

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

**Multiple Access Techniques for Broadband
Multimedia Satellite Networks**

Rocco Di Girolamo

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montréal, Québec, Canada

August 2004

© R. Di Girolamo, 2004



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-96959-2

Our file Notre référence

ISBN: 0-612-96959-2

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

Multiple Access Techniques for Broadband
Multimedia Satellite Networks

Rocco Di Girolamo, Ph.D.
Concordia University, 2004

Access to multimedia services is currently being considered by both wireless and wireline operators. Although satellites cannot compete directly with other alternatives, they can provide benefits in certain niche markets due to their wide coverage area and distance insensitivity. It is the goal of these satellite operators to simultaneously maximize channel utility (and revenue) and to meet the stringent quality of service requirements of the new multimedia services. To achieve both of these objectives, an efficient multiple access (MAC) protocol is required.

Satellite MAC protocols have traditionally been based either on fixed assignment (for voice and video) and/or random assignment (for data). The unique feature of the newer multimedia services is that they generate a mix of these three traffic types, and moreover the traffic is much burstier than originally expected. As a result, the traditional protocols fail to provide the required performance. This work addressed this issue by proposing two new schemes – one based on combined free/demand assignment multiple access (CFDAMA) with a multi-frequency time-division multiple access (MF-TDMA) frame, and the other based on flow controlled random access with a multi-code time-division multiple access (MC-TDMA) frame. Performance results (for jitter-tolerant packet delay and real-time loss probability) show that both techniques perform well in a bent-pipe environment. For an on-board processing system, both protocols can be adapted to deal with possible congestion at the output of the on-board baseband switch. Performance of the modified CFDAMA scheme is superior to that of a rate-based technique

proposed in the literature. For the proposed MC-TDMA system, soft blocking and congestion are both handled by controlling packet transmissions through a simple flow control parameter.

Performance evaluation of satellite multimedia networks is also a very important topic. Event-driven simulations (with OPNET) prove invaluable to evaluate protocol performance, but these have limitations with regards to flexibility and accuracy (especially in evaluating real-time loss probability). As a result, certain key measures were obtained through analysis. In particular, the Matrix Geometric approach was applied to the CFDAMA scheme to determine real-time loss probability and jitter-tolerant packet delay.

ACKNOWLEDGEMENTS

I wish to gratefully acknowledge my mentor and supervisor, Dr. Le-Ngoc, for his expertise, professional support, and friendship over the years. As well, I extend my gratitude to my colleagues in the Department of Electrical & Computer Engineering at Concordia University, for whose counsel I am indebted. A special thank you to Dr. Mehmet-Ali and Dr. Ahmad for their constant encouragement over the last few months.

TABLE OF CONTENTS

LIST OF FIGURES.....	VIII
LIST OF TABLES.....	IX
LIST OF ACRONYMS/ABBREVIATIONS	X
1 INTRODUCTION	1
1.1 FUTURE BROADBAND ACCESS NETWORKS	1
1.2 GENERIC SATELLITE NETWORK	3
1.2.1 <i>Satellite Primer [7][8]</i>	5
1.3 SCOPE OF RESEARCH AND CONTRIBUTIONS	8
1.3.1 <i>Contributions of Research</i>	9
1.4 THESIS OUTLINE.....	10
2 BACKGROUND INFORMATION	13
2.1 QUALITY OF SERVICE FOR BROADBAND MULTIMEDIA NETWORKS.....	13
2.1.1 <i>Integrated Services (IntServ) [17][18][19][20]</i>	14
2.1.2 <i>Differentiated Services (DiffServ) [21][22][23]</i>	14
2.1.3 <i>Multiprotocol Label Switching (MPLS) [24][25]</i>	15
2.1.4 <i>Integration of Broadband Multimedia QoS Framework and Satellite Access Network</i>	16
2.2 MULTIPLE ACCESS PROTOCOLS	17
2.2.1 <i>Channelization</i>	17
2.2.2 <i>Channel Access</i>	19
2.3 STANDARD DEVELOPMENT	37
2.3.1 <i>DVB-RCS [14]</i>	37
2.3.2 <i>DOCSIS [15]</i>	38
2.3.3 <i>Standards Summary</i>	40
2.4 TECHNIQUES FOR PERFORMANCE EVALUATION	40
2.4.1 <i>Event Driven Simulation [57]</i>	40
2.4.2 <i>Analytical Performance Evaluation</i>	41
3 BENT-PIPE SYSTEM	44
3.1 INTRODUCTION.....	44
3.2 TRAFFIC MIXES AND SIMULATIONS CONSIDERED	45
3.2.1 <i>Parameter Matching</i>	50
3.3 MF-TDMA WITH CFDAMA	56
3.3.1 <i>System Description</i>	56
3.3.2 <i>Performance Results</i>	61
3.4 MC-TDMA WITH RANDOM SLOT SELECTION.....	82
3.4.1 <i>System Description</i>	82
3.4.2 <i>Types of Flow Control</i>	88
3.4.3 <i>Performance Results</i>	90
4 ON-BOARD PROCESSING SYSTEM	94
4.1 INTRODUCTION.....	94
4.2 HOW TO DEAL WITH CONGESTION	96
4.3 PROPOSED SOLUTIONS	98
4.3.1 <i>MF-TDMA with CFDAMA</i>	98
4.3.2 <i>MC-TDMA with Random Slot Selection</i>	110
5 CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH	115
5.1 SUGGESTIONS FOR FURTHER RESEARCH	116
A REVIEW OF CHANNELIZATION TECHNIQUES	118

A.1	CHANNELIZATION TECHNIQUES	119
A.1.1	<i>Radio Capacity</i>	121
A.1.2	<i>Information Capacity</i>	125

LIST OF FIGURES

FIGURE 1-1: ACCESS NUMBERS FOR THE US MARKET [1]	2
FIGURE 1-2: FUTURE BROADBAND NETWORK WITH SATELLITE LINK	5
FIGURE 1-3: SATELLITE ORBITS [7].....	7
FIGURE 2-1: SHARING OF AVAILABLE RESOURCES	17
FIGURE 2-2: CHANNELIZATION TECHNIQUES	18
FIGURE 2-3: CLASSIFICATION OF CHANNEL ACCESS TECHNIQUES.....	20
FIGURE 2-4: COLLISIONS IN SPREAD-SLOTTED ALOHA.....	22
FIGURE 2-5: CLASSIFICATION OF RESERVATION ACCESS SCHEMES	25
FIGURE 2-6: FRAME STRUCTURE FOR ALOHA RESERVATION & TDMA RESERVATION	25
FIGURE 2-7: FODA-TDMA FRAME.....	26
FIGURE 2-8: DESIRED JITTER-TOLERANT PACKET DELAY PERFORMANCE.....	28
FIGURE 2-9: FRAME STRUCTURE FOR MOVEABLE BOUNDARY RANDOM/DAMA.....	30
FIGURE 2-10: BASIC OPERATION OF COMBINED FREE/DEMAND ASSIGNMENT SCHEMES	32
FIGURE 2-11: MC-CDMA TRANSMITTER (BASED ON [54])	35
FIGURE 3-1: MODEL OF BENT-PIPE SYSTEM	44
FIGURE 3-2: MAPPING OF ONE VIDEO SOURCE TO A 2-STATE MMPP.....	51
FIGURE 3-3: FLOWCHART FOR PARAMETER MATCHING ALGORITHM	54
FIGURE 3-4: TYPICAL EARTH STATION FOR A VOICE DOMINANT CASE	55
FIGURE 3-5: IDC CURVES FOR A TYPICAL EARTH STATION TRAFFIC	56
FIGURE 3-6: MF-TDMA FRAME FORMAT	57
FIGURE 3-7: EXAMPLE OF A BURST TIME PLAN (BTP).....	59
FIGURE 3-8: QUEUING MODEL OF EARTH STATION.....	62
FIGURE 3-9: AVERAGE JT PACKET DELAY	63
FIGURE 3-10: AVERAGE JT PACKET DELAY: EFFECT OF DATA TRAFFIC MODEL.....	63
FIGURE 3-11: REAL-TIME LOSS PROBABILITY (VIDEO/VOICE/DATA DOMINANT CASES)	68
FIGURE 3-12: COMPARISON OF PROTOCOL WITH VARIOUS MAXRBDC	69
FIGURE 3-13: EMBEDDING POINTS FOR SYSTEM	70
FIGURE 3-14: V_n CALCULATION FOR CASE 1	72
FIGURE 3-15: GRAPHICAL REPRESENTATION OF P MATRIX.....	76
FIGURE 3-16: PERFORMANCE OF MF-TDMA CFDMA SCHEME	82
FIGURE 3-17: EARTH STATION SLOT SELECTION	83
FIGURE 3-18: HARD AND SOFT BLOCKING	85
FIGURE 3-19: DISTRIBUTION OF NUMBER OF TRANSMITTED CELLS IN A TIME-SLOT	88
FIGURE 3-20: EXAMPLES OF FLOW CONTROL PROCEDURE	90
FIGURE 3-21: DATA DELAY VS LOAD	91
FIGURE 3-22: EFFECT OF PROCESSING GAIN ON SOFT BLOCKING	92
FIGURE 3-23: DATA DELAY & Pr[SOFT BLOCKING] vs p_{fc}	93
FIGURE 4-1: MODEL OF ON-BOARD PROCESSING SYSTEM.....	96
FIGURE 4-2: ARRIVALS TO SATELLITE IN RESPONSE TO RESERVATIONS	99
FIGURE 4-3: PROCEDURE AT EARTH STATIONS	101
FIGURE 4-4: PROCEDURE AT SCHEDULER.....	102
FIGURE 4-5: LOADING SCENARIO FOR 4-BEAM NETWORK.....	105
FIGURE 4-6: INSTANTANEOUS DL QUEUE SIZE COMPARISON: DL BEAM 0	107
FIGURE 4-7: QUEUE SURVIVOR FUNCTION COMPARISON: DL BEAM 0	108
FIGURE 4-8: INSTANTANEOUS EARTH STATION QUEUE SIZE (FOR EARTH STATION SENDING TO DL BEAM 0)	110
FIGURE 4-9: PROCEDURE AT SCHEDULER.....	113
FIGURE 4-10: PROCEDURE AT EARTH STATIONS	114
FIGURE A-1: CHANNELIZATION TECHNIQUES	120
FIGURE A-2: FREQUENCY REUSE FOR A MULTIBEAM ENVIRONMENT.....	123
FIGURE A-3: RADIO CAPACITY COMPARISON (TDMA, FDMA, CDMA)	124
FIGURE A-4: MODELING OF MULTIPLE ACCESS COMMUNICATIONS SYSTEM	125
FIGURE A-5: INFORMATION CAPACITY COMPARISON (TDMA, FDMA, CDMA)	127

LIST OF TABLES

TABLE 1-1: COMPARISON OF SATELLITE ORBITS	7
TABLE 1-2: MAJOR SATELLITE BANDS.....	7
TABLE 3-1: SATELLITE SYSTEM PARAMETERS	46
TABLE 3-2: TRAFFIC SCENARIOS CONSIDERED	46
TABLE 3-3: PARAMETERS FOR INDIVIDUAL ON-OFF SOURCES.....	47
TABLE 3-4: NUMBER OF VOICE AND VIDEO SOURCES PER EARTH STATION PER TRAFFIC MIX.....	50
TABLE 3-5: MAPPING OF TRAFFIC TO CAPACITY TYPE	58
TABLE 3-6: EFFECT OF MAXRBDC	67
TABLE 3-7: COMBINATIONS TO PRODUCE TRANSITION FROM $(0, J, L) \rightarrow (I, V, W)$	77
TABLE 4-1: ASSIGNMENT MATRIX (INFORMATION PER UPLINK/DOWNLINK BEAM).....	99
TABLE 4-2: LOAD MATRIX (INFORMATION PER UPLINK/DOWNLINK BEAM)	112

LIST OF ACRONYMS/ABBREVIATIONS

AF	Assured Forwarding
ATM	Asynchronous Transfer Model
AVBDC	Absolute VBDC
BE	Best Effort
BER	Bit Error Rate
BoD	Bandwidth on Demand
BTP	Burst Time Plan
CAC	Call Admission Control
CC	Congestion Control
CDMA	Code Division Multiple Access
CGR	Cell Generation Rate
CFDAMA	Combined Free/Demand Assignment Multiple Access
CIR	Committed Information Rate
CMTS	Cable Modem Termination System
C-PODA	Contention Priority Ordered Demand Assignment
CREIR	Combined Reservation with Explicit then Implicit Reservation
CRA	Collision Resolution Algorithm
CRA	Continuous Rate Assignment (correct form should be clear from context)
DCA	Dynamic channel allocation
DE	Default
DiffServ	Differentiated Services
DOCSIS	Data-Over-Cable Service Interface Specifications
D-RRAA	Dynamic Random-Reservation Adaptive Assignment
DSL	Digital Subscriber Line
DVB-RCS	Digital Video Broadcasting – Return Channel System
EF	Expedited Forwarding
ES	Earth Station
FCA	Free Capacity Assignment
FCFS	First Come First Service
FDMA	Frequency Division Multiple Access
FEC	Forward Equivalence Class
FEC	Forward Error Correction (correct form should be clear from context)
FODA-TDMA	FIFO Ordered Demand Assignment TDMA
GEO	GEOSynchronous
HEO	High Earth Orbit
IBR	In-Band Requesting
IDC	Index of Dispersion of Counts
IntServ	Integrated Services
IP	Internet Protocol
JT	Jitter-Tolerant
LEO	Low Earth Orbit
MAC	Multiple Access Controller
MC-TDMA	Multi-Code TDMA
MEO	Medium Earth Orbit
MF-TDMA	Multi-Frequency TDMA
MMPP	Markov Modulated Poisson Process
MPEG	Moving Picture Experts Group
MPLS	Multiprotocol Label Switching
NCC	Network Control Center
OBP	On-Board Processing

OBR	Out-of-Band Requesting
OPNET	Optimized Network Simulator
PCS	Personal Communication Service
PHB	Per Hop Behavior
PSD	Power Spectral Density
QoS	Quality of Service
RBDC	Rate Based Dynamic Capacity
rtPS	Real-Time Polling Service
RTT	Round Trip Time (= 240 msec for a GEO satellite)
SIT	Satellite Interactive Terminal
SNR	Signal-to-Noise Ratio
SOHO	Small Office Home Office
SREJ Aloha	Selective Reject Aloha
TD/FD	Time Division/Frequency Division
TDMA	Time Division Multiple Access
TOS	Type of Service
UGS-AD	Unsolicited Grant Service with Activity Detection
UGS	Unsolicited Grant Service
VBDC	Volume Based Dynamic Capacity
VPI/VC	Virtual Path Indicator/Virtual Channel Indicator
WFBOD	Weighted Fair BoD

Chapter 1

Introduction

1.1 Future Broadband Access Networks

As the popularity of high bandwidth applications increases, so too does the need to build a network infrastructure that is capable of providing these applications to end users. In today's networks, one is likely to find applications ranging from the traditional telephony, video broadcasting, and data transfer, to the newer Internet based applications such as web browsing, and video/audio streaming. These will soon be joined by future applications including video conferencing, video on demand, tele-education/tele-medicine, and remote storage. The Internet is emerging as the network of choice for providing these services to the commercial and enterprise markets. The main problem is providing **access** to this network. Over the last couple of years, the dial-up networking solution has been supplanted by cable modem and digital subscriber line (DSL) implementations -- as provided by the cable operators and telephone operators. Although not as successful as their cable and telephony counterparts, the wireless industry has also begun to enter the growing access market. An indication of the phenomenal growth of the US access market is provided in Figure 1-1 [1]. At the end of this year, the number of subscribers to these services is expected to double from the 2002 numbers, with wireless access (and satellite included) accounting for 14% of this total. Generally, it is agreed that satellite systems cannot compete directly with their terrestrial based counterparts, but can complement these in a number of ways. This, in large part, is due to the following advantages:

- i. Distance and location insensitivity: for a terrestrial network, the location of some users will make broadband access very difficult (lack of infrastructure, difficult terrain, etc.). In

contrast, a satellite in a high earth orbit can give access to all terminals in its coverage area.¹

- ii. Provides full mesh interconnectivity: all users in a satellite coverage area can communicate with each other. That is, an end user can communicate with any other end user.
- iii. Provides star connectivity: An on-ground hub (or gateway) can act as a central distribution center. This may be the first type of implementation for a satellite based network [2]. This point-to-multipoint connectivity is inherent in satellites, as the downlink is naturally broadcast.
- iv. Provides a certain degree of mobility: the system studied in this thesis is intended for fixed users and consequently mobility is not considered.

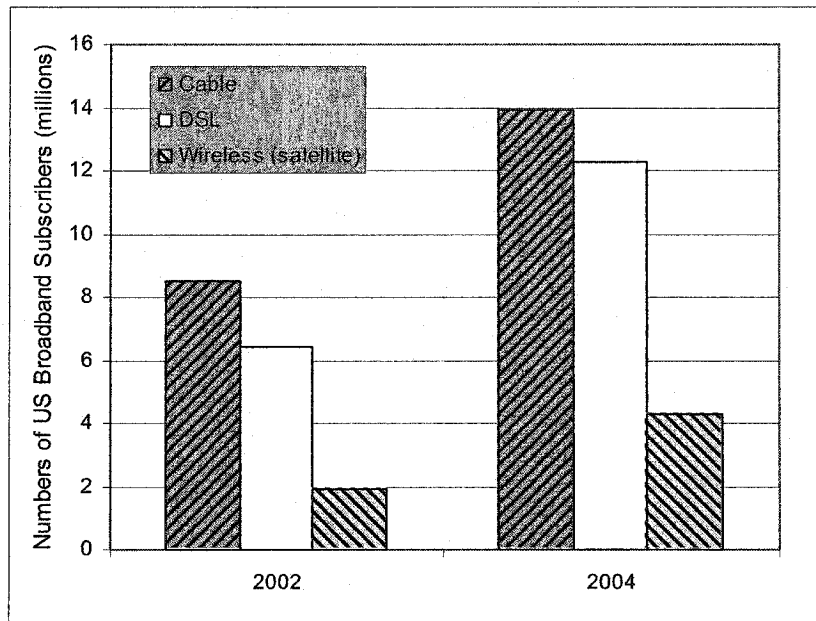


Figure 1-1: Access Numbers for the US Market [1]

Although the cable modem and DSL operators have a commanding lead in terms of user penetration, some in the satellite industry contend that this large lead is not necessarily negative. In fact, the two competitors have established a minimum performance level that is deemed acceptable, and satellites are the only solution for those users in rural areas who are underserved by the terrestrial alternatives [1].

¹ The coverage area of a satellite refers to the area on the ground from where terminals are able to communicate with the satellite antenna.

The satellite industry, as a whole, has begun to emerge from the huge downturn in the telecom sector of a few years ago. Most of the revolutionary projections of the early 90's have been replaced with a more subtle view of how and where the satellite industry should be going. A number of proposed systems have either been totally abandoned, or have been put on hold while the companies try to find additional sources of funding. Of these we can mention the CDMA based Voicespan system [3], the low earth orbit constellation from Teledesic [4], and the geosynchronous Astrolink system [5]². These systems all had major backers (AT&T, Motorola, Boeing, etc.) but in hindsight may have been overly ambitious. These systems tended to be global in scope, requiring constellations with a large number of satellites. The satellites themselves were extremely advanced, even by today's standards, and often made use of proprietary technology. This is in contrast to the evolutionary approaches taken by a number of smaller companies, who decided to leverage the existing infrastructure and to establish standards for industry adoption. In addition, their markets were decidedly more regional. Companies such as SES Global are already offering this network **access**, while others such as WildBlue are very close to starting operation (a very good summary of these systems can be found at [6]).

1.2 Generic Satellite Network

A typical configuration of a possible future satellite access network is shown in Figure 1-2. The network consists of five main components: satellite interactive terminals (SITs), gateways, switching nodes, a network control center (NCC), and the satellite. Each of these is briefly described below:

SITs: represent the access points to the network. Typically, these stations would contain multiplexers that could combine the voice, video, and data type traffic streams coming from a single location. Alternatively, these stations could be LAN routers, and the like. What these terminals share, is the need to communicate with each other.

² Information about satellite constellations and CDMA is provided in the main body of this thesis.

Gateways/Hubs: These mark the boundary between the satellite network and the terrestrial public network.

Switching nodes: lie within the terrestrial public network, and are connected by a mesh of physical links. Some switching nodes may also have the capability to communicate through the satellite.

Network Control Center: These earth stations are responsible for the maintenance of the network, and can also be used to monitor SIT usage and to provide for billing.

Satellite: Satellites can be in any number of orbits, can use any number of frequency bands, and can be classified as either transparent or regenerative. When dealing with satellite or wireless networks, it is important to understand the subtle difference between forward/reverse and uplink/downlink paths. A forward path is typically from gateway station to SIT, whereas a reverse path is from SIT to gateway. Each of these paths has an uplink (path toward the satellite) and downlink (path from satellite). Additional information on the aspects mentioned above can be found in the next section.

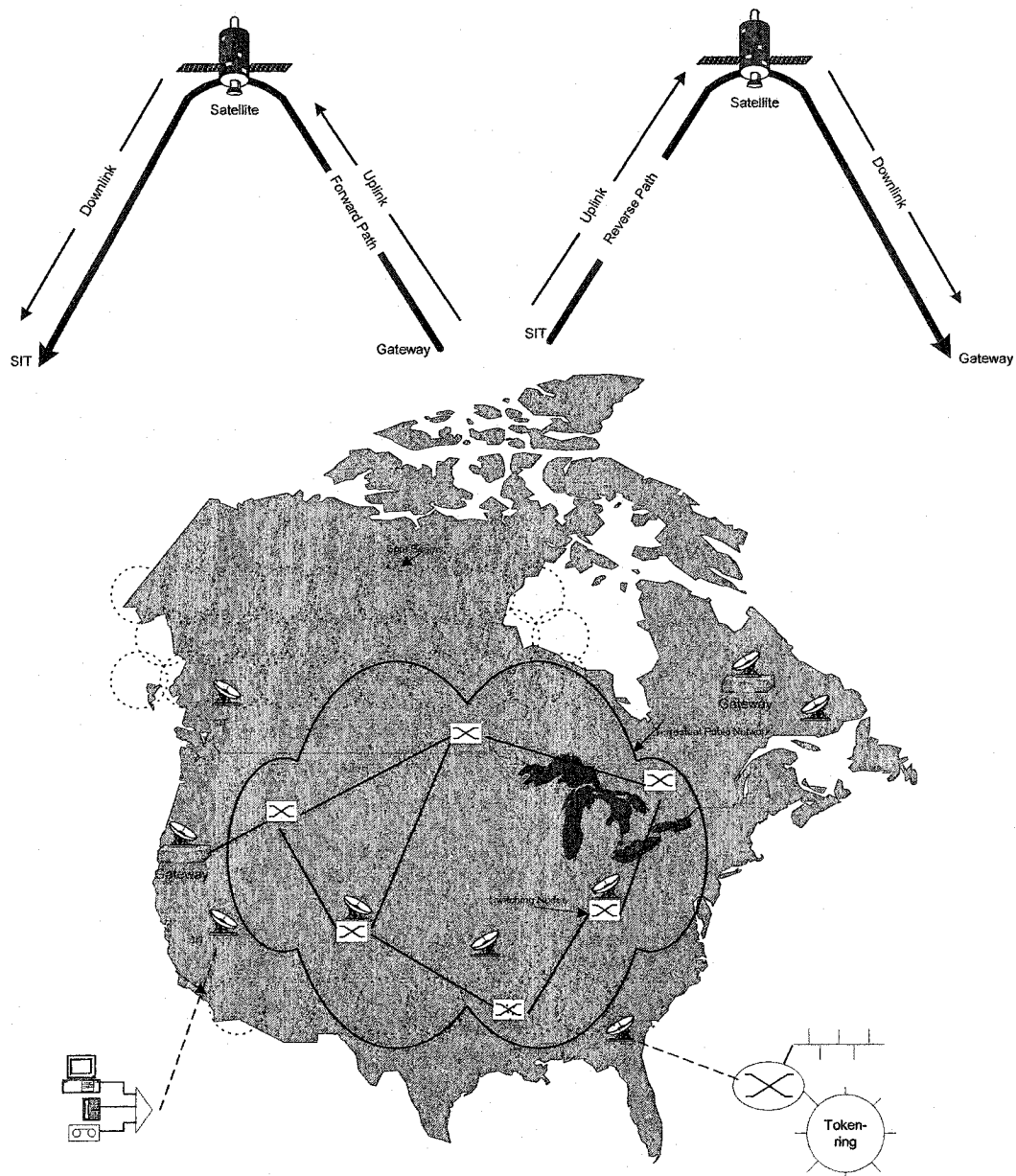


Figure 1-2: Future Broadband Network with Satellite Link

1.2.1 Satellite Primer [7][8]

Satellites can be in any one of three major orbits, distinguished by altitude (and the corresponding transmission propagation delay) and coverage area. Low earth orbit (LEO) satellite altitudes range from 500 to 2000 km above the ground. Satellites in this orbit have a period of rotation anywhere from 90 minutes to 120 minutes and a 1 to 7 msec propagation delay (on either

the uplink or the downlink). As the satellites are relatively close, in comparison to other orbits, many such satellites would be required for complete coverage of an area such as North America. Furthermore, as the satellites move quickly with respect to the SITs, the system may need some sort of tracking or steering mechanism for the antennas. As the transmission delay is very low, these orbits were originally thought of as being ideal for voice communications (as proposed in the failed Globalstar and Iridium voice communication networks). For a medium earth orbit (MEO), satellites have an altitude in the order of 10000 km, resulting in an uplink (or downlink) transmission propagation delay of 33 msec. With this higher altitude, the coverage area is larger than for a LEO satellite, but the transmission delay is also higher. For this thesis, we will assume a satellite in geosynchronous (GEO) orbit (35786 Km above the earth's equator). Satellites in GEO orbit have a huge coverage area (actually three such satellites can provide global coverage). Satellites in a GEO orbit appear stationary to SITs and gateways, and so do not require complicated algorithms for antenna pointing. As will be explained in a later section, the coverage area will actually be divided into a number of smaller spot beams, which allows for better management of the bandwidth. The main drawbacks of the GEO orbit are the large transmission propagation delay (~120 msec), and the limited orbital slots.³ This information is summarized in Figure 1-3 and Table 1-1. Note that Figure 1-3 shows a fourth, highly elliptical, orbit type (High earth orbit (HEO)) – this orbit has excellent coverage for the northern hemisphere.

