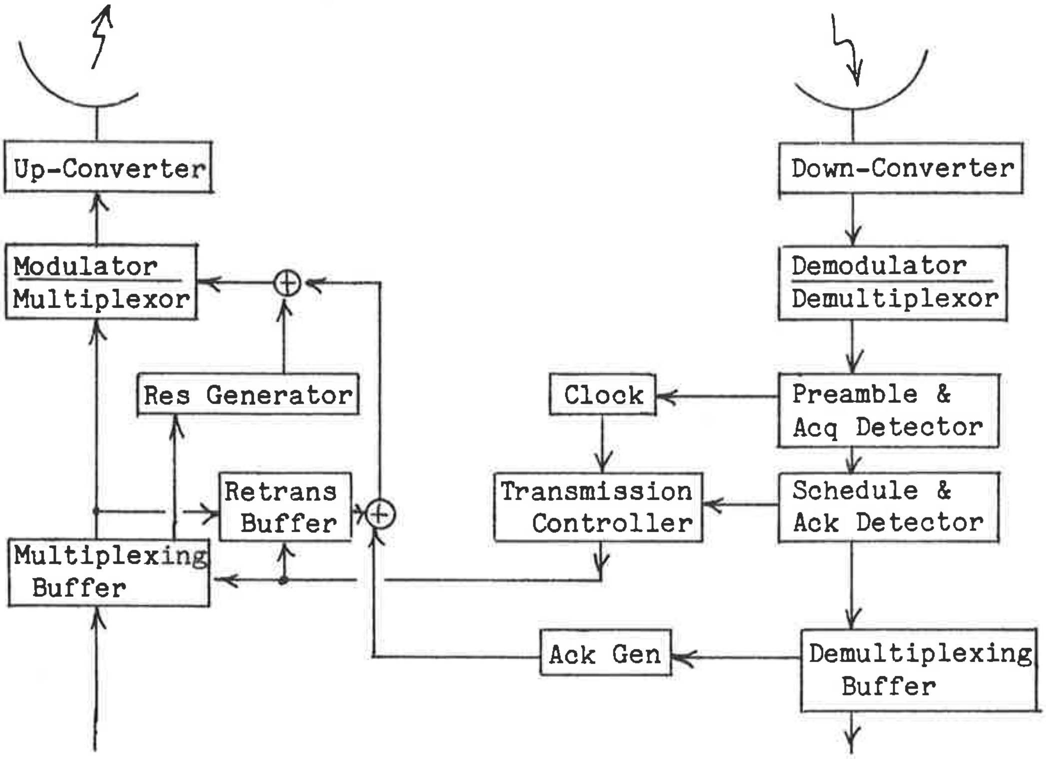


Fig.5.1.7 Earth station functional block diagram for integrated traffic using the Acknowledgement DA Protocol

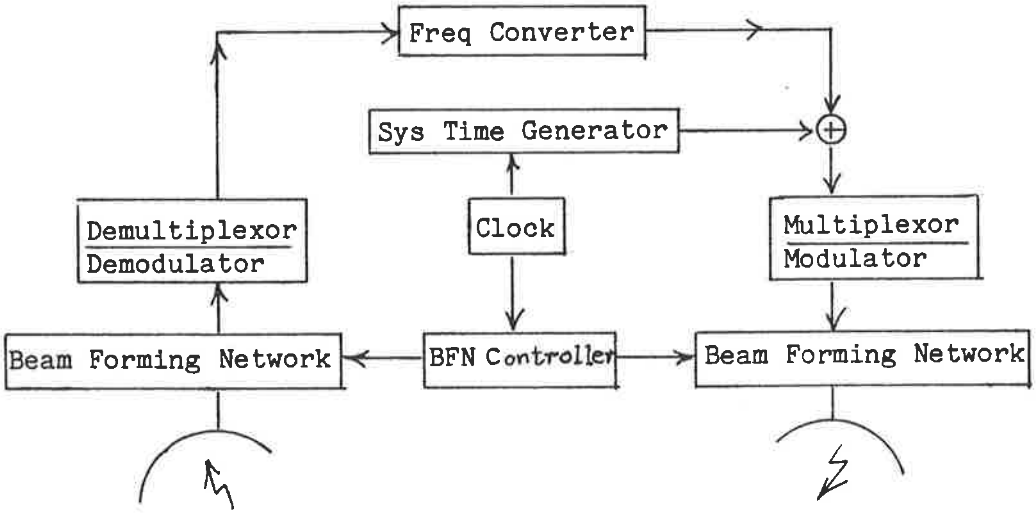


Fig.5.1.8 On-board controller functional block diagram for integrated traffic using the Distributed Control DA-FA Protocol

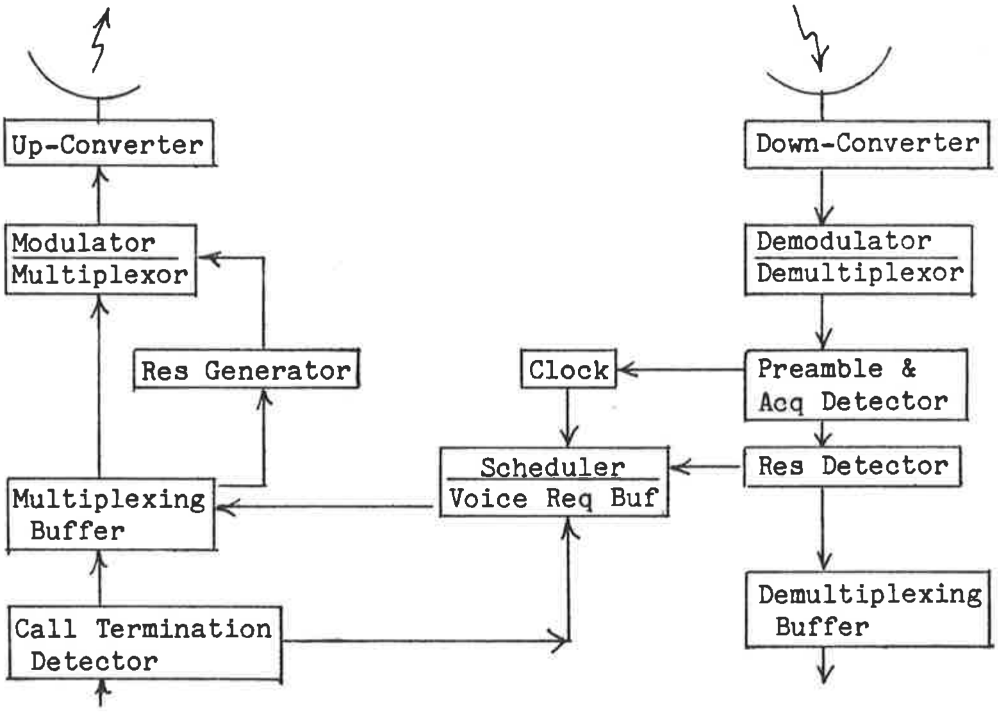


Fig.5.1.9 Earth station functional block diagram for integrated traffic using the Distributed Control DA-FA Protocol

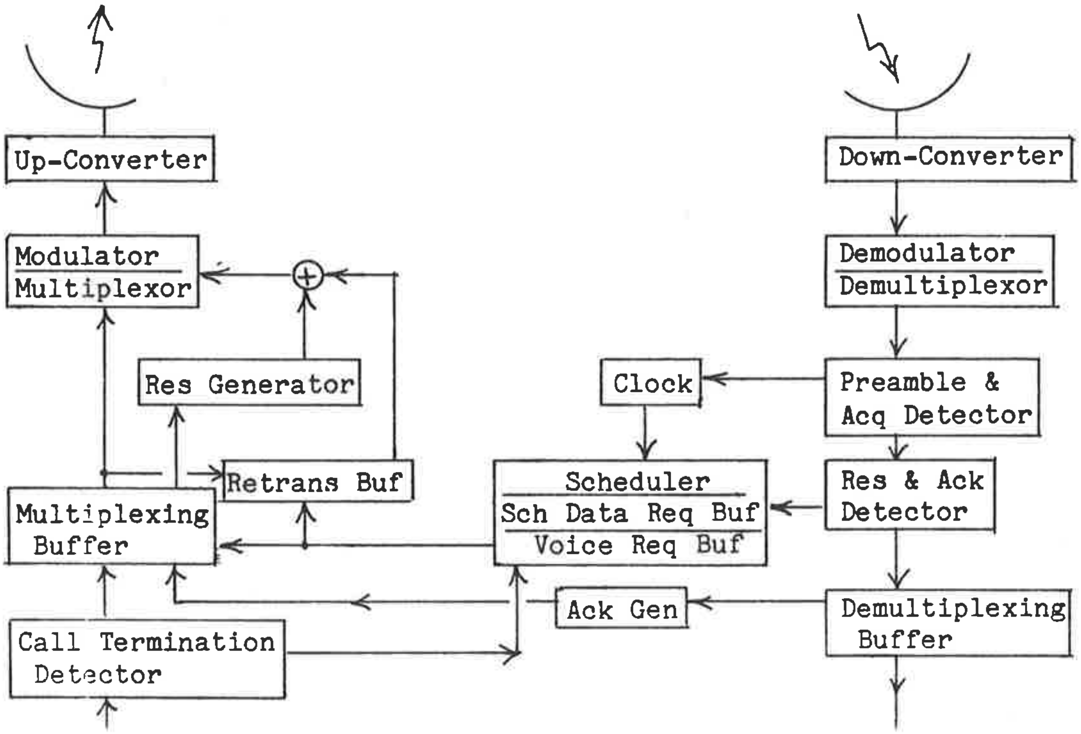


Fig.5.1.10 Earth station functional block diagram for integrated traffic using the Acknowledgement Distributed Control DA-FA Protocol

requiring data message acknowledgement, a buffer is required to store the scheduled requests until acknowledgements are received by the controller. The blocking probability of the scheduled data request buffer is given by the loss component of the M/M/1/Q queueing model and is discussed in Section 5.4.

For a system without the requirement of acknowledgement; multiplexing and demultiplexing buffers are needed in the earth station for the temporary storage of the incoming and outgoing traffic. Their buffer analyses are detailed in Section 5.3 for the FA, DA and DA-FA Protocols. A retransmisson buffer is added to the earth station for an acknowledgement system. The transmitted message is stored in the retransmisson buffer until an acknowledgement is received; the message is then released. The blocking probability for such a buffer is given in Section 5.4.

Since the voice traffic utilises circuit switching, the various blocking probabilities have been extensively studied [SYS 58]. The work here on the data traffic multilplexing buffer analyses for FA Protocol is a generalisation of [CHU 70a] and [CHU 70b], and the DA Protocol is an extension of [MAR 78]. However, it is believed that the modelling of the following topics have not received any previous attention :

1. the data multiplexing buffer analysis for the DA-FA Protocol (Section 5.3.1),

2. the scheduled data request buffer analysis (Section 5.4),

3. the data retransmisson buffer analysis (Section 5.4).

5.2 VOICE TRAFFIC BUFFERING REQUIREMENTS

The voice traffic using the SDMA protocols is operating in a circuit switching mode. The call state transition diagram is shown in Fig.5.2.1. There exist three seperate queues with the following blocking probabilities :-

1. sending station blocking probability (B_{vi}),
2. voice request buffer blocking probability (B_{vr}),
3. receiving station blocking probability (B_{vj})

Results and assumptions for the above probability studies are detailed in the following sections.

Since B_{vi}, B_{vr} and B_{vj} are statistically independent, the total blocking probability (B_v) for a call is given by

$$B_v = 1 - (1-B_{vi})(1-B_{vr})(1-B_{vj}) \qquad (5.2.1)$$

If B_{vi}, B_{vr} and B_{vj} are small, B_v is given approximately by the sum of the three individual blocking probabilties,

$$B_v = B_{vi} + B_{vr} + B_{vj} \qquad (5.2.2)$$

5.2.1 Blocking Probability of the Voice Traffic Request Buffer

For a First Come First Served (FCFS) discipline, a call will be lost if all the g voice circuits in the channel are occupied. However, if a voice traffic request buffer of size L is added in the traffic controller, the blocking probability (B_{vr}) due to the overflow of the voice request buffer is given by the loss component of the M/M/g/L delay-loss queueing model. Such a model can only be used if the following assumptions hold :-

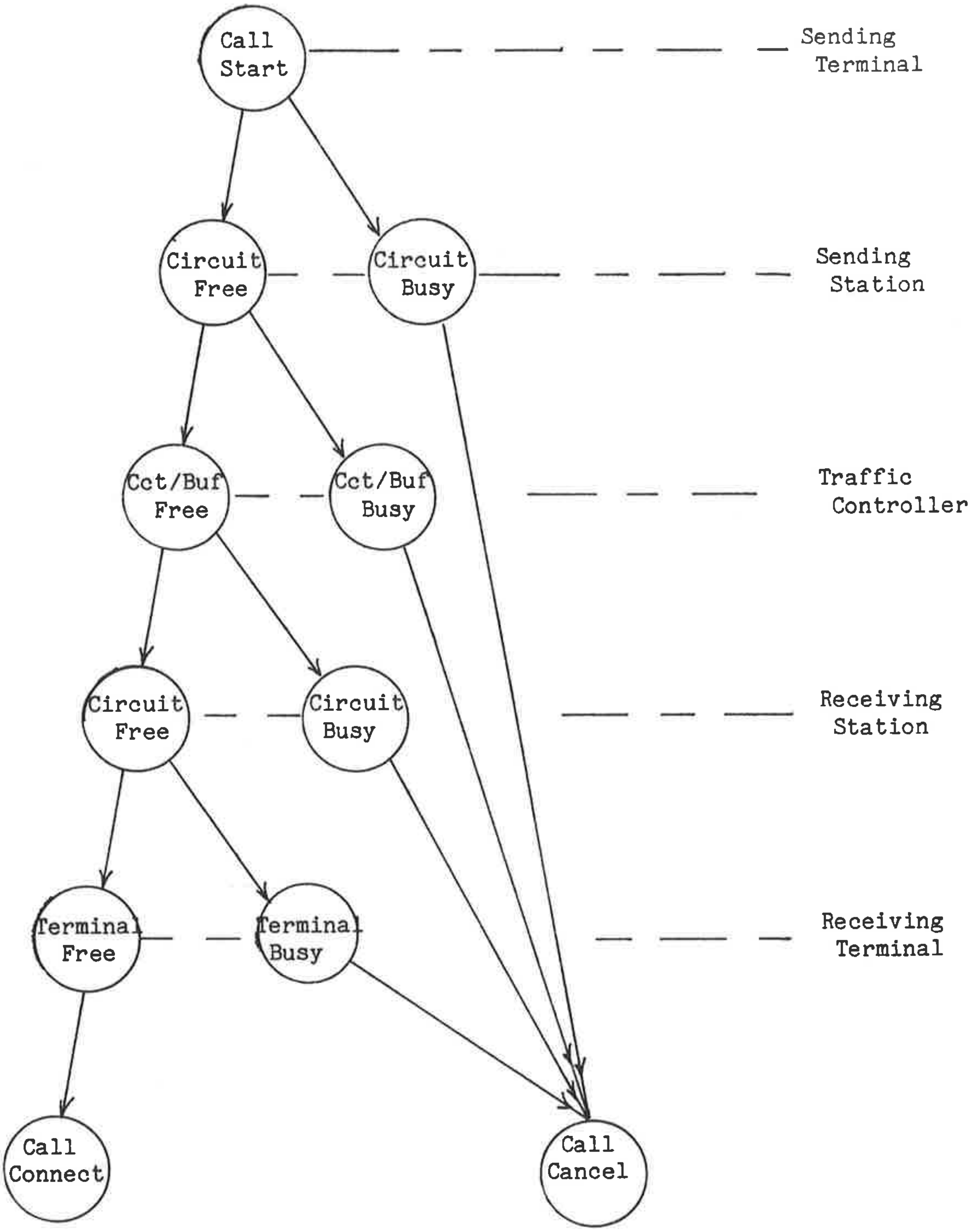


Fig.5.2.1 Voice traffic state transition diagram

1. a steady state condition has been reached,

2. calls arrive at random and the terminals abandon the call if all g circuits are busy and L request buffers are filled, and the caller does not make repeated attempts,

3. there exists a very large number of voice terminals in the network,

4. the probability of a call arriving is independent of the number of calls already in progress,

5. all g circuits and L request buffers are fully available to any incoming call,

6. call holding times are exponentially distributed.

From [TEL 78] pp.199,

$$B_{vr} = \frac{\rho^L}{E_g^{-1}(A) + \rho(\frac{1-\rho^L}{1-\rho})} \qquad \rho\neq 1$$

$$= [E_g^{-1}(A) + L]^{-1} \qquad \rho=1 \qquad (5.2.3)$$

where $E_g(A)= \frac{A^g/g!}{\sum_{j=0}^{g} A^j/j!}$ Erlang Loss Formula

$A=v\omega/\mu$ offered traffic

$\rho=A/g$ offered traffic per voice circuit

v=total number of voice terminals in the network

ω=average number of calls generated by each terminal/sec

$1/\mu$=average call holding time

For the case of no voice traffic request buffer, i.e. L=0

$$B_{vr} = E_g(A) \qquad (5.2.4)$$

See Fig.5.2.2 for a plot of B_{vr} versus buffer size with offered traffic as a parameter and a total of fifty circuits, and Fig.5.2.3 shows the same plot for a totol of five circuits and one circuit.

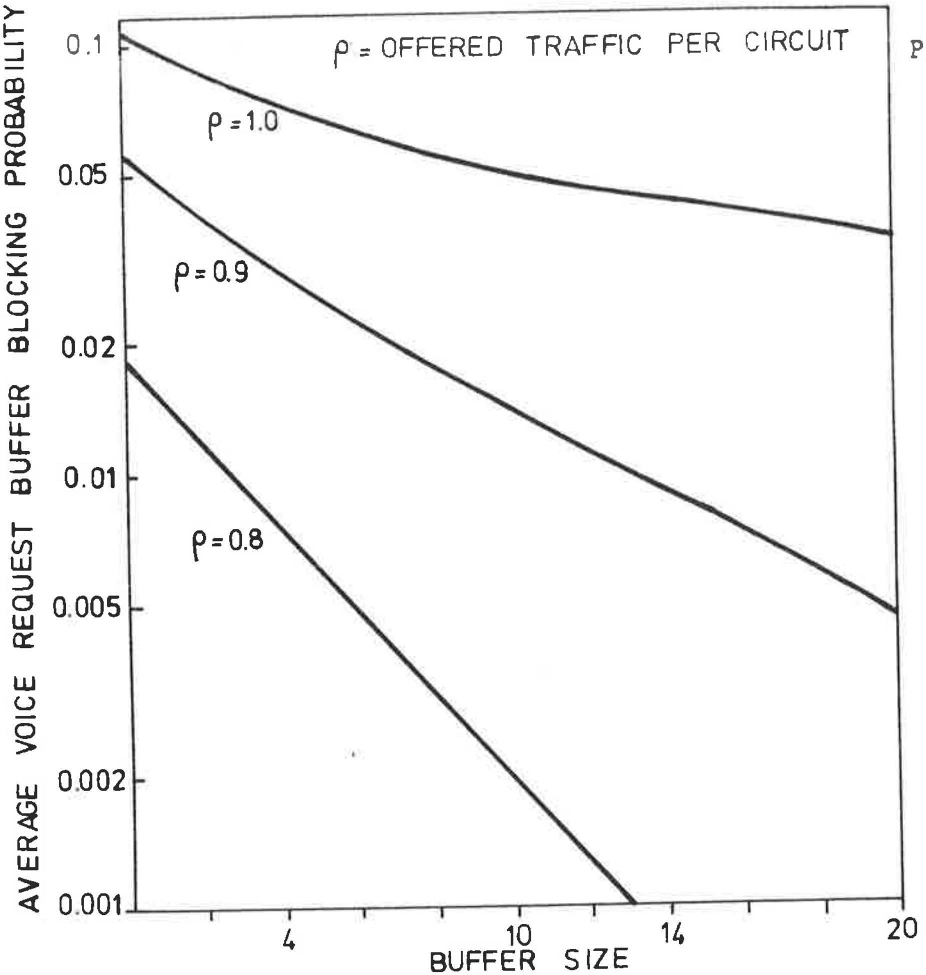


Fig.5.2.2 Blocking probability of the voice request buffer versus buffer size for 50 circuits

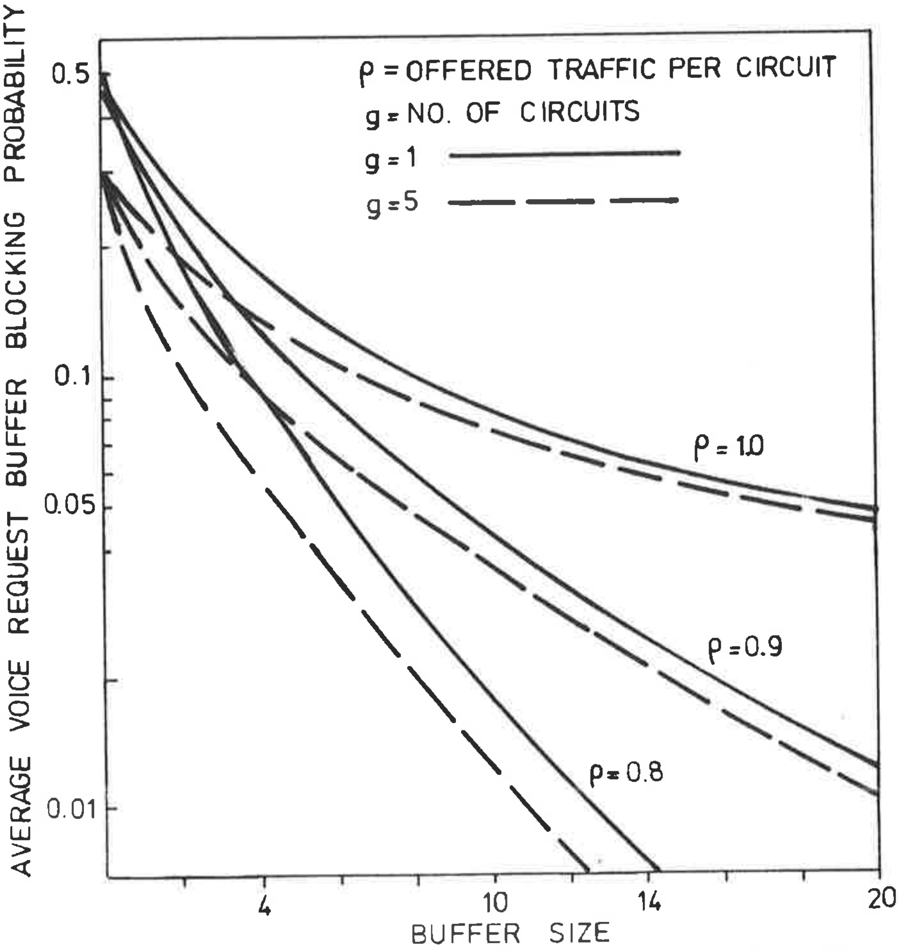


Fig.5.2.3 Blocking probability of the voice request buffer versus buffer size for 5 circuits and 1 circuit

5.2.2 Blocking Probabilities of the Sending and the Receiving Station

As a limited number of voice terminals are connected to each station, the incoming call arrival process from the network and the outgoing call to the network can no longer be assumed random, since an engaged terminal cannot be a source or a destination of a call. The blocking probability of such a limited free source and sink condition (< 200) is given by the Engset Loss Formula, ([TEL 78],pp.77). If V_i is the number of duplex voice circuits available at station i and v_i the number of voice terminals connected to station i, the sending station blocking probability (B_{vi}) is given by the no storage $M/M/V_i/0/v_i$ queueing model.

$$B_{vi} = \frac{\binom{v_i-1}{V_i}\left[\frac{\omega+\omega'}{\mu}\right]^{V_i}}{\sum_{j=0}^{v_i-1}\binom{v_i-1}{j}\left[\frac{\omega+\omega'}{\mu}\right]^{j}} \qquad (s) \qquad (5.2.5)$$

where $(\omega + \omega')/\mu$=total offered traffic

ω=average no. of calls generated/terminal/second

ω'=average no. of calls received from the network by a terminal/second

$1/\mu$=average call holding time

A recurrence formula suitable for machine calculation is detailed in [TEL 78] pp.78. A plot of B_{vi} versus offered traffic is given in Fig.5.2.4. With a uniform traffic distribution assumption :-

$$\omega' = \omega \qquad (5.2.6)$$

The same relation holds true for the receiving station blocking probability (B_{vj}), where the receiving station j has V_j duplex voice circuits and is connected to v_j voice terminals.

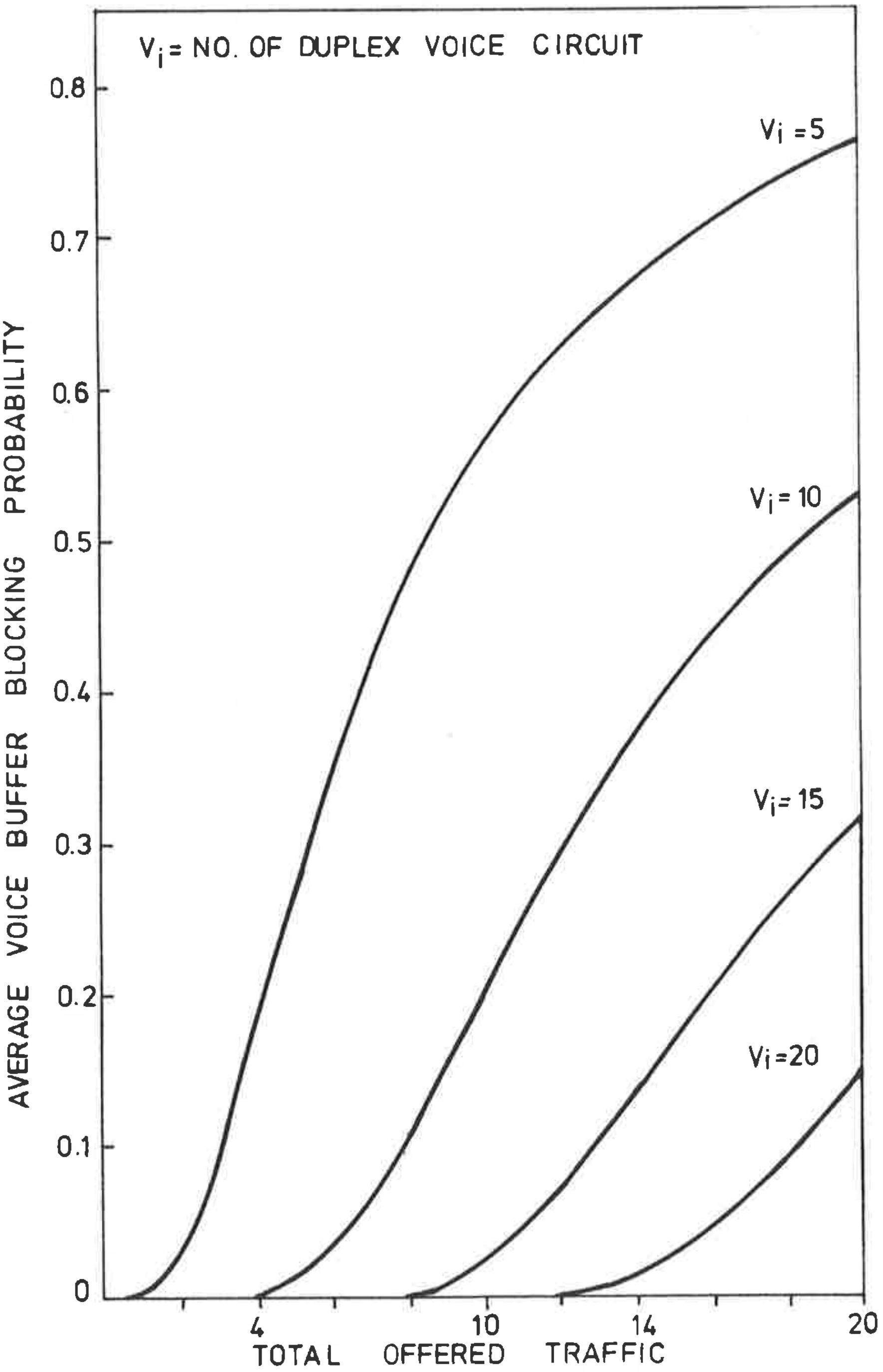


Fig.5.2.4 Blocking probability of the sending and the receiving station versus total offered traffic with 50 connecting terminals

5.3 DATA TRAFFIC BUFFERING REQUIREMENTS

For a network which does not acknowledge a received message, there exist two statistically independent blocking probabilities in the earth station :-

1. Data traffic multiplexing buffer blocking probability (B_{dd}),

2. Data traffic demultiplexing buffer blocking probability (B_{dm}).

The calculation of miltiplexing buffer blocking probability for the Fixed Assignment Protocol is similar to the problem of a multiple synchronous constant output. Chu analysed the problem of Poisson arrivals and multiple synchronous constant output [CHU 70a] and the problem of compound Poisson arrivals and single constant output [CHU 70b]. The solution in Section 5.3.1 is a generalisation of these results to an arbitrary compound arrival process. The calculation of the multiplexing buffer blocking probability for the Demand Assignment Protocol is complicated by the property of non-constant output from the sending station. The problem is partially solved by Mark and Ng [MAR 78] using a Two-Group Method for a compound Poisson input. The buffer analysis for the hybrid Demand Assignment uplink-Fixed Assignment downlink (DA-FA) Protocol has not yet been studied. The Two-Group Method is adopted in Section 5.3.1 for such a purpose. In what follows, all three studies are extended to an arbitrary compound arrival process.

The demulplexing buffer blocking probability is dependent on the amount of incoming traffic for the station. A destination function is used to relate the amount of incoming traffic for a particular station with respect to the total network traffic. For a uniform traffic destination function, the incoming traffic is proportional to the number of terminals connected to each station. A uniform traffic destination is assumed because it is simple to treat mathematically and the function is guaranteed by the law of large numbers for a network with a large number of terminals. Chu has shown that for a particular blocking probability, the uniform traffic destination function requires the smallest buffer size compared with other destination functions [CHU 72]. However, the methodology for calculating the demultiplexing blocking probability is similar to all traffic destination functions.

The analysis for the above two blocking probabilities are given in the next two sections. The total data traffic blocking probability (B_d) is given by

$$B_d = 1 - (1-B_{dm})(1-B_{dd}) \qquad (5.3.1)$$

If B_{dm} and B_{dd} are small enough, B_d is given approximately by

$$B_d = B_{dm} + B_{dd} \qquad (5.3.2)$$

5.3.1 Blocking Probability of the Data Traffic Multiplexing Buffer

The data traffic multiplexing buffer problem can be stated as follows; given a set of network parameters, traffic input and channel output characteristics, one wishes to find the buffer requirement such that its overflow probability is acceptable. This section is further divided into three subsections for the buffer analysis of Fixed Assignment, Demand Assignment and the hybrid Demand Assignment uplink-Fixed Assignment downlink Protocols.

A. Fixed Assignment Protocol

The Fixed Assignment Protocol has a multiple synchronous constant output, i.e. a fixed number of data units are taken out from the buffer at the beginning of each frame. The procedure used in the calculation of multiplexing buffer blocking probability is as follows :-

1. calculate the probability of certain arrivals into a station within a frame,
2. derive the buffer state probabilities given the buffer size and the available number of output slots,
3. calculate the average output,
4. obtain the overflow probability.