³ As the GEO orbit is over the equator, there are a limited number of positions available to “park” the satellites. Satellites using the same frequency bands are typically spaced 2-3° apart.

Table 1-1: Comparison of Satellite Orbits

	LEO	MEO	GEO
Altitude [km]	~1000	~10000	35786
Propagation delay [msec]	~3	30	120
Coverage Area	Low	Average	High
Proposed/Discontinued Systems	Teledesic, Skybridge, ...	West, Starlynx, ...	Spaceway, SATLYNX, ...

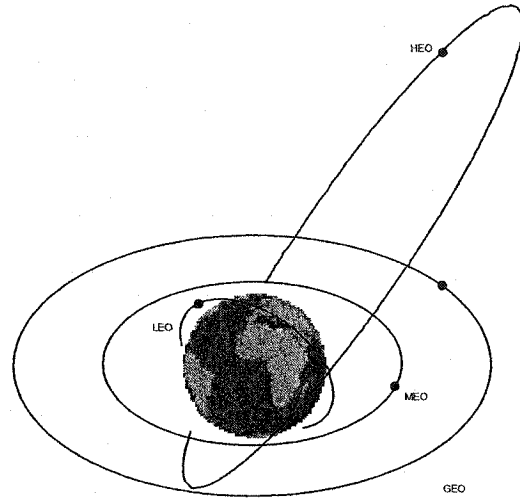


Figure 1-3: Satellite Orbits [7]

Satellites can use any number of frequency bands. Higher frequencies are generally preferred as the resulting antenna gain (gain is proportional to the square of the operating frequency) can be traded off against the antenna size. Smaller antennas are a very important consideration for the small office home office (SOHO) market. As an added benefit, the higher bands are typically wider, and are not as populated as the lower bands, resulting in less inter-satellite interference. Table 1-2 shows the frequency range definitions of the main satellite bands, separated according to uplink and downlink.

Table 1-2: Major Satellite Bands

	Downlink [GHz]	Uplink [GHz]
C Band	3.4-4.2 & 4.5-4.8	5.725-7.075
Ku Band	10.7-12.2	12.5-13.25 & 13.75-14.5
Ka Band	17.7-21.2	27-31

As far as frequency bands are concerned, the results in this thesis will assume the Ka band. Owing to the large bandwidth available, it is the band of choice for most satellite networks attempting to provide access (although the Ku band is still being considered as a first step). The last major classification of satellites is between *transparent* and *regenerative*. A transparent satellite acts like giant repeater in the sky. It receives data from the SITs and/or gateways on the

uplink, shifts this data to the downlink frequency, and broadcasts the information on the downlink. These will also be referred to as bent-pipe (BP) systems. The satellite has very little processing capability. In contrast, regenerative satellites will demodulate the received data, and then retransmit the information on the downlink. As the information is regenerated on board the satellite, the uplink noise is not propagated on the downlink. Furthermore, this will break the dependency between the uplink and downlink, and will allow independent optimization of both paths. In addition, as the satellite is recovering the data that it is forwarding, it can use this information to implement special algorithms. These satellites are said to possess on-board processing (OBP) capability. Both satellite types (BP or OBP) typically divide the coverage area into a number of smaller spot beams. These spot beams are very similar to the cells in a PCS system, and they allow for frequency reuse. As will be explained in the next section, this work addresses both satellite types – transparent and regenerative.

1.3 Scope of Research and Contributions

Within the overall satellite architecture, there are numerous networking problems to be resolved. To begin with, it must be remembered that the **access** network must be capable of handling current as well as future applications, and to provide the requisite quality of service (QoS) guarantees to these applications. At any one time, a number of SITs and gateways may be active and communicating through the satellite – these are said to be connected. As far as the satellite portion of the network is concerned, these are equivalent, and will be referred to as earth stations (ESs) or terminals. Clearly, the traffic transmitted by these earth stations is multimedia – a mixture of real-time (RT) voice and video, and jitter-tolerant (JT) data. The first issue to address is the problem of multiple access – how can geographically dispersed ESs share the limited uplink bandwidth efficiently. For a multibeam system, earth station traffic from one spot beam may be destined to another spot beam. Although the switching/routing of this traffic can be done on the ground, for a regenerative satellite it would be better to use an on-board baseband packet

switch. The study of space-qualified ATM-like packet switches is currently a major research area [10][11]. Notice that for the space portion of the access network, multiplexing occurs at three different levels:

- i. at the earth stations, where the voice, video, and data traffic are multiplexed prior to uplink transmission;
- ii. at the air interface, where the traffic from different earth stations is multiplexed;
- iii. at the output of the packet switch, where traffic from different input spot beams is multiplexed into a single output spot beam.

As a result of this multiplexing, congestion may arise at the output of the on-board switch. Traffic control, which can be either preventative (shaping, policing, call admission) or reactive (traffic marking, rate control), is therefore very important [9]. Unfortunately, the long propagation delay associated with the satellite poses a problem to the control algorithms. To complicate matters, the quality of service guarantees provided to the connections, must be maintained over the satellite hop.

1.3.1 Contributions of Research

Of the areas listed above, the main focus of this research deals particularly with the multiple access problem – i.e., the sharing of the common and limited uplink resource. This resource manifests itself as a bandwidth and power limitation. Although this type of multiple access problem has been investigated for many years, this work proposes new schemes that are better suited to deal with the type of multimedia traffic considered. Namely, the real-time traffic is assumed to be of variable bit rate (and not constant), while the JT traffic is taken to be self-similar. As a result, this research has answered the following questions:

1. Can a spread spectrum approach be used in a satellite network serving connections with multimedia traffic and still provide QoS guarantees? Narrowband spread spectrum techniques are currently being used in terrestrial cellular systems (CDMAOne) [12], and have shown promise in this environment [13]. This research shows that the advantages do carry forward to a satellite architecture. The proposed scheme is based on an MC-TDMA frame.

2. How can a dynamic channel allocation (DCA) scheme be tailored to handle the multimedia traffic considered? With a dynamic allocation strategy, uplink capacity is assigned to earth stations based on their instantaneous needs. The scheme proposed performs well for all traffic types and is compliant with current standards.
3. For most of the regenerative systems being proposed, congestion control and uplink access are treated as separate processes. Basically, the uplink access is concerned only with maximizing the uplink utility. If congestion occurs in a downlink queue of the on-board switch, congestion information will be sent to the earth stations contributing to this congestion. This information will attempt to restrict transmissions, until the congestion is cleared. The congestion control can be thought of as overriding the uplink access protocol. However, as mentioned earlier, the effectiveness of the congestion control algorithm is often affected by the long round trip propagation delay. The question is then whether the uplink access protocols can be modified to simultaneously maximize uplink utility and to minimize both average packet delay and downlink congestion. We have proposed two versions of *smart* uplink access protocols, one based on the MF-TDMA CFDMA system, and the other based on the MC-TDMA system.
4. What type of performance can be expected for the multimedia traffic using a satellite as the access network? This performance was measured in terms of numerous parameters. The most important being the average jitter-tolerant packet delay, the real-time loss probability, and the channel utility. In addition, for the spread spectrum system, soft blocking was investigated. Packets that are soft blocked experience higher than expected bit error rates (BER). In these performance evaluations, the novelty of the work is in the modeling of the system.

The simulation model itself can also be considered a minor contribution of the thesis work. The model was implemented using the OPNET simulation tool, and was designed to be very flexible in terms of network layout (number of earth stations and their makeup).

1.4 Thesis Outline

In this section, the outline of this document is discussed. The next chapter provides the necessary background information for the rest of the material that follows. The chapter focuses on four main areas:

1. How QoS is implemented in today's Internet, and how this traffic is mapped in the satellite access network.
2. A review of basic multiple access strategies, and a classification of existing techniques according to method of channelization and method of channel access. Each of the channelization techniques is evaluated in terms of two criteria – information capacity and radio capacity. The method of accessing these channels is then considered. Numerous multiple access techniques are discussed, with a specific emphasis on the limitations of these techniques. The criteria for their evaluation is in terms of suitability for the multimedia traffic being considered, satellite uplink bandwidth efficiency, type of control, number of earth stations that can be supported, and overall performance.
3. The basic techniques for the queuing analysis of such networks.
4. The current architectures proposed by two competing standards DVB-RCS [14] and DOCSIS [15]. The section includes a brief description of how both techniques plan to offer QoS, and the justification for focusing the research on the former.

Chapter 3 provides the performance evaluation for two multiple access schemes suitable for a transparent (or bent-pipe) satellite architecture. One technique is based on a demand assignment approach (MF-TDMA and CFDMA) while the second relies on spread spectrum and is from the MC-TDMA family. Particular emphasis is placed on the spread spectrum technique, and its performance evaluation. As will be explained, these systems are interference limited, and therefore their queuing performance will depend on many physical layer problems. The chapter also includes a discussion about the traffic types and models used, as well as the loading scenarios considered. The chapter also shows how the access schemes are mapped to the classes proposed within the DVB-RCS standard. Note that as the system uses a BP satellite, no baseband

on-board switching is performed. It is assumed that all the traffic from an uplink spot beam is frequency shifted, and broadcast on the downlink.⁴

Chapter 4 expands on the results presented for a bent-pipe system, and provides the performance evaluation for two multiple access schemes suitable for a regenerative satellite architecture with OBP capability. These techniques fall under the class of *smart* uplink access techniques, as they make use of the on board queue information to reduce the possibility of congestion. The two schemes are direct extensions of those presented in Chapter 3 – a demand assignment approach and a spread spectrum approach.

Conclusions and suggestions for further research are provided in the final chapter.

⁴ In reality this is not entirely correct, as satellite switched TDMA (SS-TDMA) systems can be implemented for systems using BP satellites. However this form of switching is more per carrier and is not classified as baseband switching.

Chapter 2

Background Information

2.1 Quality of Service for Broadband Multimedia Networks

Today's networks are required to transmit a host of different traffic types. For this thesis, we classify these into three broad categories: real-time video, real-time voice, and jitter-tolerant data⁵. The voice and video have a temporal requirement and cannot tolerate any jitter, but can endure some loss of information. In contrast, the data is extremely loss sensitive but is referred to as being elastic. In other words it is not affected appreciably by jitter and delay. Any network (or combination of networks) should provide some form of guarantee to the traffic, so that the individual needs of each traffic type are met. The network provides a *service* (in terms of the performance guarantee) to ensure that the *quality*, for the application generating the traffic, is maintained. This is referred to as Quality of Service (QoS). How networks provide this QoS has been an ongoing research topic for many years, and the community is generally divided along two camps – those favoring an Asynchronous Transfer Mode (ATM) approach and those favoring an Internet Protocol (IP) approach. Although ATM was initially intended to be a true end-to-end protocol with a rich set of QoS features, the protocol has never been able to penetrate into the access market, where most users are more familiar with IP. In addition to its advantages in terms of QoS, ATM technology also allows for high-speed switching of data. As a result, ATM is currently being used in the carrier backbone networks [16]. Although initially designed as a best-effort network, many of the concepts of ATM are now slowly beginning to migrate into the IP world. To implement QoS, three different IP approaches are being studied, as described in the next sections.

2.1.1 Integrated Services (IntServ) [17][18][19][20]

The integrated services approach proposes two service classes in addition to the best effort service. Each service class is associated with a traffic type requiring a specific type of performance guarantee (service) from the network. In addition to best-effort, we have:

Guaranteed Service: This service is for applications requiring a specific fixed delay bound.

Controlled Load Service: Intended for those applications that require an enhanced best-effort service. That is, for those applications that work well in today's best-effort network when it is lightly loaded. Regardless of true network conditions, applications in this service class will always believe that the network is lightly loaded.

To implement these new service classes, the applications must set-up a path through the network in order to reserve resources. To achieve this, the following should be considered:

1. There must be a way to signal the application requirements to reserve the resources in each network device or router (RSVP is considered as the most probable signaling protocol).
2. Each router must decide whether or not the resource can be reserved (admission control).
3. The traffic has to be marked so that the router knows which of the reserved resources belong to which application. The traffic from each application is called a *flow*, and each flow is treated independently.
4. The routers must be smart enough to deal with each flow independently – this can mean that the flows are queued according to class.
5. The routers have to maintain information for each flow that they handle. This state information must be refreshed periodically. One of the major problems is scalability – as the Internet grows, more and more state information will be required.

2.1.2 Differentiated Services (DiffServ) [21][22][23]

The second approach to QoS is to aggregate applications into a small number of classes of traffic. Individual flows are mapped to one of these classes. The network treats each class

⁵ Note that this classification is mainly for the results presented in Chapters 3 & 4 and is typical of most media access studies. In reality, different applications may have substantially different requirements for their voice, video, and data traffic.

differently, but is not concerned with the individual flows that constitute the traffic class. Every packet in the network is associated with one of these classes. The association between flow and class is specified in one of the fields of the IP packet (namely the Type of Service (TOS) byte). Routers need not keep per-flow information as in IntServ. Rather, the routers are configured to *behave* in a certain way depending on the traffic class of the incoming packet. This is referred to as a Per Hop Behavior (PHB). The most common PHBs are briefly explained below:

Expedited Forwarding (EF): This PHB is reserved for traffic that has strict requirements in terms of delay, loss, and jitter. Any traffic that is marked as being in the EF class should be given the highest priority from the network routers.

Assured Forwarding (AF): This class is suited for traffic that requires some performance guarantees over best-effort. The class is actually divided into sub-classes, named AF1, AF2, AF3, and AF4, with each sub-class having 3 different levels of drop precedence. Depending on the loss and delay requirements of a traffic, it can be mapped into one of these 12 AF PHBs.

Default (DE): Basically reserved for traffic that can be dealt with through a best-effort service. This is the service provided by today's Internet.

2.1.3 Multiprotocol Label Switching (MPLS) [24][25]

MPLS is often considered a convergence of ATM and IP. Initially, the main purpose of the technique was to provide fast switching of packets. Routing decisions are typically made based on the IP destination address stored within the packet. Each router maintains a routing table and cross-references the destination address to determine where to forward the IP packet. As the Internet grows, the routing tables become very large and their updating becomes difficult. Even more significant, the search through the routing table is a very time consuming task – this may pose significant limitations to certain traffic types. A simple remedy to this problem is to prepend a label to each IP packet. This label could be used as an index to the routing table to determine where to forward the packet. With MPLS, routes through the network are configured and

maintained for all traffic belonging to the same Forward Equivalence Class (FEC). Traffic entering the network is assigned to an FEC and a label is prepended. The labels themselves only have local significance. The router will look at the label information, find the correct forwarding path (configured based on the FEC), rewrite the label, and forward the packet. Essentially, this is very similar to the VPI/VCI labels in the ATM taxonomy. More recently, the usefulness of MPLS as a mechanism for “traffic engineering” has been recognized. The research community has begun to focus on the criteria by which traffic is mapped to FECs, and has considered a mapping based on QoS requirements. That is, we can assign traffic that is delay and loss sensitive to one FEC and best effort traffic to another FEC. Since these would be treated independently by the network and its routers, MPLS can be used to provide certain QoS guarantees.

2.1.4 Integration of Broadband Multimedia QoS Framework and Satellite Access Network

In order for the satellite access network to be a viable alternative to its terrestrial counterparts, it must be capable of integrating well with current and future QoS frameworks. Most studies of satellite networks in the mid-90’s focused primarily on ATM QoS. In particular, see the works by Hung [26] and Baiocchi [27]. The main focus of these papers was to show how ATM traffic classes could be mapped to multiple access protocols, and how the limitations of the satellite network could be overcome. Recently, similar results have begun to appear for IP QoS – namely for DiffServ and MPLS (please refer to [28][29]). The former shows a clever mapping of DiffServ PHBs to multiple access strategies and provides performance results for the satellite system. The multiple access scheme they consider is based on the DVB-RCS standard for the return channel. The latter paper shows a similar mapping for the MPLS approach. A good summary is also provided in [30].

As the work in this thesis focuses mostly on the media access protocol, an attempt was made to keep the results very general, so that they can be applied to any of the QoS frameworks. In fact,

all that is needed is the mapping between traffic class and media access protocol. As an illustrative example, in Chapters 3 and 4, we use the mapping proposed in [28].

2.2 Multiple Access Protocols

The multiple access problem has received considerable attention in the last 30 years, as is evidenced by the numerous papers and texts written on the subject (see [31][32] for example). As discussed in Chapter 1, the main focus of this research is on multiple access (MAC) techniques suitable for a geosynchronous satellite – both bent-pipe and on-board processing realizations – and capable of providing the access network to today's terrestrial network. Typically, the multiple access problem can be divided into methods of channelization and methods of channel access.

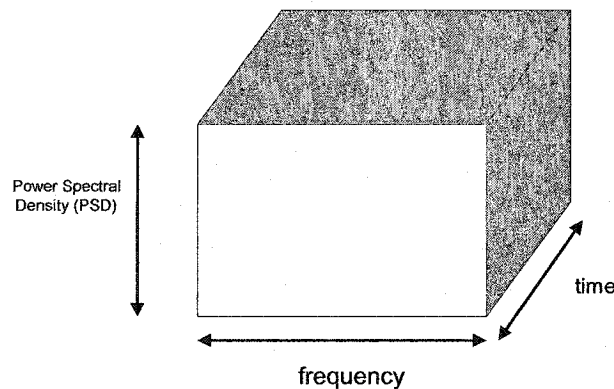


Figure 2-1: Sharing of Available Resources

2.2.1 Channelization

Channelization refers to the technique by which the available resource is divided into channels, each capable of supporting a packet. The resource can be viewed as a three-dimensional block, as shown in Figure 2-1. This block is made up of a time, frequency and power spectral

density (PSD) limitation, and so the division can be made in terms of any of these three axis: time, frequency, PSD (via multiple spreading codes), or hybrids of these (see Figure 2-2).

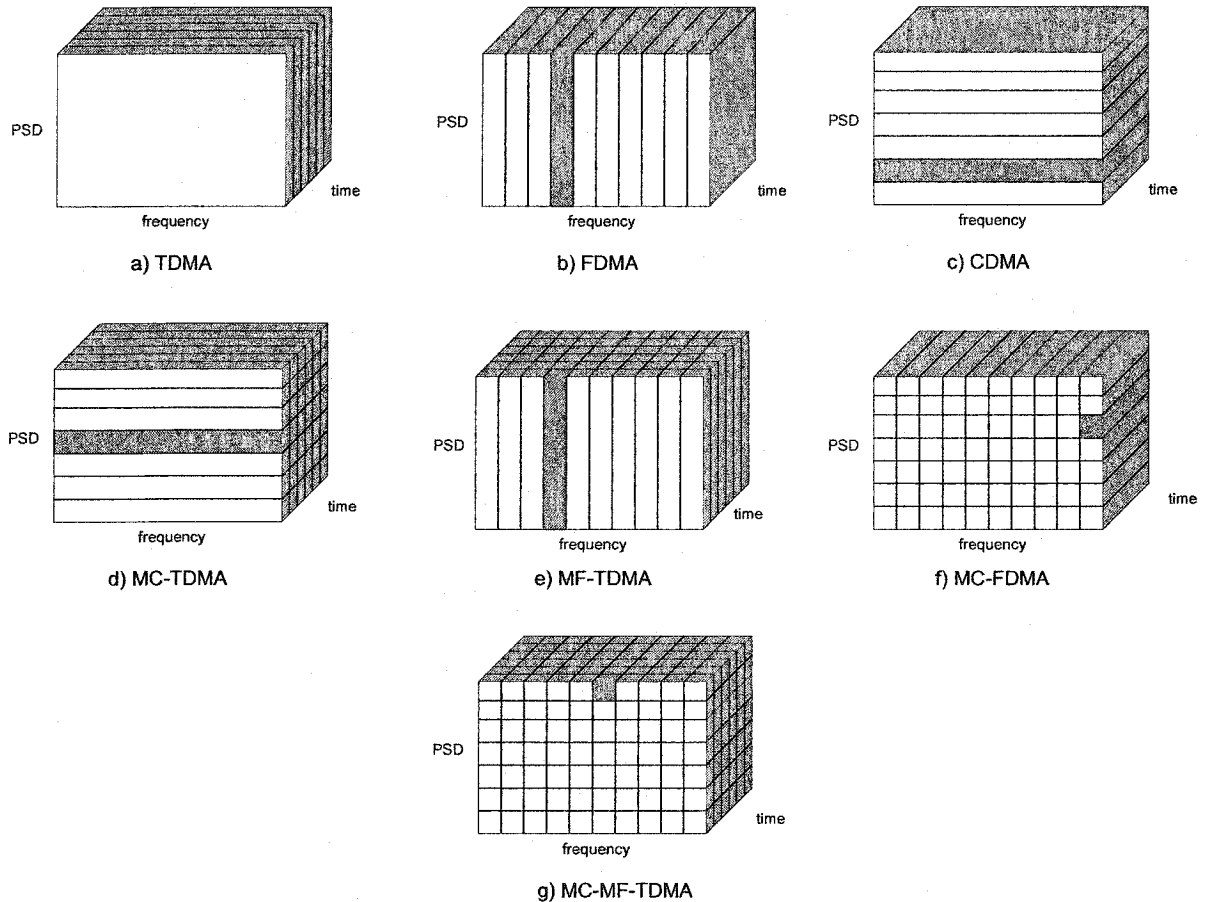


Figure 2-2: Channelization Techniques

Appendix A shows the evaluation of these methods in terms of radio capacity and information capacity. The highlights of this evaluation are repeated below:

Radio Capacity: Refers to the maximum number of constant bit rate users (R_b) that can be supported in a given bandwidth (B_w). Results show (See Figure A-3 in Appendix A) that the radio capacity of CDMA based techniques can exceed that of the time division/frequency division (TD/FD) based techniques, if the former employs very powerful low rate convolutional codes for forward error correction (FEC). Furthermore, results also show that

using multiple spot-beams, as opposed to a single beam, maximizes the radio capacity of all channelization techniques.

Information Capacity: This represents the maximum amount of information that can be transmitted every time a channel is used. The results for a typical scenario are shown in Figure A-5 (in Appendix A). Notice that the information capacity of CDMA is very close to that of TDMA and FDMA (within 1 bit/channel use), despite the fact that the channels for CDMA are not orthogonal.

The above results clearly suggest that CDMA based schemes should not be ignored as possible channelization techniques. In fact, under certain conditions (use of low rate convolution FEC code, use of multi-user detector over single user detectors, etc), the capacity of these techniques can exceed that of TD/FD based techniques.

2.2.2 Channel Access

Having divided the available resource into channels, the next issue to address is the method of assigning these channels to earth stations. The basic techniques can be classified into one of the five categories listed below:

- 1.fixed assigned,
- 2.random assigned (contention based),
- 3.demand assigned (reservation based),
- 4.free assigned,
- 5.hybrid combinations.

Within each of these categories, a large number of techniques have been proposed. A representative sampling of these techniques is given in this section. Figure 2-3 shows a classification tree based on the above categories, along with schemes falling under each of these classes. These are the schemes that will be discussed in this section.

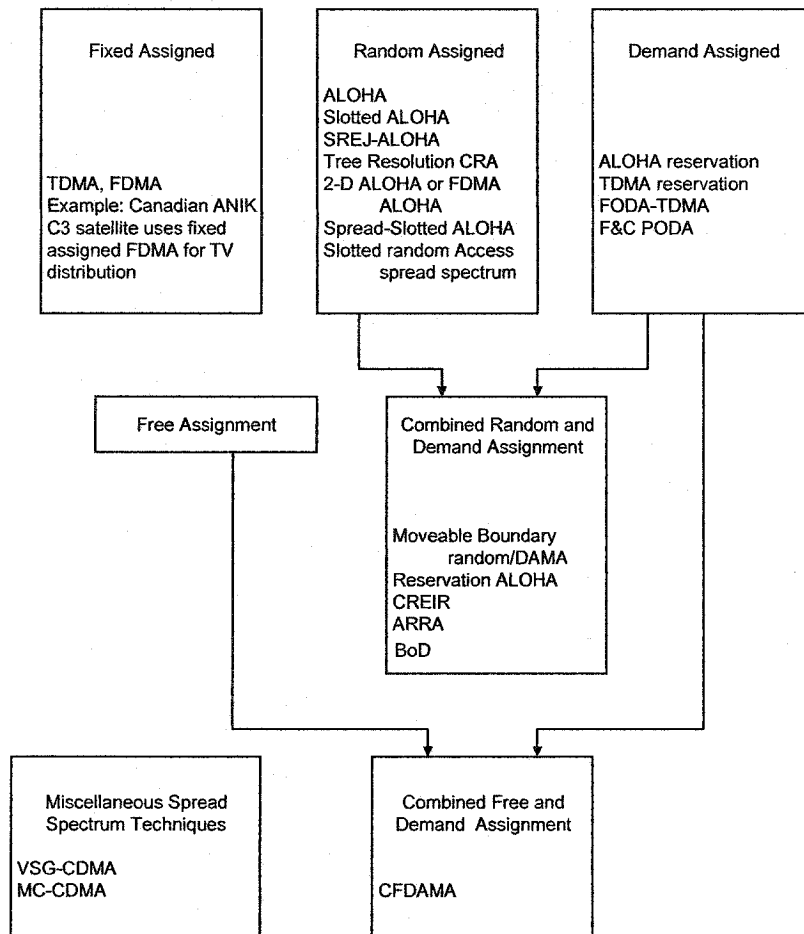


Figure 2-3: Classification of Channel Access Techniques

2.2.2.1 Fixed Assigned Schemes

In the fixed assigned schemes, the channels are assigned to earth stations on a permanent basis. Typically, these systems employ either TDMA or FDMA channelization. The most basic problem with fixed assignment, is the obvious lack of flexibility, and efficiency. If one channel is not being used, this capacity would be wasted, as no other earth station could transmit in its place. Furthermore, it does not lend itself readily to variable bit rate terminals. Consequently, fixed assignment, as a sole means of providing access, is not considered as a viable candidate for our environment.

2.2.2.2 Random Assignment Schemes

In contrast to the rigid structure imposed by fixed assignment, random access lies at the other extreme. In fact, terminals are allowed to transmit randomly in any of the available channels. As a basis, all random access schemes share the principles of the ALOHA protocol [33]. However, as a result of the slotted nature of the channelization technique, neither ALOHA nor Selective Reject SREJ-ALOHA [34], which are asynchronous, are appropriate for our application. The most basic synchronous random access technique is slotted ALOHA. The most common channelization methods used are either TDMA [35] or MF-TDMA - the latter method is referred to as FDMA/ALOHA [36]. Upon receiving a packet, a terminal immediately sends it in an empty channel. It then retains a copy of this packet in its terminal queue, while it waits for an acknowledgement (ACK). If another terminal has transmitted in the same channel, these packets collide, and neither terminal will receive an ACK. If a transmitting terminal concludes that a collision has occurred, it randomly reschedules a new transmission for the collided packet in another channel. This new transmission is again chosen randomly, to ensure that repeated collisions do not occur.

One of the major problems with this approach is that as the load per terminal (or the number of terminals) increases, the probability of collision also increases, resulting in instability. In fact, the maximum throughput of slotted ALOHA is limited to 36%. This maximum can be increased somewhat, by using a form of synchronous SREJ-Slotted ALOHA. In this case, a terminal continually sends packets, and only retransmits those that are not acknowledged. This has the advantage of reducing packet delay, as a terminal need not wait for an ACK before sending another packet. However, as only collided packets are retransmitted, these may eventually arrive at the receiving terminals out of sequence – not an acceptable situation for stream-type traffic that is delay sensitive.

Algorithms such as tree resolution CRA (Collision Resolution Algorithm) resolve the instability associated with the successive collisions of retransmitted packets [37]. The

retransmission is achieved by consecutive partitioning of the collided packets until only one is left. This packet is then transmitted without collision. The best collision resolution algorithms can provide a maximum capacity of 50%, but still lack the flexibility to meet the stringent delay requirements of real-time traffic.

Recently, a great deal of attention has been focused on techniques based on Spread Slotted ALOHA [38]. Essentially, this technique can be thought of as employing random access with an MC-TDMA channelization scheme. Each terminal is assigned a set of spreading codes from which to choose. These codes can be unique, or time-shifted versions of a single long code. Upon receiving a packet, a terminal immediately transmits this packet in the next available slot and by spreading on one of the assigned spreading codes. A collision occurs if:

1. Two or more terminals with traffic, randomly select the same spreading code for their packets. In this case, these packets experience a hard collision, and will require retransmission.
2. The number of terminals transmitting on the same slot results in a high level of multiple access interference. This condition is referred to as soft blocking.

These two types of collisions are shown schematically in Figure 2-4. Clearly, the hard collisions will result in an unacceptable performance for real-time traffic. Another important point to note is that for the system to outperform FDMA/ALOHA, it requires that the modulation, coding, etc, be

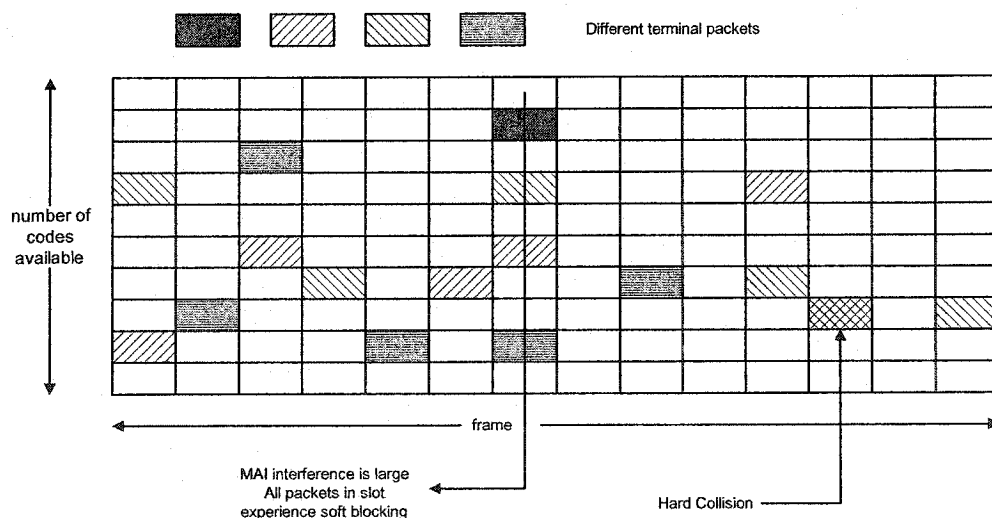


Figure 2-4: Collisions in Spread-Slotted ALOHA

such that the radio capacity of MC-TDMA exceed that of MF-TDMA.

An alternate approach would be to assign a single spreading code to each active terminal, and have the terminals randomly transmit their packets in slots [39]. This will definitely eliminate the first type of collision above, but may still lead to soft blocking. The maximum throughput for these types of systems can be shown to far exceed that of slotted ALOHA. However, as with slotted ALOHA, random access CDMA results in a performance that degrades rapidly when loaded beyond a certain threshold.

2.2.2.3 Reservation (Demand) Assignment Schemes

Demand assignment eliminates packet collisions, by having terminals make explicit requests (reservations) for capacity, and by having a scheduling algorithm assign capacity based on these requests. We say that a scheduler *honors* the requests.

The reservations can be made on three time scales (call, burst, or cell). At the call scale, channels are permanently assigned to terminals, for their exclusive use, for the duration of their call or connection. When the call ends, these channels are released and are free to be re-assigned by the scheduler. In contrast, at the cell scale, channels are assigned to earth stations based on their instantaneous demands. This is very dynamic and allows for very efficient use of capacity. In between these two scales lies the burst regime. Initially, terminals make explicit requests for capacity, and the scheduler implicitly holds these requests until they are released by the terminals. This burst scale has an advantage for stream-type traffic, since it eliminates the need for making requests every frame.

For each of these reservation scales, the demand assigned schemes can employ two forms of scheduling:

1. Distributed: All terminals run the same scheduling algorithm, and keep track of requests made by all other terminals. Although this technique is robust, and does not suffer from single points of failure, it is not appropriate for OBP GEO satellite applications. In fact, a distributed algorithm would only be appropriate for

environments where every terminal can be made aware of all requests made by every other terminal.

2. Centralized: The servicing of requests is made at a central location (either on-board the satellite in an OBP environment or in a dedicated ground station (gateway) in a bent-pipe environment). The latter requires two round-trip delays for the requests to be honored, as opposed to one, if the scheduler is on-board the satellite.

Lastly, as the demand assigned techniques require a separate reservation subchannel, there is also an access problem in this reservation subchannel. The three forms of reservation subchannel access are:

1. random: uses a form of slotted ALOHA. This scheme is especially useful for cases with a large number of terminals.
2. fixed: terminals have fixed reservation subchannels, and can make requests regularly. This is specifically appropriate for networks with a small number of terminals. As the terminal population increases, the reservation subchannel has to either be increased to satisfy these users, or the terminals have to wait a longer time before sending their reservations.
3. piggybacked: terminals append requests for more capacity to packets transmitted in already assigned channels. The only problem with this technique is the form of initial access. That is, how does a terminal piggyback its request if it has no channel in which to transmit?

Methods 1 and 2 are sometimes referred to as out-of-band requesting (OBR), while method 3 is sometimes called in-band requesting (IBR).

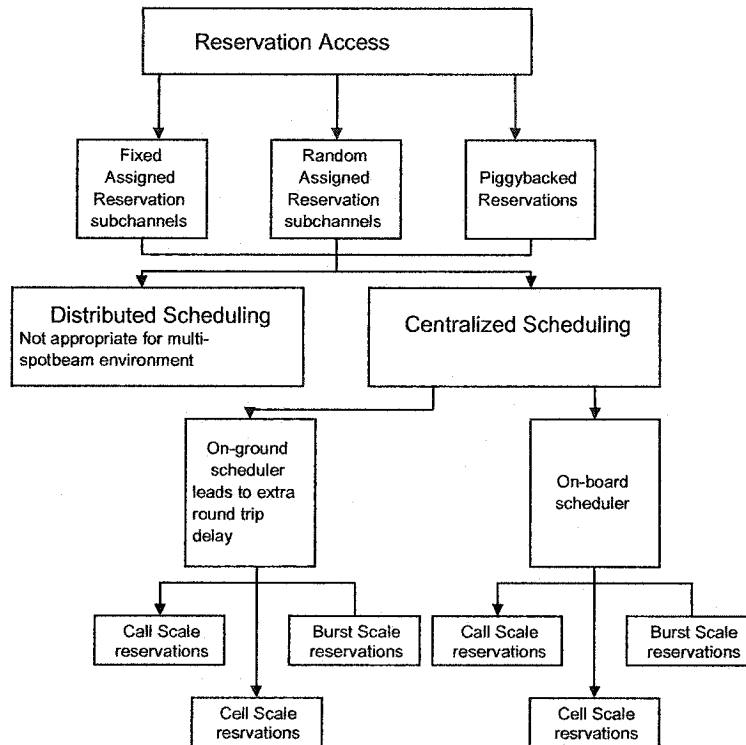
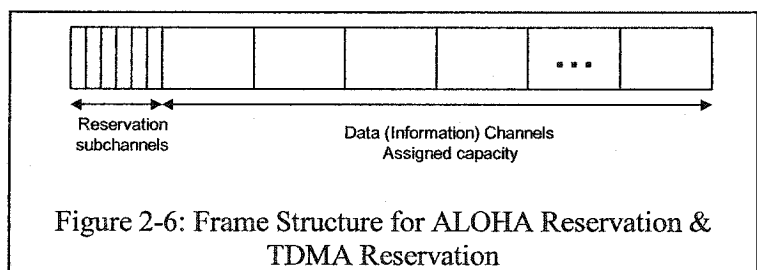


Figure 2-5: Classification of Reservation Access Schemes

The above classifications are highlighted in Figure 2-5.

ALOHA Reservation [40] and TDMA Reservation.

The frame structure for ALOHA Reservation is shown in Figure 2-6. Terminals request capacity by randomly sending their requests in one of the



reservation subchannels within a frame. The frame eventually arrives at the scheduler after the appropriate delay (D) (the value of D depends on the location of the scheduler). If this request is alone in the subchannel, it will be interpreted correctly by the scheduler, which will assign the

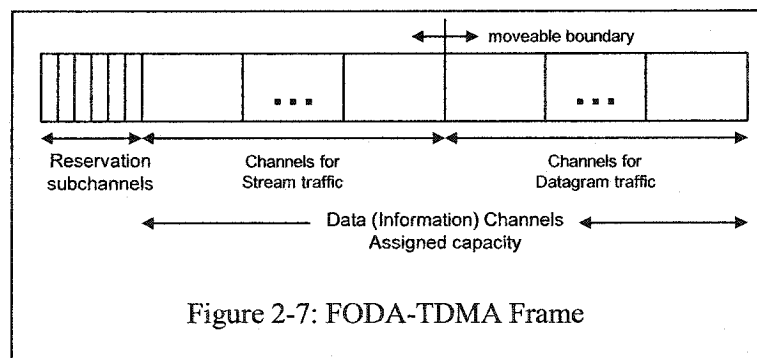
necessary channels based on the number requested. The assignment reaches the terminal after another propagation delay (D). Note that the scheduler assigns channels based on the number of requests it receives in the current frame and on the number of 'un'honored requests from previous frames. At low loads, it is possible that some of the information carrying channels remain empty. These empty channels will go unused, even though traffic may have arrived at the terminals. Furthermore, note that as in regular slotted ALOHA, instability may result as the number of requests increases.

TDMA reservation has the same frame structure as ALOHA reservation (Figure 2-6), with the exception that the reservation subchannels are permanently assigned to terminals. This version is ideally suited for cases with a small number of terminals, but suffers from increased delay as this number increases. Consider a case where there are N subchannels and M active terminals ($M \gg N$). For this simple case, a terminal, having just missed its request subchannel, would be forced to wait $\lceil M/N \rceil^6$ frames before making a request. During this time, the terminal may not transmit traffic, even though channels may be empty.

FIFO Ordered Demand Assignment TDMA (FODA-TDMA) [41]

Unlike the ALOHA reservation frame, the FODA-TDMA frame has three subsections - reservation subchannels, stream channels (for stream type traffic), and datagram channels (for jitter-tolerant traffic) (See Figure 2-7).

The reservation subchannel is shared by fixed assignment. Requests are made for both stream type traffic and jitter-tolerant traffic. The stream type reservations are made at the call



scale, while the datagram-type reservations (for jitter-tolerant traffic) are made at the cell scale

(i.e., every frame). If there is no stream-type traffic, then all the channels can be used to carry jitter-tolerant traffic (concept of moveable boundary). FODA-TDMA also has the provision for the piggybacking of requests within the transmitted channels.

Contention Priority Ordered Demand Assignment (C-PODA) [42] and Fixed Priority Ordered Demand Assignment (F-PODA) [42]

PODA was initially conceived as a general purpose MAC protocol for satellites, capable of handling both data and voice traffic, and with a priority mechanism built into the scheduler. The only major difference between C-PODA and F-PODA is in the requesting method used in the reservation subchannels - random for the former and fixed for the latter. The schemes have a great deal in common with FODA-TDMA. This includes the frame structure, the piggybacking of requests, and the scale of reservation mechanism (call for stream traffic and cell for datagram traffic). The schemes differ, however, in the manner requests are serviced at the scheduler. For the PODA techniques, this servicing can be done on a priority basis (i.e., requests can be prioritized). As a result, a newer request with a higher priority may be serviced by the scheduler, prior to an older request with a lower priority.

⁶ $\lceil x \rceil$ denotes the integer greater than or equal to x .

2.2.2.4 Combined Random/Reservation Techniques

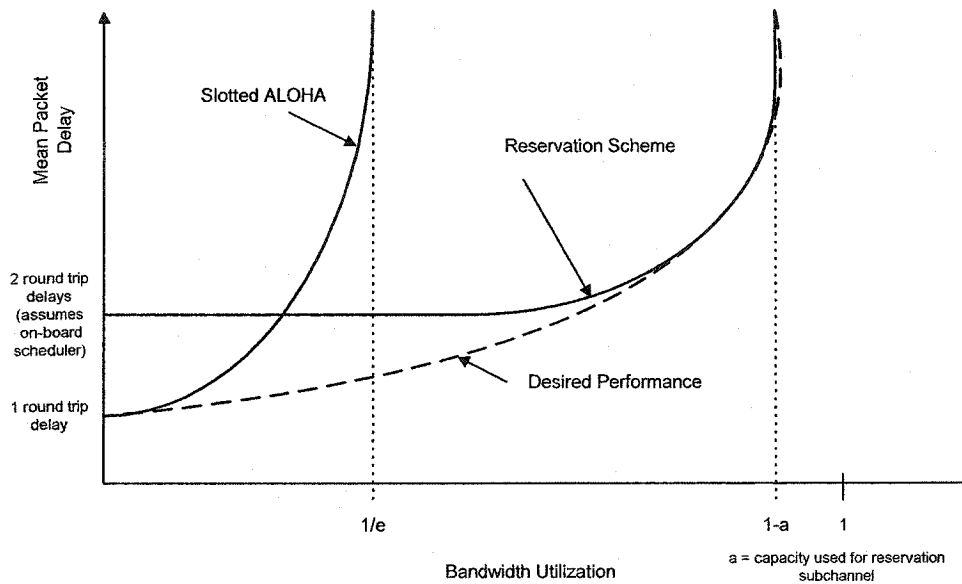


Figure 2-8: Desired Jitter-Tolerant Packet Delay Performance

The basic idea behind all hybrid schemes is to have a performance approaching that of random access at low loads, and approaching that of demand access at high loads (See Figure 2-8). To achieve this, one approach is to combine both random access and reservation access. A few of the more relevant techniques are described below.

Reservation ALOHA [43]

In Reservation ALOHA, terminals contend for channels in a frame on a random access basis. If a transmission is successful, the scheduler will implicitly hold (keep the reservation active) the assigned channels in subsequent frames. Performance is improved over slotted ALOHA, as the collision probability is reduced since fewer packets are transmitted in contention mode. Resolution of collided packets can be achieved by any of the techniques used for slotted ALOHA. Each channel in a frame can be in one of two modes – used (reserved) or unused (open for contention). Reserved channels are released by the owning terminal. This indicates to the scheduler that it no longer needs the capacity. Note that for the system to be stable, the frame size must be larger than the round trip propagation delay. If not, the terminals will not be able to know

which channels in the frame are reserved, and which are not. Performance results in the literature show that reservation ALOHA has a maximum throughput of about 70%. Interestingly, in terms of protocol dynamics, the scheme is similar to Packet Reservation Multiple Access (PRMA), proposed by Goodman [43], for terrestrial wireless communications.

Combined Random with Explicit then Implicit Reservation (CREIR) [44]

The CREIR scheme employs a frame structure with TDMA channelization. The frame is divided into two sections – one for reservations and one for user traffic. Reservations are made via slotted ALOHA in the reservation section. The data traffic channels can be in one of two states – reserved or unreserved. An unreserved channel is open to all stations via random access, whereas a reserved channel is the property of the terminal that issued the original request. To this basic premise, the concept of explicit-then-implicit reservation is added, whereby after the scheduler assigns a channel in honor of an initial request, it maintains this assignment from frame to frame. These terminals receive assigned channels without needing to make a request. A small field in the data channel is used to notify the scheduler if the implicit reservation is no longer needed. This pre-releasing ensures that no channels will go unused, while the scheduler attempts to determine whether a terminal still needs its implicit reservation. Although CREIR was originally studied for jitter-tolerant traffic, it can be extended rather easily to real-time traffic. To do this, the terminal must make the initial reservation at peak rate – otherwise jitter would be introduced in the stream arriving at the destination. Unfortunately, reserving at peak leads to inefficiency for variable bit rate traffic.

Moveable Boundary Random/DAMA [45]

This scheme is specifically designed to support various traffic types, with different QoS requirements, for a multibeam satellite environment. The TDMA frame is divided into three subframes, as shown in Figure 2-9 – a reservation subframe, an interactive data subframe, and a voice, video, high priority data subframe. Low priority data users access the channels in the interactive data subframe via slotted ALOHA. The voice, video and high priority data users,

access channels by demand assignment. Reservations are sent in reservation mini-slots of the reservation subframe. The access to these mini-slots is by slotted ALOHA. If the request is successful, it will be honored by the scheduler. This assignment information is then broadcast to the terminals. If, on the other hand, the earth station requests collide, then the scheduler notifies the earth stations of this occurrence. A typical collision resolution algorithm is used to handle the retransmissions. The scheduler also broadcasts information regarding these channels in the voice, video, high priority data subframe that are not assigned. These join a common pool, and can be used by the low priority data traffic (concept of moveable boundary). Performance results in [45] show that this MAC protocol significantly improves the throughput for the low priority data traffic.

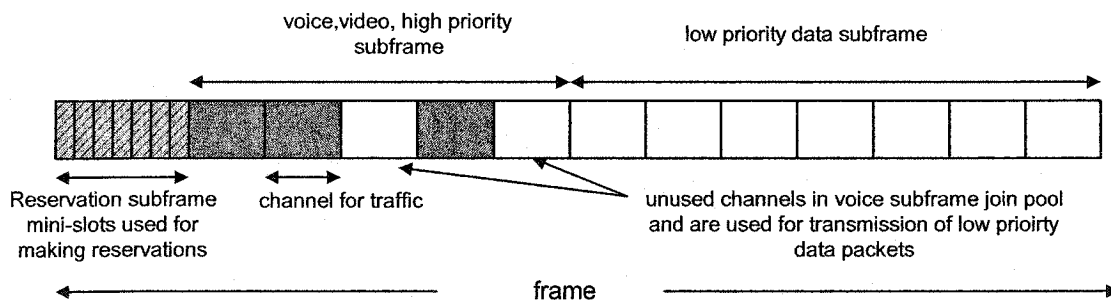


Figure 2-9: Frame Structure for Moveable Boundary Random/DAMA

Random Reservation Adaptive Assignment (RRAA) [46]

A TDMA frame is divided into four subframes – reservation, control, data, and random access. The reservation subframe is used by earth stations to send their initial requests to the scheduler (the technique proposed is slotted Aloha). Once an earth station's initial request has been accepted, it is polled by the scheduler, and sends future requests in the control subframe. The scheduler assigns capacity from the data subframe (these are the reserved slots). In addition, the random access subframe is available for jitter-tolerant traffic to use in a slotted Aloha mode. The boundary between the data and random access subframes is moveable. One of the main innovations of this proposed MAC algorithm was the prioritization of requests at the scheduler,

based on the traffic issuing the request. In addition, the work clearly showed how this priority can be used to meet QoS requirements. The technique was tweaked somewhat by Iera, Molinaro, and Marano [47], so that it was better at handling real-time variable bit rate traffic.

Bandwidth on Demand (BoD) Process [48]

BoD is an elaborate protocol, which is integrally tied to the Call Admission Control (CAC) process. The underlying frame is very similar to other combined random/demand schemes, providing different MAC strategies – random, call scale reservation & cell scale reservation. However, what is unique about BoD is that it did not map a connection to a single MAC strategy. For instance, consider a video connection that requires a guaranteed minimum transfer rate in order to produce an acceptable image but whose performance can be improved if the transfer rate is increased. It is possible then to provide the minimum transfer rate with a call scale demand assignment, and to provide the *extra* capacity using random access. The end user will always perceive an acceptable image quality thanks to the demand assignment, but when the network is lightly loaded, the extra traffic would also be transmitted through random access, and the image quality would be significantly improved.

2.2.2.5 Combined Free/Demand Assignment Techniques

The second hybrid combination strives to achieve the same performance as the combined random/reservation techniques (Figure 2-8), but it achieves this by eliminating the contention based transmission of packets (contention may still be used, but only for the reservation subchannels). This implies that the low priority data traffic does not suffer from collisions, and does not undergo retransmissions. The basic operation of combined free/demand schemes, is shown in Figure 2-10. The scheme is devised to take advantage of the long round-trip propagation delay inherent in a GEO environment. At time t_o , earth stations with traffic in their terminal queues, make requests for channels. These requests arrive at the scheduler, after a propagation

delay (K seconds), and are inserted into the scheduler queue, along with reservations that have not been honored from all previous frames. The value of K depends on the location of the scheduler. K equals 1 round trip time (RTT) for an on-ground scheduler (bent-pipe system) and $\frac{1}{2}$ RTT for an on-board scheduler (OBP system). The scheduler then services these requests based on some algorithm (typically first come first service (FCFS)). If after honoring the maximum number of requests, some channels remain unassigned, these are freely given to earth stations. The manner of distributing these free channels distinguishes the different versions of this technique. These free channels arrive at the user stations at time $t_0 + 2K$. If a terminal has received traffic in the time $t_0 < t < t_0 + 2K$, then these packets can use these free assignments. Essentially, this traffic is transmitted without having made a request (in a manner similar to random access). However, unlike random access, there is no possibility of collision.

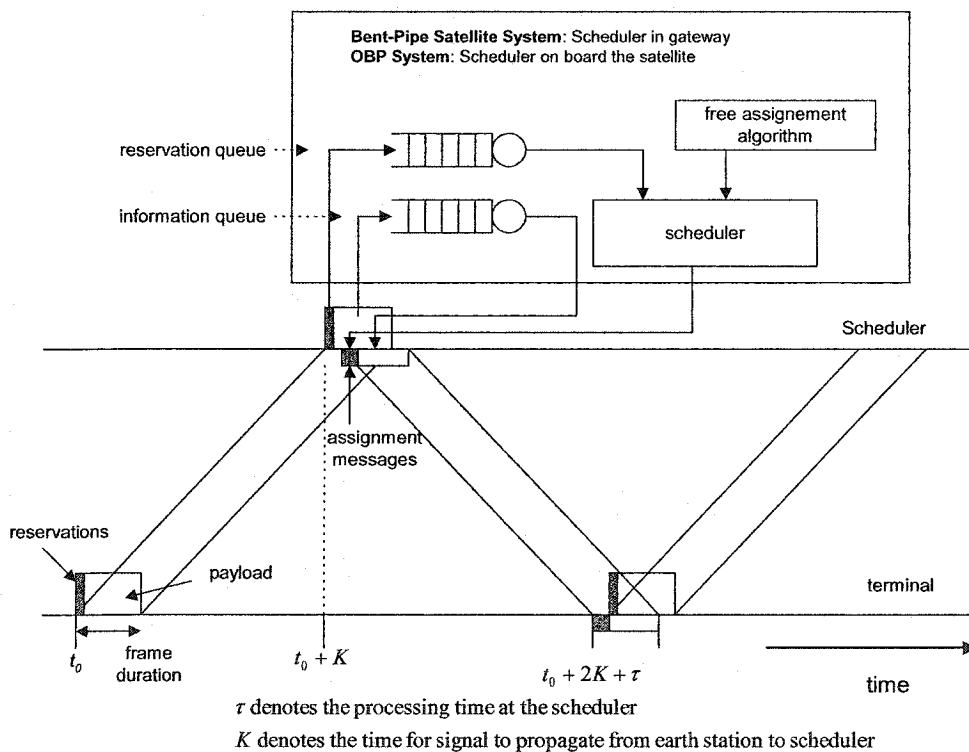


Figure 2-10: Basic Operation of Combined Free/Demand Assignment Schemes

The basic technique was studied in [49][50]. There is some flexibility associated with the form of requesting. This can be either fixed access, random access, piggybacking, or some combination of these. Results show that the optimum method depends on the number of terminals – a large number of terminals is best handled by random access, whereas a small number of terminals is best handled by fixed access of reservation subchannels.