To calculate the probability of the number of arrivals within a frame, let $\tilde{a}(k)$ be the probability of arrival of k data units into a station within a frame, and $\Phi(s)$ its characteristic function. The characteristic function is defined as

$$\Phi(s) = \sum_{k=0}^{R-1} \tilde{a}(k)\exp(jsk) \tag{5.3.3}$$

$\tilde{a}(k)$ can be obtained from $\Phi(s)$ by the Inverse Discrete Fourier Transform (IDFT) Formula :-

$$\tilde{a}(k) = \frac{1}{R}\sum_{r=0}^{R-1} \Phi(-2\pi r/R)\exp(j2\pi rk/R) \tag{5.3.4}$$

$$k=0,1,2,\ldots,R-1$$

where R is the total number of sample points used to represent $\Phi(s)$. That is, if the characteristic function of the input process is known, the probability of the number of arrivals can be easily obtained through an Inverse Fast Fourier Transform Routine for any arrival function. For the particular case of a compound Poisson arrival process with geometrical message length distribution, $\tilde{a}(k)$ has the following expression :-

$$\tilde{a}(k) = \sum_{l=1}^{k} \binom{k-1}{l-1} (\lambda_i \theta)^l \; (1-\theta)^l \; \exp(-\lambda_i)/l! \qquad k=1,2,..$$

$$= \exp(-\lambda_i) \qquad k=0 \qquad (5.3.5)$$

the corresponding characteristic function is

$$\Phi(s) = \exp\{-\lambda_i[(1-\Phi_M(s)]\} \qquad (5.3.6)$$

where $$\Phi_M(s) = \frac{\theta \exp(js)}{1-(1-\theta)\exp(js)} \qquad (5.3.7)$$

λ_i=average no of message arrivals into station i/frame

θ=average message length in number of slots

Let N be the number of assigned data slots to be used by station i for its arrivals, and B(x) be the buffer state probability of x slots in the buffer at the beginning of the frame. For a given buffer size (q), and N<q :-

$$B(x) = \sum_{j=0}^{N} B(j)\tilde{a}(x) + \sum_{j=N+1}^{q} B(j)\tilde{a}(x-j+N)u(x-j+N) \qquad (5.3.8a)$$

$$x=0,1,2,..,q-1$$

where $u(x)=1 \qquad x \geq 0$

$=0 \qquad x<0$

The first term represents the case where the number of slots in the buffer at the beginning of the frame is less than or equal to the number of assigned slots per frame. The second term gives the case where the state of the buffer is greater than the number of available slots and the number of new arrivals make up the difference, together with

$$\sum_{x=0}^{q} B(x) = 1 \qquad (5.3.8b)$$

These linear equations of (5.3.8) can be solved for B(x).

In the case that the number of available slots per frame per station is greater than or equal to the buffer size ($N \geq q$), the values of $B(x)$ will be equal to the probability of the arrivals of x units of data :-

For $N \geq q$,

$$\begin{aligned} B(x) &= \tilde{a}(x) && x=0,1,2,..,q-1 \\ &= 1 - \sum_{k=0}^{q-1} \tilde{a}(k) && x=q \end{aligned} \qquad (5.3.9)$$

The average carried traffic can be computed from the buffer state probabilities

$$\begin{aligned} E(\alpha) &= \sum_{j=1}^{N} jB(j) + N \sum_{j=N+1}^{q} B(j) && N<q \\ &= \sum_{j=1}^{q} jB(j) && N \geq q \end{aligned} \qquad (5.3.10)$$

The data multiplexing buffer blocking probability is given by

$$B_{dm} = 1 - E(\alpha)/E(\beta) \qquad (5.3.11)$$

where $E(\beta)$ is the average offered traffic of station i.

With a uniform traffic distribution assumption, the average message arrival rate and the offered traffic of station i can be simplified as :-

$$\lambda_i = \frac{\text{total network traffic arrivals}}{\text{total number of stations in the network}} = \frac{\lambda}{s}$$

$$E(\beta) = \lambda_i/\theta = \lambda/(s\theta)$$

B. Demand Assignment Protocol

The calculation of the multiplexing buffer blocking probability for the Demand Assignment Protocol is complicated by the property of non-constant output. The Two-Group Method of Mark and Ng [MAR 78] is used for the calculation.

The network is divided into two groups :-

1. Group I be the station under consideration,

2. Group II be the rest of the stations in the network.

The overflow probability can be obtained by considering the imbedded Markov Chain formed by the buffer state probabilities at the beginning of each frame. The procedure adopted is as follows :-

1. calculate the probability of certain arrivals within a frame,

2. derive the buffer state probabilities given a buffer size and a number of available slots per frame,

3. obtain the average output for a given number of available slots,

4. calculate the average output or the carried load, and

5. obtain the overflow probability.

To calculate the probability of the number of arrivals within a frame, let $\tilde{a}_I(k)$ and $\tilde{a}_{II}(k)$ be the probabilities of arrival of k data slots in a frame at Group I and II respectively and let their characteristic functions be $\Phi_I(s)$ and $\Phi_{II}(s)$ respectively. $\tilde{a}_I(k)$ and $\tilde{a}_{II}(k)$ can be easily obtained from $\Phi_I(s)$ and $\Phi_{II}(s)$ by the Inverse Fast Fourier Transform as described in the buffer analysis of the Fixed Assignment Protocol.

Let $B_i(x)$ be the buffer state probability of x slots in the buffer at the beginning of the frame, with i available service slots per frame. For mathematical tractability, the event $B_0(x)$ is set to zero, i.e. no service to Group I is not permitted to occur. Such an assumption is justified for a stable system, since the total traffic is less than or equal to the network capacity. However, it leads to certain inaccuracy as indicated by the simulation results of Chapter 7. For a given buffer size (q) and a particular available service unit (i) and number of available data slots per frame (N), where i=1,2,...,N. For i<q,

$$B_i(x) = \sum_{j=0}^{i} B_i(j)\tilde{a}_I(x) + \sum_{j=i+1}^{q} B_i(j)\tilde{a}_I(x-j+i)u(x-j+i) \qquad (5.3.12a)$$

$$x=0,1,2,\ldots,q-1$$

$$\text{where } u(x)=1 \qquad x\geq 0$$

$$=0 \qquad x<0$$

This is similar to (5.3.8a), together with

$$\sum_{x=0}^{q} B_i(x) = 1 \qquad (5.3.12b)$$

These linear equations of (5.3.12) can be solved for $B_i(x)$ for a particular i. In the case that the number of available slots per frame is greater than or equal to the buffer size (i≥q), the values of $B_i(x)$ will be equal to the probability of the arrivals of x unit of data :-

For i≥q,

$$B_i(x) = \tilde{a}_I(x) \qquad x=0,1,2,\ldots,q-1$$

$$= 1 - \sum_{k=0}^{q-1} \tilde{a}_I(k) \qquad x=q \qquad (5.3.13)$$

Once the buffer state probabilities $B_i(x)$ are calculated for a given buffer size and the number of available service units, the average output for i available service units is given by

$$\alpha(i) = \sum_{x=1}^{i} xB_i(x) + i\{ \sum_{x=i+1}^{q} B_i(x) \} \qquad i=1,2,..,q-1$$

$$= \sum_{x=1}^{q} xB_i(x) \qquad i \geq q \qquad (5.3.14)$$

The carried load is the average output over all values of i,

$$E(\alpha) = \frac{1}{P} \sum_{i=1}^{N} \alpha(i)\tilde{a}_{II}(N-i) \qquad (5.3.15)$$

where P is a normalisation factor, and

$$P = Pr(\text{some service available to Group I}) = \sum_{k=0}^{N-1} \tilde{a}_{II}(k) \quad (5.3.16)$$

The data multiplexing buffer blocking probability is given by

$$B_{dm} = 1 - E(\alpha)/E(\tilde{A}_I) \qquad (5.3.17)$$

where $E(\tilde{A}_I)$ is the average offered traffic for Group I, and it can be obtained

$$E(\tilde{A}_I) = - j\Phi_I^{(1)}(0) \qquad (5.3.18)$$

With a uniform traffic distribution assumption, the average input rate of Group I and II (λ_I and λ_{II}) and the average offered traffic to Group I are expressed as :-

$$\lambda_I = \frac{\text{total network traffic arrivals}}{\text{total number of stations in the network}} = \frac{\lambda}{s}$$

$$\lambda_{II} = \lambda(s-1)/s$$

$$E(\tilde{A}_I) = \lambda/(s\theta)$$

C. Demand Assignment Uplink and Fixed Assignment Downlink Protocol

The hybrid form of the distributed control Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol requires a new approach to the buffer analysis. The uplink capacity is contested via the reservation channel but the downlink is scanning the network in a fixed sequence of pre-determined duration. The protocol requires each station to schedule their messages within the alloocted interval.

There exists two possible buffer arrangements in each of the station. First, the network traffic is divided into K^2 groups, where K is the number of zones in the network. For the outgoing traffic to zone j, each station in the i-th zone contests for the capacity allocated for zone i to zone j. With this approach, the buffer blocking probability is calculated for one of the K separated queues, one for each zone. It also requires the assumption that the traffic between the station and zone j has reached an equilibrium state.

However a greater degree of statistical multiplexing is possible if all the traffic within zone i is assumed to have reached a state of equilibrium. In particular, the assigned capacity for zone i to zone j is greater than the offered traffic. Under this assumption, the uplink can be regarded as statistically multiplexed amongst all the traffic from the entire zone. With this approach, the buffer blocking probability is calculated for a buffer shared amongst all the outgoing traffic.

Due to the non-constant nature of the outgoing traffic from each of the stations within the zone, Two-Group Method is again used to calculate the buffer blocking probability. With a uniform traffic distribution assumption, the average input rate of Group I and II (λ_I and λ_{II}) and the average offered traffic of Group I are as following :-

$$\lambda_I = \frac{\text{total network message arrivals}}{\text{total number of stations in the network}} = \frac{\lambda}{s}$$

$$\lambda_{II} = \lambda(s-K)/(Ks) \doteq \lambda/K \qquad \text{for large } s$$

$$E(\beta) = \lambda/(s\theta)$$

and the value of N is now referred to the capacity available to each zone,

$$N = \text{total data capacity per frame/no. of zones}$$

5.3.2 Blocking Probability of the Data Traffic Demultiplexing Buffer

In order to describe the average incoming traffic $E(\tilde{Z}_i)$ for station i in relation to the offered traffic in steady state, a traffic destination function $f_d(i)$ is used :-

$$f_d(i) = E(\tilde{Z}_i)/E(\tilde{A}) \qquad i=1,2,..,s \qquad (5.3.19)$$

and $\sum_{i=1}^{s} f_d(i) = 1$

Such a traffic destination function can only be obtained by traffic measurement. Chu found that for a given average traffic level, the uniform destination function requires the smallest buffer size compared with the linear, step and geometrical destination functions [CHU 72]. Within a uniform traffic destination function, the fraction of incoming traffic for station i is proportional to the ratio of the number of terminals connected to station i with respect to the total number of terminals in the network.

$$E(\tilde{Z}_i) = \varepsilon_i E(\tilde{A}) \qquad (5.3.20)$$

With other destination functions, the ratio $f_d(i)$ is dependent on the function as well as the terminal number.

Knowing the average amount of incoming traffic into station i - $E(\tilde{Z}_i)$, the data demultiplexing buffer blocking probability (B_{dd}) can be calculated if $E(\tilde{X}_{i,d})$ is known. $\tilde{X}_{i,d}$ is the received traffic of station i for a multiplexing buffer of size d and a demultiplexing channel rate of c_d slots per frame.

$$B_{dd} = 1 - E(\tilde{X}_{i,d})/E(\tilde{Z}_i) \qquad (5.3.21)$$

For a uniform traffic distribution :-

$$B_{dd} = 1 - E(\tilde{X}_{i,d})/[\varepsilon_i E(\tilde{A})] \qquad (5.3.22)$$

Since the demultiplexing channel has a constant output rate (c_d), the derivation of $E(\tilde{X}_{i,d})$ is similar to the derivation of $E(\alpha)$ in the Fixed Assignment Protocol using Chu's method (Section 5.3.1, Subsection A). That is, the calculation of data demultiplexing buffer blocking probability is dependent on the arrival rate and the buffer size, and independent of the protocol.

5.4 Buffering Requirements for an Acknowledgement System

For any of the SDMA protocols requiring acknowledgements of the received messages, two additional buffers are needed :-

1. scheduled data request buffer in the traffic controller, and

2. data retransmisson buffer in the earth station.

The capacity request is stored in the request buffer after its allocation by the traffic controller. At the allocated time, the message is transmitted and loaded into the retransmission buffer. The destination station of the message will send back an acknowledgement after a correct reception. The request is then released from the request buffer, and the retransmission buffer in the sending station.

With the assumption of reliable acknowledgement transmission, achievable through redundancy and forward error correction coding, any retransmission is assumed to be caused by errors during the message transmission. Hence, correct message reception into the controller is transmission channel dependent. For an arbitrary message and acknowledgement arrival process, the steady state buffer behaviour can be obtained from the GI/G/1 queueing model [KOB 77].

For the specific case of Poisson message and acknowledgement arrival process with λ and $\delta\lambda$ ($\delta \leq 1$) average arrivals per frame respectively, the steady state data request buffer blocking probability (B_{dr}) can be obtained by the loss component of M/M/1/Q delay-loss queueing model [TEL 78], where Q is the size of the data request buffer.

$$B_{dr} = \frac{1 - \delta}{1 - \delta^{Q+2}} \qquad \delta < 1$$

$$= \frac{1}{Q + 2} \qquad \delta = 1 \qquad (5.4.1)$$

δ can be interpreted as the fraction of message received correctly by the receiving station. As expected, B_{dr} approaches zero as δ approaches one and B_{dr} approaches $1-\delta$ as Q tends to infinity. Hence, B_{dr} is transmission channel dependent.

If station i has an average input of λ_i messages per frame and an average correct transmission of $\delta_i\lambda_i$ messages, $1-\delta_i$ of the stored messages are required for retransmission. Under such condition, the average of the correct transmission is the determining factors of the data retransmission buffer blocking probability ($B_{d\ ret}$). Knowing the data multiplexing buffer blocking probability (B_{dm}), the blocking probability of the data retransmission buffer of size $1 - B_{d\ ret}(1)$ is equivalent to the B_{dm} of a reduced size $1/(2-\delta_i)$,or

$$B_{d\ ret}(1) = B_{dm}\{1/(2-\delta_i)\} \qquad (5.4.2)$$

The total data blocking probability for an acknowledgement system is given by

$$B_d = 1 - (1-B_{dm})(1-B_{dr})(1-B_{dd})(1-B_{d\ ret}) \qquad (5.4.3)$$

where B_{dm}= multiplexing buffer blocking probability

B_{dd}= demultiplexing buffer blocking probability

both of which have been discussed in Section 5.3. If B_{dm}, B_{dr}, B_{dd} and $B_{d\ ret}$ is sufficiently small, B_d approaches the following approximation :-

$$B_d = B_{dm} + B_{dr} + B_{dd} + B_{d\ ret} \qquad (5.4.4)$$

5.5 CONCLUDING REMARKS

In this chapter, the methods of calculation of various buffer performances for both voice and data traffic with and without acknowledgement in any of the Space Division Multiple Access protocols are presented.

In the case of voice traffic, the request buffer blocking probability is given by the loss component of a Markovian delay-loss queueing model. As a limited number of voice terminals is connected to each station, the incoming call arrival process from the network and the outgoing call to the network can no longer be assumed random. Under such limited free source and sink conditions, the voice circuit sending and receiving station blocking probabilties are given by Engset Loss Formula.

For data traffic, the multiplexing buffer blocking probability of the Fixed Assignment (FA) Protocol can be obtained by a multiple synchronous constant output model. The probabiltiy of certain arrivals into a station from an arbitrary compound arrival process can be calculated by the Inverse Discrete Fourier Transform Formula of the characteristic function.

The multiplexing buffer blocking probability of the Demand Assignment (DA) Protocol is complicated by the non-constant output. The Two-Group Method is used for this calculation. The multiplexing buffer blocking probability of the hybrid form of the distributed control Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol is calculated for a buffer shared amongst all the outgoing traffic. Due to the non-constant nature of the outgoing traffic, the Two-Group Method is again used.

Once the average incoming traffic into the station is known, the data traffic demultiplexing buffer blocking probabilities is a function of the demultiplexing buffer size and the demultiplexing channel rate. The traffic destination function is used to describe the average incoming traffic for a particular station in relation to the total offered traffic. In practice, such a traffic destination function can only be obtained by monitoring the traffic.

Two additional buffers are required for a system of data traffic with acknowledgement. They are the scheduled data request buffer in the traffic controller and the data retransmission buffer in the earth station. The capacity request is stored in the request buffer after its allocation by the traffic controller. At the allocated time, the message is transmitted and loaded into the retransmission buffer. After correct reception, the request and the message are released from their appropriate buffers. For the specific case of Poisson message and acknowledgement arrival processes, the data request buffer blocking probability can be obtained by the loss component of a Markovian delay-loss queueing model, while the data retransmission buffer can be modelled by the Two-Group Method with a modified buffer size.

6.PERFORMANCE COMPARISONS

In this chapter, the delay and buffer performances of the three proposed Space Division Multiple Access (SDMA) protocols are evaluated, The subsections in this chapter contain the following topics :-

1. A comparison of the delay performances of the three SDMA protocols, in order to obtain some insights into the sensitivities of the delay performance with respect to the variation in traffic characteristics and protocol and network parameters.
2. A comparison of the SDMA protocols with other multiple access protocols operating in a global beam network, in order to obtain some perspectives for the delay performance of the SDMA protocols.
3. A comparison of the buffer requirements of the three SDMA protocols.

To illustrate the calculation methods and formulate a basis for comparisons, the assumption that the origin and the destination of the traffic are uniformly distributed amongst all the stations and zones is made.

6.1 DELAY-THROUGHPUT COMPARISONS OF THE SDMA PROTOCOLS

In this section, the definitions and the stability criteria of the channel and the control channel throughputs of the three proposed Space Division Multiple Access protocols are discussed. Then the investigations of the delay performance of Demand Assignment (DA), Fixed Assignment (FA) and the hybrid Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol follow. The sensitivity of the delay performance in relation to the traffic characteristics, protocol and network parameters for each of the protocols is highlighted.

6.1.1 Channel and Control Channel Throughput

Let the traffic intensity (S) be the normalised traffic arrival; under steady state conditions, S is also the normalised channel throughput.

$$S = \text{network arrival} / \text{channel capacity} \qquad (6.1.1)$$

In the case of a DA Protocol,

$$S = \frac{\lambda/\theta}{N/\delta} = \frac{\lambda}{\theta Ff} \qquad (6.1.2)$$

where λ=average number of message arrivals from the network/frame

$1/\theta$=average message length (slots)

δ=data channel bandwidth assignment factor

F=data channel bandwidth (slots/s)

N=number of data slots per frame

f=framelength (s)

S is given in terms of the network parameters, since both the uplink and the downlink are allocated to the entire network on demand.

The bandwidth assignment (δ) plays an important role in the delay-throughput calculation. It is limited by the control channel capacity used in the network.

$$\delta = 1 - \frac{ks}{Ffv} \qquad (6.1.3)$$

where k=number of control slots required per station (protocol dependent)

s=number of stations in the network

v=ratio of number of control slots per data slot

The other consideration is the stability of the control channel. Let the control channel throughput be S_c, and

$$S_c = \frac{k\lambda}{(1-\delta)Ffvu}$$

where u is the maximum number of control (request or schedule) packets allowed to be carried by each control burst. In order to maintain the stability of the control channel,

$$u > \lambda_i \qquad (6.1.4)$$

where λ_i is the average requests arrivals per frame into station i. Let the data section throughput be S_d, and

$$S_d = \frac{\lambda}{\theta N} = \frac{\lambda}{\theta\delta Ff} = \frac{(1-\theta)vu}{k\theta\delta} S_c$$

For maximum utilisation of data section, $S_d=1$; and

$$S_c = \frac{k\theta\delta}{(1-\theta)vu}$$

The stability of the control channel can be assumed if $S_c < 1$, the appropriate values of u and v are chosen such that

$$uv > \frac{k\theta\delta}{1-\delta} \qquad (6.1.5)$$

For the FA Protocol, without the control channel, δ is equal to one.

$$S = \frac{\lambda}{\theta Ks} \div N_{ij} = \frac{\lambda}{\theta KsN_{ij}}$$

where K=number of zones in the network

s=number of stations in the network

The framelength for the FA Protocol f_{FA} in seconds :

$$f_{FA} = KsN_{ij} \;/\; F \tag{6.1.7}$$

For the DA-FA Protocol, only a portion of the network capacity (N_{kj}) is contested amongst all the stations in the zone. With the uniform traffic distribution assumption,

$$S = \frac{\lambda}{\theta K^2} \div \frac{N_{kj}}{\delta} = \frac{\lambda\delta}{\theta K^2 N_{kj}} \tag{6.1.8}$$

the framelength for the DA-FA Protocol in seconds is given by :

$$f_{DA-FA} = K^2 N_{kj} \;/\; (\delta F) \tag{6.1.9}$$

Similar expressions on the control channel stability can be derived by substituting Ff for $K^2 N_{kj}/\delta$.

6.1.2 Delay-Throughput Comparisons of the SDMA Protocols

The following four conditions are necessary in order to compare the three SDMA protocols :-

1. same transponder capacity,

2. same amount of offered traffic, and

for DA and DA-FA Protocols,

3. same framelength, in order to reflect the required overhead,

4. same control burst size, i.e. the same number of control packets allowed to be carried by each burst, in order to maintain parity in the value of reservation delay.

The following comparisons give the variations in the average message delay when altering :-

A. traffic characteristics - channel throughput (S) and average message length ($1/\theta$),

B. protocol parameters - framelength (f) and control burst size in terms of the number of control packets in each control burst (u),

C. network parameters - transponder capacity (F), number of spotbeam coverages or zones (K) and number of stations (s) in the network.

The following parameters are chosen for comparisons, with the view of modelling a suitable realistic network :-

a. number of zones in the network (K) = 20

b. number of stations in the network (s) = 1000

c. transponder capacity (F) = 5000 slots/s

d. slot length = 2000 bits/slot

e. ratio of the size of the control slot in relation to a data slot

$v = 10$

f. number of control slots required for every reservation slot (k)

for DA-FA Protocol k=1.0

DA Protocol k=2.0

A. Delay-throughput comparisons with varying traffic characteristics

Fig.6.1.1 gives the delay-throughput comparison of the DA, DA-FA and FA Protocols for an average message length of one slot per message; in addition the DA and DA-FA Protocols have a similar framelength (f) of 0.2 second and control burst size (u) of one set of control packets per control burst. For the FA Protocol, a minimum framelength of 4 seconds is required to obtain a full network connectivity of 1000 stations with an allocated capacity of one slot per station. Due to the long framelength, the delay-throughput performance is far inferior to both the DA and DA-FA Protocols. Hence, the FA Protocol is only suitable for a small network. The attention of this section is mainly focused on the other two SDMA protocols.

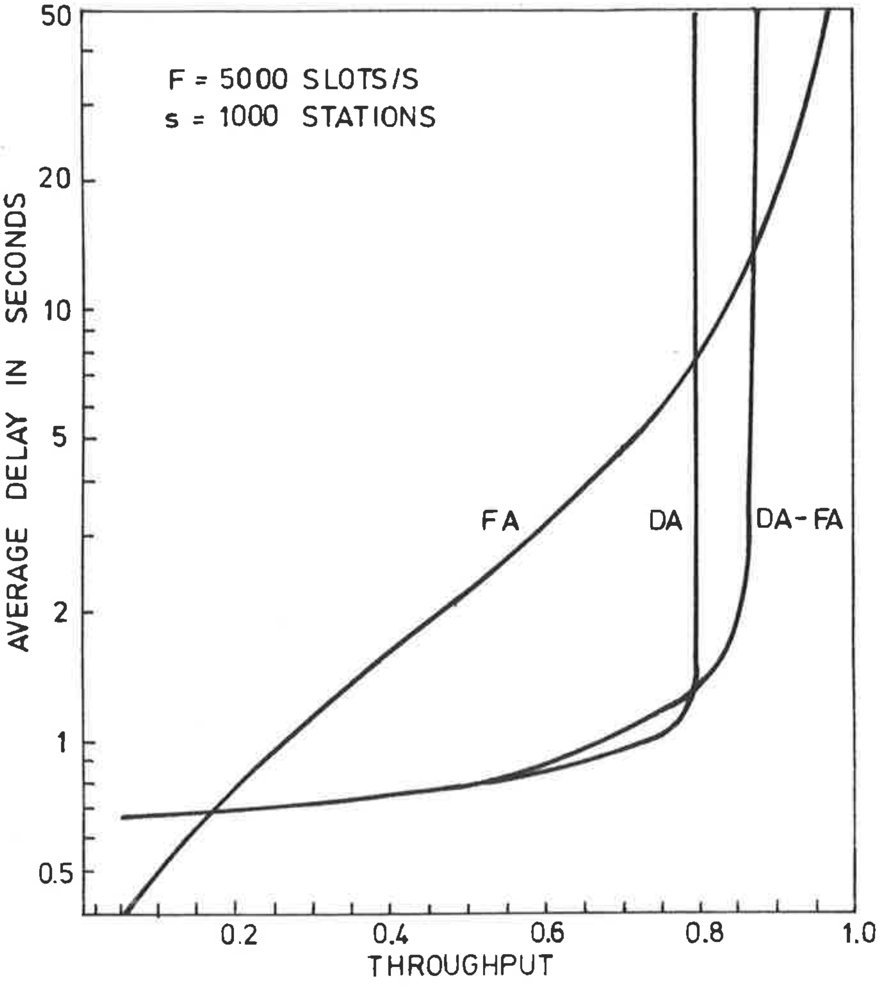


Fig.6.1.1 Delay performance comparison of the SDMA protocols for a message length of one slot

However, the price that the DA or DA-FA Protocols pay for their superior delay-throughput performance is a decrease in the data channel size given by the assignment value of δ. Since the DA Protocol requires a larger overhead than the DA-FA Protocol, it has a smaller δ value.

Fig.6.1.2 gives the delay-throughput comparison of the DA and DA-FA Protocols, for the average message lengths of one and three slots per message, with the framelength equal to 0.2 second. The performance of the DA Protocol improves while the DA-FA Protocol deteriorates as the average message length increases. The folowing two figures explain such opposing trends. Fig.6.1.3 shows that the reservation delay decreases as the average message length increases for both the DA and DA-FA Protocol. This is due to the definition of the channel throughput. For the same value of throughput, the number of message arrivals decreases as the message length increases. However, Fig.6.1.4 indicates the scheduling delay of the DA Protocol is essentially independent to the message length; while the scheduling delay of the DA-FA Protocol is message length dependent.

The difference is due to the fact that the message length dependent term is negligible for the DA Protocol since the number of data slots available for allocation is large. While the corresponding term in the DA-FA Protocol is much smaller and therefore more sensitive to the variation in message length. For the DA-FA Protocol, the increase in scheduling delay is larger than the decrease in reservation delay as the average message length increases. Hence, the delay performance of the DA-FA Protocol deteriorates as the message length increases. In other words, the DA Protocol is more suitable for traffic which has long message lengths.

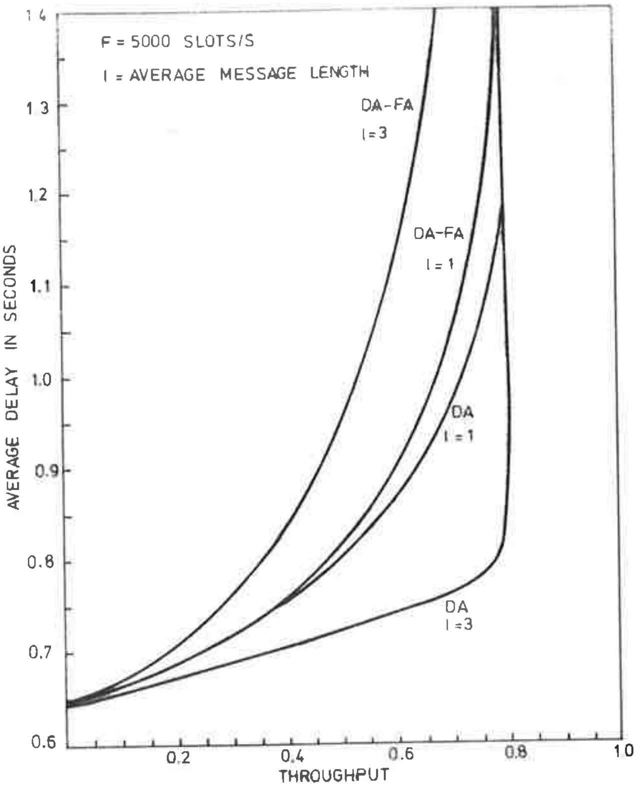


Fig.6.1.2 Delay performance comparison between the DA and DA-FA Protocols

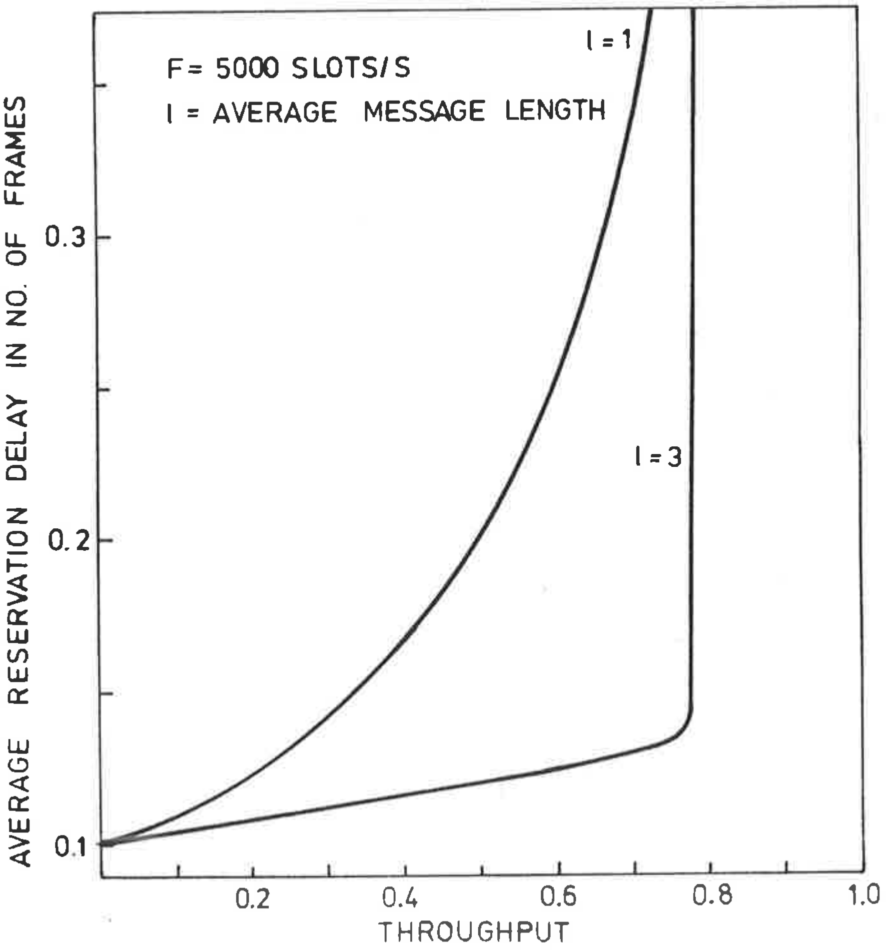


Fig.6.1.3 Reservation delay performance for both the DA and DA-FA Protocols

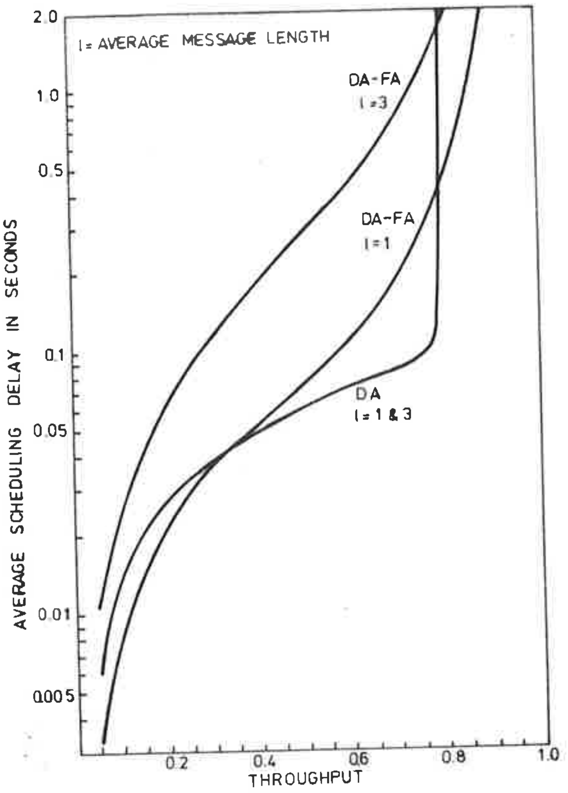


Fig.6.1.4 Scheduling delay performance for both the DA and DA-FA Protocols

It is obvious that both the DA-FA and FA Protocols have a monotonically increasing delay performance with respect to an increasing average message length. However, the delay performance of the DA Protocol as shown in Fig.6.1.5 is non-monotonic. The explanation of this behaviour is that the scheduling delay increases as the average message length increases. In addition, due to the definition of channel throughput, the average message arrivals decrease in order to maintain the same throughput value. Therefore, the reservation delay is reduced. That is, as the average message length increases, the total delay for a particular throughput decreases until an optimum message length is reached, after which any increase in message length causes an increase in scheduling delay which more than offsets the reduction in reservation delay.

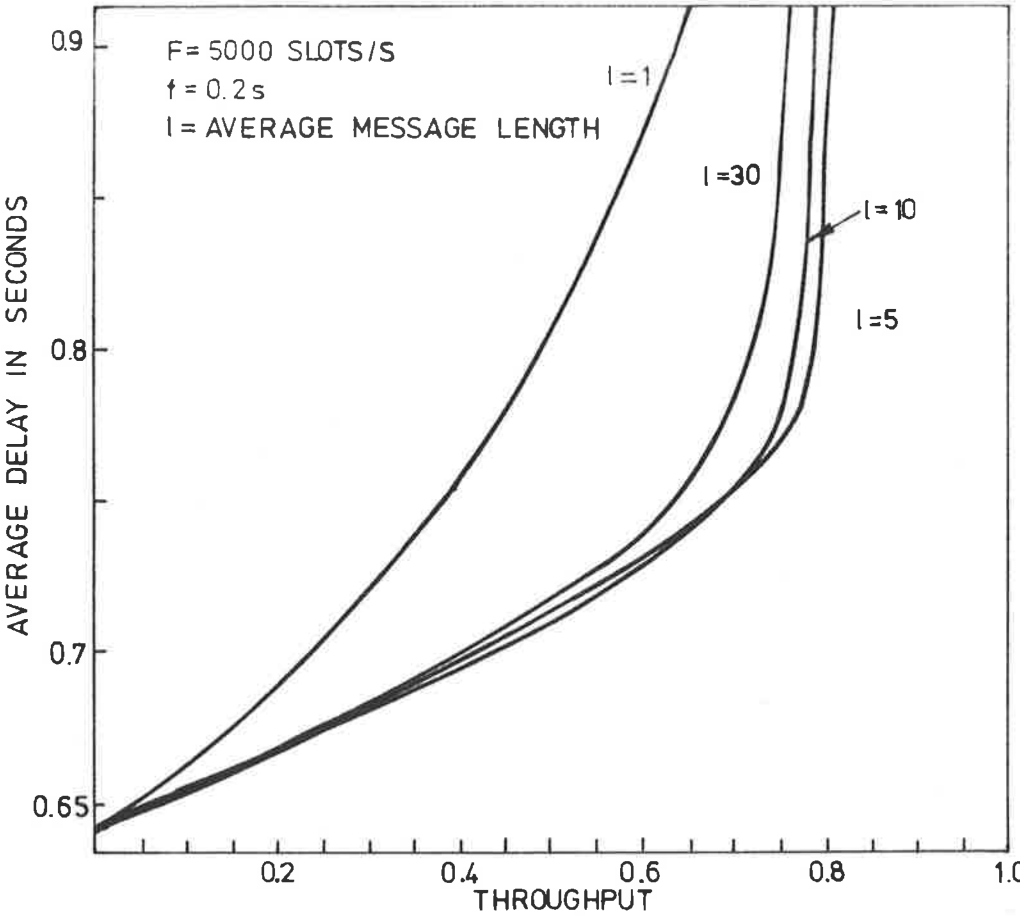


Fig.6.1.5 Delay performance comparison of the DA Protocol with different average message lengths

The total delay, reservation delay and scheduling delay versus throughput with the average message length as a parameter for a framelength of 0.4 second are shown in Fig.6.1.6, 6.1.7 and 6.1.8 respectively. The size of the control burst is increased to two , in order to avoid excessive reservation delay due to the doubling of the framelength. Despite such compensation, the reservation delay is greater than the case of framelength equal to 0.2 second because of the intrinsic average reservation delay of half a framelength. A careful study of the three figures gives similar trends as for the case of framelength equal to 0.2 second. Hence, the change in framelength does not affect the trends in any of the delay versus throughput comparisons.

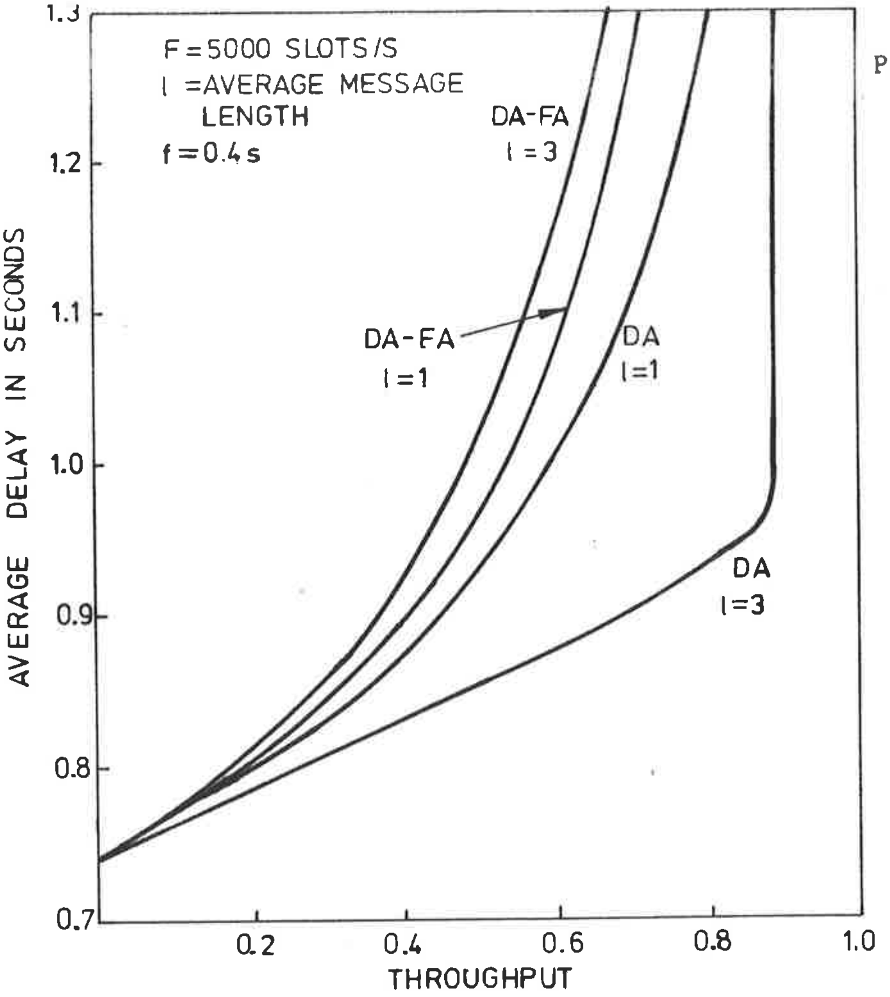


Fig.6.1.6 Delay performance comparison between the DA and DA-FA Protocols for a framelength of 0.4 second

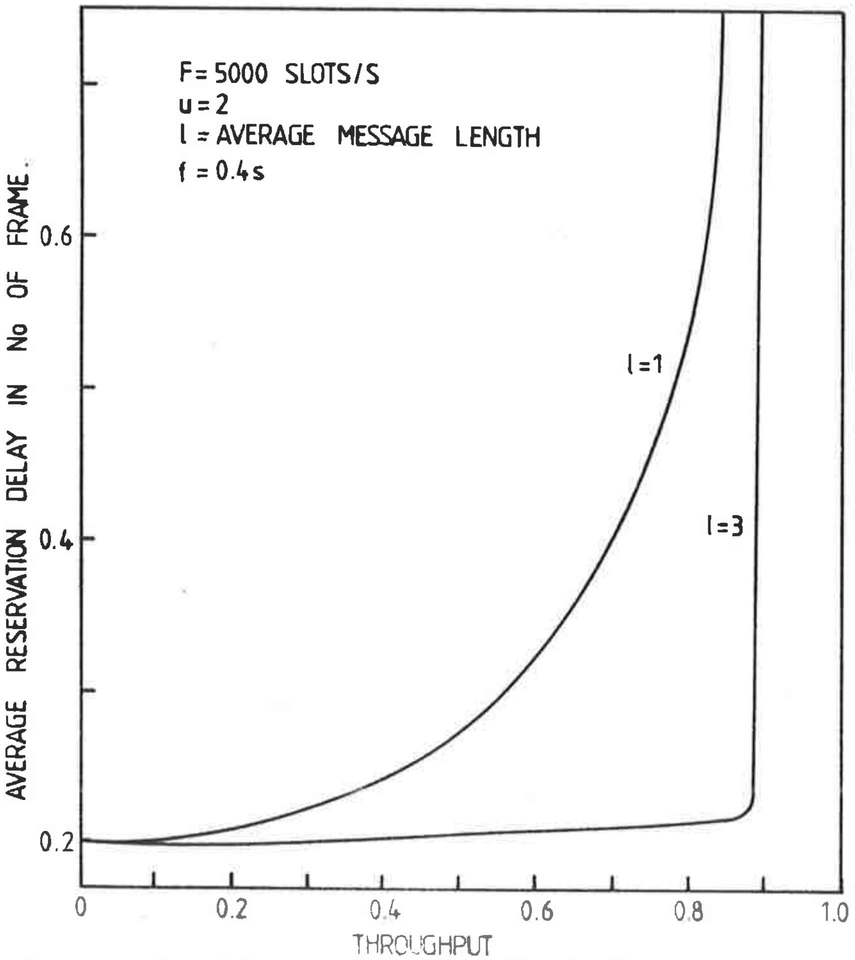


Fig.6.1.7 Reservation delay performance for both the DA and DA-FA Protocols for a framelength of 0.4 second

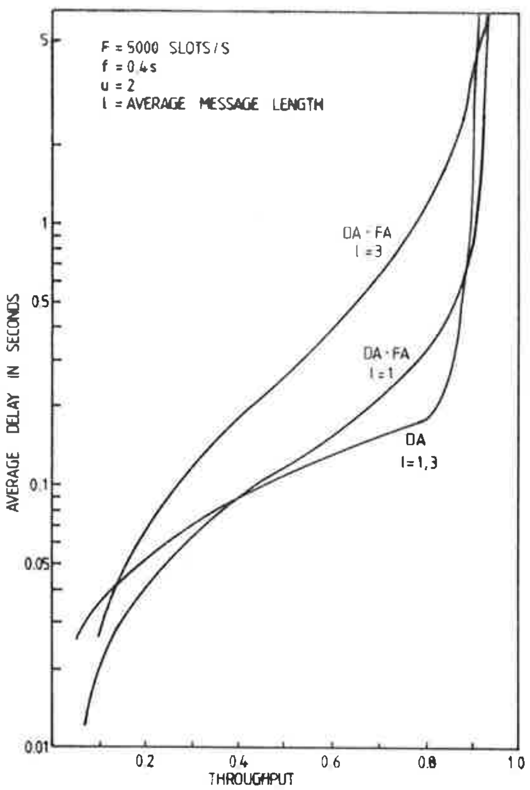


Fig.6.1.8 Scheduling delay performance for both the DA and DA-FA Protocols for a framelength of 0.4 second

B. Delay-throughput comparisons with varying protocol parameters

The sensitivity of the delay-throughput performance with variation in protocol parameters such as control burst size (u) and framelength (f) are studied in this section. Fig.6.1.9 gives the reservation delay versus throughput with the size of the control burst (u) as parameter. As expected, the higher the value of u, the smaller the reservation delay. Hence, one could reduce the total average message delay to half a framelength by increasng the size of the control burst. Recalling Equation (6.1.5), the stability of the control channel can also be improved by an increase in control burst size. The penalty one pays for such an improvement is a reduction in the capacity of the information channel. However, due to the large overhead involved, the control burst size is not linearly proportional to the control burst length, i.e. by doubling the burst length, one could have more than a twofold increase in control burst size.

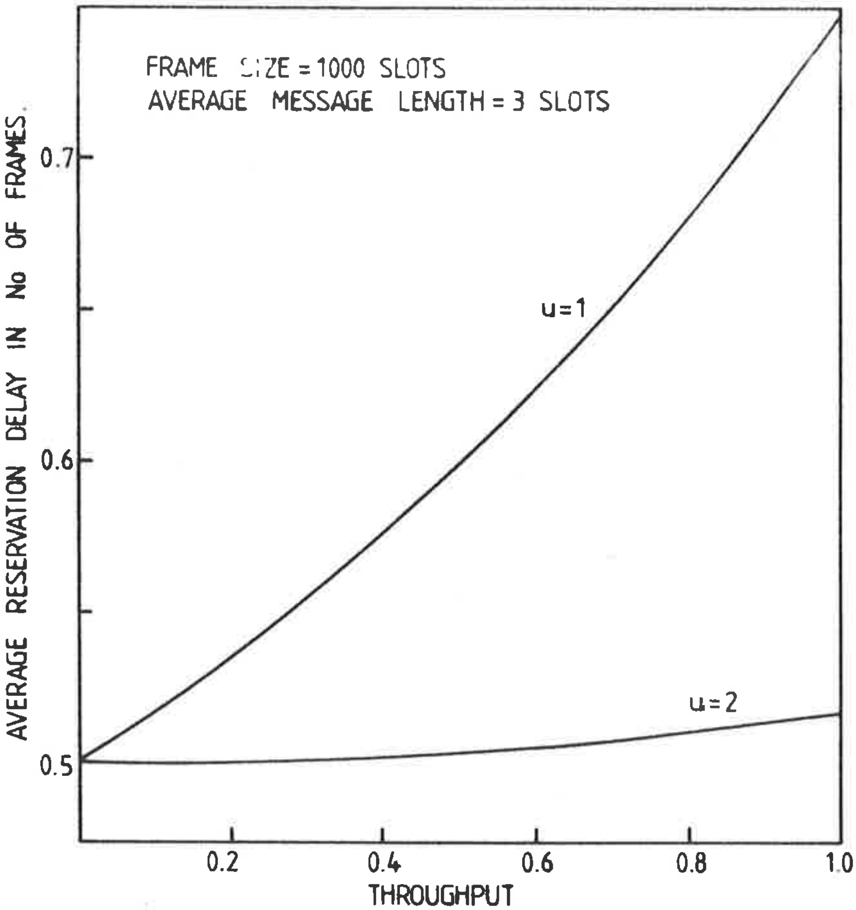


Fig.6.1.9 Reservation delay performance with the burst size (u) as a parameter

The effect of changing the framelength in the DA Protocol is shown in Fig.6.1.10. With the delay performance for the case of short framelength far outperformed the long one. In the case of the DA-FA Protocol, the longer framelength has a superior delay performance in the region of high throughput. To analyse the performance in greater depth, Fig.6.1.11 shows the two contributing delay components versus throughput with their framelength (f) as parameter.

For the reservation delay, the case of framelength equal to 0.2 second has a far superior performance than the case of 0.4 second, despite the compensating increase of control burst size from one to two. This is due to the intrinsic minimum of half a framelength reservation delay. For the random variable portion of the scheduling delay, the case of longer framelength is superior in delay performance for the DA-FA Protocol due to the larger allocated capacity. With the DA Protocol, the opposite trend occurs due to the longer framelength caused a greater amount of offered

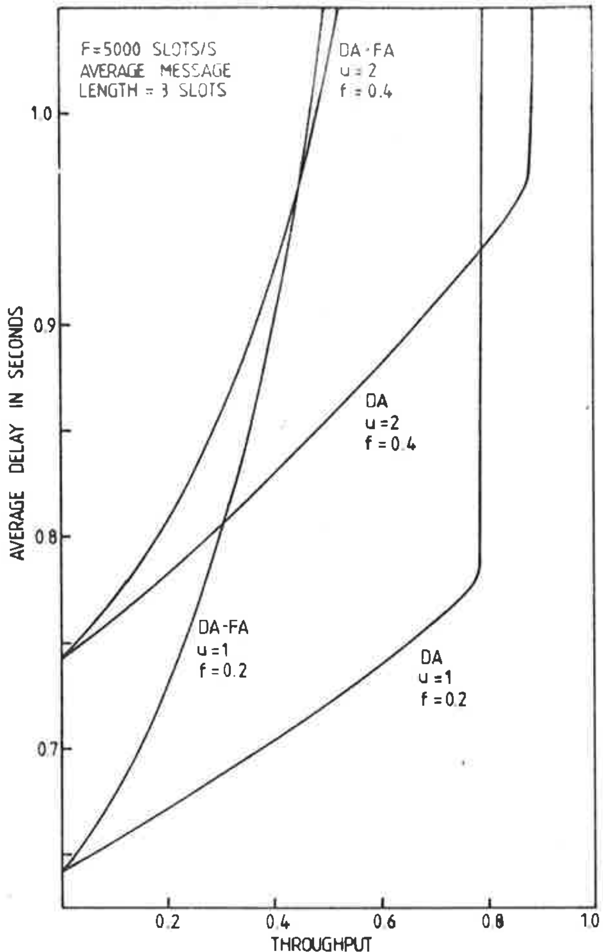


Fig.6.1.10 Delay performance comparison between the DA and DA-FA Protocols with two different framelengths and an average message length of three slots

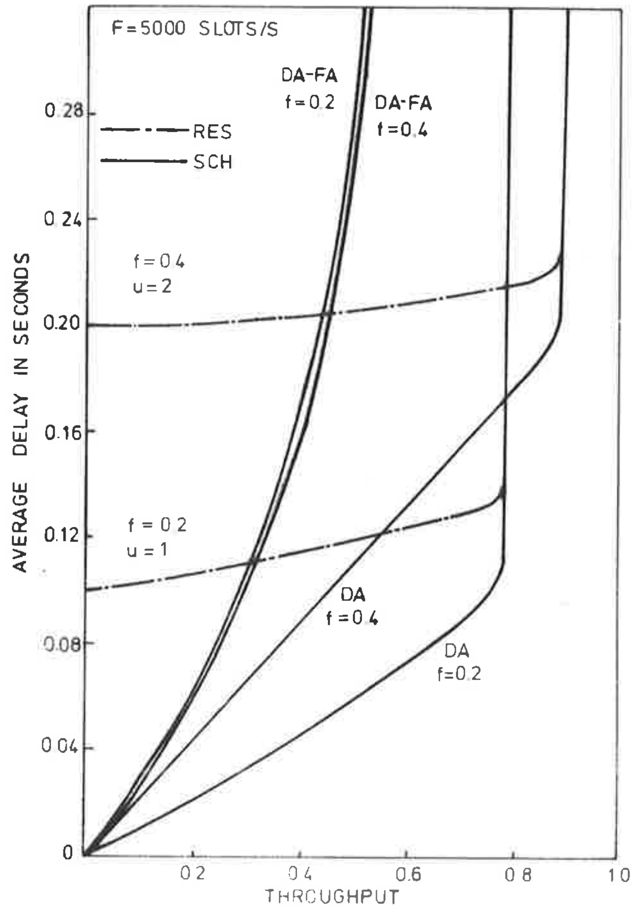


Fig.6.1.11 Reservation and scheduling delay comparisons between the DA and DA-FA Protocols with two different framelengths and an average message length of three slots

traffic, and thus longer scheduling delay. However, the penalty one pays for the superior delay performance achieved by shorter framelength is a smaller data channel capacity. A further reduction in framelengh is not advisable due to the large control overhead required.

C. Delay-throughput comparisons with varying network parameters

The change in network parameters such as the number of stations (s) and zones (K) in the network, and the transponder capacity (F) influence the delay-throughput performance. Fig.6.1.12 shows the relationship between the message delay and the number of zones for a fixed throughput and number of stations. In the case of the DA Protocol, the average message delay is independent of the number of zones. For the DA-FA Protocol, the average message delay increases with increasing number of zones. As the number of zones increases the reservation delay remains constant but the scheduling delay increases due to the reduction in assigned capacity. Finally the performance of the FA Protocol deteriorates as the number of zones increases, since the framelength (f) is directly proportional to the number of zones and the longer the framelength, the longer the corresponding scheduling delay.

Variation in the number of stations (s) for the DA Protocol is made with a fixed amount of network traffic. If the number of stations is increased, which leads to a corresponding increase in control channel capacity, it introduces two opposing effects. First, the number of message arrivals into each individual station (λ_i) is reduced. Hence the average reservation delay is reduced, as the assumption of equal traffic distribution amongst all the stations is used. Second, the increase in control channel capacity gives rise to a reduction in data channel capacity, hence the average scheduling delay increases. Fig.6.1.13 illustrates the two opposing trends and indicates the existence of a range

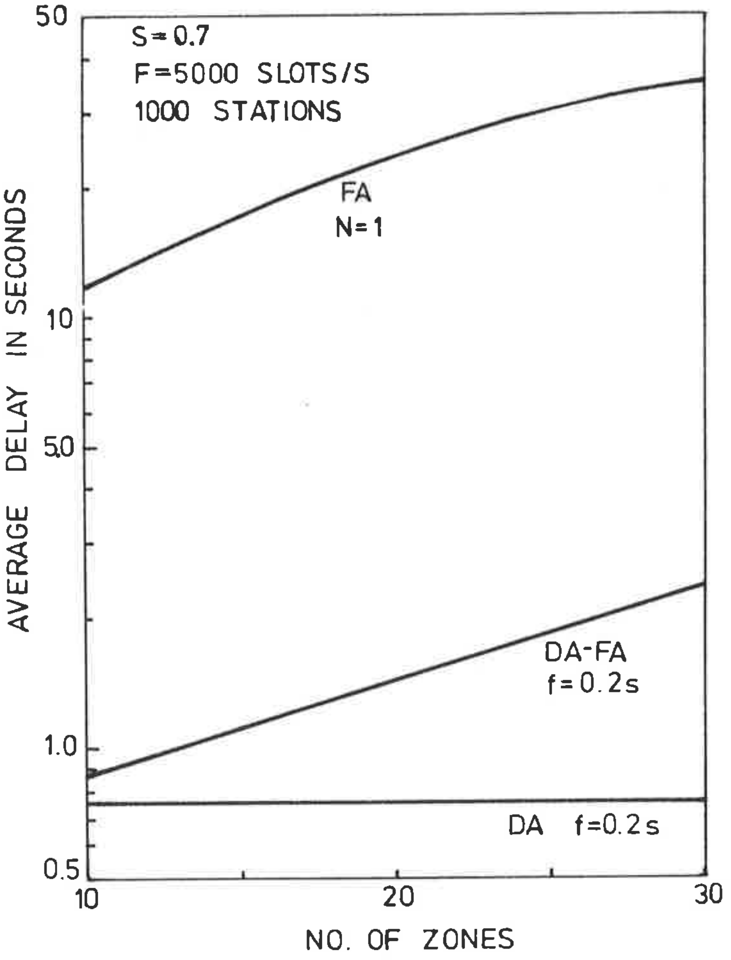


Fig.6.1.12 Delay versus number of zones for the SDMA protocols

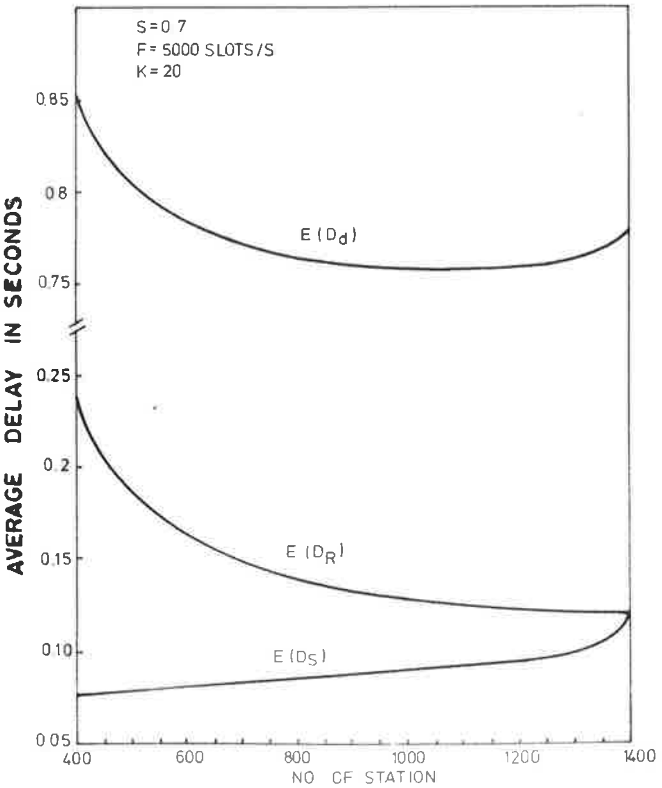


Fig.6.1.13 Delay versus number of stations for the DA Protocol

of station numbers for which the summation of the two delay components is a minimum.

A similar explanation can be used for Fig.6.1.14 which shows delay versus station numbers for the DA-FA Protocol. In this case, the average scheduling delay has a larger gradient than the previous example. The trough now exists in the lower range of station numbers. For the FA Protocol, any alteration in the number of stations has a proportional effect on the total delay, since the framelength is directly related to the number of stations in the network.

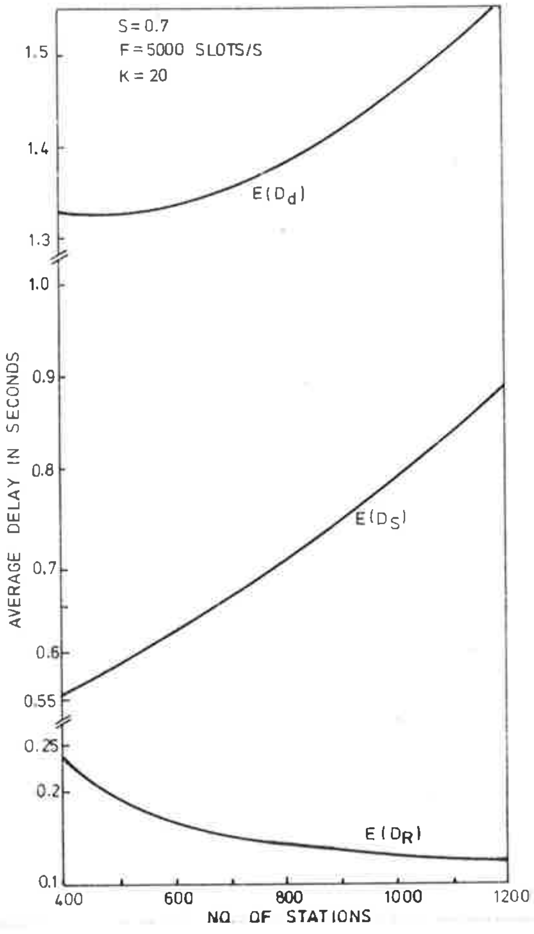


Fig.6.1.14 Delay versus number of stations for the DA-FA Protocol

Knowing the throughput (S)

$$S = \lambda/(\theta Ff)$$

There exist two possible effects as the transponder capacity F (slots/s) is varied. First, for a network with a fixed number of arrivals, the throughput decreases as the capacity increases. The impact on the delay performance gives the case of variation of capacity on an established terrestrial network. Second, for a network of fixed throughput, the increase in network arrivals is balanced by a corresponding increase in transponder capacity. This is used to investigate the impact on delay performance as the capacity is upgraded to compensate for the traffic growth.

Fig.6.1.15 gives the delay versus transponder capacity (F), for a terrestrial network with a fixed number of arrivals. For all three SDMA protocols, the delay decreases as the transponder capacity increases. The scheduling delay of the DA Protocol decreases as the throughput decreases. The improvement for the DA-FA Protocol is due to the following two factors :-

1. decrease in throughput;
2. increase in the capacity of assigned capacity as the transponder capacity increases.

Finally, in the case of the FA Protocol, the delay decreases as the required framelength to connect every station decreases.

Fig.6.1.16 depicts delay versus transponder capacity (F) for a fixed throughput, i.e. the increase in capacity is balanced by a similar increase in message arrivals. With the assumption of uniform traffic distribution, the required reservation delay increases as the number of arrivals into each station increases. For the DA Protocol, there exists an optimal transponder capacity for a fixed set of parameters. This is due to the combined effects of two opposing trends. They are an increase in the number of message arrivals, and subsequent reservation delay, due to the increases in capacity; and the decrease in scheduling delay due to an increase in data channel capacity. With the specified set of parameters, the optimal range of transponder capacity for an average message length of

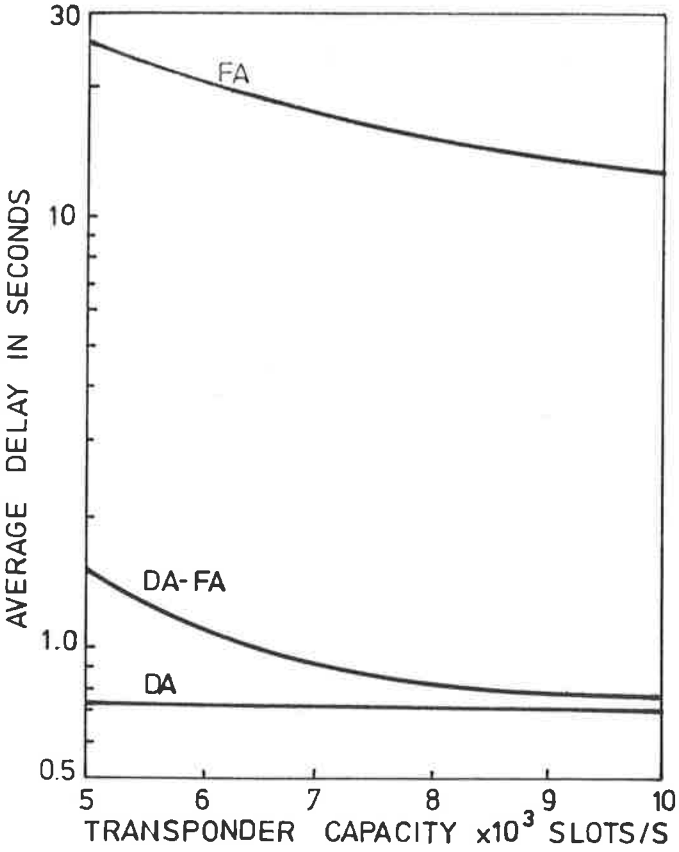


Fig.6.1.15 Delay versus transponder capacity with a fixed input

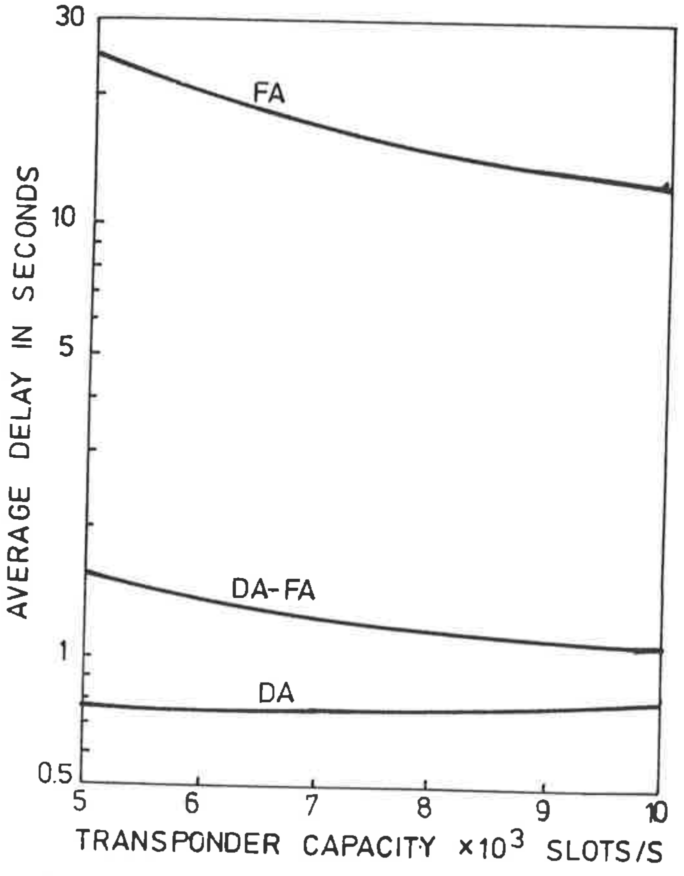


Fig.6.1.16 Delay versus transponder capacity with a fixed throughput

three slots per message is in the vicinity of 5000 to 8000 slots/s. In the case of the DA-FA Protocol, the increase in reservation delay is more than compensated by the decrease in scheduling delay. For the FA Protocol, the total delay again decreases as the framelength is reduced.