In the original scheme [49], the frame structure was rather loosely defined, with reservation mini-slots appended to each data slot. In the modified version presented in [50], a strict frame was imposed in order to accommodate voice traffic. The voice traffic employed an explicit-then-implicit reservation strategy, and the reserved channels were maintained from frame-to-frame. It was assumed that the traffic was of constant bit rate, having a constant requirement for the entire duration of the call.

Recently, researchers have been working on optimizing CFDMA so that it is better suited for a multimedia network. Although the concept of free assignment has not changed much from the original protocol, they did expand the demand assignment to include all three scales – call scale, cell scale, and burst scale. They also allowed individual connections to use any combination of the three scales (as was done by BoD). This leads to a great deal of flexibility, and integrates well with the CAC process. Results are shown in [51].

Research is also ongoing as to the best method to distribute the free assignment channels. Jiang, Li, and Leung [52] have proposed a technique that distributes the free assignment based on the trend in the arrival traffic to the earth stations. Their technique was tested with traffic that was very bursty and results show that there is in fact less waste in the free assignment.

2.2.2.6 Miscellaneous Spread Spectrum Approaches

With the acceptance of the IS-95 standard, there has been a number of spread spectrum based techniques that have been proposed, which do not fall into any one category discussed thus far. These techniques rely heavily on the features of CDMA. Two are described briefly below:

Variable Spreading Gain CDMA (VSG-CDMA) [53]

In this approach, sources with variable rates are handled by variable spreading codes. Roughly, the scheme fits under the umbrella of a MF-CDMA channelization scheme, with the transmitted chip rates of all terminals being equal. The value of this chip rate is dependent on the bandwidth available for spreading. Earth stations make reservations at the call scale, and these reservations are honored by assigning terminals to a particular frequency band within the MF-CDMA frame. Assuming a bandwidth B_w , and three services having bit rates R_{bi} ($i = 1,2,3$), the processing gain for each service is approximately given by $P_{Gi} = B_w/R_{bi}$. In making the assignments of stations to frequency band, the scheduler takes into account the amount of MAI generated by the requesting service, and assigns the call to the band that can best support this additional interference. The three major drawbacks of this technique are listed below:

1. After despreading, the user bit rates for each service will be different. Consequently, a data formatter will be necessary to equalize the bit rates prior to forwarding the traffic.
2. The management of spreading codes is very difficult.
3. With all other things equal, the services with the highest bit rates will have the lowest processing gain. This may place these services at a disadvantage in combating adjacent beam interference and the like. Note that in existing CDMA based networks, the lower processing gain is often countered by having these services use a higher transmitted power [12].

Multi-Code CDMA (MC-CDMA) [54]

This technique is also designed for variable bit rate services, but achieves this flexibility by having terminals transmit on multiple spreading codes. This will ensure that all services have the same processing gain. Each terminal is given a basic rate. When an earth station needs to transmit at M times the basic rate, it divides its stream into M parallel streams, each of which will be spread by a spreading code, and subsequently added, prior to transmission. This is shown schematically in Figure 2-11.

An active earth station is assigned a primary spreading code c . To generate the additional spreading codes, MC-CDMA uses subcode concatenation. This is a process by which the additional codes are derived from c by multiplying with a set of deterministic and orthogonal codes $\{d_j\}$. As a result, the codes used by a single user are perfectly orthogonal to each other. That is,

$$c_i = c \bullet d_i$$

$$c_i \bullet c_j = d_i \bullet d_j = 0 \text{ for } i \neq j$$

This results in two simplifications:

1. The transmissions from a single user on multiple spreading codes are orthogonal and therefore do not contribute to the MAI of each other. The orthogonality is maintained at the receiver, since all parallel signals are transmitted simultaneously, and follow identical paths.
2. The synchronization of the spreading codes is based only on synchronizing of the primary code. Therefore, the use of multiple codes will not increase the on-board synchronization circuitry, which is major concern for spread-spectrum systems.

It should be noted that this technique assumes that the system is not affected by multipath interference.

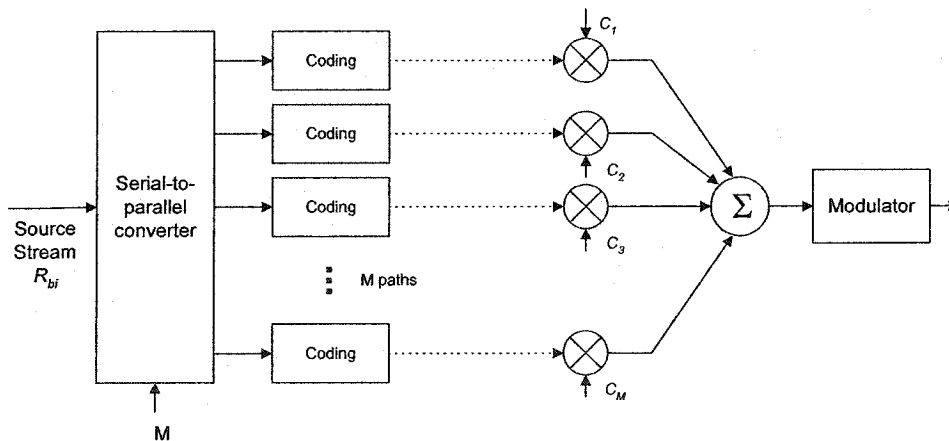


Figure 2-11: MC-CDMA Transmitter (based on [54])

2.2.2.7 General Conclusions

Based on the discussions in this section, the following conclusions can be made:

1. The capacity (radio and information) of CDMA based schemes is comparable to that of TDMA and FDMA. Furthermore, this capacity can be improved by using powerful FEC codes and/or multi-user detectors at the receiver.
2. To accommodate variable bit rate sources, an earth station requires a varying number of channels from frame to frame. In terms of channelization, this is best achieved by means of a TDMA hybrid. An earth station would vary its transmission rate by using more or less time-slots within a frame. The maximum number of time-slots would dictate the maximum station burst rate. FDMA is not considered, as it would require that a terminal have the capability to transmit simultaneously on two or more frequency carriers. Owing to the nonlinear power amplifiers used at the earth station transmitters, the multiple carriers would result in intermodulation (IM) noise and reduce the overall signal-to-noise ratio (SNR). Simple TDMA is also not considered viable, as it would require the largest transmitted power of any of the channelization techniques (transmitted power is proportional to the maximum burst rate, which is highest for the case of pure TDMA). Of the remaining schemes (CDMA, MC-TDMA, MF-TDMA, and MF-MC-TDMA), MC-TDMA and MF-TDMA are the most appropriate. Although pure CDMA can be used for variable bit rate (VBR) sources, as with VSG-CDMA and MC-CDMA, these techniques suffer from unequal protection against interference and limited variability, respectively. Lastly, the double hybrid MF-MC-TDMA approach provides no significant gain over MC-TDMA – in fact, it will have the net effect of reducing both the available bandwidth for spreading (and hence the processing gain) and the maximum earth station burst rate.
3. In terms of channel access, the most suitable approach for a GEO satellite network seems to be a hybrid combination. In order to eliminate collisions altogether, we have focused on the combined free/demand assignment scheme. As shown in [51], the scheme can be tailored to MF-TDMA, and can handle traffic with varied requirements. However, owing to its simplicity, we have also proposed a MC-TDMA scheme with random assignment. The scheme results in a very simple scheduler and can be tailored to multimedia traffic.

2.3 Standard Development

As mentioned in Chapter 1, two approaches have been standardized in order to provide broadband satellite service – one leverages the work done by DAVIC on digital video broadcasting (DVB) while the other is an extension of the popular DOCSIS standard used in cable modems. Although the MAC protocol is a part of both standards, it should be understood that the standards encompass the physical layer as well. As a result, much of these standards are concerned with techniques for coding, modulation, and synchronization. For this section, we will focus only the MAC portion of the standards.

2.3.1 DVB-RCS [14]

The DVB-RCS standard assumes that the underlying frame structure is MF-TDMA based. One small addition over the simple channelization described in the previous section, is that the MF-TDMA frame can be dynamic – that is, the time slots need not be uniform. This allows having earth stations with different maximum burst rates. The standard uses fixed length packets that are either ATM based (53 bytes) or MPEG2-TS based (188 bytes). The channels in the MF-TDMA frame are assigned by a scheduler (located on-ground) based on a combined demand/free assigned scheme. Assigned channels can be any one of five categories:

Continuous Rate assignment (CRA): These are channels that are assigned according to a call-scale reservation scheme. At the start of a connection, the demand for capacity is made, and the scheduler assigns this capacity every frame for the duration of connection.

Rate-Based Dynamic Capacity (RBDC): These are channels that are assigned according to a burst-scale reservation mechanism. As the name implies, earth stations request a rate (in terms of a number channels per frame) as opposed to an absolute number of channels. The scheduler will assign these channels every frame until the earth station alters its RBDC request, or until the request times-out. The time-out mechanism is needed to ensure that non-responding earth stations are not assigned capacity they cannot use. RBDC can also be configured with a maximum maxRBDC and this maximum can be used for certain QoS implementations.

Volume Based Dynamic Capacity (VBDC): These are channels assigned based on a cell-scale reservation scheme. The requests from the earth stations are cumulative and are stored in the scheduler. An effort should be made by the scheduler to guarantee that the channels assigned as a result of VBDC are fairly distributed to earth stations.

Absolute Volume Based Dynamic Capacity (AVBDC): These channels are assigned in a manner similar to VBDC. The only difference is that an AVBDC request from an earth station will overwrite the previous VBDC requests stored in the scheduler. The technique can be used to flush the VBDC requests from time to time.

Free Capacity Assignment (FCA): These are channels that are assigned to earth stations freely (those channels left unassigned after the scheduler has assigned all capacity for CRA, RBDC, VBDC).

The form of requesting is a combination of fixed or random in request mini-slots, and piggybacking on regular packet transfers. The scheduler broadcasts the assignment information in a burst time plan (BTP). The earth stations receive this information and determine the time slots that have been assigned to them.

What the standard does not specify is the specific implementation of the scheduler, and how the traffic is mapped to the 5 capacity types – these are implementation details left to the provider. Many of the results in the literature propose typical mappings that can be used. For instance, [51] suggests how to map the ATM service categories into the 5 request mechanisms.

DVB-RCS systems are currently deployed all over the globe. In mid-03, there were some 8500 DVB-RCS terminals in operation. Adoption by SES Global (the largest global satellite operator) has definitely helped in the acceptance of the standard [55].

2.3.2 DOCSIS [15]

The DOCSIS standards were developed by the cable industry, as a means to provide multimedia access. Since the coaxial cable distribution network is a shared medium, there has been some speculation that the standard could easily cross over to the satellite industry – with some minor modifications. At the time of writing, the exact details of the DOCSIS over satellite standard have still not been finalized. However, WildBlue communications recently announced that it will deliver its broadband **access** service using a DOCSIS based satellite modem developed (or to be

developed) by Viasat [56]. As this will be the first major deployment of a DOCSIS based system, the jury is still out as to the viability of the approach.

The MAC protocol for the DOCSIS over satellite standard is assumed to be very close to the current DOCSIS 2.0 version. The system uses a variable frame that is controlled by the scheduler (referred to as the Cable Modem Termination System (CMTS) in the standard). The upstream frame (from user to scheduler) is divided into a number of mini-slots, each of which is 8 bytes wide. A mini-slot can be in one of three main modes: request, request/data, or data. A mini-slot in request mode can be used by stations to make requests. These can be either open for contention, or fixed to a particular station. A request/data mini-slot can be used either for sending requests or small packets of data in contention mode. Typically these mini-slots will only be used for lightly loaded networks. The final mini-slot mode is used for transporting data. These will be assigned by the scheduler based on requests received in previous upstreams. The data mini-slots can also carry piggybacked requests.

To implement QoS, the DOCSIS community has defined service flows, and has required that the scheduler implement priority in its assignment of mini-slots. Six service flows have been defined:

Unsolicited Grant Service (UGS): Used for service flows that require a guaranteed bandwidth in the upstream. The request is only sent once at the start of the connection, and is very similar to demand assignment at the call scale.

Unsolicited Grant Service with Activity Detection (UGS-AD): Very similar to UGS, with the added feature that the scheduler monitors the upstream and if the connection is inactive, rather than assigning data mini-slots, it will only assign request mini-slots.

Real-Time Polling Service (rtPS): This service flow is suited for real-time VBR traffic. It is very similar to demand assignment with burst scale requesting. Periodically, the station will be given request mini-slots.

Non-Real-Time Polling Service (nrtPS): Service similar to rtPS, except that the number of request mini-slots assigned is reduced during periods when the network is heavily loaded. Used for service flows that require a guaranteed bandwidth in the upstream.

Best Effort (BE): This service model is used for best-effort traffic. Stations will normally send traffic either through request/data mini-slots or through regular requests sent over request mini-slots.

Committed Information Rate (CIR): Service to be defined by vendors.

2.3.3 Standards Summary

In addition to DOCSIS and DVB-RCS, a number of satellite operators are proposing their own proprietary standards. For this thesis, much of the work has focused on the DVB-RCS standard, as it has been field tested, and currently has the largest subscriber base. We do however, propose our own proprietary standard for both bent-pipe and OBP satellite systems based on CDMA.

2.4 Techniques for Performance Evaluation

2.4.1 Event Driven Simulation [57]

As mentioned in the introduction, most of the simulation results for this thesis were implemented with the OPNET simulation package. This tool is currently the de-facto standard for network simulations. Network modeling is based on hierarchical layers (very similar to the OSI layers). The main advantages of the tool are:

1. The network is developed hierarchically – and so the tool inherently leads to models that are re-useable.
2. Network entry is very simple through a graphical user interface.
3. There is a large user community that has contributed models to OPNET, and these have been incorporated in the vast standard library. Some of these models have been contributed by the designers of the actual hardware (Cisco, Nortel,...). These models are configurable, and implement algorithms from ATM QoS, routing, DNS,...
4. The tool can be used to simulate at many different levels (bit level to packet level). The choice of level is left to the designer.

The network is built at three main levels: network, node, process. At the first level, the network is specified. The network provides information as to how the individual nodes within the network are connected. At the second level, the nodes are specified. A node can be either a processing device and/or a queuing device. All algorithms (media access in our case) are implemented in one

or more of these nodes. The last level (the process level) details the actual C code that implements the algorithms within the nodes. The code is usually entered as a state machine.

2.4.2 Analytical Performance Evaluation

As will be shown in the next chapter, the major metric in performance analysis of any multiple access problem, is the mean queuing delay experienced by the jitter-tolerant traffic. The real-time loss probability is more mathematically tractable, as will be shown. In most cases, after embedding a Markov chain at frame boundaries, the system reduces to finding the solution of an MMPP/G/c queue (Markov Modulated Poisson Process (MMPP) arrivals to a c-server queue with a general service distribution). Numerous queuing analysis techniques have been proposed in the literature to deal with problems of this type – three of these are briefly described in this section:

Fluid Flow Approach: This approach, pioneered by Anick, Mitra, and Sondhi, in their paper [58], liken packets to a fluid, and derives a time-varying fluid equation for the fluid level in the queue. This technique is not applicable to our queuing model for two main reasons. First, it assumes that the service rate of the queue is a constant, which is not the case in our model. Secondly, since it treats packets as a fluid, it relies on the fact that the service rate and capacity are high so that small discrete changes in queue level are negligible. For our case, not only is the capacity rather small, but more importantly, the service rate can go down to 0.

Probability Generating Function (PGF) Approach: This is the most classical approach to solving embedded queuing problems (see for example [59]). For problems of our type, the matrix equation for the embedded Markov chain would be of the form:

$$\mathbf{y} = \mathbf{y} \mathbf{P}_{MG} \quad ^7 \tag{2-1}$$

where

⁷ Note that all matrices will be shown as boldface variables.

$$\begin{aligned}
\mathbf{y} &= [y_0 \ y_1 \ y_2 \ y_3 \ \dots] \\
\mathbf{y}_i &= [y_{i0} \ y_{i1} \ y_{i2} \ y_{i3} \ \dots \ y_{iL}] \\
y_{ij} &= \text{Pr}[\text{Number in queue} = i, \text{underlying phase process in state } j] \\
L &= \text{number of states of the underlying phase process.}
\end{aligned}$$

The phase process depends on both the arrival process and the service process.

P_{MG} = transition matrix of form :

$$= \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & \dots \\ A_0 & A_1 & A_2 & A_3 & \dots \\ 0 & A_0 & A_1 & A_2 & \dots \\ 0 & 0 & A_0 & A_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

The matrices B_i and A_i represent a grouping of the transition probability vectors into $L \times L$ submatrices. Equation (2.1) can be converted to the z-domain in terms of probability generating functions. After some manipulation, the following result can be obtained:

$$Q(z)[Iz - A(z)] = y_0[zB(z) - A(z)] \quad (2-2)$$

where $Q(z)$, $A(z)$, and $B(z)$ are the matrix probability generating functions of y_i , A_i , and B_i , respectively, and I is the identity matrix. From equation (2-2), the only unknown is y_0 , which represents a vector of initial queue occupancy distributions. These unknown constants are normally determined by noting that $Q(z)$ is a vector of marginal probability generating functions, and is therefore bounded within the unit disk ($|z| < 1$). Once $Q(z)$ is known, the queue occupancy can be found using transform inversion. Typically this inversion process is not a simple procedure.

Matrix Geometric Approach ([60][61]): This technique also has as a starting point the matrix equation (2-1), but rather than resorting to the transform domain, all results are found directly in the probability domain. As with the PGF approach, the only unknown is y_0 . To determine this vector, an algorithmic procedure has been found (detailed in [61]). The

procedure is amenable to software implementation, and since all the intermediate matrices have a probabilistic interpretation, the algorithm is said to be self-checking.

Although both the PFG and Matrix Geometric approaches seem appropriate for our analysis, the latter technique was selected for determining the performance results presented in the next chapter. It was much easier to implement in software, and did not require a complicated transform inversion step.

Chapter 3

Bent-Pipe System

3.1 Introduction

The model of the multibeam BP satellite system considered is shown on Figure 3-1. In order to employ frequency reuse, the coverage area of the satellite is divided into a number of spot beams. Each earth station (or user), generates multimedia traffic made up of variable bit rate real-time voice and video, and long-range dependent data. This information is packetized into fixed sized cells. To be consistent with the DVB-RCS standard, we will assume ATM type cells (48 byte payload and 5 byte header). The information is sent through the satellite to the scheduler, located in the Gateway/Hub. The scheduler performs the assignment process. Connection to the terrestrial network is provided through the Gateway.

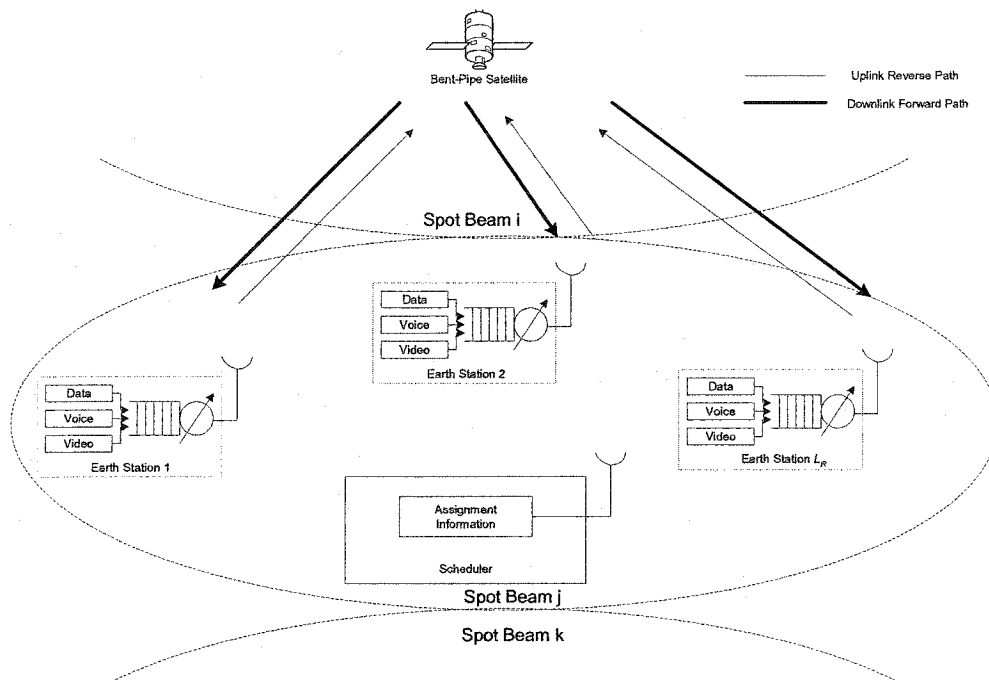


Figure 3-1: Model of Bent-Pipe System

The remainder of this chapter is divided as follows. In the next section, the network scenario and traffic mixes used in the performance evaluations are presented. A discussion of the Markov Modulated Poisson Process (MMPP) and Pareto Modulated Poisson Process (PMPP) is included to show the difference between them in terms of index of dispersion for counts (IDC). In Section 3.3, the basic operation of our proposed CFDMA with MF-TDMA implementation is explained, and the terminal queuing analysis is given. A sampling of the available simulation results is also included. The performance measures include the jitter-tolerant packet delay and the real-time loss probability, both arising due to a lack of uplink capacity. Note that we assumed an infinite buffer size at the earth stations. This assumption is carried over to the MC-TDMA case, which is considered in section 3.4. For MC-TDMA, only simulation results are presented as queuing analysis of CDMA systems are complicated by the need to consider physical layer impairments. In CDMA systems, a phenomenon known as soft blocking can occur, whereby a packet may experience higher than expected interference (caused by the multiple simultaneous user earth station transmissions). Analysis of soft blocking would involve consideration of physical layer impairments such as power control and synchronization.

3.2 Traffic Mixes and Simulations Considered

The basic parameters for the system were established based on a review of current and past proposed satellite architectures from the industrial community (see [2] [3] [75] [90]). The parameters related to the multiple access protocol are listed in Table 3-1.

Table 3-1: Satellite System Parameters

Parameter	Value
Uplink (and downlink) frame duration	24 msec.
Round trip delay	270 msec.
Packet or cell size	48 bytes (+ 5 byte header)
Capacity per uplink beam	4096 packets/frame = 65.536 Mbps
Maximum capacity per earth station (station burst rate)	128 packets/frame = 2.048 Mbps
Total Available bandwidth	960 MHz

Furthermore, for all results, it was assumed that the load per earth station was fixed, and that the total system load was varied by increasing/decreasing the number of active earth stations. Since one of the major objectives of the study was to investigate the effect of traffic mix on overall performance, four different scenarios were considered: voice dominant, video dominant, equally dominant, and data dominant. The actual percentage of each type of traffic, for each scenario, is detailed in Table 3-2. For each mix, the voice was modeled as an on-off source [62], while the video was modeled as a collection of independent mini on-off sources (each video source was taken to be 6 such on-off sources) [63]. The parameters for the individual voice and video sources are shown in Table 3-3. Furthermore, based on recent results from Ethernet LAN traffic studies [64][65], it has been argued that data traffic is bimodal and possesses self-similar characteristics. We therefore chose to model the data traffic from each earth station as a two state PMPP source, with a fixed Hurst parameter.

Table 3-2: Traffic Scenarios Considered

	Scenario 1 Voice Dominant	Scenario 2 Video Dominant	Scenario 3 Equally Dominant	Scenario 4 Data Dominant
% Voice	70	10	33.3	20
% Video	10	70	33.3	10
% Data	20	20	33.3	70

Table 3-3: Parameters for Individual on-off Sources

	Peak-to-Average Ratio	Peak Rate [Kbps]	Average Rate [kbps]	Average Rate [Packets/frame]	Average on Time [msec]
Voice	2.5	64	25.6	1.6	135
Video	5	384	76.8	4.8	N/A
Video (mini on-off sources)	5	64	12.8	0.8	265

The original parameters considered, specified that the maximum number of earth stations per spot beam be set to 8192. This corresponded to an average cell generation rate (AVE_{CGR}) per earth station of

$$\begin{aligned}
 AVE_{CGR} &= \frac{\text{total capacity}}{\text{total number of earth stations}} \\
 &= \frac{4096 \text{ packets/frame}}{8192 \text{ earth stations}} \\
 &= 0.5 \frac{\text{packets/frame}}{\text{earth station}} \\
 &= 8 \text{ Kbps/ES}
 \end{aligned}$$

This implies that over a long enough time, an earth station would transmit 1 packet every 2 frames. However, not all earth stations are active at any one time. Those not active, must first make a call admission request. For the first round of simulations, we tried to model the call admission problem, by assuming all 8192 earth stations as active, and by fixing the load of each to 0.5 packets/frame. As a direct result, each earth station was made up of a fractional number of individual video and voice sources. For example, for a 50% load, voice dominant scenario (70% voice/10% video/20% data), the number of active earth stations is $(0.5)(8192)=4096$, each of which generates, on average, 0.5 cells/frame. Based on the average cell generation rate for an individual voice and video source (Table 3-3), it is a simple matter to calculate the number of voice and video sources that correspond to this load. That is,

$$\begin{aligned}
N_{voice} &= \text{total number of voice sources} \\
&= \left\lceil \frac{(\% \text{voice})(\text{load})(\text{capacity})}{AVE_{CGR} \text{ of a voice source}} \right\rceil \\
&= \left\lceil \frac{(0.7)(0.5)(4096 \text{ cells/frame})}{1.6 \text{ cells/frame}} \right\rceil \\
&= 896
\end{aligned}$$

$$\begin{aligned}
N_{video} &= \text{total number of video sources} \\
&= \left\lceil \frac{(\% \text{video})(\text{load})(\text{capacity})}{AVE_{CGR} \text{ of a voice source}} \right\rceil \\
&= \left\lceil \frac{(0.1)(0.5)(4096 \text{ cells/frame})}{4.8 \text{ cells/frame}} \right\rceil \\
&= 42
\end{aligned}$$

where $\lceil x \rceil$ denotes the integer part of x , and N_{voice} , N_{video} denote the number of multiplexed voice and video sources for each beam. If each of the 4096 individual earth stations is homogeneous, this then implies that each is made up of 0.2188 voice sources, and 0.01025 video sources. Unfortunately, it is very difficult to gauge performance results with such fractional numbers of sources. To compensate, we mapped the 896 voice on-off sources to a 2 state MMPP, and the $(42)(6)=252$ video mini on-off sources to a 2 state MMPP, and then scaled these MMPP sources by the number of active users (4096). The simulation results obtained for MF-TDMA and MC-TDMA using this type of mapping procedure were presented in [66][67].