D. Concluding remarks on the comparisons of the SDMA protocols

In summary, all three SDMA protocols exhibit a monotonically increasing delay with respect to an increasing channel throughput. The FA Protocol was found to be unsuitable for large networks, because of its requirement of a long framelength in order to maintain full connectivity. With the two demand assignment protocols, if the size of the reservation burst is increased, the reservation delay decreases to the intrinsic value of half a framelength. However reduction in data channel capacity occurs. The DA Protocol has the superior delay performance for traffic with a long average message length, while the DA-FA Protocol is best with a short average message length. An increase in framelength will lead to an inferior delay performance in both DA and FA Protocols. However, it brings an improvement in the DA-FA Protocol in the region of high throughput due to an increase in assigned data channel capacity.

For the investigation of the sensitivity of the delay performance with respect to the variation in network parameters, the uniform traffic distribution assumption is adopted. If the number of zones in the network increased, a corresponding increase in delay for both the DA-FA and FA Protocols resulted, while the performance of the DA Protocol is independent of the number of zones. The DA-FA Protocol suffers because the capacity per zone is reduced, while the FA Protocol suffers because of the increase in framelength. Secondly, if the number of stations in the network is increased, an increase in framelength and delay results for the FA Protocol. However, there exists a optimal range of station numbers for a fixed set of traffic and protocol parameters with both DA and DA-FA Protocols. This is due to the combined effects of two opposing consequences of increasing the number of stations. They are the decrease in average reservation delay due to a reduction in traffic arriving into each station; and the increase in average scheduling delay due to a reduction in data capacity assignment, and with increasing capacity allocated to the control channel.

To investigate the effect of a change in transponder capacity on the delay performance with a fixed traffic environment, one looks at the delay performance with varying transponder capacity and fixed average network arrivals (λ/θ). The results are a reduction in delay for all three SDMA protocols. To investigate the effects of upgrading the transponder capacity to compensate for the traffic growth, one looks at the delay performance with varying transponder capacity and fixed channel throughput (S). This leads to an improvement for both the FA and DA-FA Protocols. However, an optimal transponder capacity exists for the DA Protocol due to the opposing trends of the reservation and scheduling delays in response to the varying transponder capacity.

Hence, the FA Protocol is more suitable for a small network with regular short message traffic. The DA Protocol operates with superior delay performance under the condition of short framelength and long message length, while the DA-FA Protocol is ideal for short message length traffic operating in a network with a small number of zones and a large transponder capacity. In both DA and DA-FA Protocols, where an optimal number of stations in the network exists, the reservation delay can be reduced by increasing the size of the control burst.

The results of these sensitivity studies are summarised in Table 6.1.1. From the network designer point of view, with a set of traffic characteristics and operating criteria such as maximum permissible message delay; the topology of the terrestrial network and certain protocol parameters can be optimized in order to reduce the message delay.

Protocol Variation	FA	DA-FA	DA
Long message length	X	X	/
Increase burst size (u)	-	/	/
Increase framelength	X	/	X
Increase no. of zones (K)	X	X	-
Increase no. of stations	X	0	0
Increase transponder capacity with fixed arrivals	/	/	/
Increase transponder capacity with fixed throughput	/	/	0

X unsuitable
/ suitable
0 suitable for an optimal range
- unaffected

Table 6.1.1 Sensitivies of the SDMA Protocols

6.2 COMPARISONS WITH OTHER GLOBAL BEAM COVERAGE MULTIPLE ACCESS PROTOCOLS

The Space Division Multiple Access (SDMA) protocols are designed for a multiple beam coverage environment. Hence, they are incompatible with protocols which operate in a global beam condition. However, in order to gauge the relative merits of the SDMA protocols, the performance of these global beam protocols are used for comparison.

In selecting global beam protocols for comparison, one protocol is selected to represent each of the major multiple access techniques (Chapter 1). They are Time Division Multiple Access (TDMA), Slotted Aloha, Reservation Aloha and Split-channel Reservation Multiple Access (SRMA). The expressions for delay-throughput for each of these protocols can be found in Appendix 2. The Demand Assignment uplink and downlink (DA) Protocol is chosen to represent the technique of SDMA because of its superior delay-throughput performance. Again, the average message delay for each protocol is investigated with varying message length ($1/\theta$ slots), number of stations (s) in the network and the transponder capacity (F slots/s).

6.2.1 Delay-Throughput Comparisons

The following set of parameters are chosen for the comparisons

a. number of control slots in a data slot $v = 10$

b. number of zones $K = 20$

c. number of stations s = 1000

d. transponder capacity F = 5000 slots/s

For each of the individual protocols, the following parameters are used :-

1. DA framelength f = 0.2 s

number of required control slot/station = 2.0

control burst size u = 1, 2

2. TDMA framelength f = 0.2 s

3. SRMA and Slotted Aloha

retransmission maximum randomized delay = 5 slots

4. Reservation Aloha framelength f = 0.272 s

retransmission maximum randomize delay = 5 slots

The delay-throughput performance for all five protocols is shown in Fig.6.2.1, 6.2.2 and 6.2.3 for a message length of one, three and ten slots per message respectively. The maximum data channel assignment (δ) for the DA Protocol is related to the number of stations (s) in the network; i.e. the larger the control channel, the smaller the data channel. Two curves are drawn for the DA Protocol in each figure; for a control burst size (u) of one and two control packets per burst. In the case of the Reservation Aloha and SRMA Protocols, the value of maximum data channel capacity is determined by the stability criteria of the control channel. For the TDMA Protocol, a framelength of 0.2 second is required such that each of the one thousand stations has one slot per frame each.

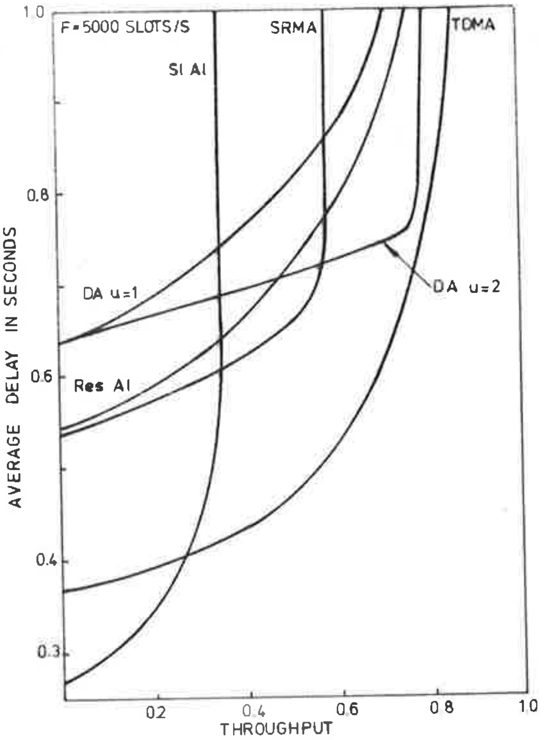


Fig.6.2.1 Delay performance comparison with the global beam protocols for a message length of one slot

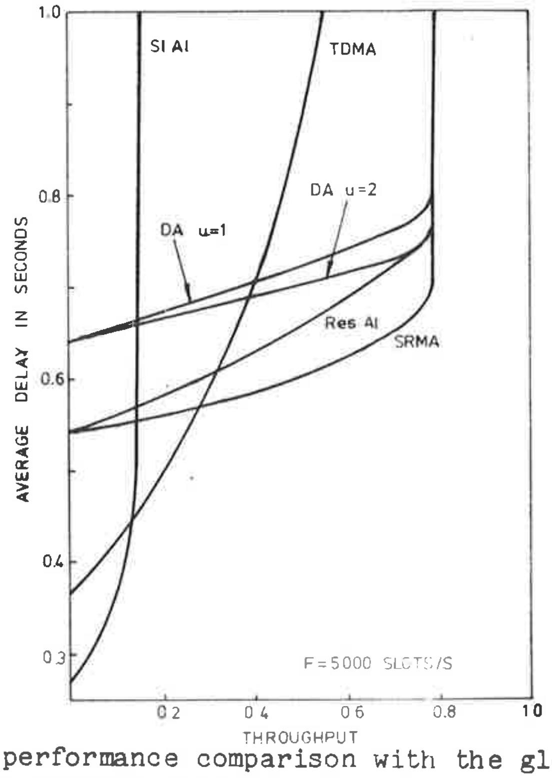


Fig.6.2.2 Delay performance comparison with the global beam protocols for an average message length of three slots

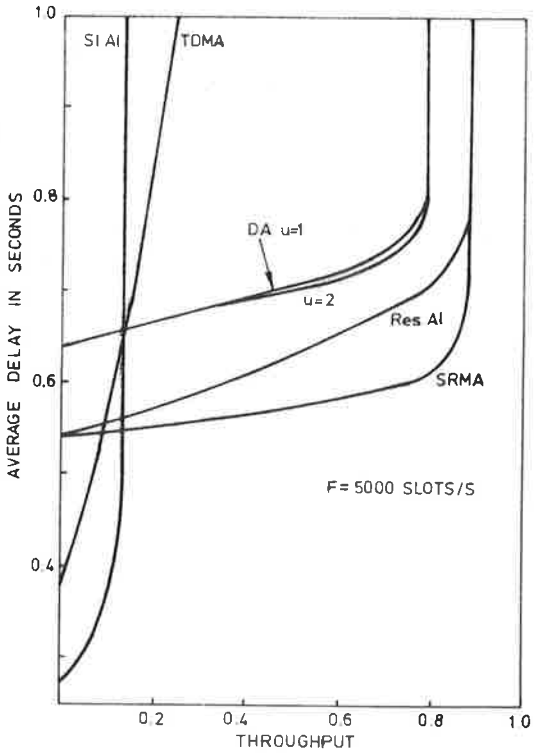


Fig.6.2.3 Delay performance comparison with the global beam protocols for an average message length of ten slots

The following oberservations can be made from the above three figures. The performance of the DA Protocol with control burst size of one is inferior to the Reservation Aloha Protocol which is in turn inferior to SRMA Protocol. However, with a small message length, the DA Protocol with control burst of two performs better than or comparable with Reservation Aloha Protocol at the region of high throughput. The superiority of the DA Protocol erodes as the message length increases. Since the performance of Reservation Aloha and SRMA Protocols improves more dramatically as the number of message arrivals decreases and the number of collisions and retransmissions of requests are reduced.

The Slotted Aloha Protocol performs extremely well in the region of low throughput ($S < 1/e$). The delay approaches infinity as the throughput approaches the throughput limit which is message length dependent. The performance of the TDMA Protocol is again message length dependent; i.e. the longer the message, the greater the delay. For a single slot message, TDMA is superior to DA, Reservation Aloha and SRMA Protocols.

In the case of varying station numbers for a fixed throughput of 0.7 and a fixed average message length of three slots per message (Fig.6.2.4), Slotted Aloha, Reservation Aloha and SRMA Protocols are independent of the number of stations in the network. The performance of Slotted Aloha is not shown in the graphs as it cannot reach a throughput of 0.7. On the other hand, the required framelength of the TDMA Protocol is directly proportional to the number of stations and therefore the average delay per message is station dependent. In the case of the DA Protocol, the reservation delay decreases as the number of message arrivals into each station decreases, since the total amount of network traffic is fixed. If the number of stations increases further, the capacity of the control channel also needs to increase. Consequently, the available bandwidth for the data channel reduces and the scheduling delay increases. Hence, there exists an optimal range of station numbers for which the total delay is minimum, as shown in Section 6.1.

The transponder capacity can be varied either with a fixed number of arrivals (λ/θ) or a fixed throughput (S). These implications are explained in Part C of Section 6.1.2. With the case of fixed arrivals, Fig.6.2.5 gives the performance comparison for various protocols as the transponder capacity is varied. The comparison is made with a throughput value of 0.7 and a bandwidth assignment of 0.8 for Reservation Aloha and SRMA Protocols. For all protocols, the average delay per message decreases as the capacity increases. In the case of TDMA Protocol, the decrease in delay is due to the reduction in framelength. For Reservation Aloha and SRMA Protocols,

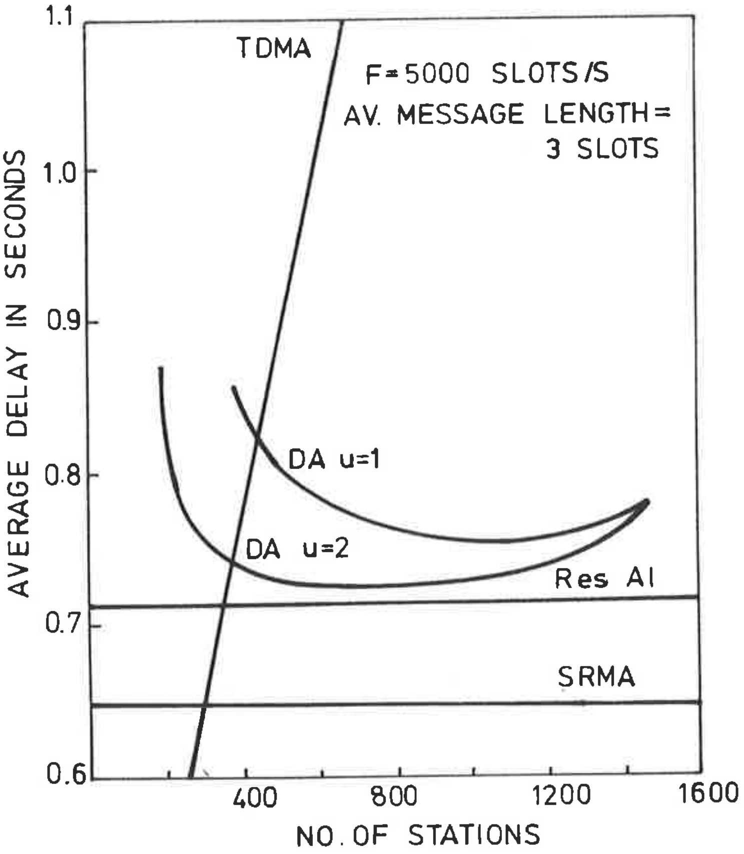


Fig.6.2.4 Delay versus number of stations

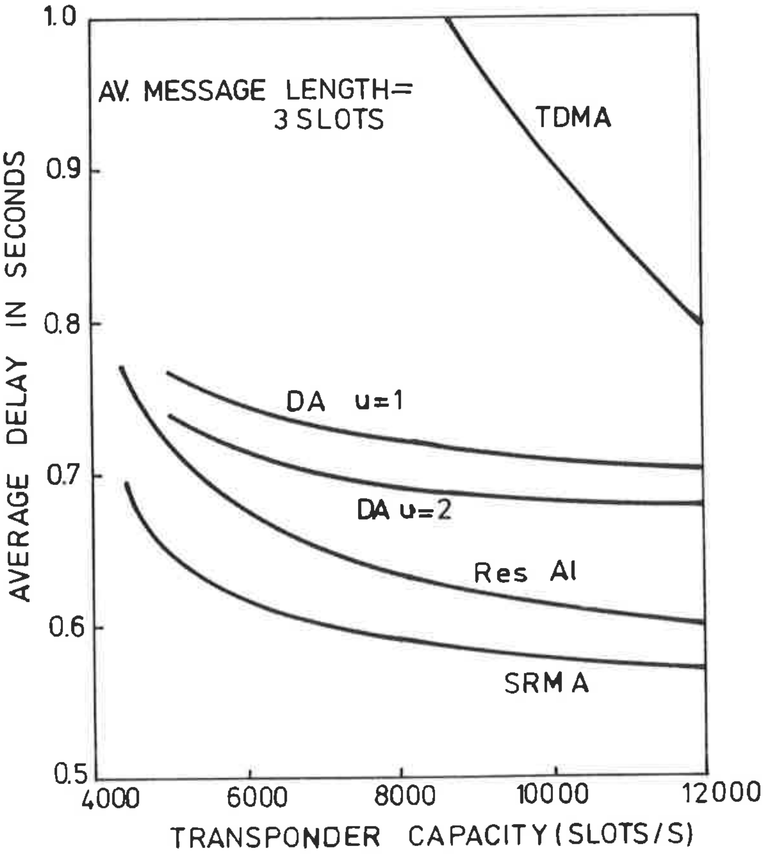


Fig.6.2.5 Delay versus transponder capacity with a fixed input

both the reservation and the scheduling delay decreases as the capacity increases. In the case of the DA Protocol, the reservation delay remains constant as the number of message arrivals is constant, only the scheduling delay decreases as the bandwidth increases.

For the other condition of fixed throughput values, TDMA, Reservation Aloha and SRMA Protocol behave similarly as in the previous condition of fixed number of arrivals (Fig.6.2.6). However for the case of the DA Protocol, there are now two opposing trends :- First, the message arrivals and subsequent reservation delay are directly proportional to the change in transponder capacity. Second, the scheduling delay decreases as the data channel capacity increases. Such trends create an optimal range of transponder capacity for a fixed set of conditions.

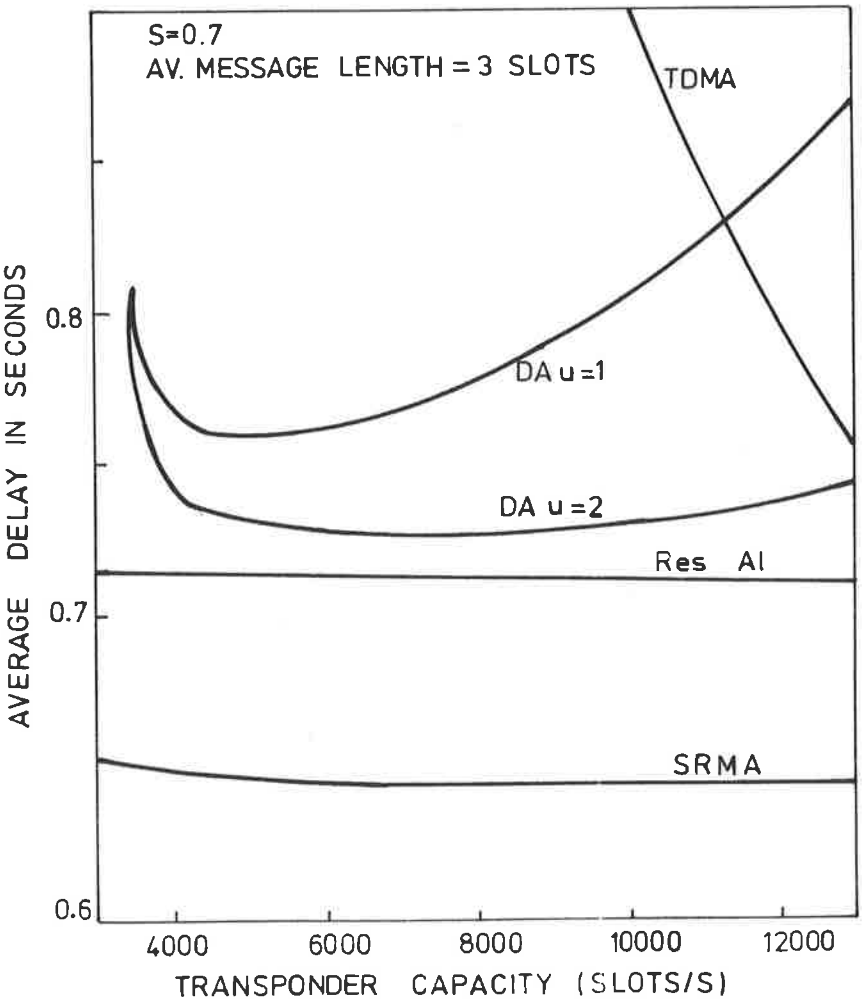


Fig.6.2.6 Delay versus transponder capacity with a fixed throughput

To sum up, a Demand Assignment Protocol with an ability to carry multiple sets of information by each control burst has a delay performance comparable with other distributed control or centrally controlled reservation protocols designed for the global beam application, especially with short message length traffic. While most other protocols are independent of the number of stations in the network or the change in transponder capacity, there exists an optimal range of station numbers and transponder capacity for the DA Protocol.

6.3 BUFFER PERFORMANCES AND COMPARISONS OF THE SDMA PROTOCOLS

Once the relations between the data multiplexing and demultiplexing buffer blocking probabilities and various system parameters are established, the changes in blocking probabilities with respect to the variations in parameters can be investigated. These parameters include :

buffer size = q slots
frame capacity = Ff slots
data channel capacity = N slots/frame
average network message arrivals = λ /frame
average message length = $1/\theta$ slots/message
channel throughput = S
total no. of spotbeam coverages = K
total number of stations = s

It is assumed that the traffic is uniformly distributed amongst all the stations and zones.

In this section, the buffer performances of the three Space Division multiple Access (SDMA) protocols are investigated and compared. The buffer analyses are detailed in Section 5.3.1, where the distinctive characteristics of each of the protocols is highlighted. In the following examples on data buffer performances, the entire information channel is used for the data traffic.

6.3.1 Buffer Performances of the SDMA Protocols

With a Fixed Assignment (FA) Protocol, each earth station divides its offered traffic into K separate queues according to the destination addresses, since there are K zones in the network. At their allocated instant, a predetermined number of data units are sent from the station in synchronization with the satellite beam switching. Since no overhead is required in the FA Protocol the entire frame capacity is allocated to data channel.

For full connectivity, each station must maintain one or more slots for each zone per frame, i.e. the framelength is proportional to both K and s.

Number of slots assigned to traffic originated from station i and destinated to station j $= N_{ij}$

Frame capacity $Ff = N_{ij}Ks$

As the average number of arrivals per frame (λ) is directly proportional to the framelength, it is also proportional to the product of K and s .

λ = constant x Ks

The buffer blocking probability (B_{dm}) is primarily a function of the buffer size (q) per zone and channel throughput (S), where

$S = \lambda/(\theta Ff) = \lambda/(\theta N_{ij}Ks) = \text{constant}/(\theta N_{ij})$

Hence, in the case of the FA Protocol, S is independent of the number of zones and stations, but the buffer size per zone and hence B_{dm} is a function of the total number of zones.

The steady state buffer performance of the Demand Assignment (DA) Protocol and the distributed control Demand Assignment uplink and Fixed Assignent downlink (DA-FA) Protocol is obtained by the Two-Group Method [MAR 78]. With the required overhead taken into account, the DA-FA Protocol buffer analysis can be regarded as a scaled down version of the DA Protocol. However, the performance of the DA Protocol displays certain distinct characteristics, which are described in the following paragraphs.

With the DA Protocol, the entire data capacity is distributed amongst all the network traffic. The on-board traffic controller allocates capacity to each of the requesting stations and at the assigned instant, the stored message is sent from the station. The amount of offered traffic into a station and hence data multilplexing buffer blocking probability (B_{dm}) are dependent on the number of stations and independent of the number of zones.

With the distributed control DA-FA Protocol, the uplink capacity is divided into K different segments, one for each zone. Within each segment, the downlink scans the K zones, i.e. every zone is connected to all other zones in the network. The uplink capacity is distributed amongst all the stations within the zone through the reservation channel. The stations then schedule the requests using the same algorithm. The stored messages are sent at the appropriate instants in synchronism with the connecting beam switching. With a uniform traffic assumption, the amount of offered traffic into a station or a zone is inversely proportional to the total number of stations (s) and zones (K) respectively. Furthermore, the amount of uplink capacity assigned to each zone and distributed amongst all the stations within the zone is inversely proportional to the number of zones (K). Hence, in the case of the DA-FA Protocol, B_{dm} depends on the number of zones as well as the number of stations.

A certain amount of capacity overhead is used for the periodic assessment of the required capacity of each station and the notification of the capacity allocation. The overhead is proportional to s and given by ks/v, where k is the number of control slots required per station and is protocol dependent, and v is the ratio of the number of control slots per data slot. In the following calculations, v is assumed to have arbitrary value of ten and k equal to two for the DA Protocol and one for the DA-FA

Protocol. This is due to the fact that the DA Protocol requires a reservation and a scheduling slot for each station, while the DA-FA Protocol requires the reservation slot only.

The relationship between the data multiplexing blocking probability (B_{dm}) and the buffer size (q) with a particular average message length, frame capacity and a network configuration, is given in Fig.6.3.1, 6.3.2 and 6.3.3 for the FA, DA and DA-FA Protocols respectively. B_{dm} increases as the channel utilisation increases; in addition for the case of the DA and DA-FA Protocols, B_{dm} increases as the number of stations decreases. In both cases, the number of message arrivals into a station is increased. The graphs indicate the buffer size required to satisfy a particular blocking probability and traffic condition. The staircase function of the FA Protocol emphasises the incremental nature of the required buffer size.

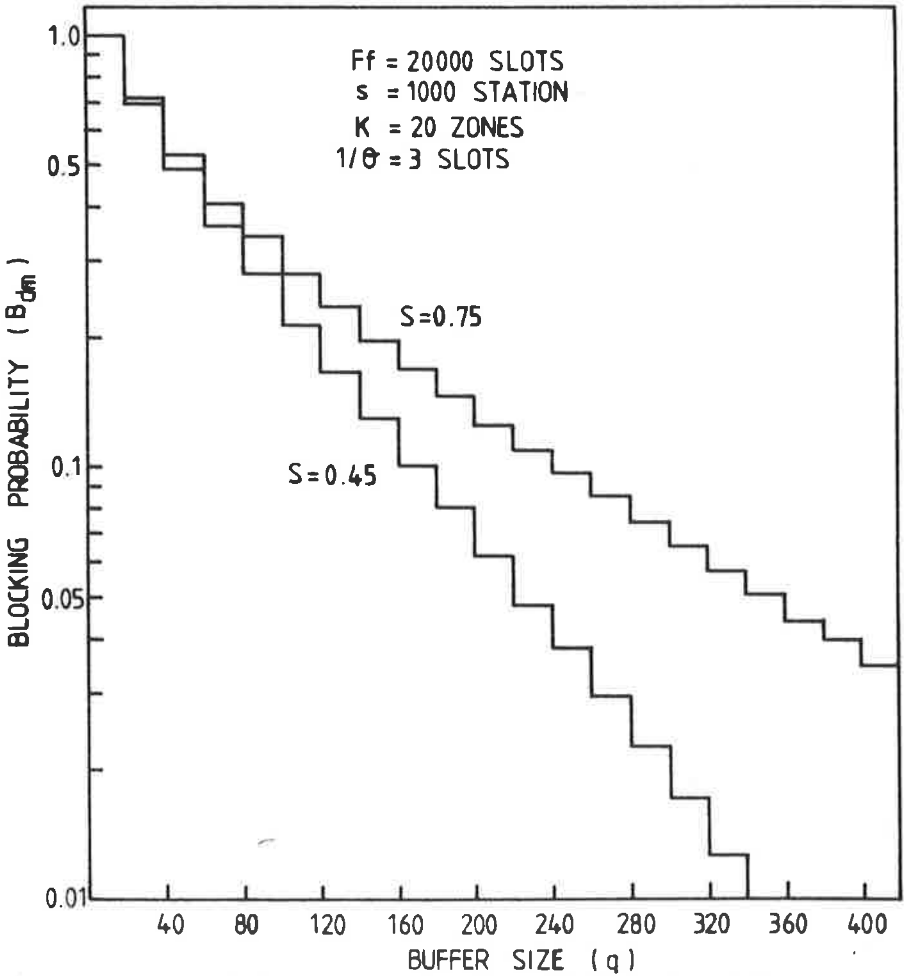


Fig.6.3.1 Data multiplexing buffer blocking probability versus buffer size for the FA Protocol with the channel throughput (S) as a parameter

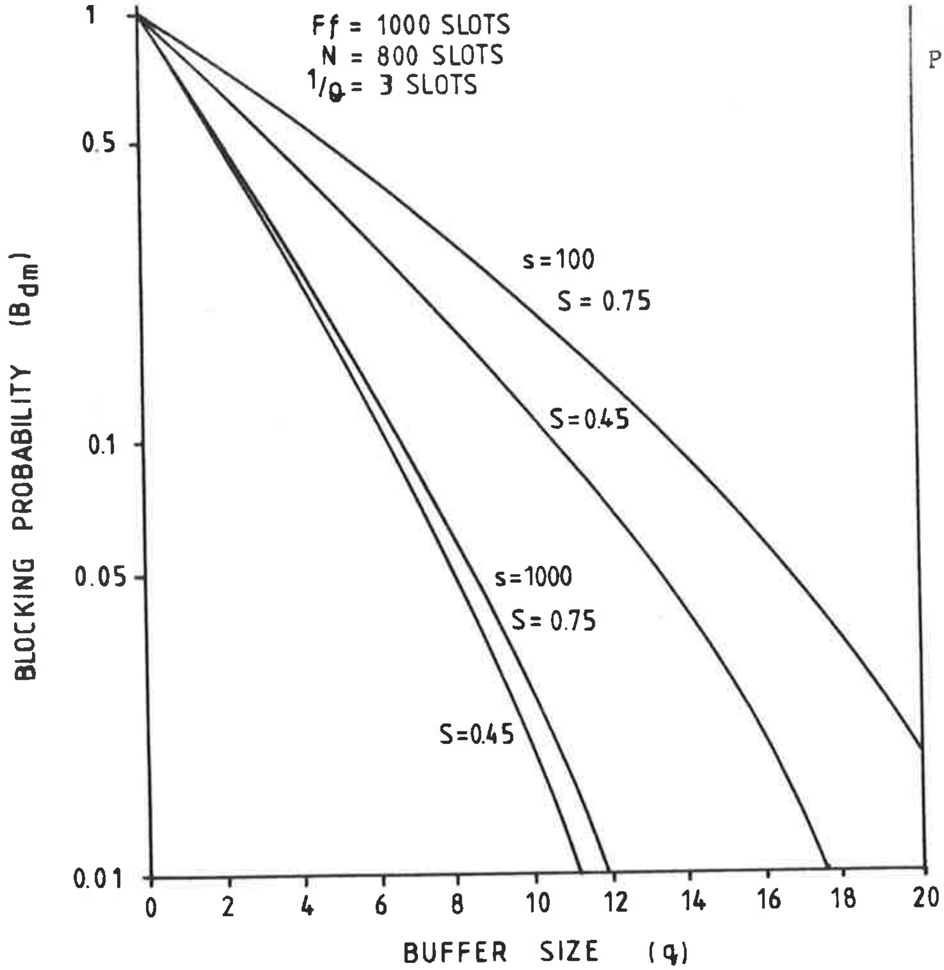


Fig.6.3.2 Data multiplexing buffer blocking probability versus buffer size for the DA Protocol with the channel throughput (S) and station number (s) as parameters

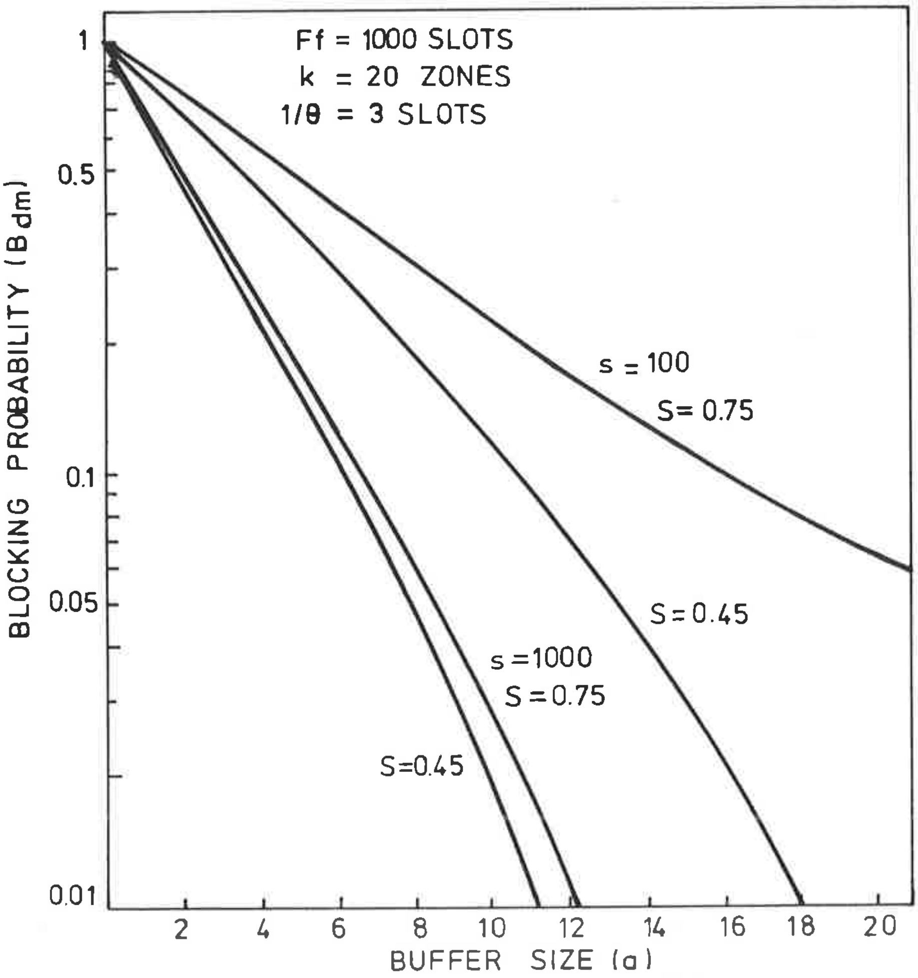


Fig.6.3.3 Data multiplexing buffer blocking probability versus buffer size for the DA-FA Protocol with the channel throughput (S) and station number (s) as parameters

More insight into the buffer behaviour can be obtained by examining the following expressions of the channel throughput :-

$$S = \text{channel throughput} = \frac{\text{total arrivals per frame}}{\text{frame capacity}} \qquad (6.3.1)$$

$$S_d = \text{data channel throughput} = \frac{\text{total arrivals per frame}}{\text{data channel bandwidth}} \qquad (6.3.2)$$

In the case of the FA Protocol, the data channel throughput is the same as the channel throughput, because no overhead is involved.

$$S = S_d = \lambda/(\theta Ff) \qquad (6.3.3)$$

However, in the case of the DA and DA-FA Protocols, the data channel throughput is given by

$$S_d = \frac{\lambda}{\theta(Ff - \text{control channel capacity})} \qquad (6.3.4)$$

Hence the channel throughput (S) and consequently the blocking probability (B_{dm}) will increase if

1. the number of message arrivals per frame (λ) increases,or

2. the average message length ($1/\theta$) increases.

In the case of the FA Protocol, full network connectivity can only be maintained through a fixed number of slots per frame. S and B_{dm} will increase if the transponder capacity (F) is reduced since this leads to a proportionate increase in the framelength (f) and the number of arrivals (λ). Besides a larger data multiplexing blocking probability, a smaller capacity transponder also leads to the penalty of longer message delay (due to the longer framelength).

For the DA and DA-FA Protocols, S and consequently B_{dm} increases under the following additional conditions :-

1. the transponder capacity (F) decreases, or

2. the framelength (f) decreases.

The effects on the blocking probability of an increase in average message length for a fixed channel utilisaton are shown in Fig.6.3.4, 6.3.5 and 6.3.6 for the FA, DA and DA-FA Protocols respectively. As the average message length increases, the blocking probability increases despite the reduction in number of message arrivals. The graphs show that a larger buffer is needed for lengthy traffic in order to maintain the same blocking probability.

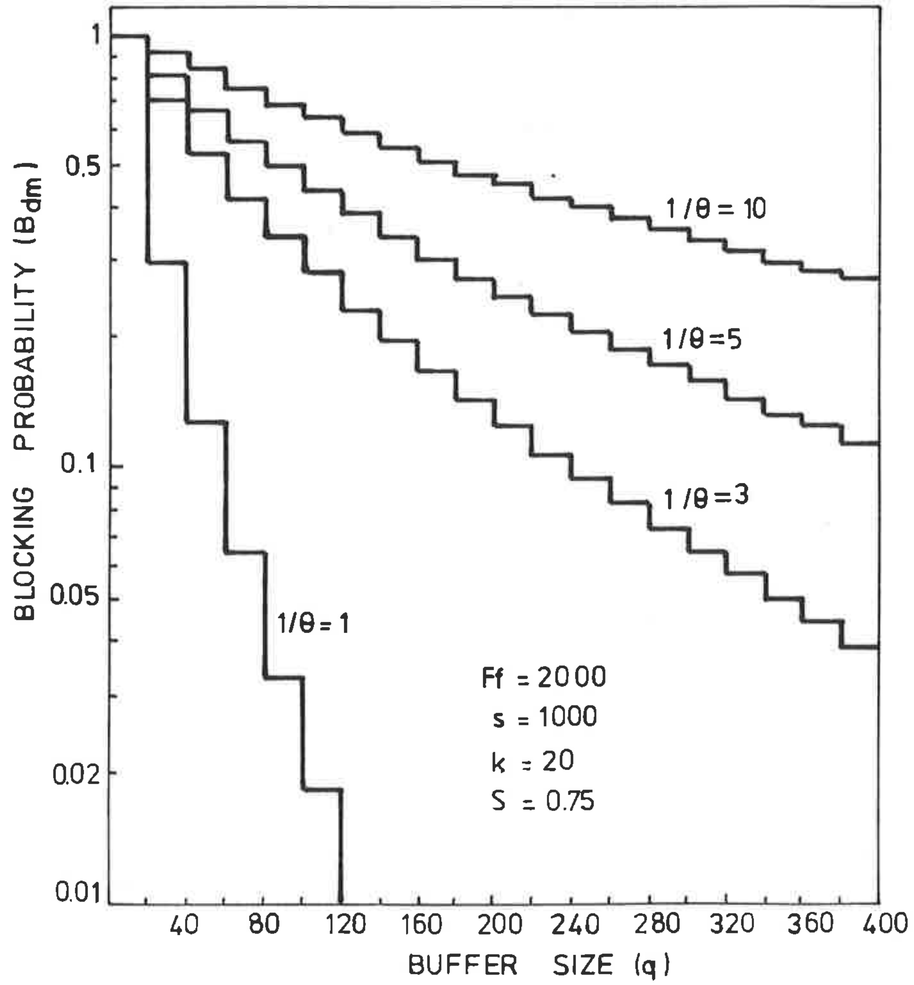


Fig.6.3.4 Data multiplexing buffer blocking probability versus buffer size for the FA Protocol with the average message length ($1/\theta$) as a parameter

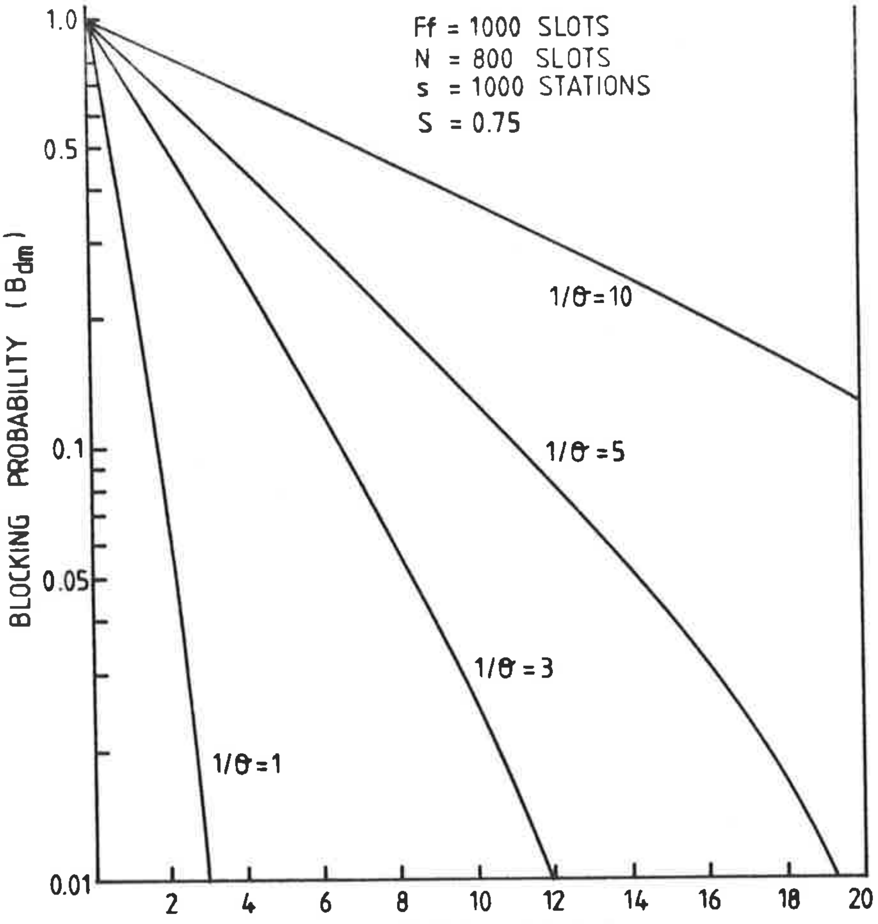


Fig.6.3.5 Data multiplexing buffer blocking probability versus buffer size for the DA Protocol with the average message length ($1/\theta$) as a parameter

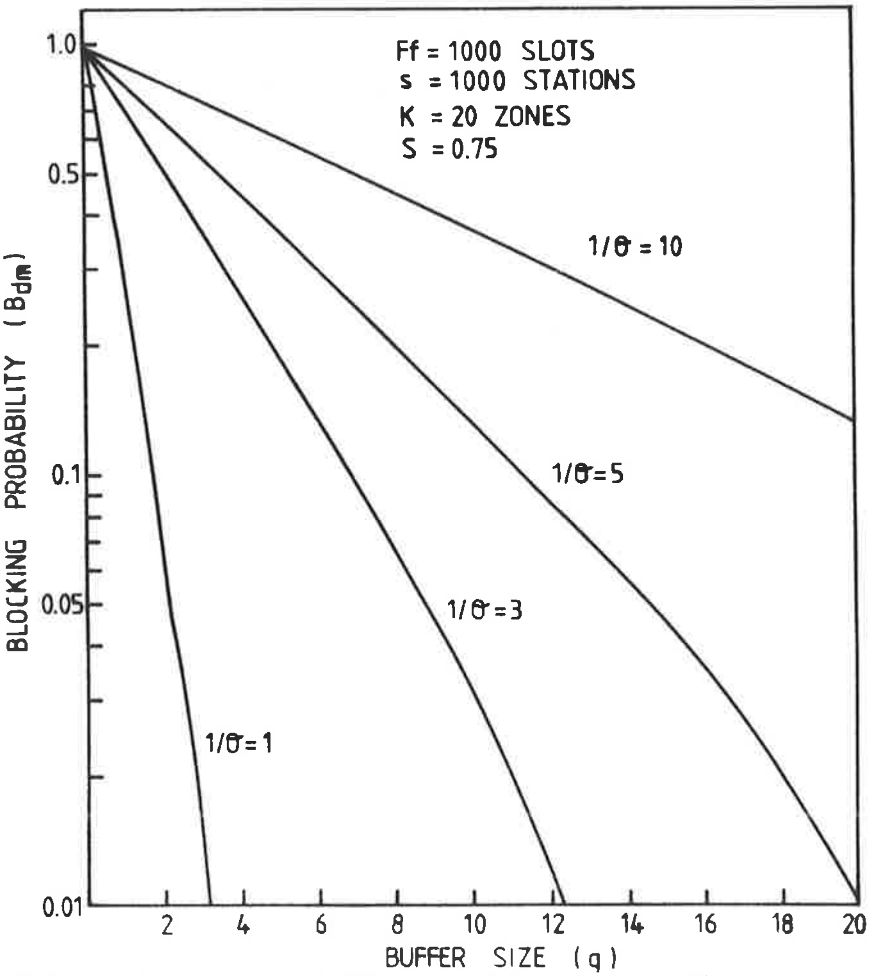


Fig.6.3.6 Data multiplexing buffer blocking probability versus buffer size for the DA-FA Protocol with the average message length ($1/\theta$) as a parameter

The relation between the terrestrial network and the blocking probability is shown in Fig.6.3.7 and 6.3.8 for the DA and DA-FA Protocols respectively. They give the buffer sizes required to satisfy a certain blocking probability for a network with s stations. When s is large, the average amount of offered traffic at a station is comparatively small if a uniform traffic distribution is assumed. The two graphs indicate a smaller buffer size can be used in each of the s stations. Hence, the terrestrial network design is a compromise between the buffer size per station and the total number of stations in the network.

The FA and DA Protocols are independent of the number of zones (K) in the network from the nature of their buffer architecture. Although the DA-FA Protocol is dependent on K, the insensitivity of the multiplexing buffer blocking probability is shown in Fig.6.3.9.

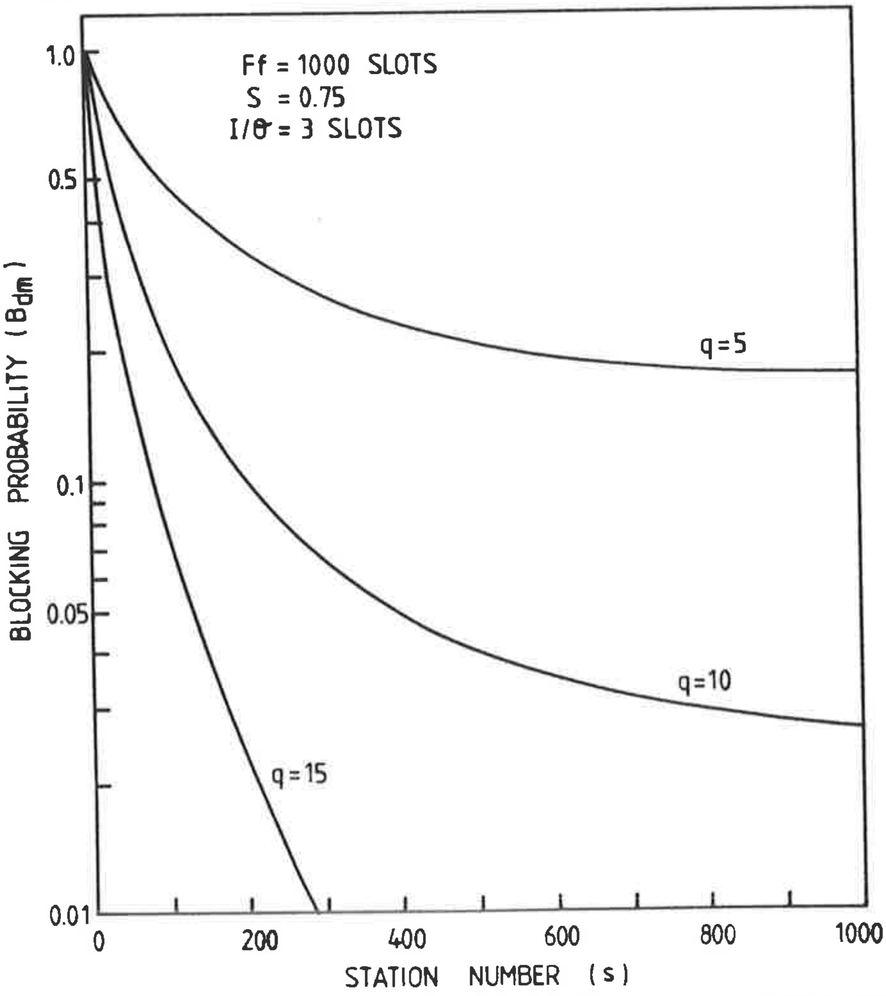


Fig.6.3.7 Data multiplexing buffer blocking probability versus number of stations (s) for the DA Protocol with the buffer size (q) as a parameter

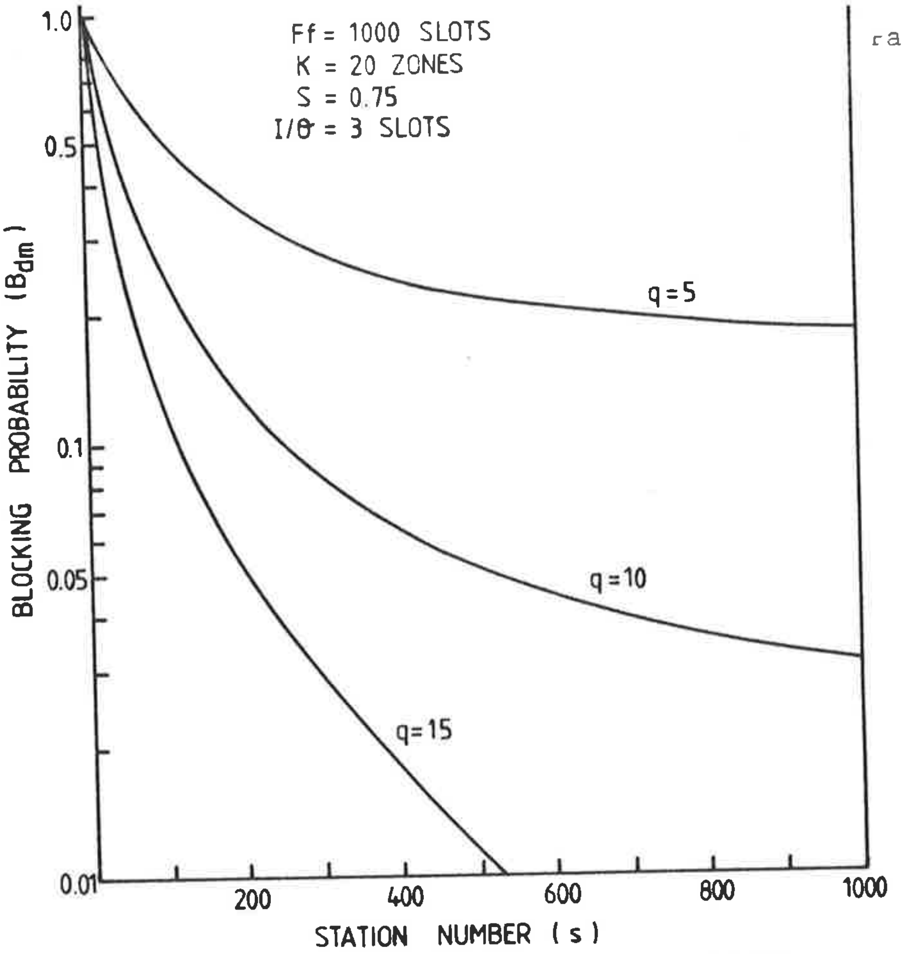


Fig.6.3.8 Data multiplexing buffer blocking probability versus number of stations (s) for the DA-FA Protocol with the buffer size (q) as a parameter

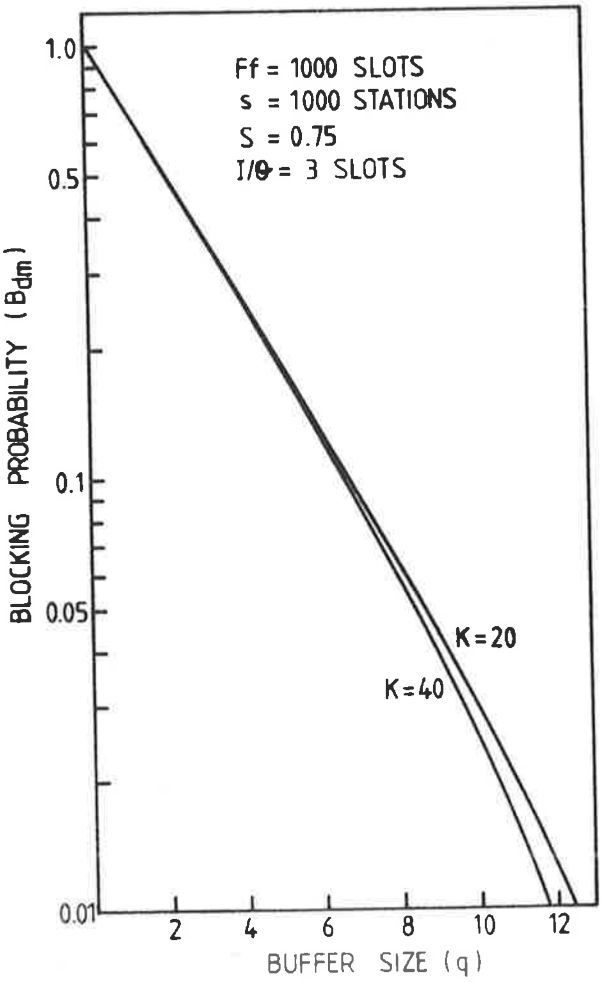


Fig.6.3.9 Data multiplexing buffer blocking probability versus buffer size for the DA-FA Protocol with the number of zones (K) as a parameter

To sum up, for a given traffic condition all three SDMA protocols exhibit a monotonically decreasing blocking probability with respect to an increasing buffer size. For a fixed buffer size, the blocking probability increases as the number of arrivals into a station increases. Equation (6.3.3) and (6.3.4) give the tradeoff between the number of arrivals and the traffic or network characteristics. To achieve a certain value of blocking probability, a larger buffer is needed for a long message. Furthermore, for demand assignment protocols, the terrestrial network design is a compromise between the buffer size per station and the total number of stations in the network. However, the number of spotbeam coverages is not a critical factor. Using the above set of figures, the network designer has the necessary information to determine the required buffer size for a particular set of blocking probability, network and traffic parameters.

6.3.2 Buffer Performance Comparisons of the SDMA Protocols

A comparison of the buffer performances of the SDMA protocols is shown in Fig.6.3.10. Since the FA Protocol requires a separated queue for the outgoing traffic to each zone, the required buffer size is proportional to the number of zones. Hence it has the worst buffer performance of the SDMA protocols considered (curve I of Fig.6.3.10). On the other hand, the DA Protocol which multiplexes a larger capacity amongst a larger population has the best buffer performance (curve II), while the distributed control DA-FA Protocol, which shared the stations buffer space amongst all its outgoing traffic, follows closely behind (curve III). Since the DA-FA Protocol has a lower overhead requirement, a larger data channel capacity can be obtained with the same frame capacity. Curve IV shows the buffer performance of the DA-FA Protocol, if the same amount of data channel capacity as the DA Protocol is permitted.

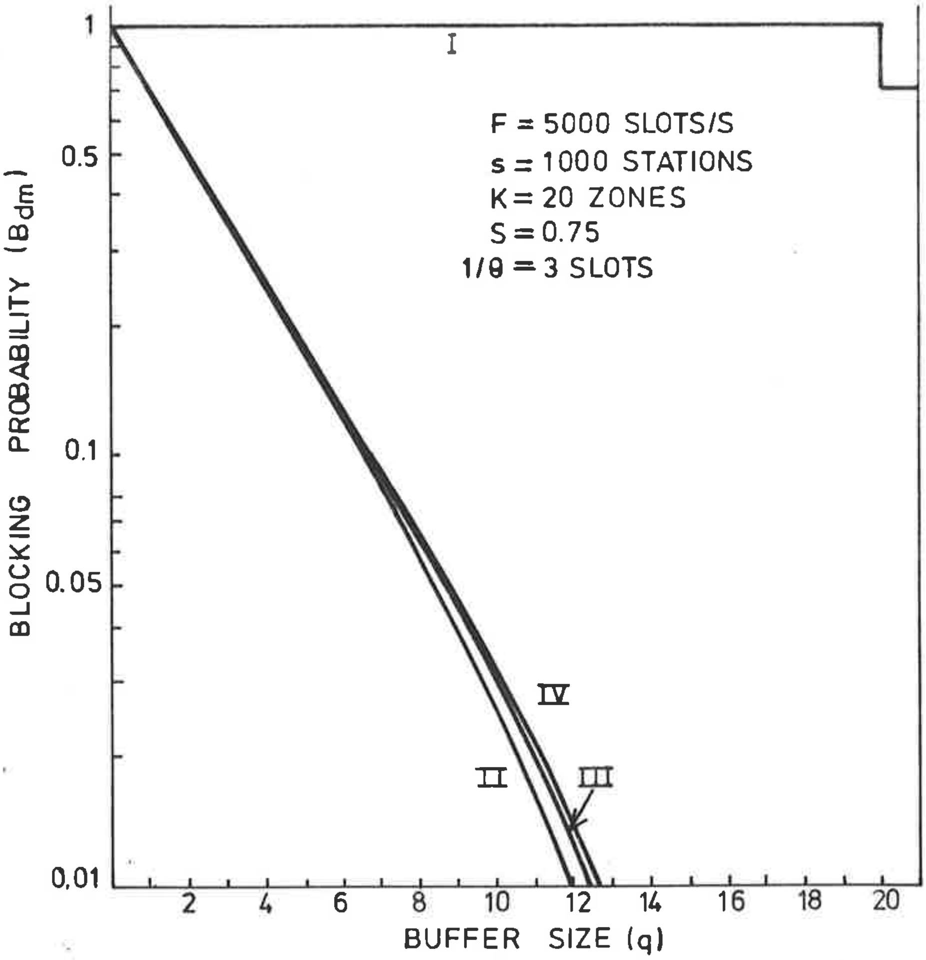


Fig.6.3.10 Comparison of data multiplexing buffer blocking probability versus buffer size for the SDMA protocols

6.3.3 Demultiplexing Buffer Performances

The data demultiplexing buffer blocking probability is a function of the average data arrival rate and buffer size. The probability is calculated by Chu's Method (see Section 5.3.1). The relations between the blocking probability and the buffer size for

1. different average arrival rates,

2. different average message length,

3. different demultiplexing channel rate

are given in Fig.6.3.11, 6.3.12 and 6.3.13 respectively. The figures indicate the blocking probability increases as the average rate or the message length increases, or the demultiplexing channel rate decreases.

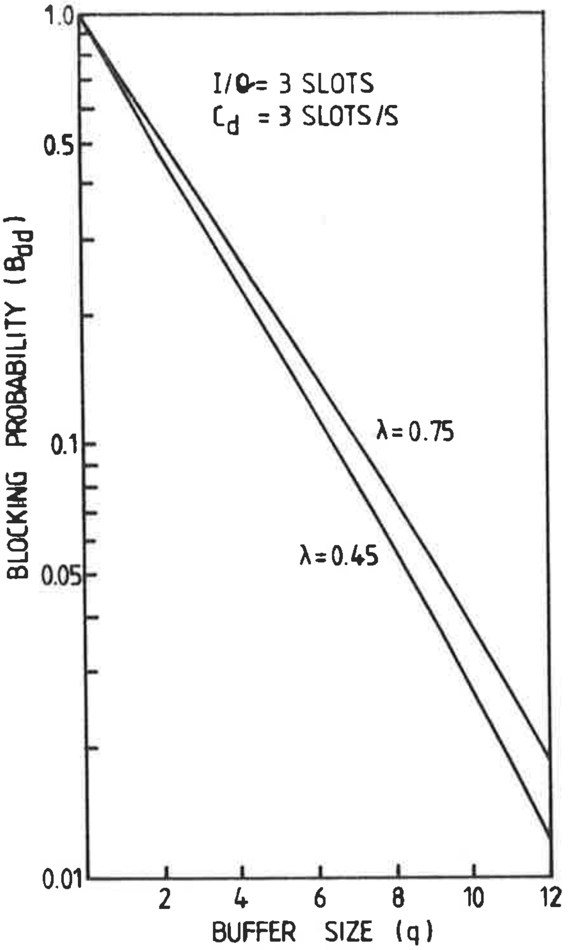


Fig.6.3.11 Data demultiplexing buffer blocking probability versus buffer size with two different arrival rates (λ)

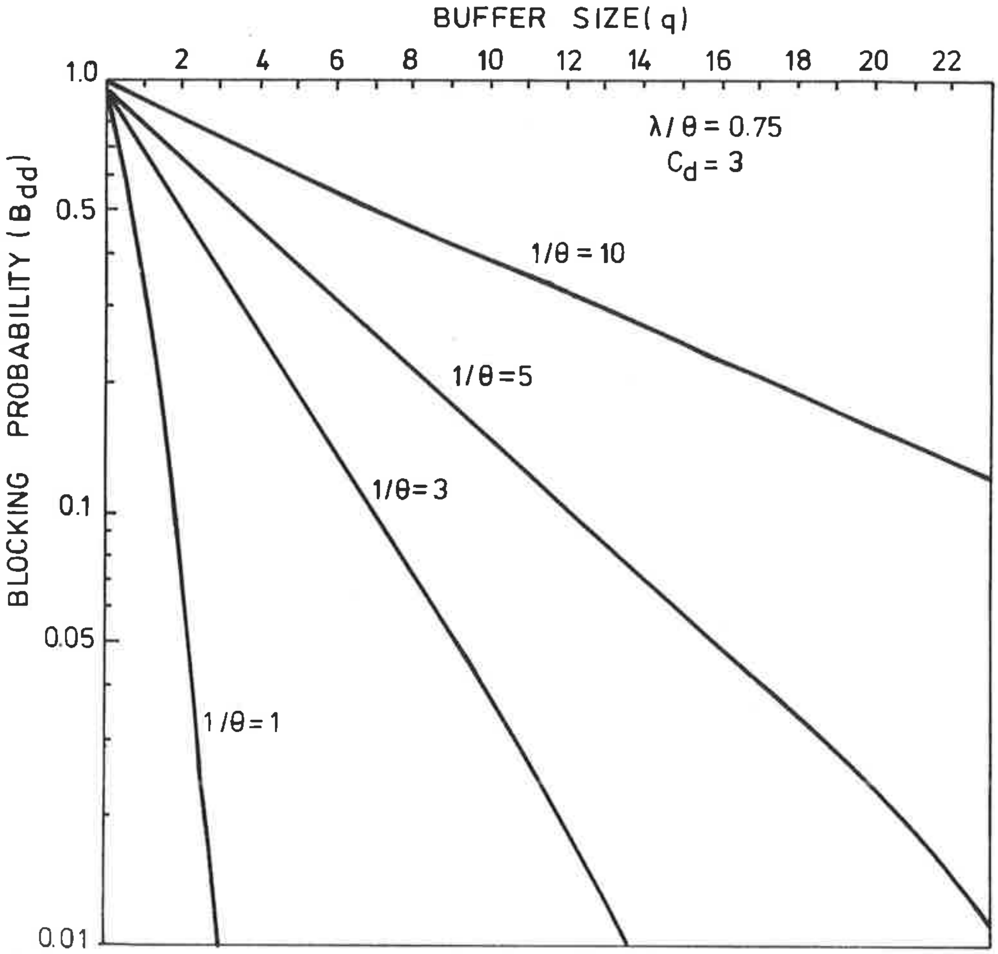


Fig.6.3.12 Data demultiplexing buffer blocking probability versus buffer size with the average message length (1/θ) as a parameter

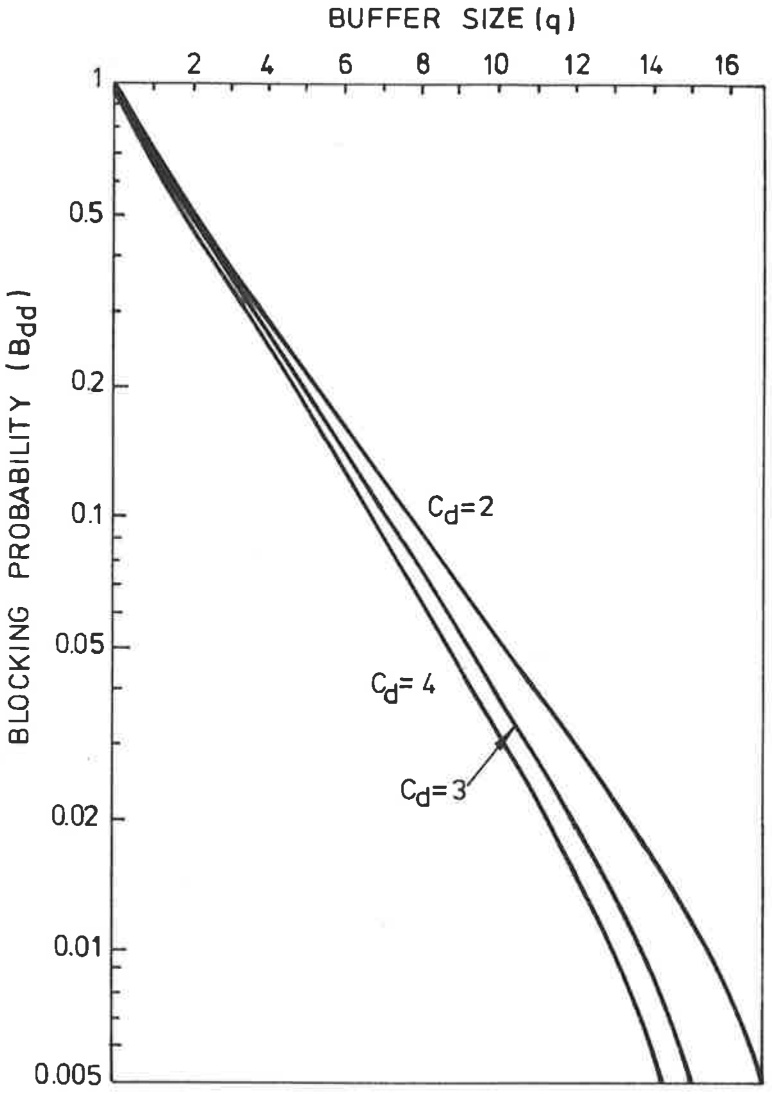


Fig.6.3.13 Data demultiplexing buffer blocking probability versus buffer size with the demultiplexing channel rate (c_d) as a parameter

6.4 CONCLUSION

In conclusion, out of the three Space Division Multiple Access (SDMA) protocols, the Demand Assignment (DA) Protocol has the best delay-throughput performance. It performs best in the region of high throughput with long message length traffic and using a small framelength. Furthermore, it is insensitive to any variation in zone partitioning. The required data channel assignment imposes an upper bound on the number of control slots and hence the maximum number of stations the protocol is able to serve; or alternatively, the number of stations in the network determines the maximum data channel assignment.