Upon close examination of the results presented in the two reports, we found that this type of modeling was not very appropriate for the performance results we wished to obtain. In fact, the scaling of the MMPP parameters by the number of active users resulted in earth stations having approximately Poisson type voice and video traffic. To counter this problem, two alternatives were considered:

1. Use of heterogeneous earth stations: If the underlying assumption is that the earth stations are not identical, then we can split up the active video and voice sources, so that each earth station gets at least 1 source. The above technique will ensure that the overall beam load is divided between 70% voice/ 10% video/ 20% data. However, because the earth

stations are not homogeneous, it is more difficult to interpret the performance measures - data delay, and real-time loss probability. Furthermore, analysis of such a heterogeneous network becomes very complicated.

2. Scale down the network: That is, use a much lower number of active users (i.e., model only active users, and ignore the call admission problem): Ideally, we would like to meet the following criteria:

- overall uplink load is divided between voice, video, and data, according to Table 3-2 (for the example considered 70% voice/10% video/20% data)
- number of active voice and video sources per earth station is an integer.

To meet the above criteria, we considered the average cell generation rates (CGR) per active voice and video sources, and found the minimum number of such sources that would produce the desired traffic mix. As an example, consider the voice dominant case, where for each earth station, the cell generation rate for voice should be 7 times that for video, and 3.5 times that for data. Assume that each earth station has 1 video source, which we know has an average CGR of 4.8 cells/frame. The cell generation rate for voice should be $(7)(4.8)=33.6$ cells/frame, in order to maintain the correct traffic mix. These 33.6 cells/frame correspond to $(33.6)/(1.6)=21$ on-off voice sources. At this point, we can also see that the average data CGR is $(33.6)/(3.5)=9.6$ cells/frame. If each earth station has a ratio of voice to video sources of 21:1, the two criteria above are met. With this information, the average CGR per earth station is:

$$\begin{aligned}
 AVE_{CGR} &= \frac{\text{average CGR of voice}}{\text{earth station}} + \frac{\text{average CGR of video}}{\text{earth station}} + \frac{\text{average CGR of data}}{\text{earth station}} \\
 &= 21(1.6) + 1(4.8) + 9.6 \text{ cells/frame} \\
 &= 48 \text{ cells/frame} \\
 &= 768 \text{ Kbps for a 48 byte cell and a 24 msec frame}
 \end{aligned}$$

The uplink can support 4096 cells/frame, which means that the total number of active users per uplink is 85.

Since it was felt that the second alternative was more realistic, this was the technique adopted. The results for each traffic scenario, calculated using the procedure above, are shown in Table 3-4. Note that for the voice, video, and data dominant cases, the average CGR per earth

station is the same - 48 cells/frame. However, no combination of voice:video:data sources gave this average for the equally dominant case.

3.2.1 Parameter Matching

As was already explained, each of the on-off sources was mapped into an equivalent 2 state MMPP. To show the procedure, we will take the voice dominant scenario as an example, and derive the 2 state MMPP parameters for the video. From Table 3-4, the first thing we notice is that for this case, each earth station has 1 video source. This video source is represented by 6 independent mini on-off sources with parameters given in Table 3-3. The overall mapping is shown in Figure 3-2.

Table 3-4: Number of Voice and Video Sources per Earth Station per Traffic Mix

Traffic Scenario	Number of Voice: Video sources per earth station	Average Cell Generation Rate per earth station (cells/frame)	Load Points Considered (in Simulations)	Number of Active Earth Stations
Voice Dominant	21:1	48	0.95	81
			0.8	68
			0.5	43
			0.2	17
Video Dominant	3:7	48	0.95	81
			0.8	68
			0.5	43
			0.2	17
Equally Dominant	3:1	144	0.95	270
			0.8	227
			0.5	142
			0.2	57
Data Dominant	6:1	48	0.95	81
			0.8	68
			0.5	43
			0.2	17

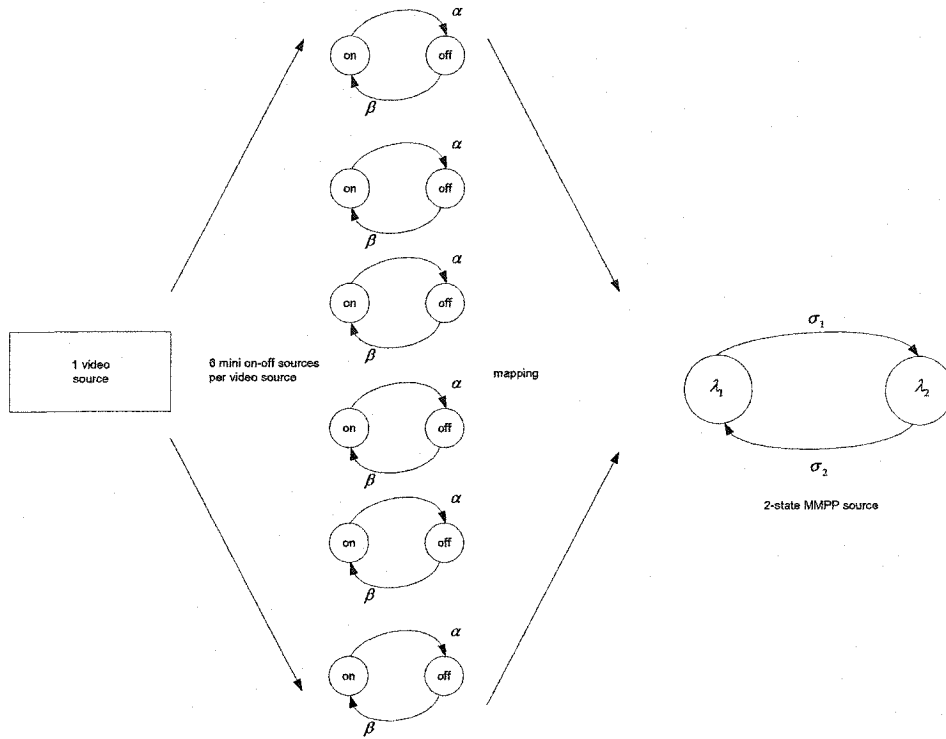


Figure 3-2: Mapping of One Video Source to a 2-state MMPP

A two state MMPP is a doubly stochastic Poisson process with arrival rates λ_1 and λ_2 [pkts/sec], and exponential sojourn times σ_1^{-1} and σ_2^{-1} [sec], for each of the two states. The mapping used is based on matching:

1. the average arrival rates
2. the variance of the arrival rates, and
3. the asymptotic IDC

with the condition that $\sigma_1 = \sigma_2$. This latter condition states that the mean sojourn times in each of the 2 states is equal. In an underload/overload [68] interpretation, this is very close to choosing the number of on-off sources in the underload region equal to $(N)PAR^{-1}$ where

N = total number of on - off sources

PAR^{-1} = Probability that the on - off source is in on state

As will be shown shortly, the choice of $\sigma_1 = \sigma_2$ simplifies the matching considerably. The average arrival rate and variance of arrival rate for N on-off sources with peak rate A , and peak-to-average ratio PAR are shown below. Also included, are the corresponding quantities for a 2 state MMPP [69].

$$\begin{aligned}
\bar{\lambda}_{N-on-off} &= E[\text{arrival rate for } N \text{ on - off sources}] \\
&= NA(PAR^{-1}) \\
\\
VAR_{N-on-off} &= VAR[\text{arrival rate for } N \text{ on - off sources}] \\
&= NA^2(1 - PAR^{-1})PAR^{-1} \\
\\
\bar{\lambda}_{MMPP} &= \frac{\lambda_1 + \lambda_2}{2} \\
\\
VAR_{MMPP} &= \frac{(\lambda_1 - \lambda_2)^2}{4}.
\end{aligned} \tag{3-1}$$

Furthermore, the asymptotic IDC for a 2 state MMPP source and for N on-off sources, are derived in [70], and the results for the case $\sigma_1 = \sigma_2 = \sigma$ are

$$\begin{aligned}
IDC_{MMPP}(\infty) &= \frac{1 + (\lambda_1 - \lambda_2)^2}{4\sigma\lambda_{MMPP}} \\
IDC_{N-on-off}(\infty) &= (1 - PAR^{-1})^2 \left[\frac{2A}{\alpha} - 1 \right]
\end{aligned} \tag{3-2}$$

where $1/\alpha$ is the average on time for the on-off source. Setting

$$\begin{aligned}
IDC_{MMPP}(\infty) &= IDC_{N-on-off}(\infty) = IDC, \\
\bar{\lambda}_{N-on-off} &= \bar{\lambda}_{MMPP} = \bar{\lambda}, \text{ and} \\
VAR_{N-on-off} &= VAR_{MMPP} = VAR
\end{aligned}$$

we can solve for σ , λ_1 , and λ_2 . The results are

$$\begin{aligned}
\sigma &= \frac{VAR}{\bar{\lambda}(IDC - 1)} \\
\lambda_1 &= \bar{\lambda} + \sqrt{VAR} \\
\lambda_2 &= \bar{\lambda} - \sqrt{VAR}
\end{aligned} \tag{3-3}$$

For the example we are considering, $N = 6$, $PAR = 5$, peak on rate $A = 166.67$ packets/sec, and $1/\alpha = 265$ msec. Therefore,

$$\begin{aligned}
\bar{\lambda}_{N-on-off} &= NA(PAR^{-1}) = 200.004 \text{ packets/sec} \\
VAR_{N-on-off} &= NA^2(1 - PAR^{-1})PAR^{-1} = 26667.733 \text{ (packets/sec)}^2 \\
IDC_{N-on-off}(\infty) &= (1 - PAR^{-1})^2 \left[\frac{2A}{\alpha} - 1 \right] = 55.9
\end{aligned}$$

Solving,

$$\begin{aligned}
\sigma &= 2.4287 \text{ sec}^{-1} \\
\lambda_1 &= 363.3 \text{ packets/sec} \\
\lambda_2 &= 36.7 \text{ packets/sec}
\end{aligned}$$

This same procedure is repeated for all traffic scenarios and for the matching of both the voice and video.

For the data traffic, we assumed a 2 state PMPP source for each earth station. A PMPP source is a doubly stochastic Poisson process whose sojourn time in each state is governed by a Pareto process (with parameter β). The resulting traffic is self similar with Hurst parameter [71]:

$$H = \frac{3 - \beta}{2}$$

The mapping was done by matching the required load ($\bar{\lambda}$), setting $\lambda_1 = 2.5\lambda_2$, and fixing the Hurst parameter (H) to 0.8. The resulting system of equations is shown below:

$$\begin{aligned}
\lambda_2 &= \frac{2\bar{\lambda}}{3.5} \\
\lambda_1 &= 2.5\lambda_2 \\
\sigma &= \frac{3-2H}{2-2H}
\end{aligned} \tag{3-4}$$

These values were determined from measurements on existing Ethernet LAN data traffic [71]. The overall mapping procedure is summarized in the flowchart in Figure 3-3.

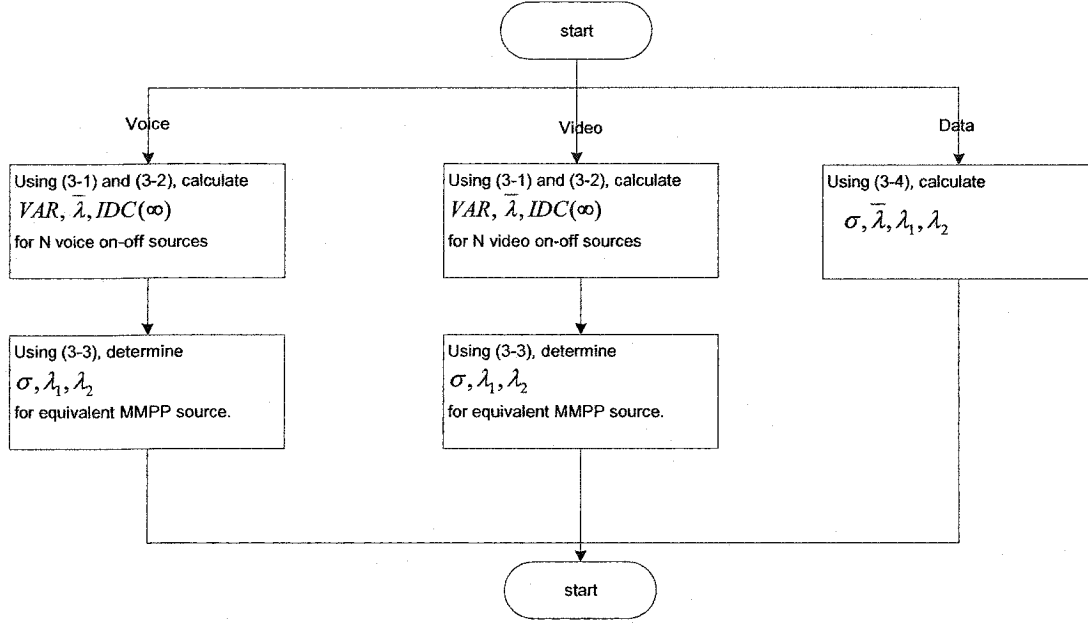


Figure 3-3: Flowchart for Parameter Matching Algorithm

As an example, a typical earth station with a voice dominant traffic mix, is shown in Figure 3-4. This earth station is comprised of a single video source (6 mini on-off sources), 21 on-off voice sources, and a self-similar data source. For this typical source, Figure 3-5 shows the index of dispersion for counts (IDC) curves for each of the three types of traffic. The IDC curve is generally regarded as a very reliable measure of variability and burstiness [72]. If the number of arrivals in a time interval t is denoted by $A(t)$, then the IDC is defined as

$$IDC(t) = \frac{\text{Var}[A(t)]}{E[A(t)]}$$

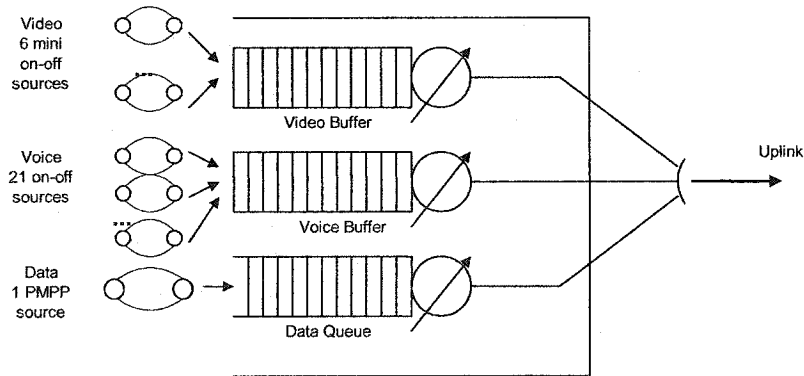


Figure 3-4: Typical Earth Station for a Voice Dominant Case

The IDC at time t (also referred to as lag t), measures the variability as we consider arrivals over larger and larger time scales. For traffic that is not self-similar, this variability will level off for longer lags (t). This does not imply that the traffic is not bursty, but rather that this burstiness has stopped increasing. In other words, at very large time scales, the traffic tends to look rather smooth and steady. In contrast for self-similar traffic, the IDC curve continues to increase regardless the lag. These types of traffic are said to experience variability across many time scales. The property that the traffic is *similar* across many time scales is characteristic of long range dependence. As a result, very regular and deterministic traffic has an IDC of 0, whereas Poisson type traffic has a constant IDC of 1 (for all t). Since the evaluation of the uplink access protocols discussed in the previous chapter is based almost entirely on Poisson and constant traffic models, it should be clear that they fail to account for the variability shown in Figure 3-5. Notice that the IDC curve for the PMPP traffic increases even for very long lags, and that the IDC curves for the voice and video, level off at 15.9 and 55.9 respectively.

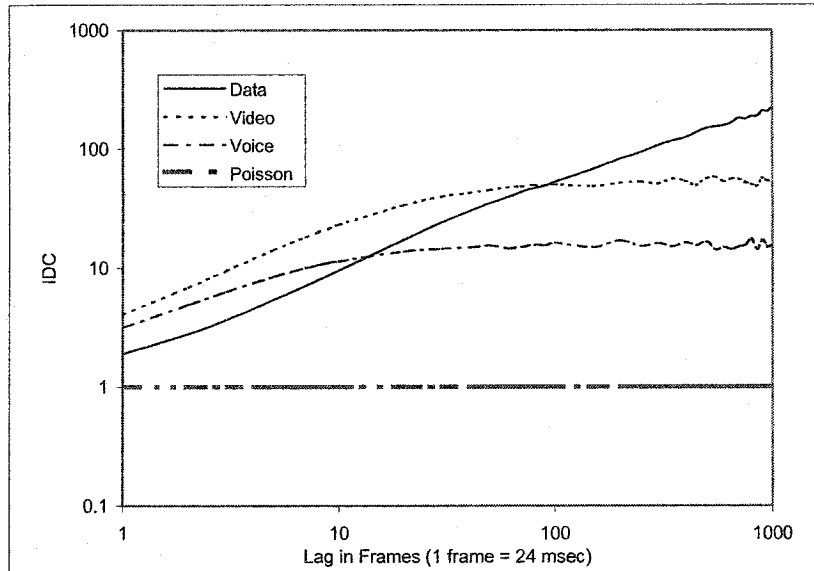


Figure 3-5: IDC Curves for a typical Earth Station Traffic
(comparison to Poisson Traffic is included)

3.3 MF-TDMA with CFDAMA

In this section, we will describe the basic operation of the bent-pipe satellite system, assuming an MF-TDMA frame format with combined free/demand assigned multiple access (CFDAMA) requesting strategy.

3.3.1 System Description

The available beam bandwidth per spot beam results in a maximum capacity of 4096 channels/frame. This capacity is divided into 32 carriers of 128 slots each. Each of these channels is assigned to users by a scheduler.

The main concept of CFDAMA is to interleave channels that are assigned based on user requests, with those that are freely assigned. Every frame is divided into two sections - a reservation section and a traffic section (See Figure 3-6). The ES's make requests in the reservation section, and are assigned channels from the traffic section. ESs can also piggyback their requests on packets sent in the traffic section.

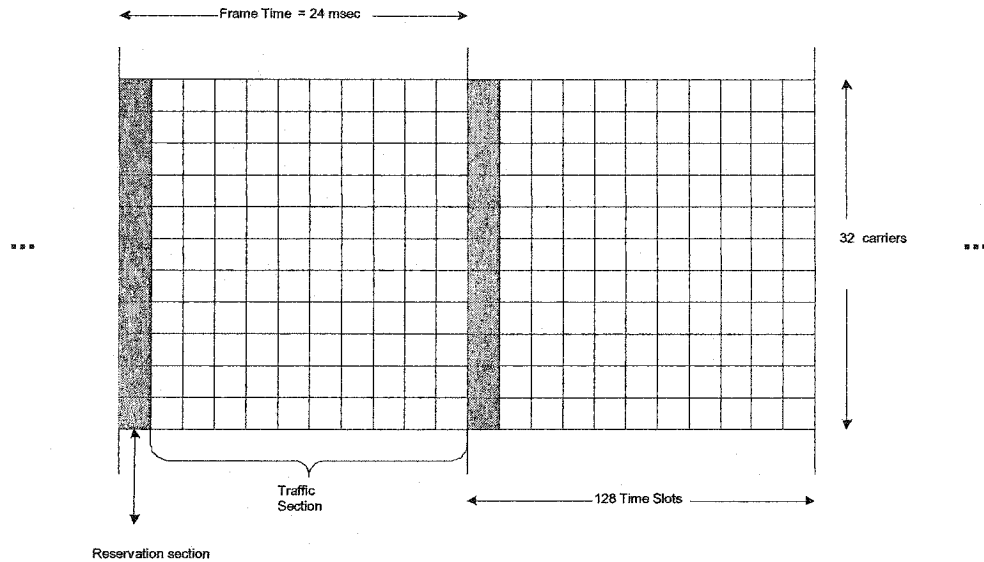


Figure 3-6: MF-TDMA Frame Format

At the earth stations, the traffic is mapped to a corresponding capacity type. Recall that the DVB-RCS standard has five such types (CRA, RBDC, VBDC, AVBDC, FCA). The actual mapping depends on the characteristic of the traffic. Also, a single connection can be mapped to more than one capacity type. For the results presented, we will assume a very simple mapping, as highlighted in Table 3-5. Notice that the voice and video are assigned to one capacity type (RBDC). An earth station will distribute its assigned channels based on the traffic type. Channels assigned in response to an RBDC request are reserved for video and voice traffic. Since we assumed that video has a higher priority than voice, it has first opportunity to use any assigned channels. Free capacity assignment is only permitted for the jitter-tolerant data traffic. For voice and video, all channel assignments are based on requests sent by the earth stations. These channels regularly arrive one round trip time after the request is made. On the other hand, free capacity is more haphazard. If real-time traffic would use this free capacity it would tend to introduce jitter in the real-time traffic stream.

Table 3-5: Mapping of Traffic to Capacity Type

Earth Station Traffic	Capacity Type	Comments
Real-Time Video	RBDC	Video and Voice data are treated together. At the earth station, video is given priority. The maximum number of channels assigned for each earth station in response to an RBDC capacity request is limited to maxRBDC.
Real-Time Voice	RBDC	
Jitter-Tolerant Data	VBDC & FCA	We will assume that requests are never lost, and as a result, AVBDC and VBDC are equivalent.

The basic system operation can be explained by considering Figure 2-10, which shows how requests are transmitted over time, and how the scheduler deals with these requests. For the moment, consider only VBDC requests. As user k receives JT packets, it stores these in its VBDC queue and sends requests to the scheduler based on the number of channels required to drain its queue. These requests arrive at the scheduler after the appropriate delay and are placed in a scheduler VBDC queue, along with the requests from all other users. These requests are then serviced at the start of every frame. When the scheduler services a request, we mean that it assigns a channel in the current frame to the user that initiated the request. If the VBDC request can not be serviced fully, then it is updated, and returned to the VBDC request queue. To be fair, the scheduler keeps track of the last earth station that has been serviced, so that in subsequent frames, it can begin servicing at the next station. The scheduler assigns all the channels in the current frame, and notifications are sent to the users in the form of a burst time plan (BTP). In its simplest form, the BTP contains a mapping that notifies the earth stations the channels that they have been assigned in the next frame. An example is shown in Figure 3-7.

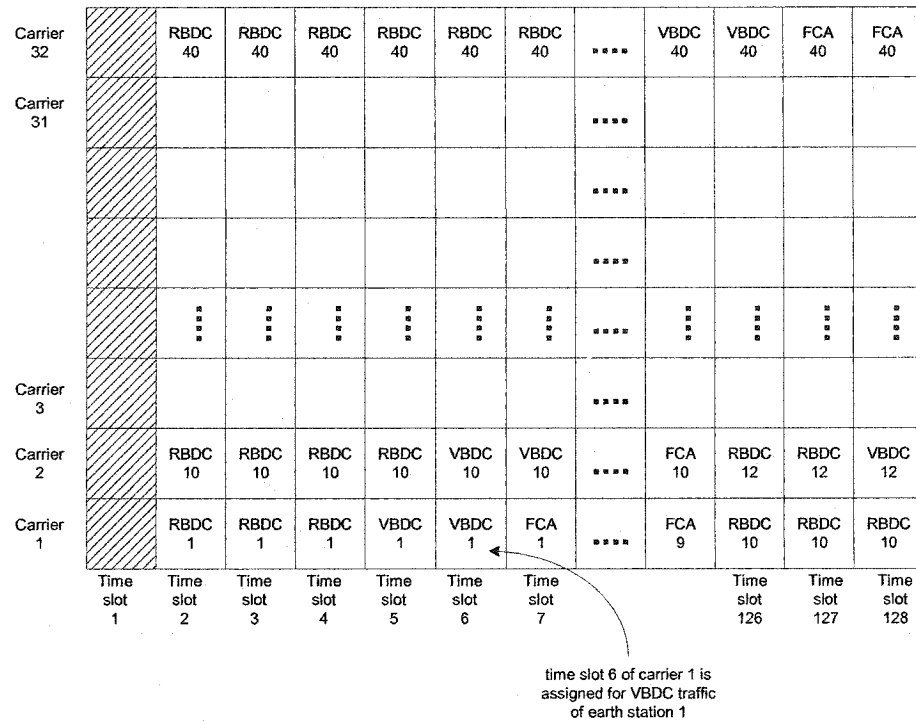


Figure 3-7: Example of a Burst Time Plan (BTP)

Here, earth station 1 has time slots 2 to 4 of carrier 1 assigned for its RBDC traffic, time slots 5 to 6 of carrier 1 assigned for its VBDC traffic, and time slot 7 of carrier 1 which is free assigned. Notice that it is possible for the scheduler to assign channels to an earth station on two different carriers (for example, consider earth station 10). As a result, the earth stations must be capable both of carrier and time hopping. The BTP arrives at the user terminals, again after a suitable propagation delay. The scheduler assigns channels with the objective of maximizing uplink utility while being fair to all earth stations.

At the scheduler, the requests are served on a first-come first serve basis. Owing to the randomness of earth station transmissions, it is possible that after assigning a certain number of channels, the scheduler queue becomes empty. The scheduler knows that there are channels available, but it has no more requests to service. However, by the time the scheduler notification arrives at the user earth stations, additional packets may have arrived. Therefore, the scheduler can free assign any unused channels to earth stations. When the notification for the free assigned channels arrives at the earth stations, and these users have VBDC packets to transmit, these can

be transmitted without having made a request. The delay experienced by these packets is therefore very small.