In comparing the DA Protocol with the other global beam multiple access protocols, the DA Protocol which has an ability to carry multiple control packets by each control burst, possesses a delay performance comparable with other distributed control or centrally controlled reservation global beam protocols. While most other global beam protocols are insensitive to variation in the number of stations and the transponder capacity, there exists an optimal range of station numbers and transponder capacity for the DA Protocol.

On the buffer performance of the SDMA protocols, for a fixed buffer size, the data multiplexing buffer blocking probability increases as the average number of arrivals into a station or the average message length increases. The DA Protocol has the best data multiplexing buffer performance for a given set of network and protocol parameters and traffic characteristics. It is followed closely by the hybrid form of Demand Assignment uplink and Fixed Assignment downlink (DA-FA) Protocol, whereas the Fixed Assignment (FA) Protocol has a far inferior performace. On the data demultiplexing buffer blocking probability with a fixed buffer size, the blocking probability increases as the average arrival rate or the average message length increases.

On selecting an appropriate SDMA protocol, one has to take the available technology, cost, reliability and system parameters into account. The other necessary considerations are complexity of the multiple access protocol and the location of its major hardware complexity. With the FA and DA-FA Protocols, the uplink and the downlink are switched through a fixed sequence of pre-determined duration, whereas the DA Protocol requires the alternation of the switching sequence and duration to suit the instantaneous traffic requirements. The FA Protocol requires the minimum hardware complexity in both the space and the earth segments. The DA Protocol locates the majority of its hardware complexity on-board the satellite and uses a relatively simple earth station, whereas the distributed control DA-FA Protocol has the reverse strategy.

Given the three SDMA protocols have comparable cost structure and reliability, the factors in selecting the appropriate multiple access protocols include permissible message delay and buffer blocking probabilities with fixed buffer sizes; with the optimization of protocol parameters under the constraints of certain traffic characteristics and network parameters. It is hoped that this chapter presents the network designer some insights into the delay and buffer performances of the SDMA protocols.

7. SIMULATION

The aims of the simulation of a satellite network using a Space Division Multiple Access protocol are threefold :-

1. to validate the analytical methods;
2. to determine the accuracy of the analytical results in cases where simplifying assumptions are made;
3. to investigate the effects of variation of certain parameters on the delay and buffer performance which are difficult to obtain using an analytical method.

In order to achieve these objectives, a detailed description of the network is required. The simulation program contains the following three elements :-

a. a detailed modelling of the proposed network and its protocol;

b. a procedure for generating traffic to the system;

c. a procedure for recording the desired performance characteristics.

In this chapter, a data communication network utilising a Demand Assignment (DA) Space Division Multiple Access (SDMA) Protocol is simulated. The delay and buffer performances are extracted and compared with the calculated values. Attention is focussed on the DA Protocol because it is the most versatile of the three SDMA protocols; i.e. the calculation methods used in the hybrid Demand Assignment uplink and Fixed Assignment downlink Protocol and Fixed Assignment Protocol are similar to Demand Assignment Protocol.

7.1 DISCUSSION ON SIMULATION

The simulation programs can be classified into two main catagories :-

1. Markov Chain simulation,
2. event-by-event simulation.

The former approach is restricted to a model which can be represented by a Markov Chain. The simultion proceeds from one state to another as the events are generated. The results of the simulation are not dependent on the timing but the sequence of the events. On the other hand, the event-by-event simulation method can be used on a non-Markovian model, and therefore it has a more general application. The system is studied by generating the random time intervals between events using appropriate distribution functions. The disadvantages of the event-by-event simulation approach are its programming complexity, the requirements of a larger storge and longer computation time [NAY 66].

The complexity of the programming can be reduced if a high level simulation language is used. One of these is SPURT76 [COHC 76], which is used in the simulation of the proposed network and is discussed in Section 7.1.1. Section 7.1.2 describes the network model. The accuracy of the results obtained from the simulation is discussed in Section 7.1.3.

7.1.1 The Simulation Language - SPURT76

A high level simulation programming language is used to perform the tedious task of time clock monitoring and to provide means for the output analysis. For the purpose of the network simulation, a discrete system simulation programming language package - SPURT76 [COHC 76] is used. It is an event orientated language, i.e. the clock is advanced from an event to the next event, and is comprised of a series of FORTRAN IV subroutines and functions. It enables a FORTRAN programmer to implement simulation models without having to learn the semantic and syntactic rules of a simulation programming language.

The SPURT76 provides the controlling subroutine - SCHEDULER which acts as a clock for controlling the discrete events. It also contains various stochastic generators for random number generations, a number of statistical and testing routines. Furthermore, a group of list processing subroutines is provided for array manipulations.

7.1.2 Simulation Model

PROGRAM SIM is produced to simulate the Demand Assignment Protocol. It describes the network and protocol procedure as accurately as possible. Four major events are identified :-

1. arrival of a message into an earth station,

2. transmission of the message from the earth station,

3. arrival of a message at a destination station,

4. termination of a frame, which is a convenient time for account keeping.

An event is loaded into the SCHEDULER of SPURT76 by the bootstrapping technique, i.e. when the clock reaches an event, the time and attributes of the next event are immmediately calculated. PROGRAM SIM allows the monitoring of all stations over a fixed number of frames for their reservation and scheduling delay, multiplexing and demultiplexing buffer blocking probability. In order to avoid the loading up transient of the simulation run, typical initial conditions are inserted before the delay performance is measured. In the case of buffer performance, the first portion of the simulation run is ignored in order to allow the buffer to reach a steady state . A typical run of PROGRAM SIM simulating forty frames takes 100 to 200 seconds of central processor time on a CDC Cyber 173®. The flowchart of PROGRAM SIM is shown in Fig.7.1.1.

7.1.3 Accuracy of the Simulation Results

With the introduction of pseudo random variables into a simulation, the simulated results are no more than samples of an unknown distribution. The significance of the results is estimated by a statistical confidence level. The result of each simulation X_i is an estimate of the true value μ. The central limit theorem states that if $X_1,\ldots,X_n$ are independent and identically distributed with finite mean and variance and meet some relatively weak conditions, then $\bar{X}$ (= $\sum_{s=1}^{n} X_s/n$) has approximately the normal distribution if the number of simulation (n) is large. Such approximation allows the confidence interval for the mean (m) to be derived from the unknown distribution of $X_1,X_2,\ldots,X_n$:-

$$\Pr(\bar{X} - Q\sqrt{m/n} < \mu < \bar{X} + Q\sqrt{m/n}) \sim 1 - \alpha \qquad (7.1.1a)$$

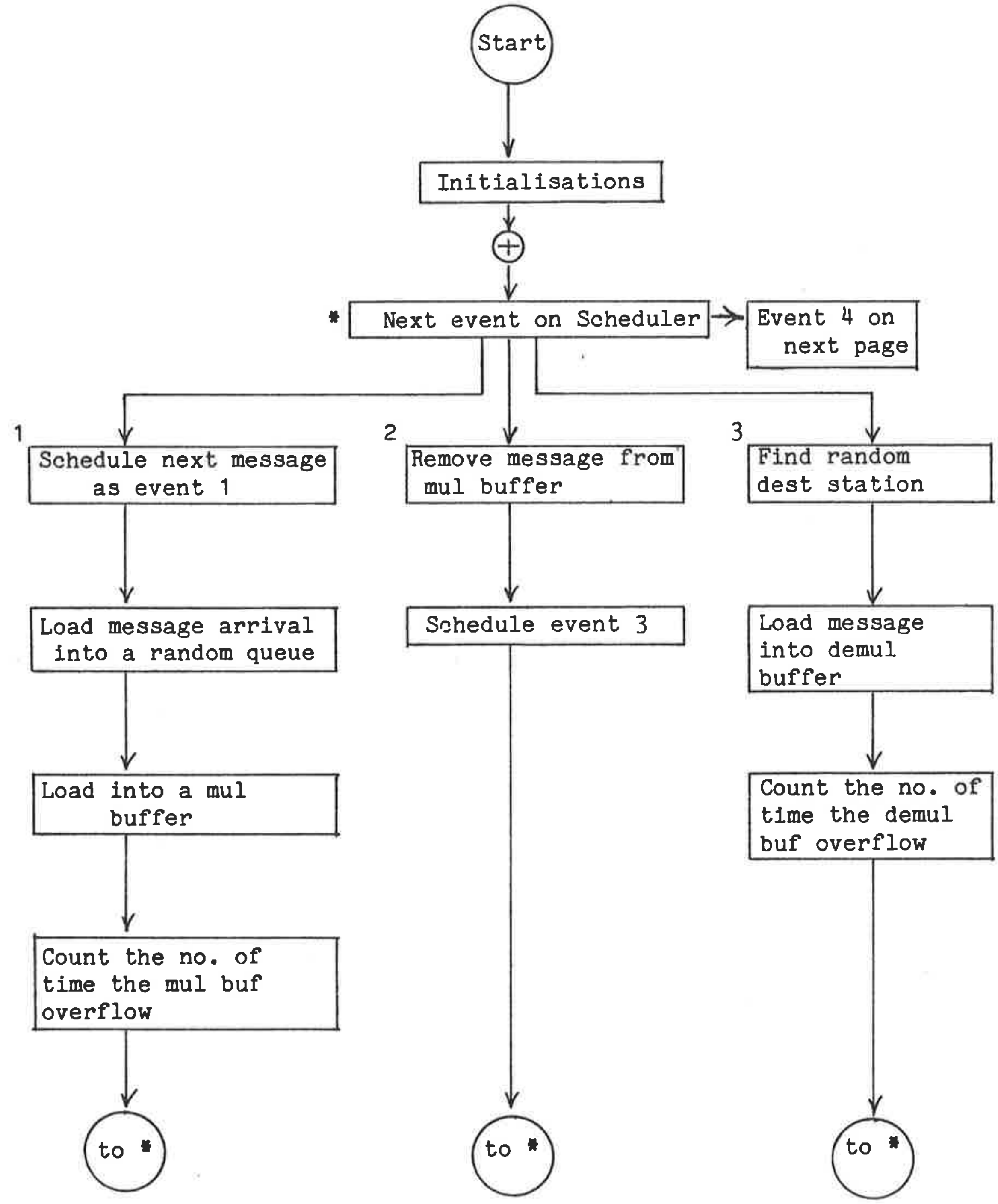


Fig.7.1.1 Flowchart of PROGRAM SIM

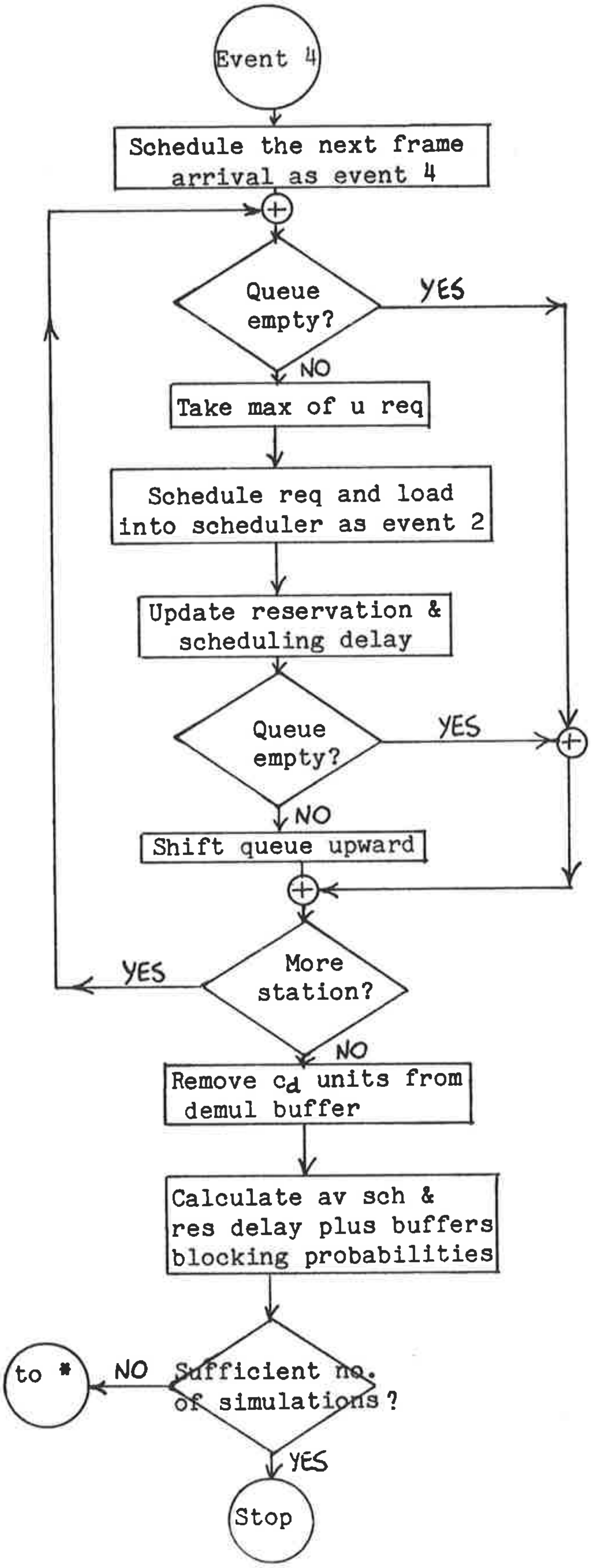


Fig.7.1.1 Flowchart of PROGRAM SIM

$$\text{where } \mathrm{var}(\bar{X}) = E[(X-\mu)^2] = m/n + O(1/n) \text{ for large } n \qquad (7.1.1b)$$

and Q is the point corresponding to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q} \exp(-z^2/2)dz = 1 - \alpha/2 \qquad (7.1.1c)$$

Many sophisticated methods have been developed to estimate var($\bar{X}$) and the confidence interval. The autoregressive method is chosen because of its accuracy. The basic assumption of the autoregressive method is that each member of the series can be predicted by a linear combination of the previous k members in the sequence

$$\sum_{s=0}^{k} a_s x_{t-s} = \varepsilon_t \qquad (7.1.2)$$

$$\text{where } x_{t-s} = X_{t-s} - \bar{X} \qquad t=1,2,\ldots,n$$

and ε_t are random, independent and normally distributed with the mean of zero and variance of σ^2. When a_o is defined as 1 and the equation is rearranged, the autoregressive nature of the relationship becomes apparent.

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_p x_{t-p} + \varepsilon_t \qquad (7.1.3)$$

After the autoregressive order p and the coefficients a_s, s=1,...,p are determined [FIS 73],[COHC 76] the autoregressive model is used to determine various statistical properties of the series, especially

$$\mathrm{var}(\bar{X}) = \frac{{\sigma_p}^2}{n\left[\sum_{s=0}^{p} a_s\right]^2} \qquad (7.1.4)$$

Using such a model, a confidence level of 95% is indicated on various simulated results in Section 7.2.

7.2 RESULTS OF THE SIMULATIONS

In order to validate and determine the accuracy of the analytical results, the comparisons of the simulated and calculated values of the reservation and scheduling delay, multiplexing and demultiplexing buffer blocking probability are presented. The separation of message delay into reservation and scheduling delay is used to highlight each of the two delay components.

7.2.1 Reservation Delay

The calculation method of the reservation delay is given in Section 4.3, the comparison between the simulated and the calculated results for an average message length of one and three slots per message are shown in Fig.7.2.1 and 7.2.2 respectively. The reservation delay is expressed in number of frames for a particular network throughput. Using the uniform traffic distribution assumption, the average message input to a particular station is the total number of network arrivals divided by the total number of stations. The comparison shows a close correlation between the simulated and the calculated results. The close correlation reinforces the proposed analytical solution for the reservation delay.

7.2.2 Scheduling Delay

The calculation of the scheduling delay is discussed in Section 4.2, and the comparison between the simulated and the calculated results for three different message lengths are shown in Fig.7.2.3, 7.2.4 and 7.2.5. The scheduling delay performance is expressed in number of data slots versus the network throughput. The data capacity is expressed in number of data slots per frame, with the rest of the capacity spent on the overhead. In all the three figures, the calculated results correlated very well with the simulated results and certainly within the 95% confidence interval.

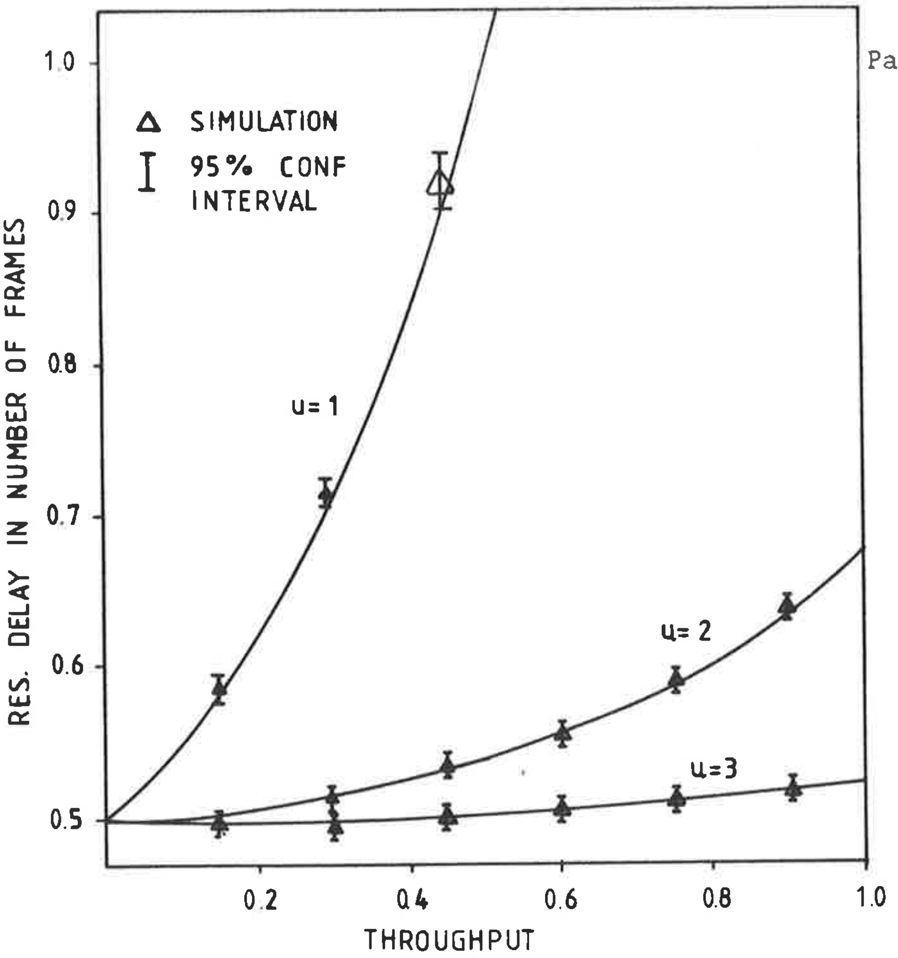


Fig.7.2.1 Comparison of simulated and calculated values of the reservation delay performance for a message length of one slot

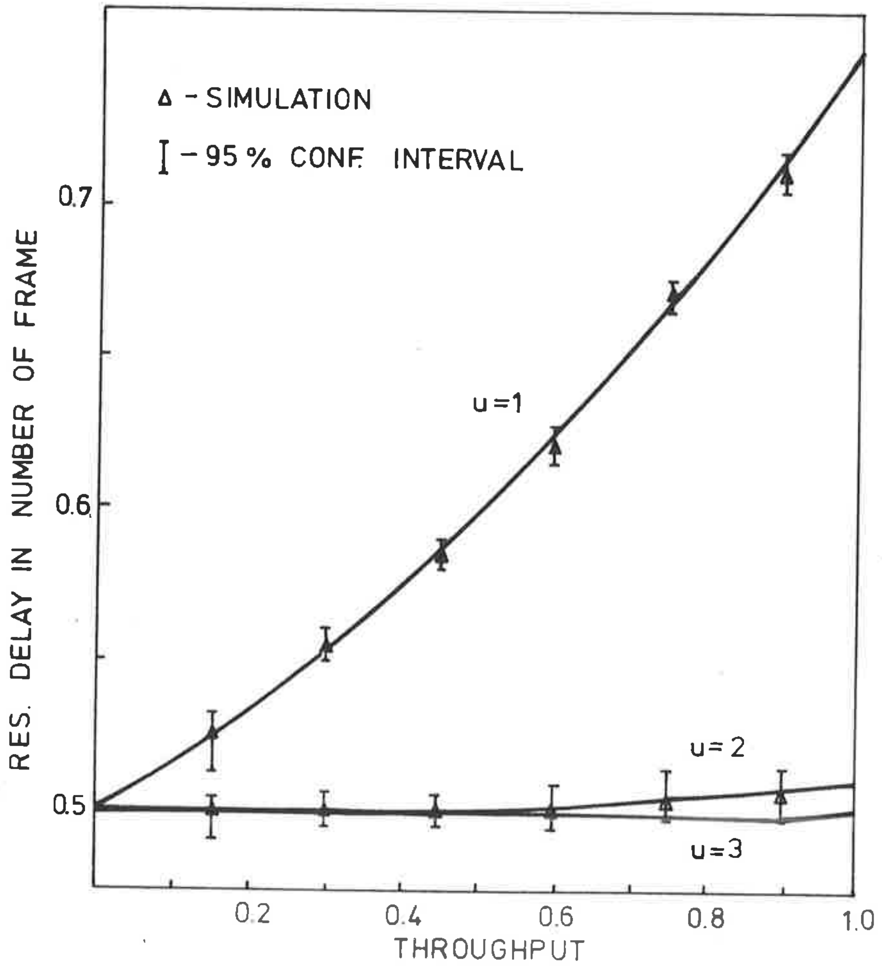


Fig.7.2.2 Comparison of simulated and calculated values of the reservation delay performance for an average message length of three slots

The Pollaczek-Khintchine approximation [COHJ 69] is also drawn on each of the three figures, where the approximation is assumed for a single continuously available channel or a M/G/1 queueing model. As shown on the figures, the approximation is an extremely poor estimate of the scheduling delay. The comparison reinforces the proposed integration solution for scheduling delay as discussed in Section 4.2.3.

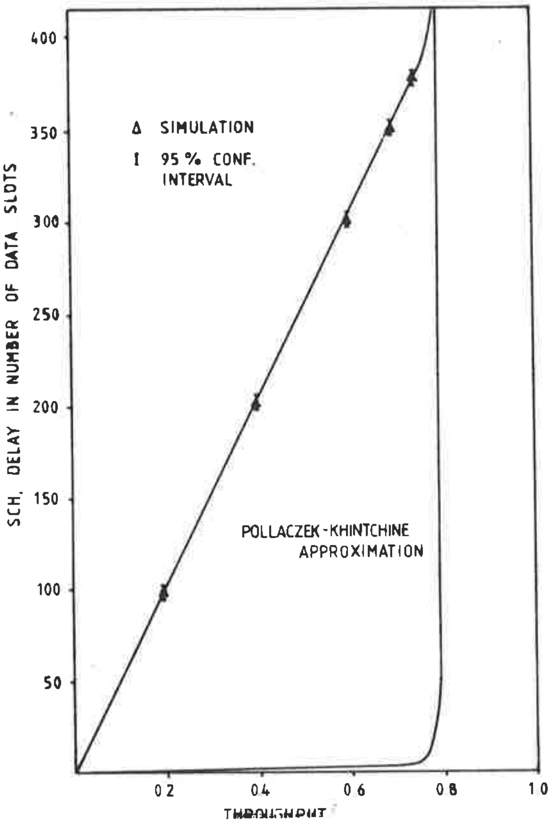


Fig.7.2.3 Comparison of simulated and calculated values of the scheduling delay performance for a message length of one slot

7.2.3 Multiplexing Buffer Blocking Probability

The comparison of the simulated and the calculated buffer performance using the Two-Group Method has been shown in [NG 77] for a simple Poisson arrival process. Fig.7.2.6 and 7.2.7 shows the data multiplexing buffer blocking probability versus the buffer size for a compound Poisson arrival with a throughput of 0.45 and 0.75 respectively, in the region of interest. In both examples, the simulated results possess similar trends as the calculated results but exhibit a higher value.

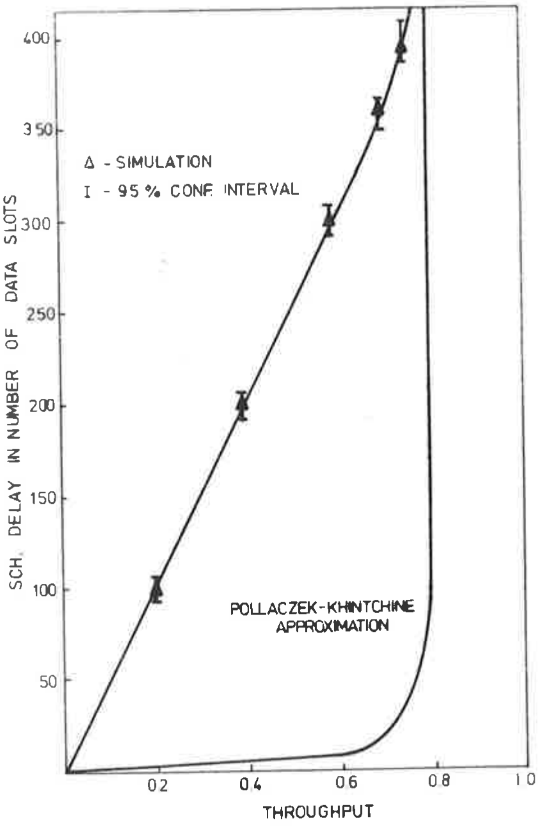


Fig.7.2.4 Comparison of simulated and calculated values of the scheduling delay performance for an average message length of three slots

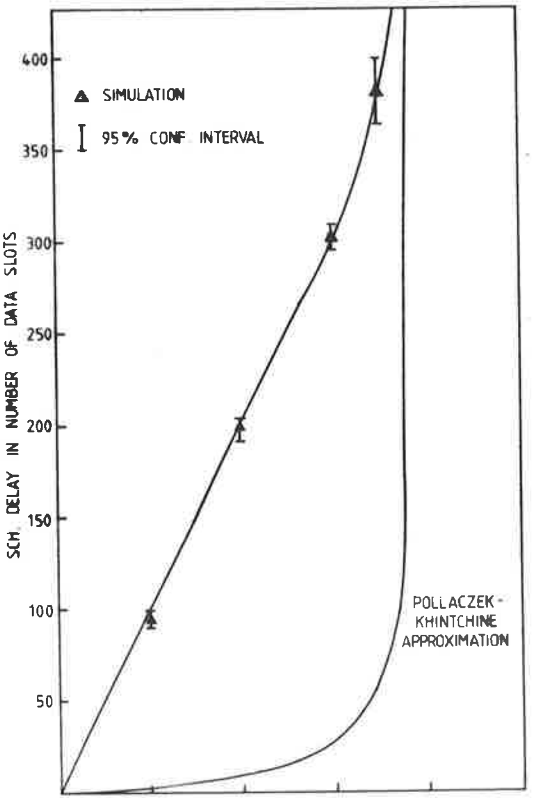


Fig.7.2.5 Comparison of simulated and calculated values of the scheduling delay performance for an average message length of ten slots

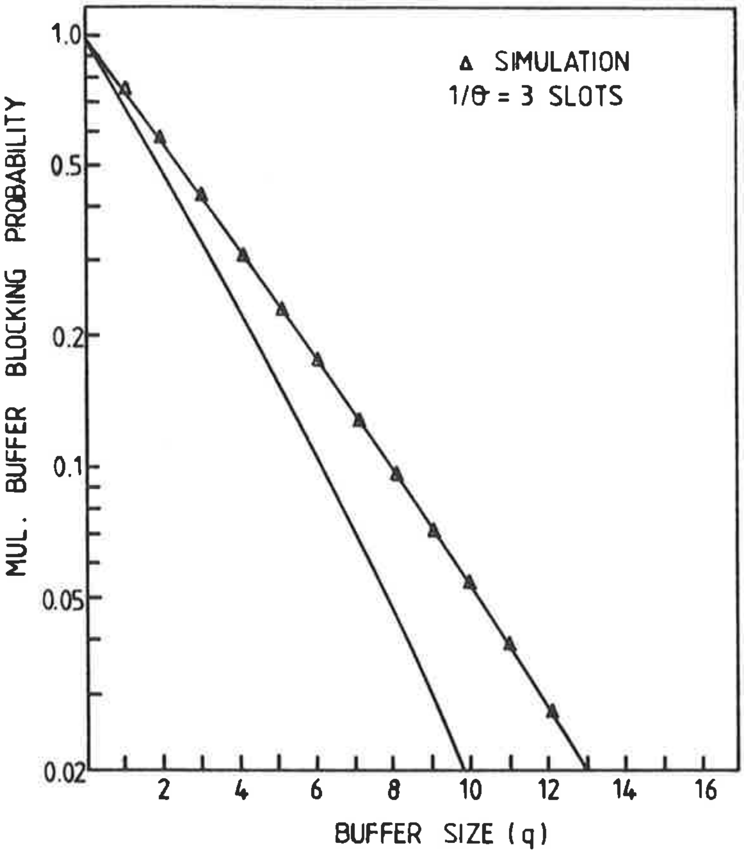


Fig.7.2.6 Comparison of simulated and calculated values of the multiplexing buffer blocking probability versus buffer size for a throughput of 0.45

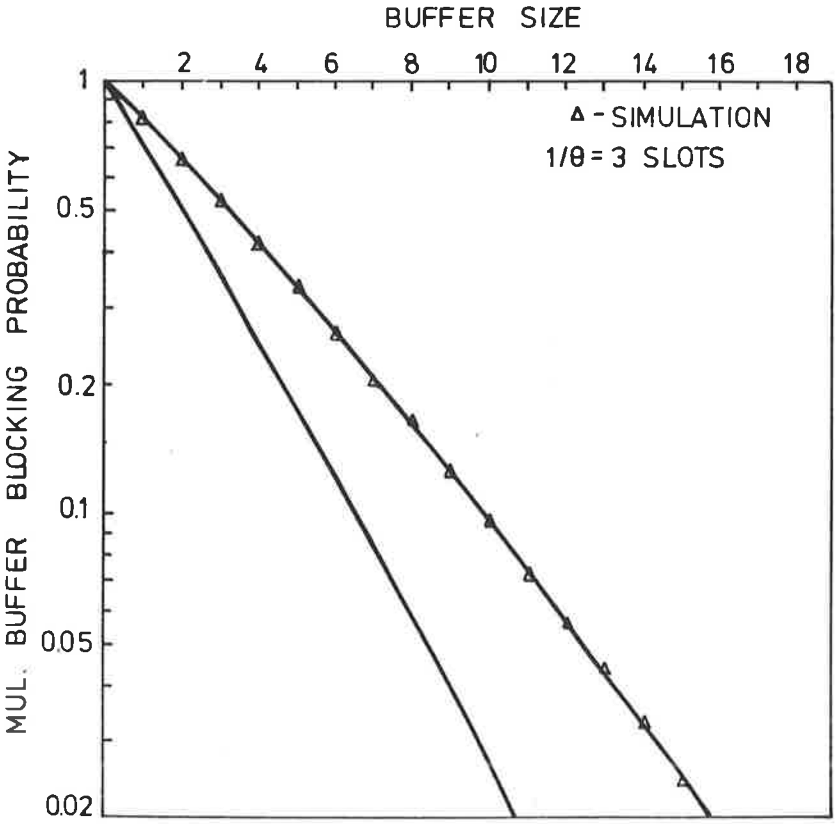


Fig.7.2.7 Comparison of simulated and calculated values of the multiplexing buffer blocking probability versus buffer size for a throughput of 0.75

The discrepancy is due to the assumption used in deriving the calculated results, i.e. Group I will always receive service from the channel (see Section 5.3.1). In the simulation and in a First Come First Served operating environment, such preferential treatment will not be given. Therefore the simulated results are more accurate and the calculated method will only provide an estimate to the order of magnitude. This is reinforced by the 95% confidence interval of a simulated multiplexing buffer blocking probability as shown in Table 7.2.1. Fig.7.2.8 shows that when the average message length increases, the simulated results exhibit the same trend as the calculated results.

	Simulated result			Calculated result
	Average	95% conf. interval		
Throughput				
0.45	0.7661	0.7542	0.7781	0.6904
0.75	0.8256	0.8173	0.8339	0.7050

Table 7.2.1 95% confidence interval of a simulated multiplexing buffer blocking probability for a buffer size of one

7.2.4 Demultiplexing Buffer Blocking Probability

The simulated buffer performance for a compound Poisson arrival process and constant output presented in this Section is believed not to have been previously published. The result shown in [CHU 72] is calculated for a buffer size greater than twenty and therefore not applicable for the purpose of a packetized message.

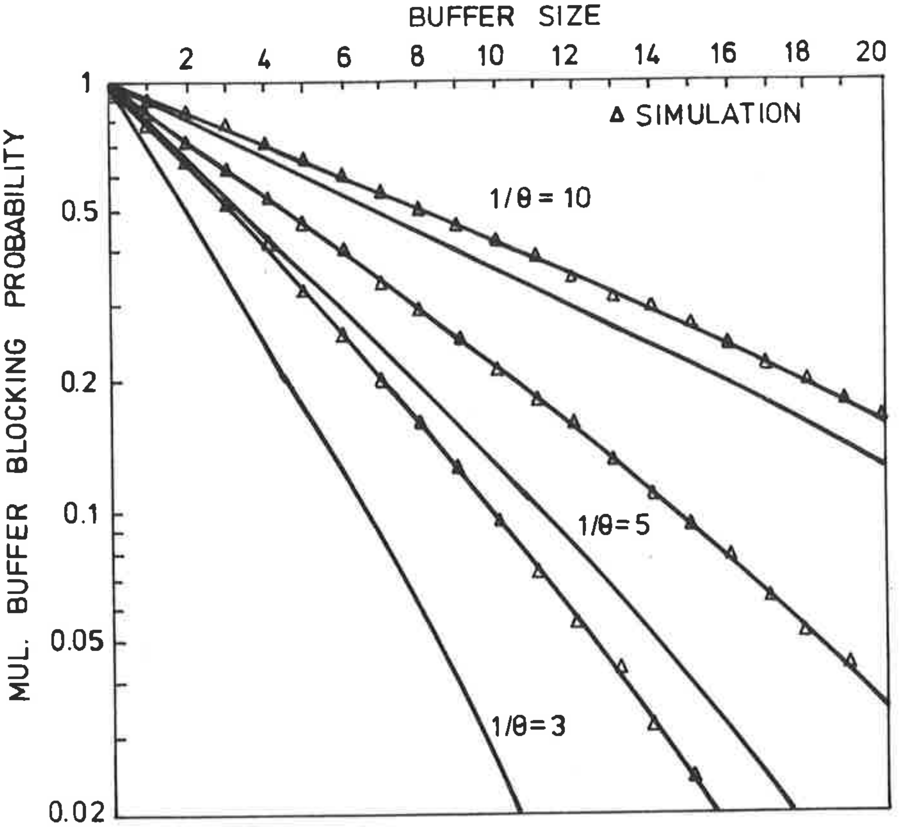


Fig.7.2.8 Comparison of simulated and calculated values of the multiplexing buffer blocking probability versus buffer size for a throughput of 0.75 but different average message length

Fig.7.2.9 and 7.2.10 show the data demultiplexing buffer blocking probability versus buffer size for a compound Poisson arrival process and constant output rate, with a network throughput of 0.45 and 0.75 respectively. Both the simulated and the calculated results assume a uniform distribution of network traffic. The simulated results follow closely but have a higher value of blocking probability than the calculated results, especially in the low throughput operating region. The 95% confidence interval of a simulated demultiplexing buffer blocking probability as shown in Table 7.2.2. does not account for the difference. The small discrepancy indicates the inadequancy of deriving the values of $\tilde{a}(k)$ by the Inverse Discrete Fourier Transform using a finite number of sample points (Section 5.3.1). Further increase in the number of sample points is prohibited by the memory space of the computing machinery. Similar to the calculated results, Fig.7.2.11 indicates that for a fixed throughput and buffer size, the buffer blocking probability increases as the average message length increases.

	Simulated result			Calculated result
	Average	95% conf. interval		
Throughput				
0.45	0.7170	0.7053	0.7287	0.6904
0.75	0.7478	0.7383	0.7572	0.7050

Table 7.2.2 95% confidence interval of a simulated demultiplexing buffer blocking probability for a buffer size of one

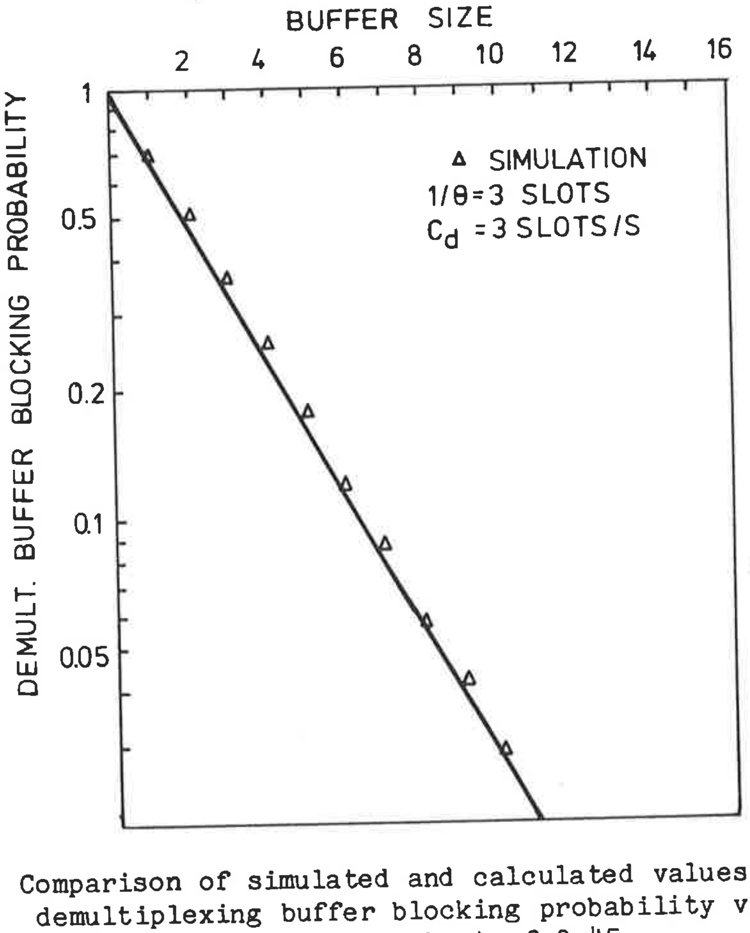


Fig.7.2.9 Comparison of simulated and calculated values of the demultiplexing buffer blocking probability versus buffer size for a throughput of 0.45

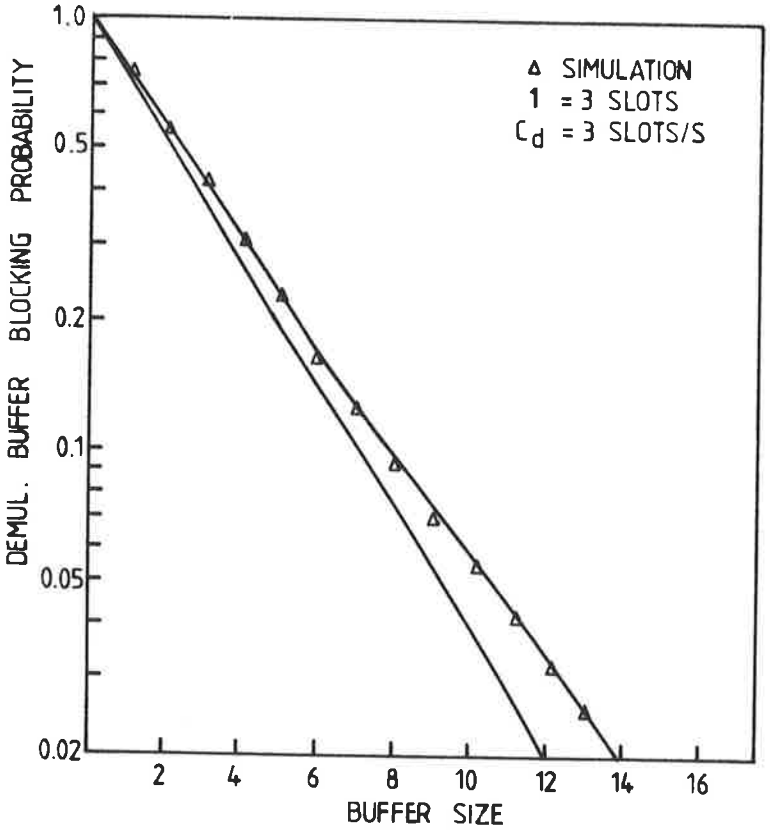


Fig.7.2.10 Comparison of simulated and calculated values of the demultiplexing buffer blocking probability versus buffer size for a throughput of 0.75

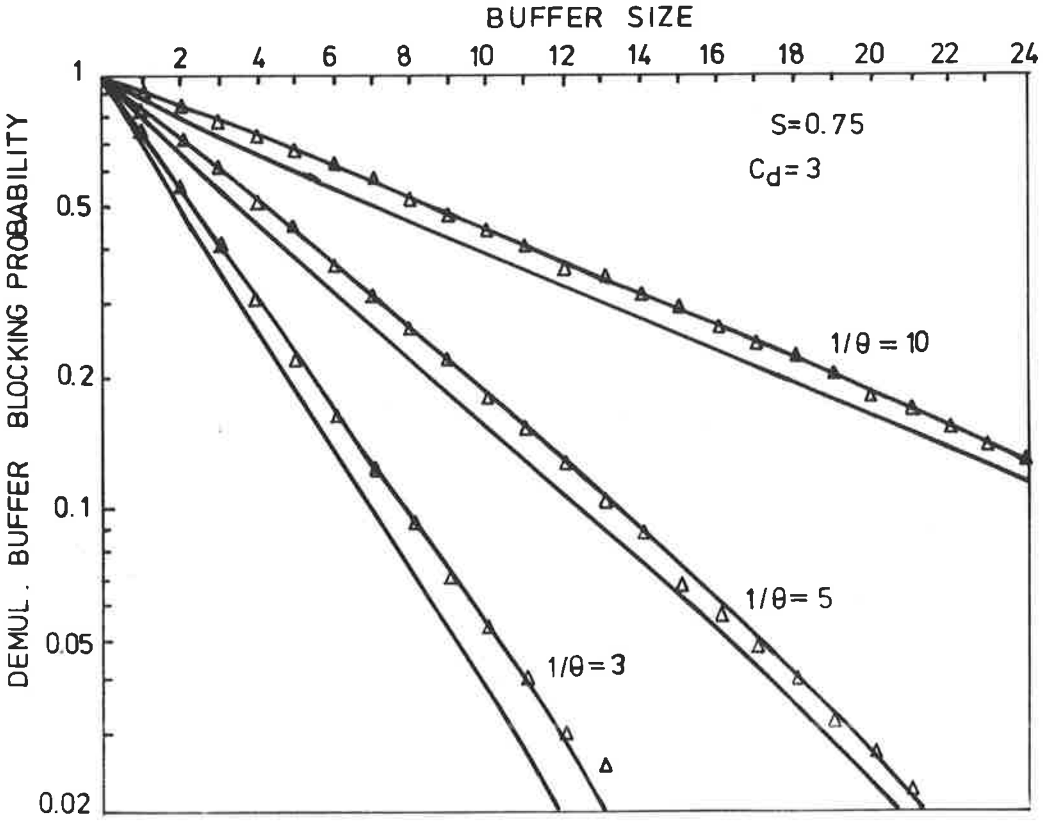


Fig.7.2.11 Comparison of simulated and calculated values of the demultiplexing buffer blocking probability versus buffer size for a throughput of 0.75 but different average message length

As well as the buffer size and the throughput, the demultiplexing buffer blocking probability is also a function of the demultiplexing channel rate. Fig.7.2.12 shows the simulated buffer performance for the two different demultiplexing channel rates.

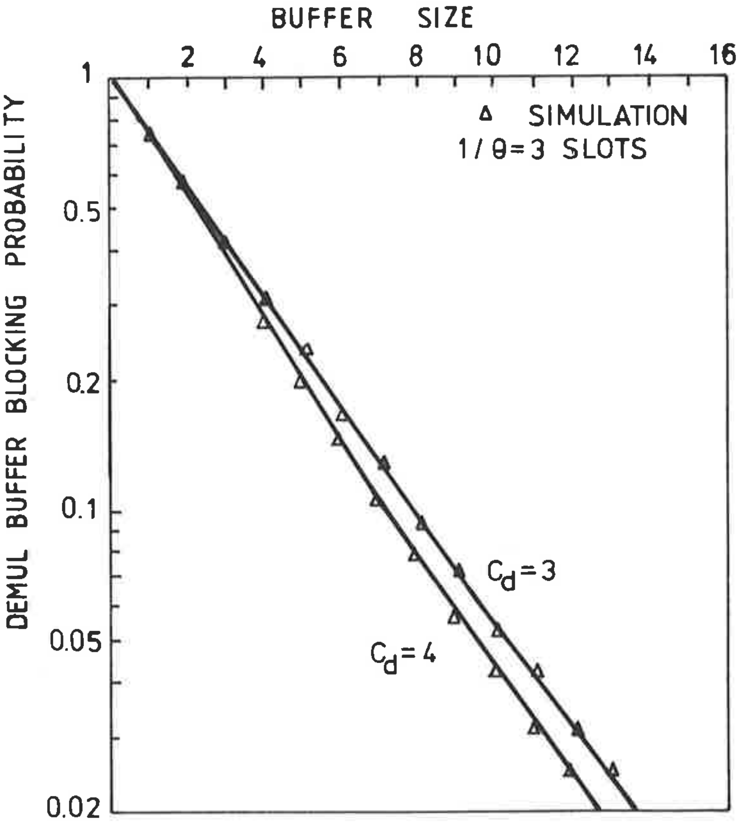


Fig.7.2.12 Simulated demultiplexing buffer blocking probability versus buffer size for different demultiplexing channel rates

7.3 RESULTS OBTAINED BY THE EXTENDED SIMULATION MODEL

Once a correct simulation model is constructed, the effects on the performance, by varying certain parameters and relaxing certain assumptions, can be easily obtained. This approach is useful for results which are difficult to obtain analytically.

7.3.1 Effects of the Arrival Process

Analytical methods are derived for the average message reservation and scheduling delay, multiplexing and demultiplexing buffer blocking probability. Since the capacity is allocated to the requests once a frame, it is believed that the average number of arrivals rather than the actual arrival process is the dominant influence on the values of delay, and buffer blocking probability. Such an hypothesis is confirmed by the results of the simulations of two other arrival processes, where the message interarrival times have Erlangian and Weibull distributions.

The probability density function of an n-phase Erlangian distribution is

$$f(t) = (\lambda t)^{n-1}\lambda \exp(-\lambda t)/(n-1)! \qquad \lambda > 0,\ t \geq 0$$

with mean $=n/\lambda$

The negative exponential distribution is a special case of the Erlangian distribution where $n=1$. The probability density function of the Weibull distribution is

$$f(t) = \lambda a(\lambda t)^{a-1}\exp -(\lambda t)^{a} \qquad \lambda \geq 0,\ a \geq 0,\ t \geq 0$$

with mean $= \Gamma(1/a + 1)/\lambda$

The negative exponential distribution is a special case of the Weibull distribution where $a=1$.

The results of the simulations show that all the delays and buffer blocking probabilities remain relatively constant for the three distributions with equal average arrivals (Table 7.3.1).

	Poisson			Erlang			Weibull		
	Ave	95% conf.int.		Ave	95% conf.int.		Ave	95% conf.int.	
D_R	.1341	.1329	.1353	.1328	.1311	.1331	.1315	.1301	.1329
D_S	.3861	.3541	.4181	.3870	.3847	.3894	.3840	.3838	.3842
$B_{dm}(1)$	.8256	.8173	.8339	.8260	.8190	.8329	.8228	.8153	.8304
$B_{dd}(1)$	.7478	.7384	.7572	.7495	.7415	.7576	.7451	.7372	.7530

Poisson	$\lambda=250$	
Erlang	$n=3$	$\lambda=250/3$
Weibull	$a=3$	$\lambda=250/\Gamma\ (1+1/3)$

Tabe 7.3.1 Comparison of simulated results between three different arrival processes having an average message length of three slots

7.3.2 Priority System

The solution of a priority queueing network is often intractable, even in its simplest form. Hence only approximate solutions are available [MOR 80], [JAI 68]. However simulation results can be more easily obtained from a validated model. A dual priority class system is simulated and its results are shown below. The model can be easily extended to any arbitrary priority queueing discipline.

Attention is focussed on the improved performance of the high priority users. The penalty paid for such improvement is the performance deterioration suffered by the lower priority users. Fig.7.3.1, 7.3.2 and 7.3.3 shown that in the case of low average message length, the average reservation delay for a high priority user can be drastically reduced compared with a user of the FCFS discipline. In the case of high average message length, the reservation delay is already close to the intrinsic value of half a framelength. A priority system brings little improvement. However for the average scheduling delay, Fig.7.3.4, 7.3.5 and 7.3.6 shown that a high priority user has superior delay performance compared with a user in a FCFS discipline irrespective of the average message length and throughput.

7.3.3 Other Observations

Two other observations of the extended simulation model are worth mentioning. First, if a Demand Assignment Protocol is used in any other application besides the satellite communication, the transmission delay will be reduced according to the distance between the transmitting station and the controller or the receiving station, but the average reservation delay and the variable portion of the scheduling delay remain unaffected. However the average multiplexing buffer blocking probability is reduced, since the message holding time is smaller.

Second, besides the expected reduction of the reservation delay when the control burst size (u) is increased; there is also a very slight reduction in average scheduling delay, multiplexing and demultiplexing buffer blocking probability. Such improvements are due to more efficient data channel utilisation, with a less efficient control channel as a penalty.

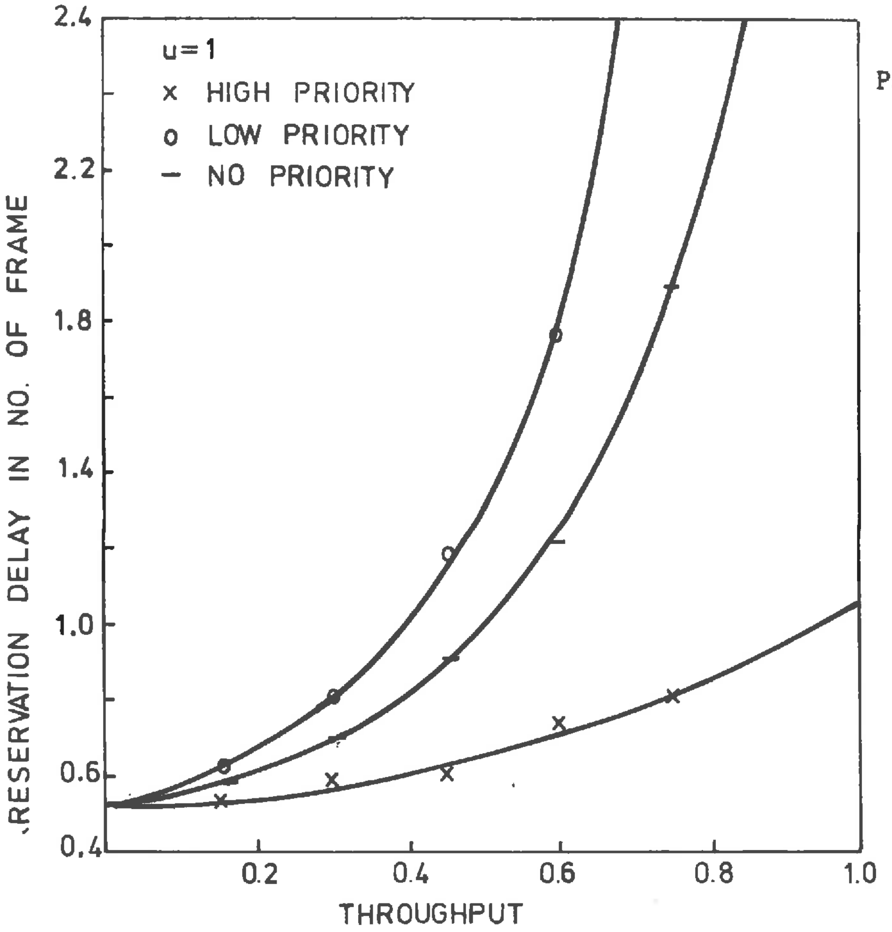


Fig.7.3.1 Simulated reservation delay performance in a dual priority system with a single slot message

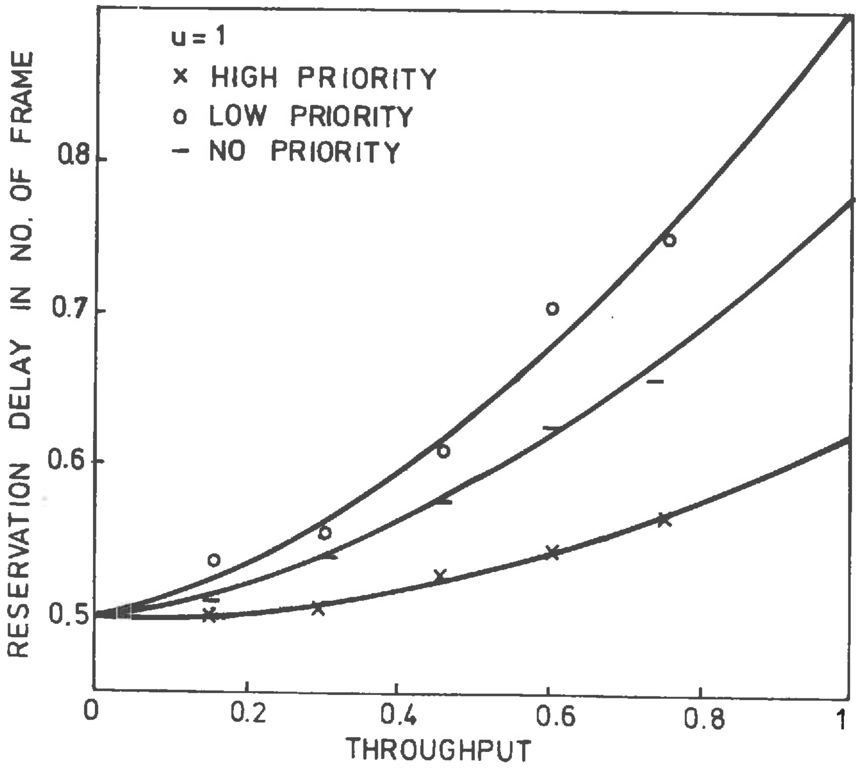


Fig.7.3.2 Simulated reservation delay performance in a dual priority system with an average message length of three slots

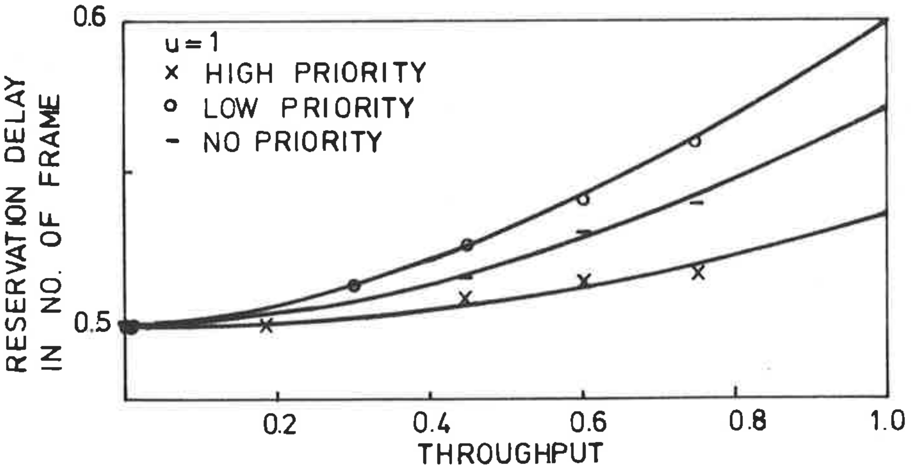


Fig.7.3.3 Simulated reservation delay performance in a dual priority system with an average message length of ten slots

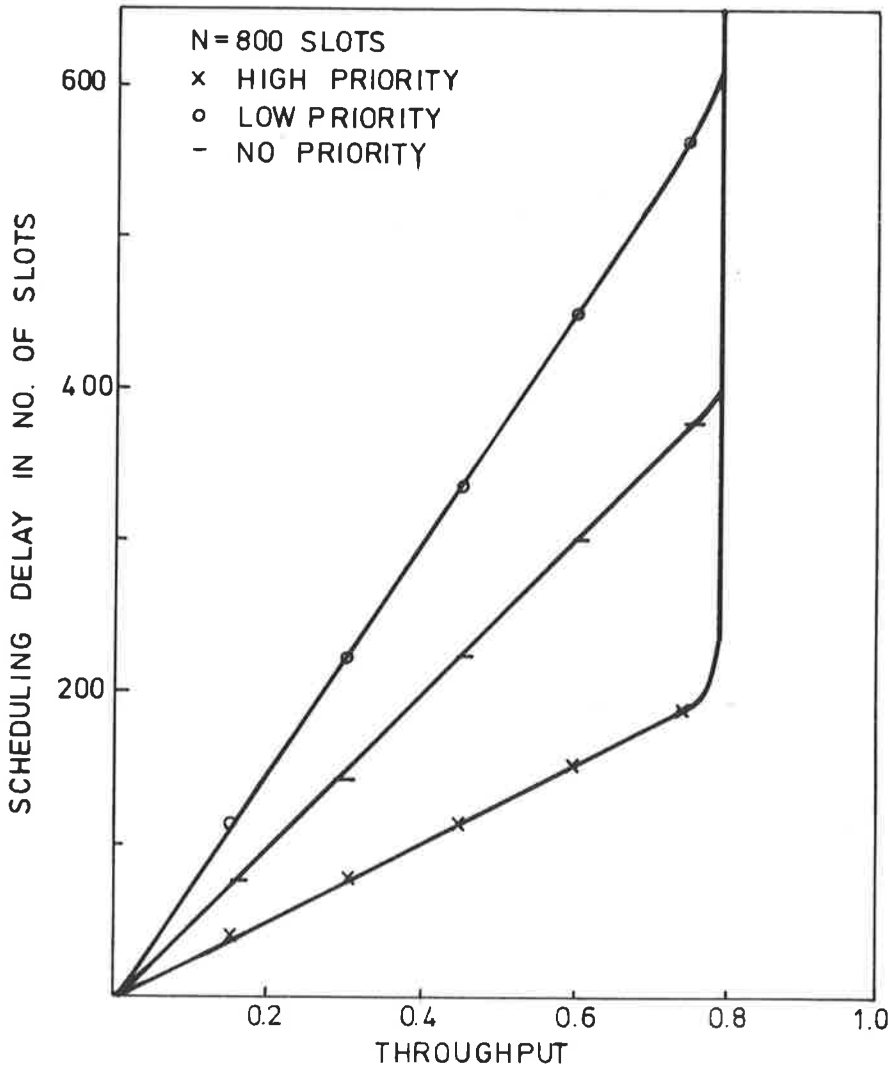


Fig.7.3.4 Simulated scheduling delay performance in a dual priority system with a single slot message

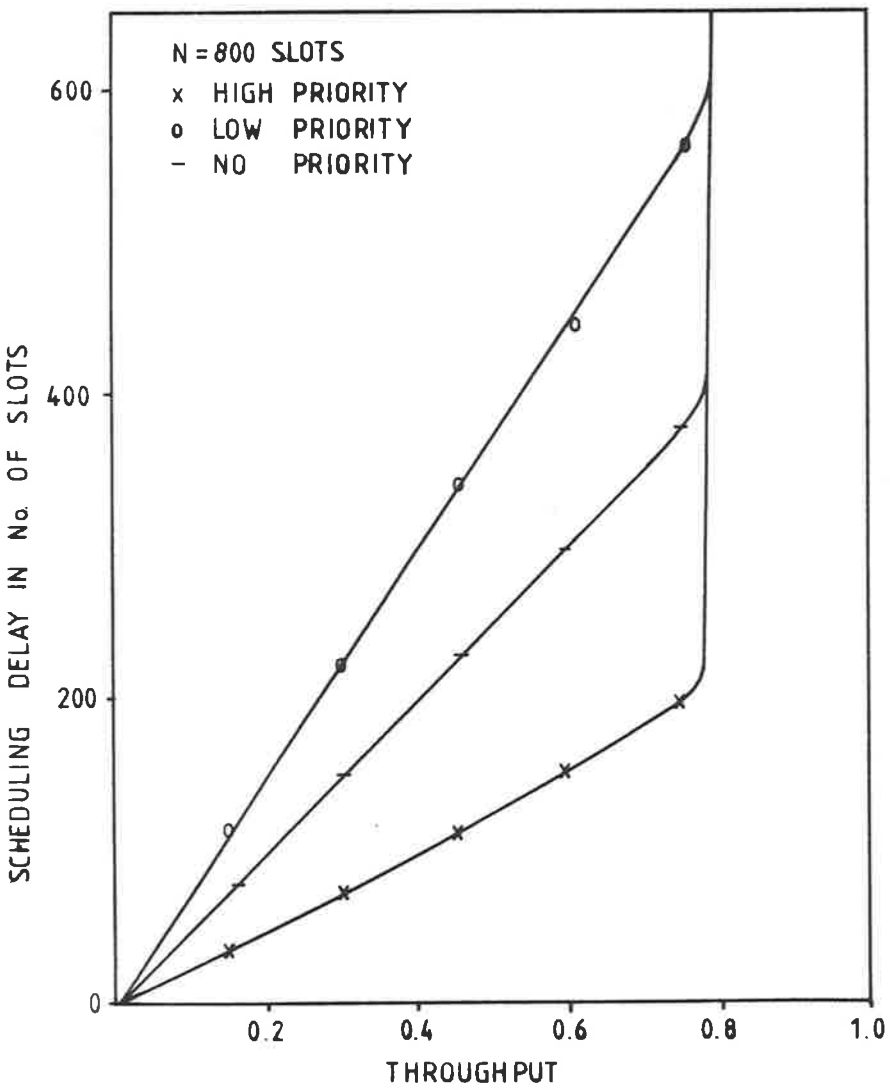


Fig.7.3.5 Simulated scheduling delay performance in a dual priority system with an average message length of three slots

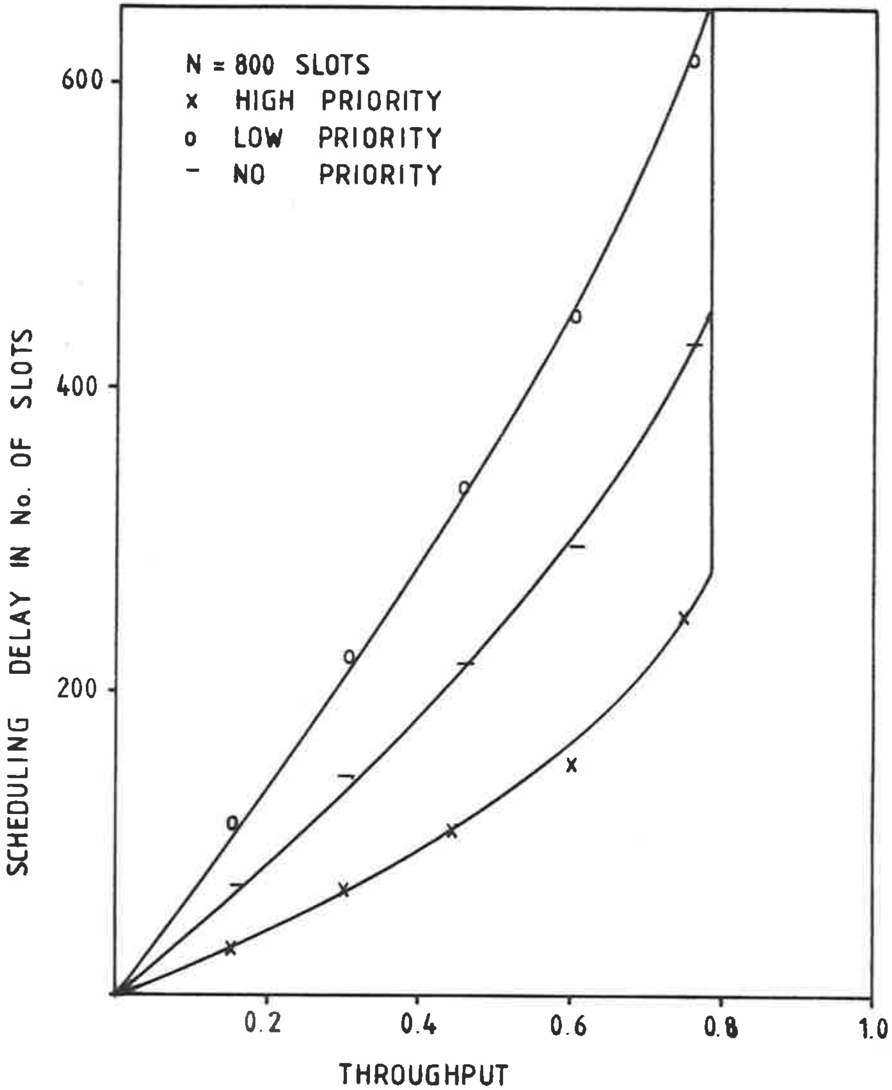


Fig.7.3.6 Simulated scheduling delay performance in a dual priority system with an average message length of ten slots

7.4 CONCLUSION

The simulations validate the analytical results for the calculation of reservation and scheduling delay. They also lend support to the method of calculating demultiplexing buffer blocking probabilities described in Section 5.3.2. However the simulated results indicate that the Two-Group Method can only give an order of magnitude accuracy for calculating the demultiplexing buffer blocking probability, due to the oversimplified assumption of capacity is always available.