The above discussion strictly applies only to VBDC traffic, which is not delay sensitive. For real-time traffic, the situation is slightly different. Basically, the voice and video traffic is mapped to the RBDC capacity type. Once an RBDC request arrives at the scheduler, it is stored in the RBDC request queue, along with requests from other earth stations. The RBDC request remains valid until it is either updated or it times-out. The scheduler assigns channels based on the requests, up to a configured maximum (maxRBDC). This maxRBDC is usually configured at call setup⁸ and limits the number of channels assigned for the voice and video traffic.

The proposed scheduler algorithm is rather simple, and is based on running four passes through the RBDC and VBDC request queues (for every frame):

Pass1: The scheduler cycles through the RBDC request queue, and determines how many RBDC channels are required for all earth stations (i.e., numb_RBDC_required). If an earth station requires more than maxRBDC, the assignment is limited to this maximum.

Pass2: If $\text{numb_RBDC_required} < C_{max}$ ($C_{max} = 4096 = \text{capacity}$), then all RBDC requests are satisfied. Otherwise, only C_{max} are assigned (the maximum), in a round robin fashion. Note that with the use of maxRBDC, the system could be configured so that RBDC capacity never exceeds C_{max} .

Pass3: If there are still empty channels after Pass2, then these are assigned to the VBDC packets whose requests are in the scheduler queue. This assignment is on a first-come first-serve basis. As mentioned earlier, for every frame the scheduler keeps track of the last earth station to be assigned VBDC channels. During the subsequent assignment process, the next earth station is serviced first.

Pass4: If empty channels remain after Pass3, these are freely assigned to earth stations in a round-robin fashion. Again in order to be fair, the scheduler will keep track of the last earth station to be awarded free assignment capacity.

It is expected that maxRBDC will have a considerable impact on the real-time loss probability.

⁸ Call setup is not considered in this thesis. It will be assumed that the maxRBDC has been set up accordingly.

3.3.2 Performance Results

Based on the typical earth station shown in Figure 3-4, the queuing model shown in Figure 3-8 has been chosen for the evaluation of the MF-TDMA/CFDAMA protocols. The model is based on the frame boundaries acting as embedding points for both the arrival and the service mechanisms. The real-time buffer is fed by a 2-state Markov-Modulated Poisson Process (MMPP) voice source and a 2-state MMPP video source. The matching of the individual on-off sources to a single 2-state MMPP is detailed in Section 3.2.1. Such a mapping is quite common in literature, as is evidenced by the papers from Lucantoni and Daigle [70][81]. In fact, simulation results for typical system parameters show that mapping the combination of mini on-off video sources and the superposition of the on-off voice sources by respective MMPPs has very little effect on the average data packet delay over time. Figure 3-9 shows the simulation results for an MF-TDMA system for a data dominant case and for a system load of 0.8. Clearly, the use of the MMPP approximation has a limited impact on the convergence results for the average data packet delay (results differ by 1 msec). Owing to this, and similar such results, it was concluded that the MMPP approximation was suitable for the evaluation of average data packet delay. In addition, we also used a 2-state MMPP to model the bursty jitter-tolerant traffic. To verify the implications of this assumption, simulations were run for typical cases, with data generated by three different traffic models: Poisson, MMPP, and PMPP. The data dominant case simulation results for one of these runs are included in Figure 3-10. Note that the use of the Poisson model to simulate bursty JT traffic severely underestimates the jitter-tolerant traffic delay. In fact, the Poisson assumption implies that the delay would remain at a minimum for loads in excess of 75%. The burstier PMPP model, on the other hand, shows that self-similar traffic leads to substantially worse results. Surprisingly, however, modeling the JT traffic by an MMPP model leads to results that are comparable to those obtained with the PMPP model, except at very high loads (over 85%). This

indicated that using the MMPP model for JT traffic would lead to fairly accurate results, and more importantly, analytic tractability (as will be shown later).

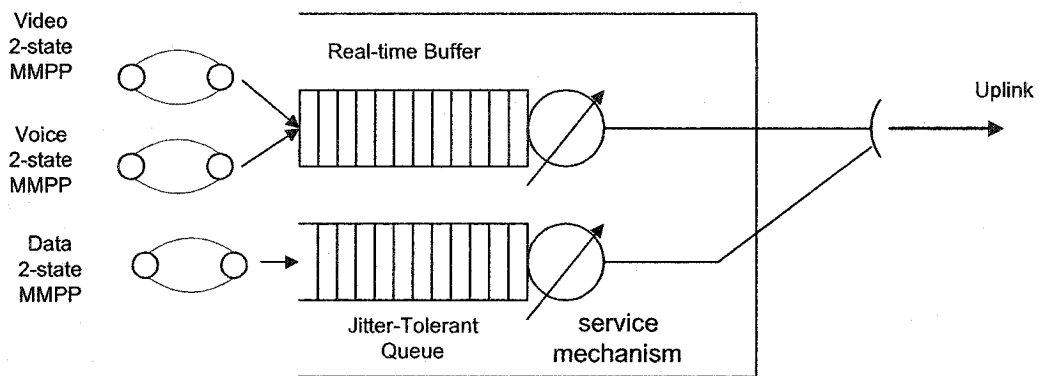


Figure 3-8: Queuing Model of Earth Station

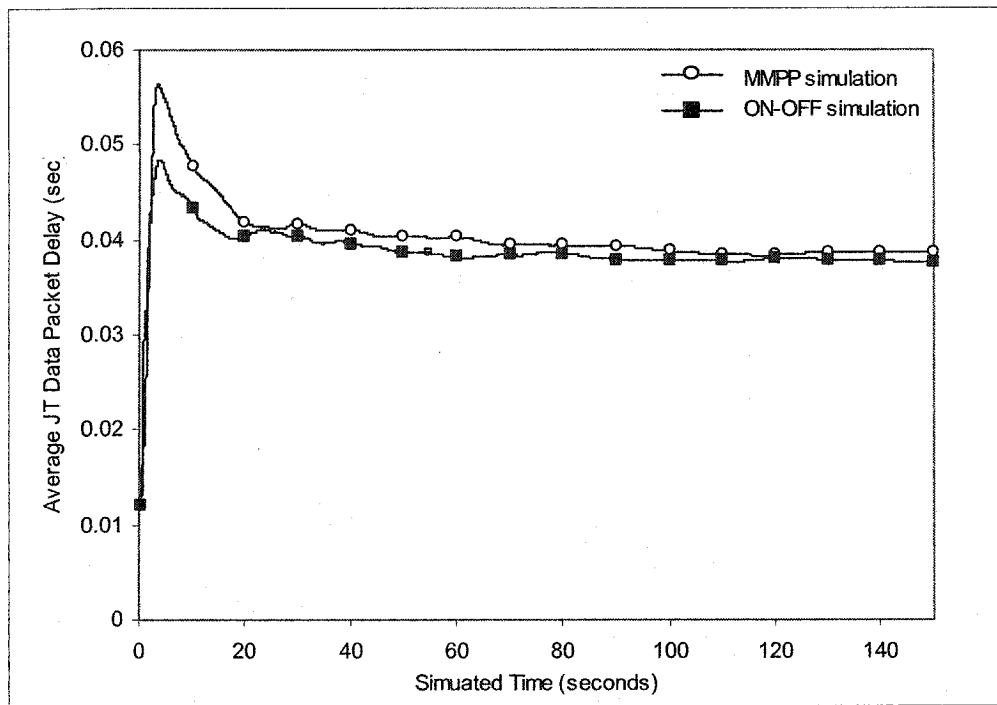


Figure 3-9: Average JT Packet Delay
Convergence results for on-off and MMPP representations of Voice and Video

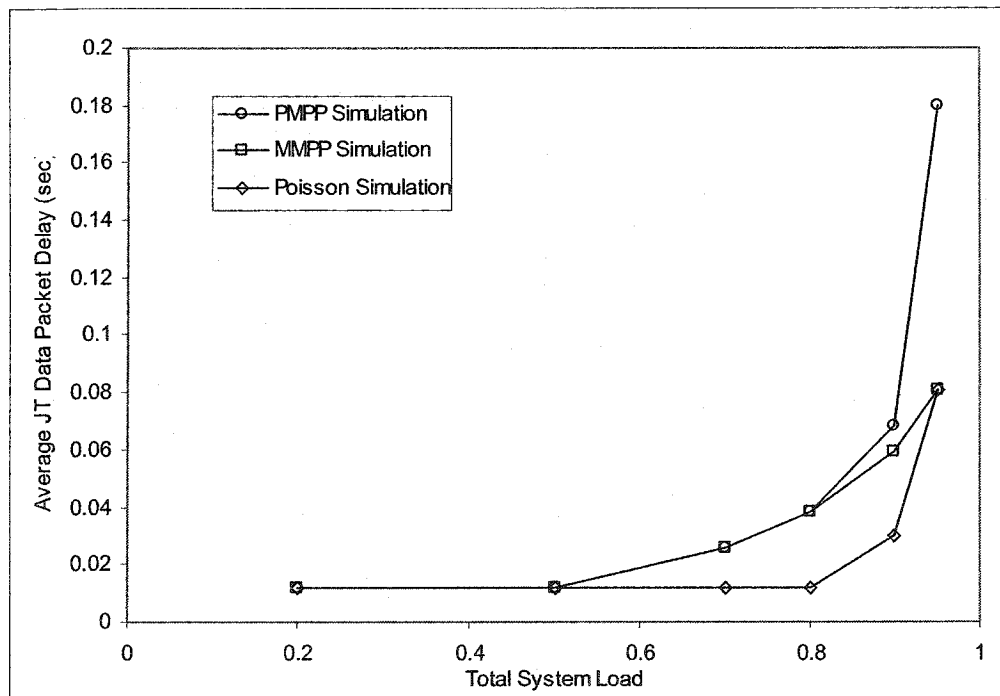


Figure 3-10: Average JT Packet Delay: Effect of Data Traffic Model

The key advantages of the queuing model depicted in Figure 3-8, are

- 1) Real-time loss probability becomes a total probability problem, which depends on the counting process of the MMPP sources (a process whose distribution is known), and the service given to the real-time buffer.
- 2) The data queue becomes a discrete time MMPP/G/c queue (if we embed the chain at frame intervals) with, albeit, a very complicated service mechanism. As discussed in Section 2.4.2, models of this type can be solved using the Matrix Geometric Technique. All that is required is to put the transition matrix in the form:

$$P_{MG} = \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ A_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (3-5)$$

3.3.2.1 Real-Time Loss Probability

The real-time traffic is the superposition of two 2-state MMPP processes. The net 4-state MMPP process has infinitesimal generator matrix R and arrival rate matrix A [61]. These are given by,

$$R = R_{vi} \oplus R_{vo} \quad A = A_{vi} \oplus A_{vo}$$

where $R_{vi} = \begin{pmatrix} -\sigma_{1,vi} & \sigma_{1,vi} \\ \sigma_{2,vi} & -\sigma_{2,vi} \end{pmatrix}$ generator matrix for 2-state video

$R_{vo} = \begin{pmatrix} -\sigma_{1,vo} & \sigma_{1,vo} \\ \sigma_{2,vo} & -\sigma_{2,vo} \end{pmatrix}$ generator matrix for 2-state voice

$A_{vi} = \begin{pmatrix} \lambda_{1,vi} & 0 \\ 0 & \lambda_{2,vi} \end{pmatrix}$ rate matrix for 2-state video

$A_{vo} = \begin{pmatrix} \lambda_{1,vo} & 0 \\ 0 & \lambda_{2,vo} \end{pmatrix}$ rate matrix for 2-state voice

$\lambda_{1,vi}$, $\lambda_{2,vi}$, $\sigma_{1,vi}$, and $\sigma_{2,vi}$ are the parameters for the 2 state MMPP video source, and $\lambda_{1,vo}$,

$\lambda_{2,vo}$, $\sigma_{1,vo}$, and $\sigma_{2,vo}$ are the corresponding parameters for the 2 state MMPP voice source.

λ_i denotes the arrival rate in state i , while σ_i^{-1} denotes the average sojourn time in state i . In the above \oplus denotes for the Kronecker sum. The Kronecker sum for 2 matrices A and B , is defined as

$$A \oplus B = (A \otimes I_B) + (I_A \otimes B)$$

In the above I_A denotes an identity matrix whose size matches that of matrix A , and \otimes denotes the Kronecker product,

$$A \otimes B = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nm} \end{bmatrix} \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix}$$

Once we know the R and A of the combined 4-state arrival process to an earth station, we can determine the steady state distribution of the number of real time arrivals. This can be found through the uniformization technique described in [82]. The main steps are highlighted below:

Step 1: Solve for π = steady state vector of the underlying 4-phase state process. Let

$$\begin{aligned} \pi &= [\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4] \\ \pi_i &= \text{Pr}[\text{phase process in state } i] \end{aligned}$$

Solve for π :

$$\begin{aligned} \pi R &= 0 \\ \pi e^T &= 1 \end{aligned}$$

where e is a unit row vector and 0 is a zero row vector⁹.

Step 2: Determine the D matrices

$$\begin{aligned} D_0 &= R - A \\ D_I &= A \end{aligned}$$

Step 3: Let

⁹ When using the vector e , we will assume that this is a unit row vector whose size corresponds to the size of the vector it multiplies (for example shown $e = [1 \ 1 \ 1 \ 1]$). When using the vector 0 , we will assume that

T_F = frame duration

$N(t)$ = number of arrivals in time t

$J(t)$ = phase of arrival process at time t (phase of 4-state MMPP)

Then,

$$P_{rt, lw}(k) = \Pr[N(t) = k, J(t) = w | J(0) = l] = \Pr[k \text{ real-time arrivals in time } t \text{ and phase process in state } w, \text{ given that phase process was in state } l \text{ at time } 0].$$

In matrix form,

$$\begin{aligned} P_{rt}(k) &= \{P_{rt, lw}(k)\} \\ &= \sum_{j=0}^{\infty} e^{-\theta t} \frac{(\theta t)^j}{j!} K_n^{(j)} \end{aligned}$$

where $\theta = \max\{(-D_0)_{ii}\}$, the maximum diagonal element of $(-D_0)$, and $K_n^{(j)}$ can be found recursively from:

$$\begin{aligned} K_0^{(0)} &= I \\ K_n^{(0)} &= 0 \quad n \neq 0 \\ K_0^{(j+1)} &= K_0^{(j)}(I + \theta^{-1}D_0) \\ K_n^{(j+1)} &= \theta^{-1} \sum_{i=0}^{n-1} K_i^{(j)} D_{n-i} + K_n^{(j)}(I + \theta^{-1}D_0) \quad n \neq 0 \end{aligned}$$

Step 4: Determine $\Pr[RT = k] = \Pr[\# \text{ real time arrivals in frame} = k]$

$\Pr[RT = k]$ is found by first using the total probability theorem and then unconditioning the result. That is,

$$\Pr[RT = k] = \sum_{l=1}^4 \left(\left(\sum_{w=1}^4 \Pr[N(T_f) = k, J(T_f) = w | J(0) = l] \right) \Pr[J(0) = l] \right)$$

or in matrix form

$$\Pr[RT = k] = \pi P_{rt}(k) e^T$$

this is a zero row vector whose size corresponds to the size of the vector it multiplies (for example shown $\mathbf{0} = [0 \ 0 \ 0 \ 0]$).

Once the probability of the arrival process is known, we can calculate the real-time loss probability by noting that real-time traffic is never queued, and that it is sent via the RBDC mechanism. As a result, a loss occurs when

1. an earth station generates more than maxRBDC (maximum RBDC request) real-time packets in a frame, or
2. the number of real-time packets generated in a frame, by all earth stations, exceeds the total capacity C .

The probability can be calculated as

$$\Pr[\text{real time loss}] \leq \Pr[RT > \max RBDC] + \Pr\left[\sum_{\text{all earth stations}} RT > C\right] \quad (3-6)$$

Based on the above analytic expression, the results for the voice, video, and data dominant cases highlighted in section 3.2 are shown in Figure 3-11 (we used a maxRBDC of 100). Simulations were run for all three traffic scenarios and no real-time losses were observed. This was expected, as the number of real-time packets transmitted in these simulation runs was of the order of 10^6 . As losses are extremely rare events (less than once in 10^{20}), evaluation of real-time loss probability by simulation is impossible without resorting to specialized simulation techniques (such as importance sampling) [78].

To check both the validity of the simulation model and the analytical result, we ran simulations for the data dominant case, with 2 different values of maxRBDC. The average cell generation rates for each traffic type, and the maxRBDC considered, are listed in Table 3-6.

Table 3-6: Effect of maxRBDC

Traffic Type	Average Cell Generation Rate [cells/frame]	Capacity Type	maxRBDC [channels/frame]	maxRBDC [channels/frame]
Voice	9.6	RBDC	20	100 ¹
Video	4.8			
Data	33.6	VBDC	N/A	N/A

Note 1: This corresponds to a value well over the peak cell generation rate for the voice and video traffic.

The results in Figure 3-12 show the average JT packet delay for both values of maxRBDC. Note that the delay is slightly lower for the case with lower maxRBDC. This can be explained since there are fewer real-time packets being transmitted when maxRBDC is configured to 20. In fact, simulation results show that 15% (analysis suggests that this is actually closer to 20%) of real-time packets are lost for maxRBDC = 20 (even though this is 30% higher than the average generation rate of the traffic). Clearly the large variability in the traffic has a significant effect on the performance of the system. With so many real-time packet losses, more channels are available for the jitter-tolerant data traffic and so these benefit from a lower average delay. As expected no voice or video packet losses were observed when maxRBDC was set to 100.

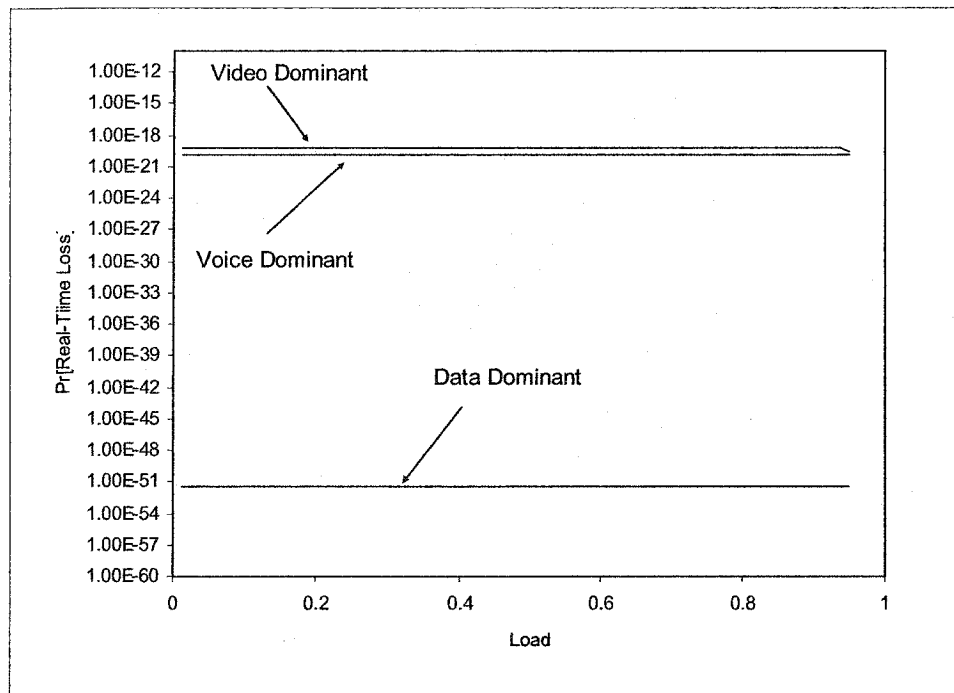


Figure 3-11: Real-time Loss Probability (Video/Voice/Data Dominant Cases)

An interesting observation about the real-time loss is that it is almost independent of load. Since the load per earth station is fixed, an increased load implies more active earth stations. Each of these earth stations has a configured maxRBDC. As far as the scheduler is concerned, the RBDC requests from each active earth station are treated independently. If we ignore the losses arising from the total RBDC exceeding capacity (second term in equation (3-6)), a loss will only

occur when an earth station's RBDC traffic exceeds its maxRBDC. As a result, it makes no difference if there are 17 active users (system load of 0.2) or 81 active users (system load of 0.8), so long as the second term in equation (3-6) is negligible. As the load increases, the number of real-time packets transmitted, as well as the number of real-time packets discarded, increase at the same rate and their ratio remains the same.

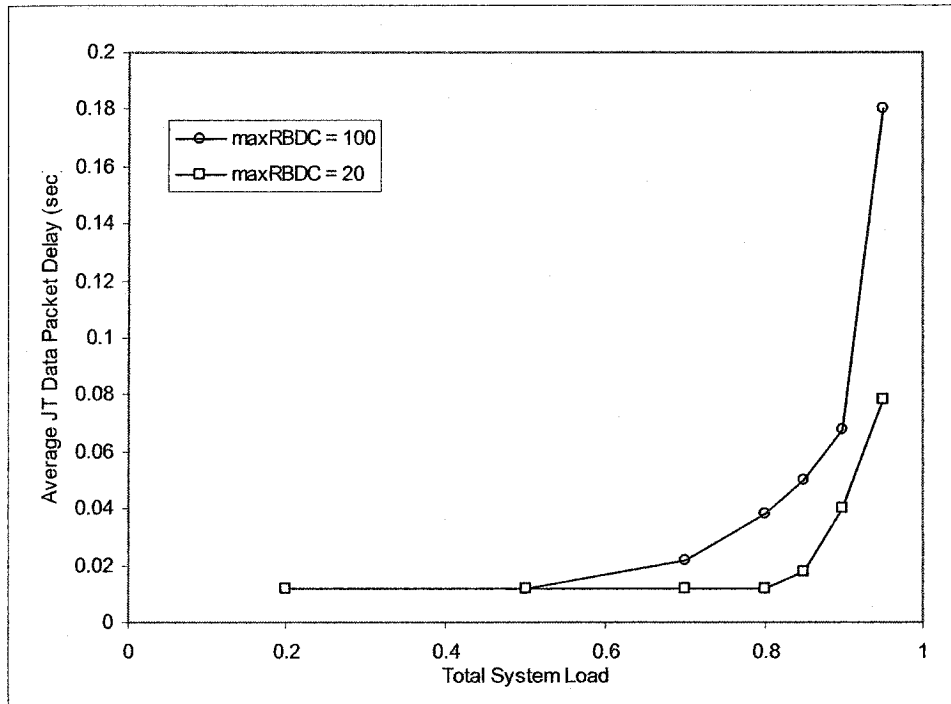


Figure 3-12: Comparison of Protocol with Various maxRBDC

3.3.2.2 Jitter-Tolerant Delay

To determine the average jitter-tolerant delay, we found the average queue occupancy at frame boundaries, and then applied Little's Theorem to determine the delay. To find the terminal queue occupancy, we embedded a Markov chain at frame boundaries (See Figure 3-13), and arrived at a system state equation.

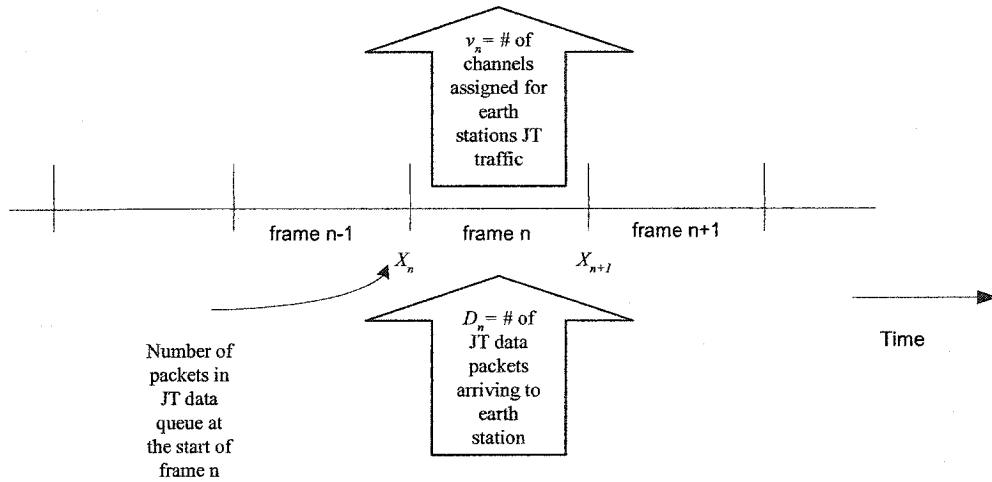


Figure 3-13: Embedding Points for System

$$X_{n+1} = \max(0, X_n - v_n) + D_n$$

where

X_n = state of system

= number of packets in the tagged earth station queue at the start of frame n

v_n = number of channels assigned to the data queue of tagged earth station during frame n

D_n = number of jitter - tolerant data packets arriving at the tagged earth station during frame n

To find the solution to this state equation, we need the distribution of v_n . This process is rather complex and depends on a number of factors:

1. the requests made by the terminals,
2. the location of the request in the scheduler queue,
3. the number of channels required for the real-time traffic (modeled as a 4-state MMPP process)
4. The free assignment mechanism at the scheduler.

These factors make it very difficult to determine v_n . As a result, we make the following simplifications. First, we define

$N =$ Number of active earth stations

$C =$ System capacity [channels/frame]

$M =$ Number of timeslots per carrier [channels]

Note that since an earth station can not transmit simultaneously on 2 carriers, the number of time slots per carrier limits the maximum earth station burst rate. That is, an earth station is allowed to transmit at most M packets/frame. Alternatively, we can say that an earth station can be assigned at most M channels/frame.

Clearly the number of carriers is given by C/M . Consider two different cases:

Case 1: $NM \leq C$

In this case, after scheduling, each earth station would be assigned the maximum number of channels (equal to M). Although these can be split across two or more carriers, the scheduler guarantees that no earth station is assigned two channels during the same time slot. This simplifies the service mechanism considerably. As a result, JT and RT traffic in the earth station must share the M channels. Since the RT traffic has priority, it can be argued that:

$$v_n = \max(0, M - RT_{n-K}) \quad (3-7)$$

where RT_{n-K} is the number of real-time packets waiting to be transmitted during frame n . (K is the scheduler latency in frames). This is a 4-state MMPP process (as defined in section 3.3.2.1). This is shown schematically in Figure 3-14. Recall that real-time packets arriving in frame $n-K$ must be transmitted in frame n . Any that cannot be transmitted are lost.

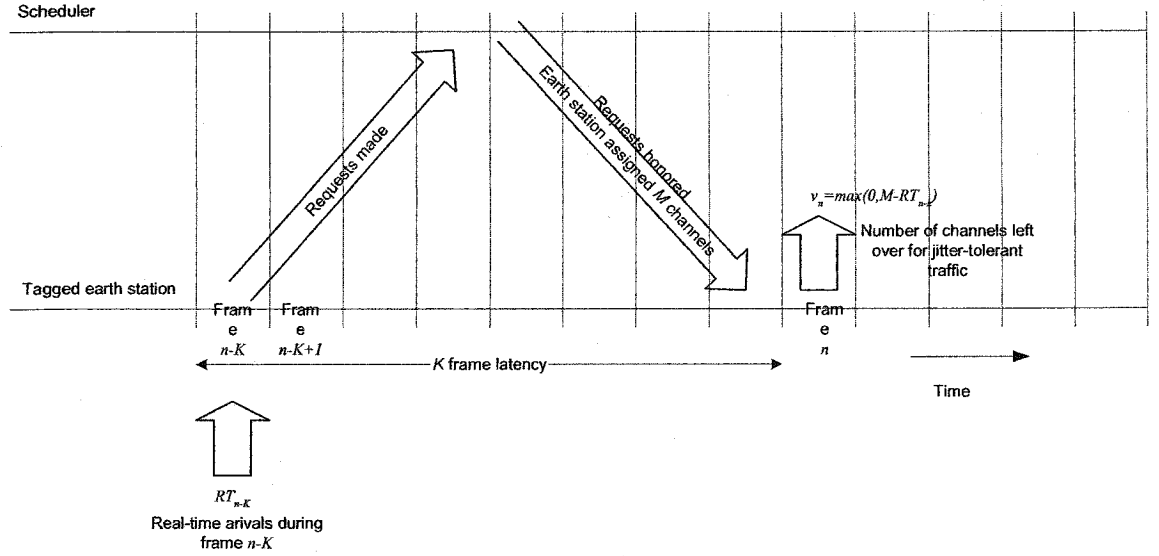


Figure 3-14: v_n calculation for Case 1

Case 2: $NM > C$

In this case, after scheduling, each earth station would get its fair share of the uplink capacity – however, fair share is very difficult to quantify. To simplify the analysis, we will make the following assumption:

- Every earth station receives an average amount of the uplink capacity. Although this fails to take into account the dynamics of the demand assignment, the results for the jitter-tolerant data packet delay should serve as an upper bound for the true delay. In fact, the performance would be equivalent to assuming a fixed channel assignment mechanism. Since the free capacity channels are distributed fairly to all earth stations, it tends to equalize the channel assignments.

Under this assumption, each station is awarded L channels, where L changes as the number of active earth stations (or load) changes. Namely $L = \lceil C/N \rceil$. These L channels are shared between an earth station's jitter-tolerant and real-time traffic. Since the real time traffic has priority, we can argue that

$$v_n = \max(0, L - RT_{n-K}) \quad (3-8)$$

where again RT_{n-K} is the number of real-time packets arriving in frame $n-K$ and awaiting transmission in frame n .

Equation (3-7) and (3-8) can be combined in a general form, yielding

$$v_n = \max(0, \min(L, M) - RT_{n-K}) \quad (3-9)$$

and so

$$X_{n+1} = \max(0, X_n - v_n) + D_n$$

The system state at time n (beginning of frame n) is given by the 3-tuple (i, j, k) . The state depends on

1. the number of packets in the jitter-tolerant data queue, denoted by $X_n = i$ ($i=0,1,2,3,\dots$)
2. the phase of the 2-state JT MMPP process, denoted by $J_{jt,n} = j$ ($j=1,2$)
3. the phase of the 4-state RT MMPP process, denoted by $J_{rt,n} = k$ ($k=1,2,3,4$).

The evolution of the embedded Markov chain is governed by the transition probability

$$P_{ijk,uvw} = \Pr[X_{n+1} = u, J_{jt,n+1} = v, J_{rt,n+1} = w \mid X_n = i, J_{jt,n} = j, J_{rt,n} = k],$$

which gives the probability of going from state (i, j, k) to state (u, v, w) .

The evolution of the embedded Markov chain is governed by the transition probability

The overall state transition matrix P is given by:

As a first step, let us group the transition matrix according to the shading in equation (3-10). The resulting matrix is of the form:

$$P = \begin{pmatrix} G_0 & G_1 & G_2 & \dots & G_{c-1} \\ U_1 & D & G_1 & \dots & G_{c-2} \\ U_2 & E_1 & B_0 & \dots & G_{c-3} \\ U_3 & E_2 & B_1 & \dots & G_{c-4} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{c-1} & E_{c-1} & E_{c-1} & \dots & D \\ U_1 & E_{c-1} & E_{c-1} & \dots & E_1 \\ 0 & U_2 & E_{c-2} & \dots & E_2 \\ 0 & 0 & 0 & \dots & E_3 \\ 0 & 0 & A_0 & \dots & E_4 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & U \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

where:

c = number of channels assigned to the tagged earth station $c = \min(L, M)$,

G_i = matrix of transition probabilities from an empty queue to a queue of size i ,

\mathbf{D} = matrix of transition probabilities from a queue of size i to a queue of size i ,

U_i = matrix of transition probabilities from a queue of size i to an empty queue ($i = 1, 2, \dots, c-1$),

E_i = matrix of transition probabilities from a queue of size j to a queue of size $j-i$ ($i = 1, 2, 3, \dots, c$).

1)

The definition of these matrices is shown graphically in Figure 3-15.

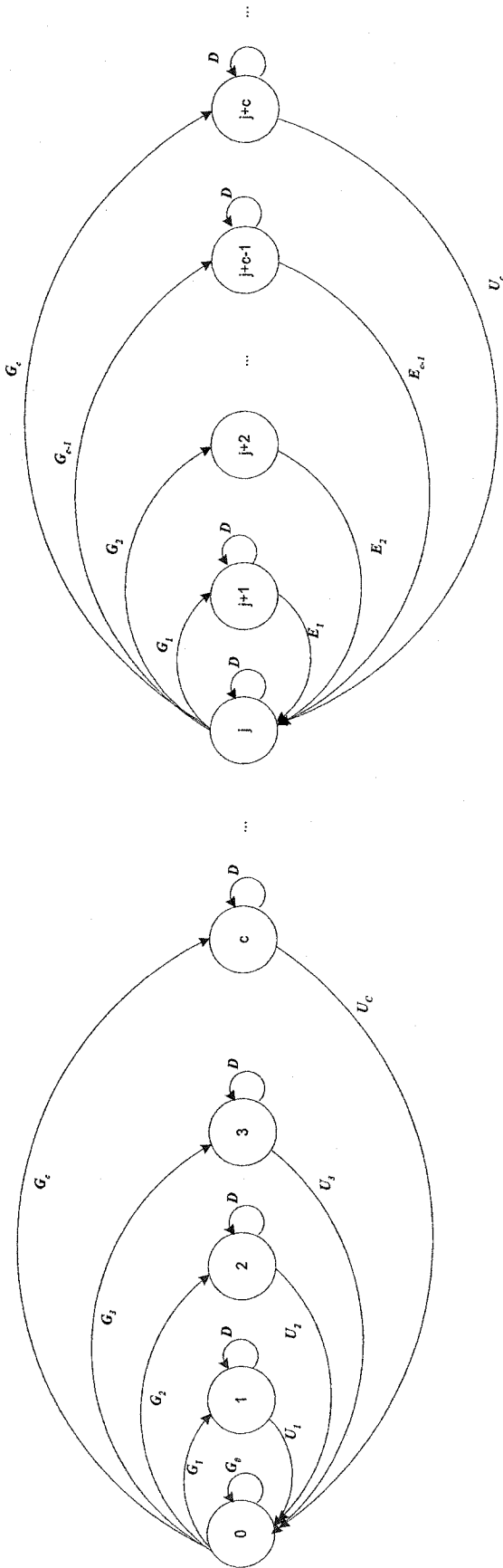


Figure 3-15: Graphical Representation of P matrix

The elements of these matrices can be found by considering the probability distribution for both the real-time and the jitter-tolerant traffic P_{rt} , P_{jt} . As an example, consider the matrix G_i , $i \geq 1$. The elements of this matrix are given by the probabilities $P_{0j,l,ivw}$. There are a number of arrival combinations that will lead to a transition from state $(0,j,l)$ to (i,v,w) . These are shown in Table 3-7. One should keep in mind that during each frame, the scheduler assigns c channels to the earth station. These c channels are used by the real-time traffic and then the jitter-tolerant data traffic. From the table, we notice that

$$P_{0j,l,ivw} = \sum_{m=0}^c P_{jt,jv}(c+i-m)P_{rt,lw}(m) + \sum_{m=c+1}^{\infty} P_{jt,jv}(i)P_{rt,lw}(m)$$

Table 3-7: Combinations to Produce Transition from $(0,j,l) \rightarrow (i,v,w)$

Initial Number in queue	Real-Time Transition		Jitter-Tolerant Transition		Final Number in queue
	Arrivals	Phase Change	Arrivals	Phase Change	
$(0,j,l)$	0	$l \rightarrow w$	$c+i$	$j \rightarrow v$	(i,v,w)
$(0,j,l)$	1	$l \rightarrow w$	$c+i-1$	$j \rightarrow v$	(i,v,w)
$(0,j,l)$	2	$l \rightarrow w$	$c+i-2$	$j \rightarrow v$	(i,v,w)
$(0,j,l)$	3	$l \rightarrow w$	$c+i-3$	$j \rightarrow v$	(i,v,w)
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(0,j,l)$	m	$l \rightarrow w$	$c+i-m$	$j \rightarrow v$	(i,v,w)
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$(0,j,l)$	c	$l \rightarrow w$	i	$j \rightarrow v$	(i,v,w)
$(0,j,l)$	$c+1$	$l \rightarrow w$	i	$j \rightarrow v$	(i,v,w)
$(0,j,l)$	$c+2$	$l \rightarrow w$	i	$j \rightarrow v$	(i,v,w)
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

The P_{rt} and P_{jt} transition probabilities are defined below:

$$P_{rt,lw}(m) = \Pr[\text{number of real time arrivals in a frame} = m, \text{phase of RT process transitions from } l \text{ to } w \text{ in frame}]$$

$$P_{jt,jv}(m) = \Pr[\text{number of jitter-tolerant arrivals in a frame} = m, \text{phase of JT process transitions from } j \text{ to } v \text{ in frame}]$$

Both of these transition probabilities can be found from the uniformization technique discussed earlier.

The last grouping step is to put the matrix in a form resembling (3-5). To do this, one final grouping is made, as shown by the shading in (3-11). The following A_i and B_i are obtained:

$$B_0 = \begin{pmatrix} G_0 & G_1 & G_2 & \dots & G_{C-1} \\ U_1 & D & G_1 & \dots & G_{C-2} \\ U_2 & E_1 & D & \dots & G_{C-3} \\ U_3 & E_2 & E_1 & \dots & G_{C-4} \\ \vdots & \vdots & \cdot & \dots & \cdot \\ \vdots & \ddots & \cdot & \dots & \cdot \\ U_{C-1} & E_{C-2} & E_{C-3} & \dots & D \end{pmatrix}, B_i = \begin{pmatrix} G_{iC} & G_{iC+1} & \dots & G_{(i+1)C-1} \\ G_{iC-1} & G_{iC} & \dots & G_{(i+1)C-2} \\ G_{iC-2} & G_{iC-1} & \dots & G_{(i+1)C-3} \\ G_{iC-3} & G_{iC-2} & \dots & G_{(i+1)C-4} \\ \vdots & \vdots & \dots & \cdot \\ \vdots & \vdots & \dots & \cdot \\ G_{(i-1)C+1} & G_{(i-1)C+2} & \dots & G_{iC} \end{pmatrix} \text{ for } i \geq 1$$

$$A_0 = \begin{pmatrix} U_C & E_{C-1} & E_{C-2} & \dots & E_1 \\ 0 & U_C & E_{C-1} & \dots & E_2 \\ 0 & 0 & U_C & \dots & E_3 \\ 0 & 0 & 0 & \dots & E_4 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & U_C \end{pmatrix}, A_i = \begin{pmatrix} D & G_1 & E_{C-2} & \dots & G_{C-1} \\ E_1 & D & E_{C-1} & \dots & G_{C-2} \\ E_2 & E_1 & U_C & \dots & G_{C-3} \\ E_3 & E_2 & 0 & \dots & G_{C-4} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ E_{C-1} & E_{C-2} & 0 & \dots & D \end{pmatrix}$$

where

$$G_0 = [P_{0,jl,0vw}] = \left[\sum_{m=0}^C \sum_{i=0}^{C-m} P_{jt,jv}(i) P_{rt,bw}(m) + \sum_{m=C+1}^{\infty} P_{jt,jv}(0) P_{rt,bw}(m) \right]$$

$$G_i = [P_{0,jl,ivw}] = \left[\sum_{m=0}^C P_{jt,jv}(C-m+i) P_{rt,bw}(m) + \sum_{m=C+1}^{\infty} P_{jt,jv}(i) P_{rt,bw}(m) \right]$$

$$D = [P_{1,jl,lvw}] = \left[\sum_{m=0}^C P_{jt,jv}(C-m) P_{rt,bw}(m) + \sum_{m=C+1}^{\infty} P_{jt,jv}(0) P_{rt,bw}(m) \right]$$

$$U_i = \left[\sum_{m=0}^{C-i} \sum_{n=0}^{C-m-i} P_{jt,jv}(n) P_{rt,bw}(m) \right]$$

$$E_i = \left[\sum_{m=0}^{C-i} P_{jt,jv}(C-m-i) P_{rt,bw}(m) \right]$$

Notice that $B_i = A_{i+1}$. Once the A_i and B_i are known, we can use the Matrix Geometric algorithm to determine the queue occupancy. The general steps of the algorithm are shown in [79] [83] and are repeated below for convenience.

Step 1: Solve for matrix G

This can be found by letting $G_0 = 0$ and solving recursively:

$$G_{k+1} = \sum_{n=0, n \neq 1}^{\infty} [I - A_1]^{-1} A_n (G_k)^n$$

until the matrix converges.

STEP 2: Find the invariant probability vector of G (namely g)

$$\begin{aligned} gG &= g \\ ge^T &= 1 \end{aligned}$$

STEP 3: Find β^*

$$\beta^* = \sum_{v=1}^{\infty} v A_v e^T$$

STEP 4: Find temporary matrices

$$\begin{aligned} A_{sum} &= \sum_{v=0}^{\infty} A_v \\ \bar{G} = e^T g &= \begin{bmatrix} g_0 & g_1 & g_2 & \cdots & g_{8c-1} \\ g_0 & g_1 & g_2 & \cdots & g_{8c-1} \\ g_0 & g_1 & g_2 & \cdots & g_{8c-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_0 & g_1 & g_2 & \cdots & g_{8c-1} \end{bmatrix} \\ \Delta(\beta^*) = diagonal(\beta^*) &= \begin{bmatrix} \beta_1 & 0 & 0 & \cdots & 0 \\ 0 & \beta_2 & 0 & \cdots & 0 \\ 0 & 0 & \beta_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_{8c} \end{bmatrix} \end{aligned}$$

STEP 5: Solve for ϕ

$$\phi = (I - G + \bar{G})(I - A_{sum} + \bar{G} - \Delta(\beta^*)\bar{G})^{-1} e^T$$

STEP 6: Find K

$$\begin{aligned} K &= \sum_{v=0}^{\infty} B_v G^v \text{ or} \\ K &= I - (A_0 G^{-1}) - A_1 + B_0 \end{aligned}$$

STEP 7: Find invariant probability vector K , namely κ .

$$\kappa K = \kappa$$

$$\kappa e^T = 1$$

STEP 8: Find κ^*

$$\kappa^* = e^T + \sum_{v=1}^{\infty} B_v \sum_{r=0}^{v-1} G^r \phi$$

STEP 9: Find x_0

$$x_0 = \frac{\kappa}{\kappa \kappa^*}$$

STEP 10: Find \overline{A}_i and \overline{B}_i

These can be found recursively by setting $\overline{A}_i = 0$ and $\overline{B}_i = 0$ for a large enough k , and

then using:

$$\begin{aligned} \overline{A}_j &= A_j + \overline{A}_{j+1} G \\ \overline{B}_j &= B_j + \overline{B}_{j+1} G \end{aligned} \quad j = k-1, k-2, \dots, 0$$

STEP 11: Solve for remaining x_i 's using Ramaswami's algorithm [84].

$$x_i = \left(x_0 \overline{A}_{i+1} + \sum_{v=1}^{i-1} x_v \overline{A}_{i+1-v} \right) (I - \overline{A}_1)^{-1} \quad i \geq 1$$

An expanded version of the matrix x_i is shown below:

$$x_i = [x_{0,1} \ x_{0,2} \ \dots \ x_{0,8c} \ x_{1,1} \ x_{1,2} \ \dots \ x_{1,8c} \ x_{2,1} \ x_{2,2} \ \dots \ x_{2,8c} \ \dots]$$

Where

$$x_{i,j} = \Pr[\text{queue occupancy} = i \text{ and phase of underlying phase process is } j]$$

Let

$$y_i = \Pr[\text{queue occupancy} = i].$$

Then,

$$y_i = \sum_{j=1}^{8c} x_{i,j}$$

When the y_i are known, the average queue occupancy (\bar{y}) can be found by a brute-force approach. That is,

$$\bar{y} = \sum_{i=0}^{\infty} i y_i$$

The average queueing delay (\bar{d}) then can be determined by applying Little's Theorem.

$$\begin{aligned}\bar{y} &= \lambda \bar{d} \\ \bar{d} &= \bar{y} / \lambda\end{aligned}$$

Where λ is the average arrival to this queue, which is given by:

$$\lambda = \frac{\lambda_1 \sigma_2 + \lambda_2 \sigma_1}{\sigma_1 + \sigma_2}.$$

The analytical results are shown in Figure 3-16 for the data dominant case. In deriving these results, we

1. terminated the convergence of Step 1 when the maximum change in matrix element (from iteration $k-1$ to k) was less than 10^{-20} ,
2. limited all infinite summations to size 50 (in steps 3,6,8). This number was based on previous experience from [79].

Also shown are the simulation results (dashed line) for the 3 different traffic mixes. As can be seen, the analytical model overestimates the jitter tolerant delay for those loads where the analysis follows Case 2 (i.e., $NM > C$). This is expected for 2 main reasons. Firstly, our analysis assumes a fixed channel assignment per frame, and fails to account for the demand reservation aspect of the scheduler. As demand reservation is better than fixed assignment at high loads, the results presented should be considered an upper bound on the true average JT data packet delay. Secondly, for the case $NM > C$, we assumed that an earth station is always assigned $L = \lceil C/N \rceil$ channels. In fact, C/N is typically a rational number and we should actually consider that some earth stations will be assigned $\lceil C/N \rceil + 1$ channels while others will be

assigned $\lceil C/N \rceil$. In addition, the simulation results show that the jitter-tolerant delay is virtually identical for the voice and video dominant cases. They only start to diverge for very high loads, and this can be attributed to the fact that video is somewhat burstier than voice (See IDC curves from Figure 3-5) and tends to leave less available capacity for the data. Note also that as the percentage of jitter-tolerant traffic in the traffic mix increases, so too does the data delay.

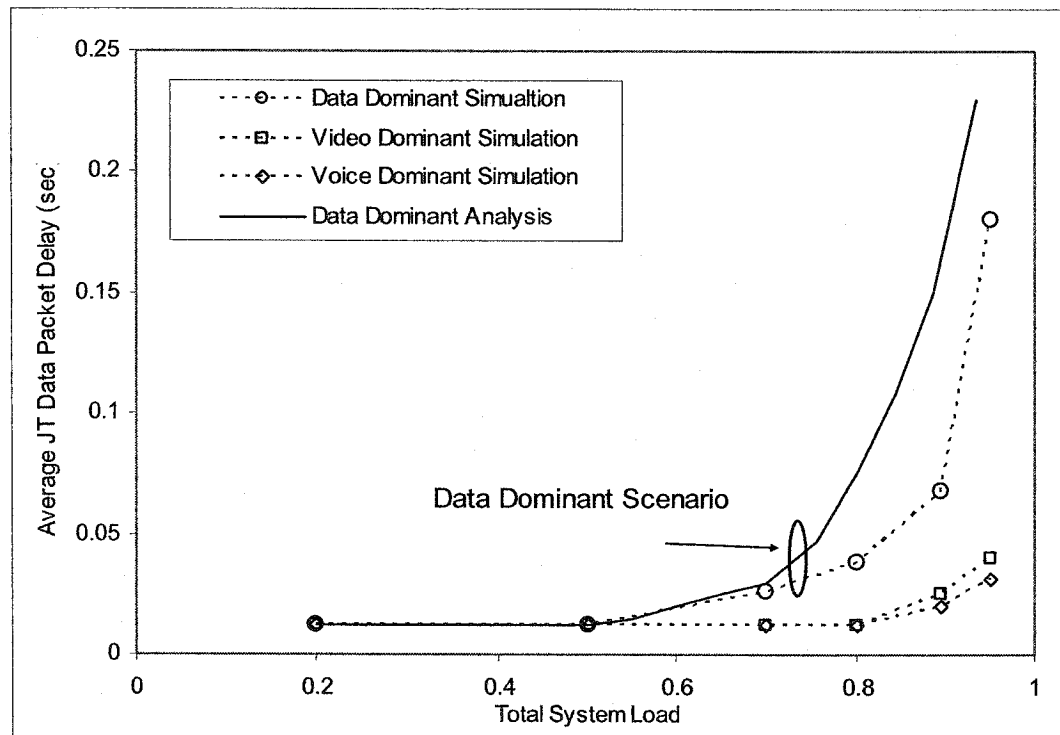


Figure 3-16: Performance of MF-TDMA CFDMA scheme

3.4 MC-TDMA with Random Slot Selection

3.4.1 System Description

This uplink access scheme does not follow any of the proposed standards and therefore can be categorized as a proprietary alternative. It is however in line with work that is being carried out in regards to the Universal Mobile Telecommunications System (UMTS) and

Wideband CDMA (W-CDMA) [80]. The main differences between this scheme and the MF-TDMA based scheme are in terms of:

- The complexity and operation of the scheduler
- How the earth stations are told which channels to use.

The value of spread spectrum as a basis for multiple access has been discussed in the previous Chapter.

The proposed channelization is based on MC-TDMA, and is shown in Figure 3-17. Each active earth station is assigned a unique spreading code¹⁰, which acts as a user ID. The set is denoted by $\{C_i\}$ $i=1,2,3,...,L_U$, where L_U is the total number of earth stations in a spot-beam. The number of earth stations that can be active at one time, is physically limited to L_R by the complexity of the resulting receiver at the gateway ($L_R \ll L_U$). To serve all L_U earth stations, the despreading codes at the gateway must be dynamically changed as stations initiate and terminate calls. One major assumption made is that L_R is greater than the maximum number of active earth stations.

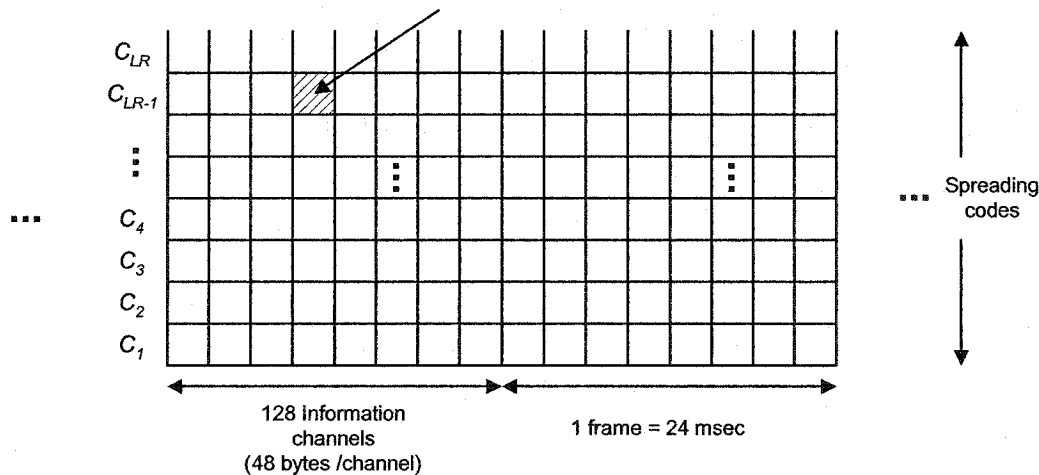


Figure 3-17: Earth Station Slot Selection

We have attempted to be fair in terms of the comparison with the MF-TDMA system. As a result, the earth stations still queue the traffic according to the 5 capacity types specified by DVB-

RCS (CRA, RBDC, VBDC, AVBDC, and FCA). Based on the discussion for MF-TDMA, one alternative for uplink access is to use a form of CFDAMA, and have the scheduler tell the earth stations which slots to use for transmission (they could then spread the information by using their User ID code). However, it should be clear that such a system would be equivalent to MF-TDMA, and many of the advantages of CDMA would be lost. Specifically, if the scheduler is to assign time slots to earth stations, then it is to do so on a frame basis. We would lose the advantage of a simple scheduler. In fact, the scheduler would still be required to make multiple passes over the scheduler queues, to generate a BTP, and to transmit the BTP to the earth stations. In addition, although the scheduler can still take advantage of some statistical multiplexing, it is no longer inherent. Statistical multiplexing refers to the ability of a MAC protocol to support more traffic than its peak would allow. Consequently, a random assignment strategy is much better suited for MC-TDMA.