The simulations also indicate that the dominant influence on the average reservation and scheduling delay, multiplexing and demultiplexing buffer blocking probabilities is the average number of arrivals rather than the actual arrival process. The simulation model can also be used to obtain the reservation and the scheduling delay of a priority queueing discipline.

8. CONCLUSION AND SUGGESTIONS FOR FUTURE RESEARCH

In this investigation, three possible multiple access protocols suitable for a space division satellite switched communication network, known collectively as the Space Division Multiple Access (SDMA) protocols, have been described in Chapter 3. New analytical methods and models have been developed for the terrestrial network design, delay and buffer performance calculations in Chapter 2, 4 and 5 respectively. Their results are further confirmed in Chapter 7 by simulations.

The study indicates that the use of multiple fine spotbeam antennas and dynamic on-board switching for satellite communication networks is a viable method of alleviating the problems of overcrowded orbital space and frequency spectrum. The spotbeam coverage allows high energy concentration and the dynamic on-board switching provides the appropriate beam connections. These techniques are particularly appropriate for future domestic satellite communications, especially for low traffic density, mobile, maritime or direct to user applications. Suitable protocols for data traffic, and integrated data and voice traffic, using a beam switched transponder have been identified. Their delay and buffer performances indicate that a space division network is a feasible proposition.

Due to the large population and infrequent usage by the users of the space division network, a Fixed Assignment Protocol seems unsuitable because of its inefficiency, large buffer requirements and its poor delay performance.

The higher performance demand assignment protocols suffer a reduction in data channel capacity, since part of the capacity is used by the control mechanism. If the size of the reservation burst is sufficiently large, a reduction of the reservation delay to the intrinsic value of half a framelength is possible. The penalty paid for such an improvement is a decrease in data channel capacity.

With the hybrid protocol of distributed control Demand Assignment uplink and Fixed Assignment downlink, the pair of connecting beams, like the Fixed Assignment Protocol, switch through a fixed sequence. The stations in each spotbeam coverage share the allocated capacity. The associated disadvantages include inflexibility in capacity assignment and added complexity in the earth segment. While the advantage lies in its simplicity in the space segment. The protocol performs best when the allocated capacity for each spotbeam coverage is large and the number of spotbeam coverages, and the average message length, are small. Hence a large capacity transponder and spotbeam coverage are recommended. Its delay and buffer performances are only slightly inferior to those of the fully Demand Assignment Protocol.

Of the three possible space division protocols considered, the fully Demand Assignment Protocol has the best delay and buffer performances as shown in Chapter 6. Its delay performance approaches those of the global beam multiple access protocols, especially in the region of high throughput. The delay and buffer performances are independent of the number of spotbeam coverages, hence a large number of highly concentrated spotbeams can be utilised. The message length is not a critical factor in the delay performance, therefore the Demand Assignment Protocol can accommodate a larger range of traffic. As the entire capacity is shared amongst all the stations in the network, a high degree of statistical multiplexing or efficiency can be achieved. However, the required hardware in the space segment and the control mechanism is relatively more complex.

For all three protocols, the buffers blocking probabilities are monotonically decreasing functions with respect to increasing buffer size, i.e. any blocking probability can be achieved if a sufficiently large buffer is installed. In addition, both the delay and the buffer blocking probabilities are monotonically increasing functions with respect to increasing throughput. That is, the greater the amount of traffic, the worse is its performance. With a fixed set of traffic characteristics in both demand assignment protocols, their protocol parameters, terrestrial network and hardware designs, can be optimized to satisfy the criteria on average message delay and buffer blocking probabilities. Their inter-relationships and their tradeoffs are discussed in Chapter 6.

However, the economic viability and the available technology will be the paramount factors in deciding a system configuration for a space division satellite switched network. It is envisaged that the simpler hybrid system of demand assignment uplink and fixed assignment downlink will be a suitable interim solution, before it is upgraded to the higher performance and more complex fully demand assignment system. Besides satellite communication, the proposed space division multiple access protocols can also be utilised in any land based centrally controlled communication network, where each region of the network has only access to the network via a central node.

Some other aspects which could be of importance in a practical space division satellite switched communication system and might form the basis of further research are :-

1. An analytical and simulation study on the delay and the buffer performances for a combined transponder and beam switching system or a multiple transmission rate system.

2. A study on the network stability if the transmission deterioration on the control information can no longer be assumed negligible.

Finally it is hoped that the work presented in this thesis will provide a set of tools for performance evaluation and methodologies in terrestrial network and protocol design for a space division satellite switched communication network.

Appendix 1. Calculation of the Average and the Variance of the Queue Size with Integration Method

From Equation (4.2.20),

$$\frac{B(z)}{\tilde{A}(z)} = \frac{-z^{-N}\sum_{j=0}^{N-1} b_j(z^j - z^N)}{1 - z^{-N}\tilde{A}(z)} \overset{\Delta}{=} \frac{\gamma(z)}{\phi(z)} \qquad \text{(A1.1)}$$

The following observations are made :-

1. $\phi(z)$ has N zeros on and inside the unit circle, including $z=1$.

2. Since $B(z)$ is analytical in and on the unit circle, $\gamma(z)$ has the same zeros as $\phi(z)$.

3. $\phi(z)$ and $\gamma(z)$ have poles of order N at $z=0$.

4. $\gamma(z)/\phi(z)$ and $B(z)/\tilde{A}(z)$ have no poles or zeros in and on the unit circle.

From (A1.1), by Cauchy's Integral,

$$\ln B(z)/\tilde{A}(z) = \ln \gamma(z) - \ln \phi(z)$$

$$= \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \gamma(x) - \ln \phi(x)}{x-z}\, dx$$

where Γ is a circle of radius $1+\varepsilon$ and $|z| \leqslant 1$. Since $\ln B(z)/\tilde{A}(z)$ is analytic in and on Γ , $\ln \gamma(x)$ and $\ln \phi(x)$ are single valued on Γ .

With Residue Theorem,

$$\ln B(z) = \ln \tilde{A}(z) + \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \gamma(x)}{x-z}\,dx - \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \phi(x)}{x-z}\,dx$$

$$= \ln \tilde{A}(z) + \ln \sum_{j=0}^{N-1} b_j - \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \phi(z)}{x-z}\,dx \qquad (A1.2)$$

Repeatedly differentiate (A1.2),

$$\frac{B^{(1)}(z)}{B(z)} = \frac{\tilde{A}^{(1)}(z)}{\tilde{A}(z)} - \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \phi(x)}{(x-z)^2}\,dx$$

$$\frac{B^{(2)}(z)}{B(z)} - \frac{B^{(1)}(z)^2}{B(z)} = \frac{\tilde{A}^{(2)}(z)}{\tilde{A}(z)} - \frac{\tilde{A}^{(1)}(z)^2}{\tilde{A}(z)^2} - \frac{1}{j\pi}\int_{\Gamma} \frac{\ln \phi(x)}{(x-z)^3}\,dx$$

Set z=1 :-

$$B^{(1)}(1) = \tilde{A}^{(1)}(1) - \frac{1}{j2\pi}\int_{\Gamma} \frac{\ln \phi(x)}{(x-1)^2}\,dx \qquad (A1.3)$$

$$B^{(2)}(1) - B^{(1)}(1)^2 = \tilde{A}^{(2)}(1) - \tilde{A}^{(1)}(1)^2 - \frac{1}{j\pi}\int_{\Gamma} \frac{\ln \phi(x)}{(x-1)^3}\,dx \qquad (A1.4)$$

The integration can be facilitated by removing the singular points and integrating the residue area. Let us define

$$\phi(x) \triangleq \psi(x)\ (x-1)/x \qquad (A1.5)$$

$$\ln \phi(x) = \ln (x-1)/x + \ln \psi(1) + \{\psi^{(1)}(1)/\psi(1)\}(x-1) + .5\{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2/\psi(1)\}\ (x-1)^2 + \ldots \qquad (A1.6)$$

Recall $\phi(x) = 1 - x^{-N}\,\tilde{A}(x)$ **(A1.7)**

Equating with (A1.5), repeatedly differentiate and set z=1, various values of ψ can be obtained :-

$$\psi(1) = N - \tilde{A}^{(1)}(1)$$

$$\psi^{(1)}(1) = -\frac{1}{2}\{\tilde{A}^{(2)}(1) - 2(N-1)\tilde{A}^{(1)}(1) + N(N-1)\}$$

$$\psi^{(2)}(1) = -\frac{1}{3}\{\tilde{A}^{(3)}(1) - 3(N-1)\tilde{A}^{(2)}(1) + 3N(N-1)\tilde{A}^{(1)}(1) - N(N^2-1)\}$$

Combining (A1.3) and (A1.6),

$$B^{(1)}(1) = \tilde{A}^{(1)}(1) - \frac{\psi^{(1)}(1)}{\psi(1)} - \frac{1}{j2\pi}\int_{\Gamma} g(x)\, dx$$

$$\text{where } g(x) = \frac{1}{(x-1)^2}\ln\left\{\frac{x}{(x-1)}\,\frac{\phi(x)}{\psi(1)}\right\} - \frac{\psi^{(1)}(1)}{(x-1)\psi(1)} \qquad x\neq 1$$

$$= .5\{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2\}/\psi(1)^2 \qquad x=1$$

Using the polar co-ordinates, x=exp(jρ),

$$B^{(1)}(1) = \tilde{A}^{(1)}(1) - \psi^{(1)}(1)/\psi(1) - \frac{1}{\pi}\int_{0}^{\pi} G(\rho)d\rho \tag{A1.8}$$

$$\text{where} \quad G(\rho) = \mathrm{Re}\{g[\exp(j\rho)]\ \exp(j\rho)\}$$

$$\text{or} \quad = \frac{-1}{2(1-\cos\rho)}\ln\left\{\frac{1}{2\psi(1)}\left[\frac{2(m^2+n^2)}{1-\cos\rho}\right]^{\frac{1}{2}}\right\} - \frac{\psi^{(1)}(1)}{2\psi(1)} \qquad \rho\neq 0$$

(A1.9)

$$= 0.5\{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2\}/\psi(1)^2 \qquad \rho=0$$

$$\text{and} \quad m+jn = \phi\{\exp(j\rho)\} = 1-\exp(-jN\rho)\tilde{A}[\exp(j\rho)] \tag{A1.10}$$

From (A1.4),

$$\mathrm{var}(B) = \mathrm{var}(\tilde{A}) - \tilde{A}^{(1)}(1) + B^{(1)}(1) - \frac{1}{j\pi}\int_{\Gamma}\frac{\ln\,\phi(x)}{(x-1)^3}\,dx \tag{A1.11}$$

Using (A1.6),

$$\mathrm{var}(B) = \mathrm{var}(\tilde{A}) - \tilde{A}^{(1)}(1) + B^{(1)}(1) - 0.5\{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2\}/\psi(1)^2 - \frac{1}{j\pi}\int_{\Gamma} h(x)\,dx$$

where $$h(x) = \frac{1}{(x-1)^3}\ln\left\{\frac{x}{x-1}\,\frac{\phi(x)}{\psi(x)}\right\} - \frac{\psi^{(1)}(1)}{\psi(1)(x-1)^2} - \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{2(x-1)\psi(1)^2} \qquad x\neq 1$$

$$= \frac{\psi^{(3)}(1)\psi(1)^2 - 3\psi^{(2)}(1)\psi^{(1)}(1)\psi(1) + 2\psi^{(1)}(1)^3}{6\psi(1)^3} \qquad x=1$$

Using polar co-ordinates,

$$\mathrm{var}(B) = \mathrm{var}(\tilde{A}) - \tilde{A}^{(1)}(1) + B^{(1)}(1) - \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{2\psi(1)^2} - \frac{2}{\pi}\int_{0}^{\pi} H(\rho)\,d\rho \qquad \text{(A1.12)}$$

where $$H(\rho) = \mathrm{Re}\{h[\exp(j\rho)]\exp(j\rho)\}$$

or $$H(\rho) = \mathrm{Re}\left\{\left[\frac{1-\cos\rho + j\sin\rho}{4(1-\cos\rho)^2}\right]\left[\ln\left\{\frac{m + jn}{2\psi(1)}\left(1 - \frac{j\sin\rho}{1-\cos\rho}\right)\right\}\right]\right\} \qquad \text{(A1.13)}$$

$$+ \frac{\psi^{(1)}(1)}{2\psi(1)(1-\cos\rho)} - \frac{\psi^{(2)}(1)\psi(1) - \psi^{(1)}(1)^2}{4\psi(1)^2} \qquad \rho\neq 0$$

$$= \frac{\psi^{(3)}(1)\psi(1)^2 - 3\psi^{(2)}(1)\psi^{(1)}(1)\psi(1) + 2\psi^{(1)}(1)^3}{6\psi(1)^3} \qquad \rho=0$$

Appendix 2 Expressions for the Average Message Delay of various Global Beam Network Protocols

The total data message delay (D_d) is defined as the duration between the message arrival into the sending station and the moment that the message is received by the destination station.

$$E(D_d) = E(D_R) + E(D_S) + rtpd \qquad (A2.1)$$

where D_R and D_S are the random variables of reservation and scheduling delay respectively, and in satellite communication the round trip propagation delay (rtpd) is approximately 0.27 second.

There are two criteria for channels stability, the number of message arrivals per frame must be smaller than the reservation channel capacity. Similarly, the data arrivals per frame must also be smaller than the assigned capacity.

The comparisons are made with Poisson message arrival process and geometrical message length distribution. Detailed derivations of the expressions of the average message delay can be found in the listed references. The symbols used have similar definitions as those in Chapter 4.

A. Time Division Multiple Access (TDMA) Protocol

Ref : [RUB 79]

1. $E(D_R) = 0$

2. $$E(D_S) = \frac{f}{2} + \frac{\lambda_i/\theta(2/\theta-1)f}{2(1-\lambda_i/\theta)} \quad (s) \qquad \text{(A2.2)}$$

Stability criterion for each station :-

$$\lambda_i/\theta < 1 \qquad \text{(A2.3)}$$

B. Slotted Aloha Protocol

Ref : [BEL 80], [SCHW 77]pp.292

Definitions :-

λ_b = total channel input rate (included retransmit messages)

τ = slot interval in second

K = maximum number of randomised delay in slots

1. $E(D_R) = 0$

2. $$E(D_S) = 1.5/F + \varepsilon\{rtpd + (K+2)/(2F)\} \quad (s) \qquad \text{(A2.4)}$$

For a single slot message,

$$\varepsilon = G/S - 1 \qquad \text{(A2.5a)}$$

G can be derived iteratively from

$$S = G\exp(-G) \qquad \text{(A2.5b)}$$

For a multiple slots message,

where

$$\varepsilon = \lambda_b/(SFf\theta) - 1 \qquad \text{(A2.6a)}$$

$$S = \frac{\lambda}{Ff\theta} = \frac{\lambda_b}{Ff\theta}\left[\frac{1-\chi\exp(\lambda_b)}{\theta\exp(\lambda_b)}\right]\exp\left[-\lambda_b\left[\frac{1}{1-\chi\exp(\lambda_b)} - \tau\right]\right] \qquad \text{(A2.6b)}$$

Using (A2.6b), λ_b can be derived iteratively.

Stability criterion :-

network arrival < 1/e X channel capacity

i.e. $\lambda/\theta < 1/e \times Ff$ (A2.6c)

C. Reservation Aloha Protocol

Ref : [ROB 73], [SCHW 77]pp.303

Definitions :-

V = number of reservation slots per data slot

δ = data channel bandwidth assignment

K = maximum number of randomised delay

and K is much smaller than the total number of reservation slots per frame.

1. $E(D_R) = D_1(1-S/\delta) + D_2S/\delta + rtpd$ (s) (A2.7)

i.e. D_R is composed of Aloha State, Reserved State and the propagation delay of the request respectively.

a. $D_1 = 1.5/(VF) + \varepsilon_1\{ rtpd + (K+2)/(2VF) \}$ (s) (A2.8a)

where $\varepsilon_1 = G_1/S_1 - 1 = \exp G_1 - 1$ (A2.8b)

and G_1 can be obtained iteratively by

$S_1 = \theta S/V = G_1 \exp -G_1$ (A2.8c)

b. $D_2=f/2+1.5/(VF)+\varepsilon_2\{\max(f,rtpd)+f/2+(K+1)/(2VF)\}$ (s) (A2.9a)

where $\varepsilon_2=G_2/S_2 - 1 = \exp G_2 - 1$ (A2.9b)

and G_2 can be obtained iteratively by

$S_2 = \theta S/(V-V\delta) = G_2 \exp -G_2$ (A2.9c)

2. $E(D_S) = \dfrac{S(2-\theta)}{2\delta F\theta(\delta-S)}$ (s) (A2.10)

Stability criteria :-

1. message arrival < 1/e X number of reservation slots

$$\lambda < 1/e \times (1-\delta)FfV \qquad (A2.11)$$

2. the maximum data bandwidth assignment (δ_{max}) in order to have a stable reservation channel :-

$$\delta_{max} = V/(V+\theta e) \qquad (A2.12)$$

3. network arrival < data channel capacity

$$\lambda/\theta < \delta Ff \qquad (A2.13)$$

D. Split-channel Reservation Multiple Access (SRMA) Protocol

Ref : [TOB 76]

Note :-

i. The reservation and data channel occupy different portions of the bandwidth. Each channel is divided into time slots.

ii. The average input rate (λ) is now defined as arrivals per second.

iii. All symbols as in Reservation Aloha Protocol.

1. $$E(D_R)=3/(VF-VF\delta)+ \varepsilon\{rtpd+(K+2)/(VF-VF\delta)\}+rtpd \quad (s) \qquad (A2.14a)$$

where $\varepsilon = G_r/S_r - 1 = \exp G_r - 1$ (A2.14b)

and G_r can be obtained iteratively by

$$S_r = 2\theta S/(V-V\delta) = G_r \exp -G_r \qquad (A2.14c)$$

2. $$E(D_S) = \frac{S(2-\theta)}{2\delta F\theta(\delta-S)} \quad (s) \qquad (A2.15)$$

Stability criteria :-

1. **message arrival < 1/e X number of reservation slots**

$$\lambda < 1/e \times (1-\delta)FV/2 \qquad \text{(A2.16)}$$

2. **the maximum data bandwidth assignment (δ_{max}) in order to have a stable reservation channel :-**

$$\delta_{max} = V/(V+2\theta e) \qquad \text{(A2.17)}$$

3. **network arrival < data channel capacity**

$$\lambda/\theta < \delta F \qquad \text{(A2.18)}$$

Appendix 3. Publication

-Journal of Australia Telecommunication Research (ATR), Vol.16, No.2, Nov. 82.

Ko, K. T. & Davis, B. (1982). Design of a demand-assignment satellite-switched space division multiple access communication network. *Australian Telecommunication Research, 16*(2), 25-38.

NOTE:

This publication is included in the print copy
of the thesis held in the University of Adelaide Library.

EPILOGUE

LAWS AND PARADOXES

During the course of work towards this dissertation, the author came to terms with one of the fundamental laws of nature :-

"Murphy was an optimist."

(see [KLI 68] for a comprehensive discussion on Murphy´s Law.)

Further the captivation of the author by life´s paradoxes, epitomized by Groucho Marx´s remark:-

"I wouldn´t join any Club which would have Me as a Member."

should hopefully not influence the opinion of the readers of this dissertation.

REFERENCES

[ABR 77] Abramson N., "The throughput of packet broadcasting channels", IEEE Trans. on Commun., Vol.COM-25, No.1, pp.117-128, Jan. 77.

[BEL 80] Bellini S. and Borgonovo F., "On the throughput of an Aloha channel with variable length packets", IEEE Trans. on Commun., Vol.Com-28, No.11, pp.1932-1935, Nov. 80.

[BOO 77] Boorstyn R.R. and Frank H., "Large-scale network topological optimization", IEEE Trans. on Commun., Vol.COM-25, No.1, pp.29-47, Jan. 77.

[CHU 70a] Chu W.W., "Buffer behavior for Poisson arrivals and multiple synchronous constant outputs", IEEE Trans. Computer, Vol.C-19, No.6, pp.530-534, Jun. 70.

[CHU 70b] Chu W.W., "Buffer behavior for batch Poisson arrivals and single constant output", IEEE Trans. Commun. Tech., Vol.COM-18, No.5, pp.613-618, Oct. 70.

[CHU 72] Chu W.W., "Demultiplexing considerations for statistical multiplexors", IEEE Trans. on Commun., Vol.COM-20, No.6 PartII, pp.603-609, Jun. 72.

[COHC 76] Cohen C. and Robinson B., "Simulation modelling and programming with SPURT-76", Northwestern University, Illinois, 1976.

[COHJ 69] Cohen J.W., "The single server queue", Armsterdam, North-Holland, 1969.

[CRO 73] Crowther W.R., Rettbert R., Walden D., Ornstein S. and Heart F., "A system for broadcast communication : Reservation-Aloha", Proc. 6th Hawaii Int. Syst. Sci. Conf., Jan. 73.

[FIS 73] Fishman G.S., "Concepts and methods in discrete event digital simulation", John Wiley and Sons, New York, 1973.

[FOR 79] Fordyce S. and Stamminger R., "The use of geostationary platforms for future U.S. domestic satellite communications", Proc. ICC, pp.49.4.1-49.4.5, 1979.

[FUC 70] Fuchs E. and Jackson P.E., "Estimates of distributions of random variables for certain computer commumications traffic models", Commun. of Assoc. for Comput. Mach., Vol.13, pp.752-757, Dec. 70.

[GAR 72] Garfinkel R.S. and Nemhauser G.L., "Integer programming", John Wiley and Sons, New York, 1972.

[JAI 68] Jaiswal N.K., "Priority queues", Academic Press, New York, 1968.

[KER 72] Kershenbaurn A. and Van Slyke R., "Computing minimum spanning tree efficiently", Proc. Assoc. for Comput. Mach. Annun. Conf., pp.518-527, Aug. 72.

[KLE 75I] Kleinrock L., "Queueing systems Vol.I - theory", John Wiley and Sons, New York, 1975.

[KLE 75II] Kleinrock L., "Queueing systems Vol.II - computer application", John Wiley and Sons, New York, 1975.

[KLI 68] Klipstein D.L., "The contributions of Edsel Murphy to the understanding of the behaviour of inaminate objects", Technicalities, pp.16-17, Jun. 68.

[KO 82] Ko K.T. and Davis B.R., "Design of a demand assignment satellite switched space division multilple access communication network", Jour. Aust. Telecommun. Res., Vol.16, No.2, pp.25-38, 1982.

[KO 83] Ko. K.T. and Davis B.R., "A space division multiple access protocol for spotbeam antenna and satellite switched communication network", to be published in IEEE Selected Areas in Commun., Jan. 83.

[KOB 77] Kobayashi H. and Konheim A.G., "Queueing models for computer communications system analysis", IEEE Trans. on Commun., Vol.COM-25, No.1, pp.2-29, Jan. 77.

[KON 72] Konheim A.G. and Meister B., "Service in a loop system", Jour. of Assoc. for Comput. Mach., Vol.19, No.1, pp.92-108, Jan. 72.

[LAM 79] Lam S.S., "Satellite packet communication - multiple access protocols and performance", IEEE Trans. on Commun., Vol.COM-27, No.10, pp.1456-1466, Oct. 79.

[MAR 78] Mark J.W. and Ng S.F.W., "Satellite packet switching with global assignments and batch Poisson arrivals", IEEE Trans. Computers, Vol.C-27, No.12, pp.1216-1221, Dec. 78.

[MCG 77] McGregor P.V. and Shen D., "Network design : an algorithm for the access facility location problem", IEEE Trans. on Commun., Vol.COM-25, No.1, pp.61-73, Jan. 77.

[MOR 80] Morris T.J., "Priority queueing network", Bell Sy. Tech. J., Vol.60, No.8, pp.1745-1769, Oct. 81.

[NAL 78] Nalos E.J., "New developments in electromagnetic energy beaming", Proc. of IEEE, Vol.66, No.3, pp.276-289, Mar. 78.

[NAY 66] Naylor T.H., Balintfy J.L., Burdich D.S. and Chu K., "Computer simulation techniques", John Wiley and Sons, New York, 1966.

[NG 77] Ng S.F.W. and Mark J.W., "A multiaccess model for packet switching with a satellite having some processing capability", IEEE Trans. on Commun., Vol.COM-25, No.1, pp.128-135, Jan. 77.

[NUS 77] Nuspl P.P., Brown K.E., Steenaart W. and Ghicopoulous B., "Synchronization methods for TDMA", Proc. of IEEE, Vol.65, No.3, pp.434-444, Mar. 77.

[PAW 81] Pawlita P.F., "Traffic measurements in data networks, recent measurement results, and some implications", IEEE Trans. on Commun., Vol.COM-29, No.4, pp.525-535, Apr. 81.

[PRI 57] Prim R.C., "Shortest connection networks and some generalizations", Bell Syst. Tech. J., Vol.36, pp.1389-1401, Nov. 57.

[RET 80] Retnadhas C., "Satellite Multiple Access Protocols", IEEE Commun. Magazine, Vol.18, No.5, pp.16-20, Sep. 80.

[REU 80] Reudnik D.O. and Yeh Y.S., "Scanning spotbeams - a new approach to satellite communications", Bell Lab. Record, Vol.58, No.2, pp.38-45, Feb. 80.

[ROB 73] Roberts L., "Dynamic allocation of satellite capacity through packet reservation", Proc. AFIPS Conf., Vol.42, Jun 73.

[ROS 82] Ross M.J. and Mowafi O.A., "Performance analysis of hybrid switching concepts for integrated voice/data communications", IEEE Trans. on Commun., Vol.COM-30, No.5,pp.1073-1087, May 82.

[ROT 69] Roth R., "Computer solutions to minimum-cover problems", Operations Res., Vol.17, pp.455-465, 1969.

[RUB 79] Rubin I., "Access-control disciplines for multi-access communication channels : reservation and TDMA schemes", Trans. on Infor. Th., Vol.IT-25, No.5, pp.516-536, Sep. 79.

[SCHM 74a] Schmidt W.G., "Satellite-switched TDMA : transponder-switched or beam-switched", AIAA 5th Com. Sat. Sy. Conf., Los Angeles, Paper 74-460, 1974.

[SCHM 74b] Schmidt W.G., "Satellite time division multiple access systems: past, present and future", Telecommunication, Vol.7, pp.21-24, Aug. 74.

[SCHW 77] Schwartz M.,"Computer-communication network design and analysis", Prentice-Hall, New Jersey, 1977.

[SMI 78] Smith K.W. and Carter C.R., "Synchronizing to the switched satellite using a microprocessor based system", Proc. ICC, pp.43.2.1-43.2.5, 1978.

[STE 80] Stern T.E., Schwartz M., Meadows H.E., Ahmadi H.K., Gadre J.G., Gopal I.S. and Matsuo K., "Next generation communications satellites : multiple access and network studies", Rep. of Dept. Electrical Eng., Columbia University, Jul 80.

[SYS 58] Syski S., "Congestion theory in telephone systems", Oliver and Boyd, London, 1958.

[TEL 78] Telecom Australia, "A course in teletraffic engineering", Traffic Engineering Section, Planning Services Branch, Headquarter, 1978.

[TOB 76] Tobagi F.A. and Kleinrock L., "Packet-switching in radio channels : Part III - Polling and (dynamic) split-channel reservation multiple access", IEEE Trans. on Commun., Vol.COM-24, No.8, pp.832-845, Aug. 76.

[TOB 80] Tobagi F.A., "Multiaccess protocols in packet communication systems", IEEE Trans. on Commun., Vol.COM-26, No.4, pp.468-488, Apr. 80.

[TOR 73] Toregas C., ReVelle C. and Swain R., "Reply to Rao´s notes on the location of emergency service facilities", Operations Res., Vol.21, pp.1262-1267, 1973.