One of the innovations of the proposed technique is to randomly spread out packets over the time-slots of a frame. Consequently, rather than having all stations transmitting packets in consecutive slots of a frame, these transmissions are spread over a longer time interval. This spreading acts to average out the number of stations transmitting in the same slot, thereby keeping multiple access interference low for all slots. The only major drawbacks of this approach are:

1. The random transmissions would result in increased jitter in the received stream. This is considered acceptable since the transmissions are spread over a frame ($T_F = 24$ msec),
2. The receiver is required to monitor multiple code channels in parallel. We assume that L_R (number of code channels) is large enough to deal with the maximum number of active earth stations. This approach would not be suitable for very low load terminals, as the maximum number of active earth stations could be too high.

¹⁰ The actual type of code is not considered. It is assumed that a sufficient number of good quality codes exist. Where necessary, it will be assumed that the codes are random.

Since the frame is slotted, each user transmits its packet by randomly selecting one of these slots. Figure 3-18 can be used to explain this in more detail. For the simple case shown, the frame is divided into 7 slots (1 is for reservations), and there are 7 active users (numbered 1 through 7). These active users have been given access to spreading codes C_1 to C_7 .

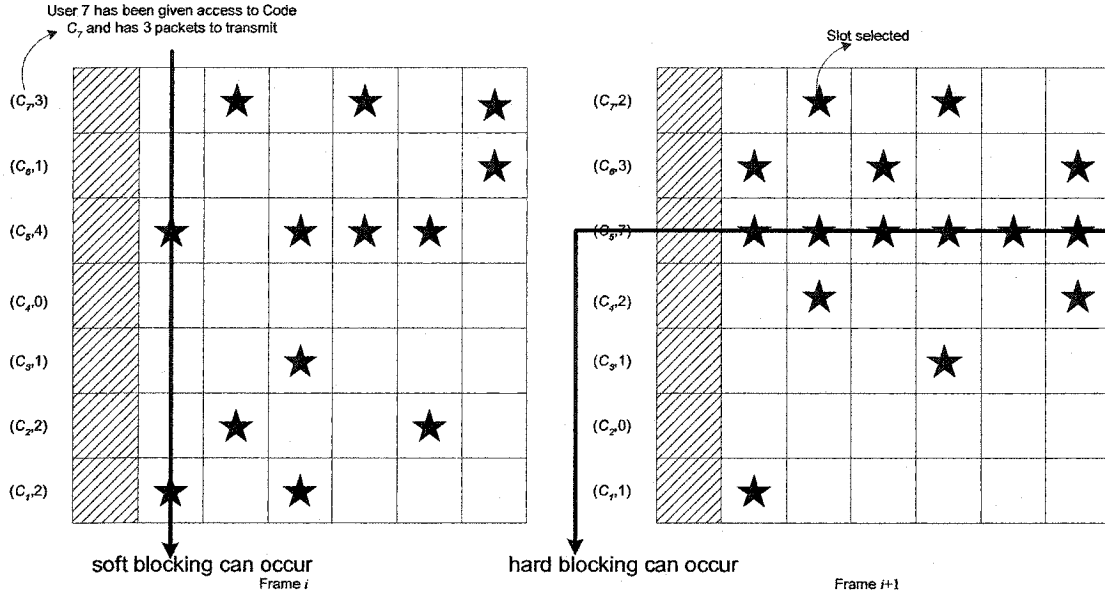


Figure 3-18: Hard and Soft Blocking

In a given slot, we have many users transmitting on different codes. These users act as interference to one another - multiple access interference ($I_{o,MAI}$). The number of packets that can be allowed to transmit simultaneously is limited by the allowable MAI, and the processing gain.

For our purposes, the frame is divided into 128 information slots (129 slots if we count the reservation slot). In each of these slots, we transmit an ATM type cell (or packet) of 53 bytes (48 byte payload and 5 byte header), yielding an instantaneous bit rate of:

$$R_b = 53 \frac{\text{bytes}}{\text{cell}} \cdot 8 \frac{\text{bits}}{\text{byte}} \cdot 129 \frac{\text{slots}}{\text{frame}} \cdot 1 \frac{\text{frame}}{24 \text{ msec}} = 2.279 \times 10^6 \text{ bps}$$

For the moment, let us assume that we spread over an available bandwidth of B_w . Using the results from [88], we see that the signal to multi-access interference ratio (SMAIR) per slot is given by:

$$\frac{E_b}{I_{0,MAI}} = \frac{\frac{E_b}{(N-1)P}}{(3B_w)} \quad (3-12)$$

where E_b = transmitted bit energy,

P = transmitted power = $E_b R_b$

B_w = Spread Bandwidth

3 -- term arising from a synchronous system

N = # of users sharing the bandwidth.

From the AT&T FCC filing [3], it is assumed that the maximum multiple-access interference is such that $E_b / I_{0,MAI} = 9dB$. Setting (3-12) to 9 dB, we get,

$$\frac{\frac{E_b}{(N_{\max}-1)E_b R_b}}{3B_w} = 10^{0.9} \quad (3-13)$$

For a spread bandwidth of B_w , the processing gain is

$$P_G = \frac{B_w}{2.279 \times 10^6 \text{ bps}}$$

Using equation (3-13) we can find an upper limit on the number of users permitted per slot, in order to ensure that the signal-to-multiple access interference ratio is larger than 9 dB. That is,

$$N_{\max} = \left\lceil \frac{3P_G}{10^{0.9}} + 1 \right\rceil \quad (3-14)$$

For the voice, video, and data dominant cases, there are a maximum of 85 active earth stations. Each of these generates, on average, 48 cells/frame. The users transmit these cells by

selecting a slot randomly from among the 128 available. The same occurs for every other user. If enough users transmit in the same slot, then a condition known as soft blocking occurs. Here, the packets are not lost, but rather, their bit error performance is degraded due to the increase in interference.

As an approximation, we can calculate the probability of soft blocking if we assume that each user makes Bernoulli attempts with probability $p = 48/128$ for every slot. In making this assumption, we remove the correlation in the arrival process. Despite this, the procedure provides insight into the distribution of the number of earth stations transmitting in a particular slot. If there are K active users, each generating traffic in a slot with probability p , then the number of users selecting a particular slot (denoted by random variable X) is binomial distributed. That is,

Let X = random variable denoting of the # of users selecting same slot

$$p = \text{probability that a slot is occupied by a packet} = \frac{48}{128}$$

$$\begin{aligned} \therefore \Pr[k \text{ users select same slot}] &= \Pr[X = k] \\ &\cong \binom{85}{k} p^k (1-p)^{85-k} \end{aligned} \quad (3-15)$$

$$\begin{aligned} \therefore \Pr[\text{soft blocking}] &\cong \Pr[X > N_{\max}] \\ &= \sum_{k=N_{\max}+1}^{85} \binom{85}{k} p^k (1-p)^{85-k} \end{aligned} \quad (3-16)$$

Figure 3-19 shows the binomial distribution of the number of transmitted cells in a slot (using approximation (3-15)). Note that in most slots, 16 to 50 cells are transmitted. Therefore, if N_{\max} is of the order of 45 (this corresponds to a processing gain of 117), $\Pr[\text{soft blocking}]$ is expected to be very small, which agrees with the simulation results of section 3.4.3.

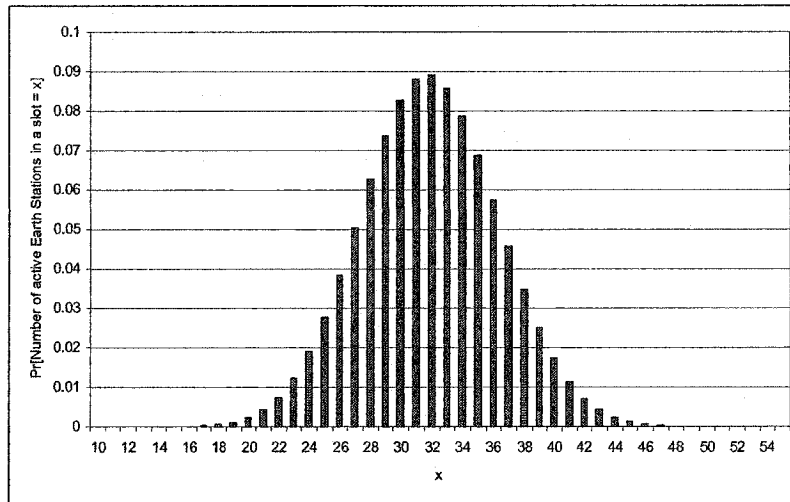


Figure 3-19: Distribution of Number of Transmitted Cells in a Time-slot

For a given user spreading code C_i , the number of cells transmitted per frame is limited by the number of slots in the frame – M (M is 6 for the example of Figure 3-18 and 128 for the parameters for our satellite environment). If a user has more packets than number of slots, then one of the following occurs:

- a) If the packets are all real-time, then up to the maximum is transmitted (M), and the remaining are lost (this is hard blocking).
- b) If the packets are a mixture of real-time and jitter-tolerant, then the real-time packets are transmitted (again up to the maximum allowed), with any excess being lost. Any unused slots are used by data packets. Those data packets that cannot be transmitted are queued. Notice that if there are unused slots for spreading code C_i , this means that the data buffer is empty for user i .

3.4.2 Types of Flow Control

From the above discussion, one should notice that at the earth stations, the MC-TDMA uplink access scheme makes no distinction between voice, video, and data traffic. Therefore, as long as slots are available, data packets will be transmitted. As one might expect, simulations confirm that the data delay is very low. However, the transmitted data packets act as interference

to the real-time packets. Maybe, it would be better to queue some of the data, in an attempt to reduce this multiple access interference? To achieve this, we use a method of flow control. To keep the advantage of a simple scheduler, we proposed a distributed algorithm that is implemented at the earth stations, and that is based on parameters agreed upon at call set-up. The basic idea can be explained by referring to Figure 3-20. An active earth station has an average cell generation rate, which we denote as AVE_{CGR} . The number of slots per frame is fixed, and depends on the maximum rate the provider wishes to allocate per earth station (M). For our case.

$$M = 128 \text{ channels/frame}$$

$$AVE_{CGR} = 48 \text{ packets/frame (for voice, video, and data dominant cases)}$$

Furthermore, consider a tagged earth station k , and for this earth station let

$$RT(n) = \text{number of RT packets in queue of earth station } k \text{ at start of frame } n$$

$$D(n) = \text{number of JT packets in queue of earth station } k \text{ at start of frame } n$$


In frame n , earth station k is allowed to transmit $\max(0, AVE_{CGR} - RT(n))$ data packets. It does so by randomly selecting from the unused slots. The remaining $D(n) - \max(0, AVE_{CGR} - RT(n))$ data packets contend for the remaining slots by making Bernoulli trials. A data packet gets a free slot with probability p_{fc} . This probability is decided on at call set-up. Figure 3-20 shows a case with $M = 48$, $AVE_{CGR} = 20$, $D(n) = 5$, $RT(n) = 18$ and with 3 different values of p_{fc} (1, 0, 0.01). The tagged earth station (station k) has been assigned spreading code C_k . For each frame, 48 channels are available. The x's mark the transmissions of the real-time packets (all 18 are transmitted). The o's mark the transmissions of jitter-tolerant packets. Note that when $p_{fc} = 1$, all data packets are transmitted (this is equivalent to having no flow control). In contrast, when $p_{fc} = 0$, a data packet is only transmitted if an earth station has transmitted less than its average. Any remaining data packets are queued, and contend for capacity in the next frame.

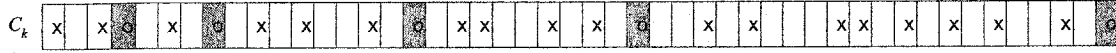
Earth Station k

Consider following example

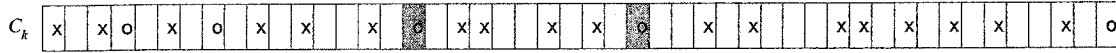
- $M = 48$ channels
- $AVE_{CGR} = 20$ channels/frame
- $D(n) = 5$ packets
- $RT(n) = 18$ packets

x : real-time packet transmission
o : jitter-tolerant packet transmission
 C_k : spreading code for earth station k

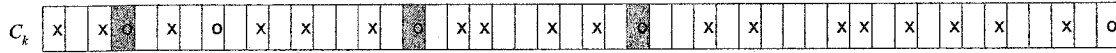
 : Bernoulli trial succeeds



a) $p_{fc} = 1$
no data packets are queued



b) $p_{fc} = 0$
3 data packets are queued



c) $p_{fc} = 0.01$
2 data packets are queued

Figure 3-20: Examples of Flow Control Procedure
(during frame n)

3.4.3 Performance Results

Case 1: Delay vs. load for different traffic mixes

In this simulation, the objective was to see the performance of the MC-TDMA scheme when $p_{fc} = 1$. To this end, we simulated the voice and data dominant cases, and the results are shown in Figure 3-21 and Figure 3-22. As expected, with $p_{fc} = 1$, the data delay is equal to half the frame size for all loads and for both traffic mixes. This was expected for the $p_{fc} = 1$ case, since all jitter-tolerant packets pass the Bernoulli trial and are transmitted. No data packets are queued.

The cost of this excellent delay performance is an increase in soft blocking. Recall that a packet is soft blocked when the number of earth stations transmitting in a particular slot exceeds N_{max} defined in equation (3-14). This maximum depends on the processing gain, and as a result, on the available bandwidth. In Figure 3-22, the $\text{Pr}[\text{soft blocking}]$ vs. Processing Gain is plotted, for the voice dominant and data dominant scenarios (four system load points are considered:

20%, 50%, 80%, 95%). Also included on these curves, is the binomial approximation to soft blocking, as was discussed in section 4.1.2. Three conclusions are obvious from these curves:

1. As the load increases, the soft blocking increases. This is expected, since the number of active users increases with load.
2. The processing gain can be used to effectively reduce $\text{Pr}[\text{soft blocking}]$.
3. The $\text{Pr}[\text{soft blocking}]$ is underestimated by using the binomial approximation (this is equivalent to assuming independent traffic). As a simple comparison, consider the 80% load case, with a processing gain of 100. The binomial approximation underestimates the soft blocking of the data dominant scenario by 1 order of magnitude (10^{-4} as opposed to 10^{-3})

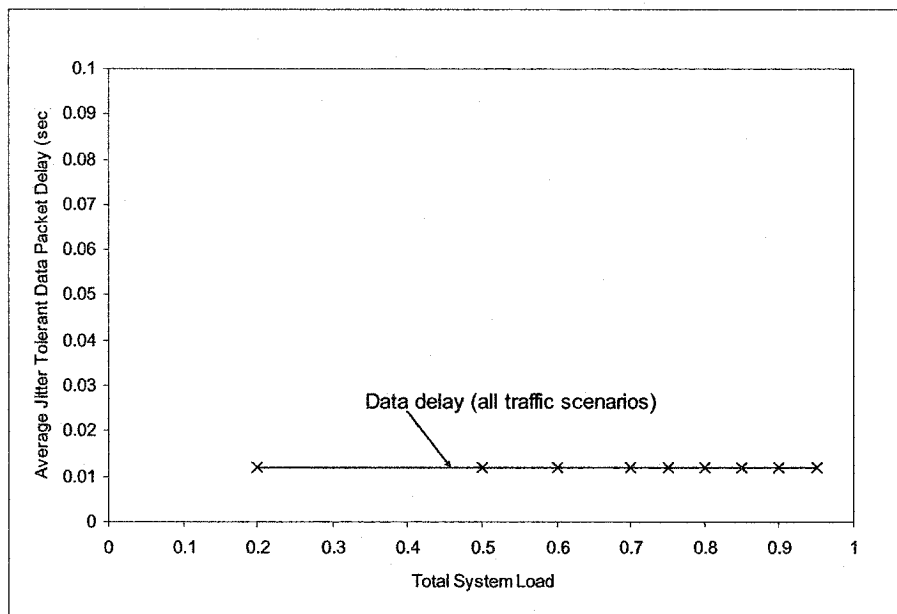


Figure 3-21: Data Delay vs Load

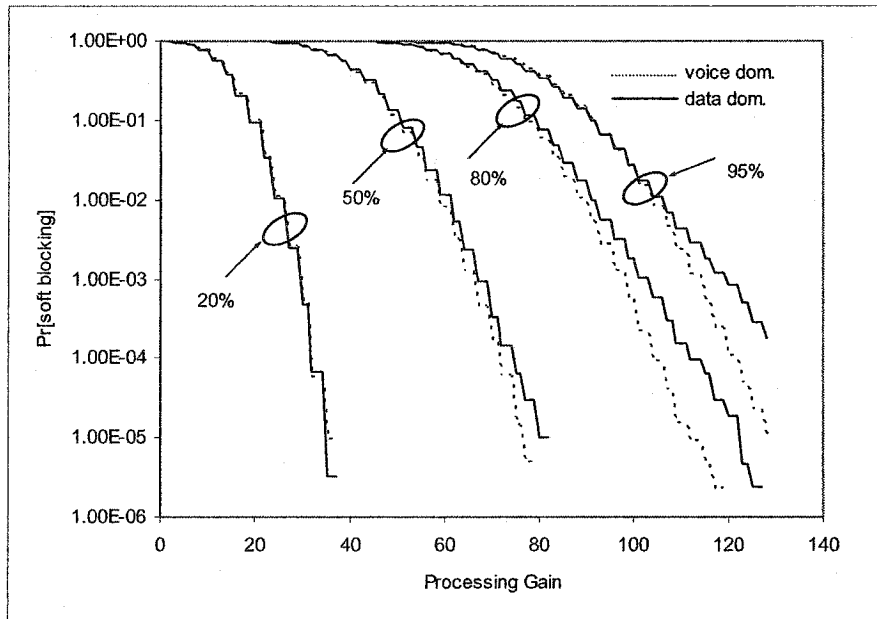


Figure 3-22: Effect of Processing Gain on Soft Blocking

Case 3: Effect of Flow Control on Performance

As was seen in the last section, the data delay is optimum for the case when $p_{fc} = 1$. However, the system does suffer from soft blocking. Can we trade some of the delay performance for a reduction in soft blocking? To investigate this, we used an equal dominant scenario, and found the delay and soft blocking for various p_{fc} for a 50% load.

Figure 3-23 shows the simulation results for delay and $\text{Pr}[\text{soft blocking}]$ vs. p_{fc} . To determine soft blocking, we considered no system imperfections, and set the processing gain to 75. It is clear from the figure that small variations in p_{fc} have a marked effect on performance. In fact, we can see that the delay changes by over an order of magnitude (0.01 sec to 3 sec) as p_{fc} decreases from 0.02 to 0. At the same time however, the probability of soft blocking is reduced from 2×10^{-4} to 1×10^{-4} . Although the reduction in soft blocking seems minor, we must remember

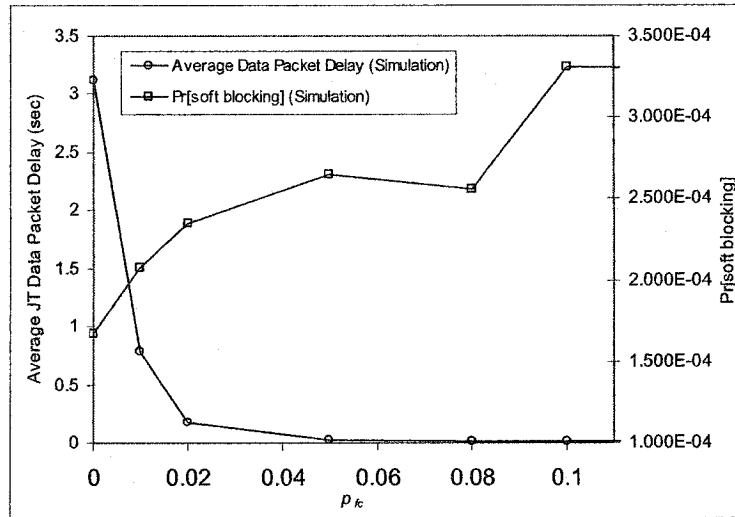


Figure 3-23: Data Delay & Pr[soft blocking] vs p_k

that in the equal dominant scenario with 50% load, the percentage of data is very small (16.5%). Therefore, for this case, the amount of flow control is limited. Despite this, the trend is still quite obvious - we can trade data delay for soft blocking. The trade-off, however, is more equal for higher loads and for scenarios with higher percentages of data.

Chapter 4

On-Board Processing System

4.1 Introduction

The basic model for an OBP multibeam satellite system is shown in Figure 4-1. The main difference between this system and the one studied in Chapter 3 is that the satellite has dedicated receivers for each spot beam, an on-board scheduler, and on-board switch. This results in three main advantages: for the physical layer, for the MAC portion of the data link layer, and for layer-2 switching.

Physical layer: As mentioned in Chapter 1, an OBP system has the advantage of splitting the uplink and downlink paths and therefore allows for independent optimization of both. Since noise on the uplink is not propagated to the downlink transmission, for a given transmitted power, the bit error rate performance is improved.

Data Link layer: The location of the scheduler has a considerable impact on the performance of reservation-based MAC schemes, where an earth station has to wait for channel assignments before it can transmit any packets. The latency is 2 RTTs for a BP system and only 1 RTT for an OBP system. For a satellite in GEO orbit having a RTT of 0.27 seconds, this savings can be considerable.

Layer 2 Switching: Since the OBP system has an on-board baseband switch, it is possible for traffic to be routed from one uplink spot-beam to another downlink spot-beam. In order to do this, the OBP satellite must look inside the received packet to find the downlink destination. This information can be either in the payload of the packet (within an MPLS or ATM header for instance) or it can be specifically added by the satellite MAC layer at the earth station. In this type of configuration, the communication between two earth stations is possible without the intervention of the terrestrial network. The system can still provide

access, but the provider has much more flexibility in offering direct services to customers (such as LAN interconnection and the like). The switched packets are stored in the downlink queues (one for each spot-beam) while waiting to be transmitted. These packets are then transmitted using the available downlink capacity.

The main issues to address in this type of network were highlighted in Chapters 1 and 2, and are repeated here. First, there must be a method to efficiently share the uplink resource between the earth stations. Whereas this can be done independently from beam to beam in a BP system, this would not be the best approach for an OBP system. Clearly, the satellite switch acts as a multiplexing point for traffic destined to a particular downlink destination. If the uplink capacities are optimized independently, it is possible that the downlink queues become congested. Although this may not be a major problem for a terrestrial switch, which typically has enough space in its output queues to handle the occasional queue build-up, the same can not be said for a switch in GEO orbit. Memory is a very expensive resource, and usually very limited [74]. If there is congestion, the satellite will have to drop packets. This will result in an increase in the real-time loss probability, and an increase in the average data delay (since the lost packets may require retransmission). To avoid/prevent this congestion, a congestion control (CC) scheme is required.

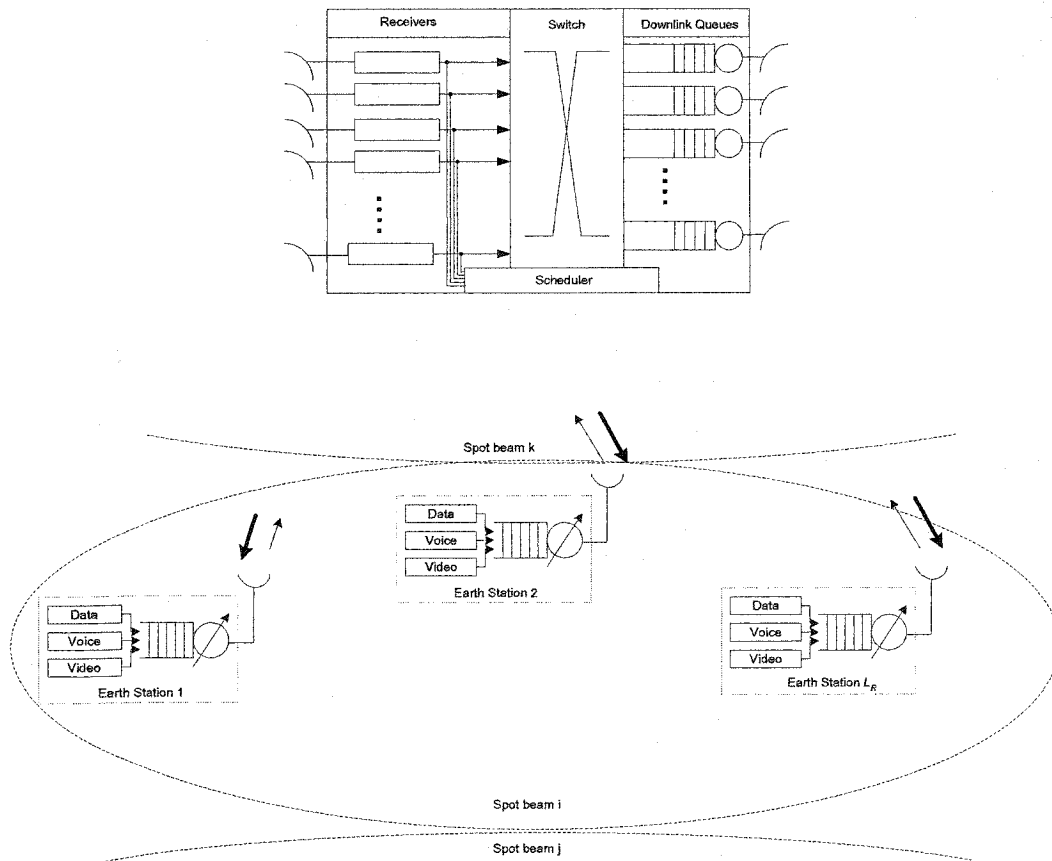


Figure 4-1: Model of On-Board Processing System

4.2 How to Deal with Congestion

Any CC scheme must take into account the characteristics of the satellite network, namely the long round trip propagation delay and the limited satellite resources (uplink bandwidth, downlink bandwidth, on-board buffer size and computational power). To deal with the limited uplink bandwidth, a MAC protocol is required. The issue then becomes how to coordinate the congestion control and MAC protocols. There are three basic approaches, which are discussed below.

CC overrides MAC: This approach is definitely the simplest to implement. Both algorithms are treated independently, with the MAC maximizing the uplink utility and the CC minimizing

the on-board congestion. As the MAC is only concerned with the uplink, it is not aware of the congestion occurring in the downlinks, and can inadvertently contribute to this congestion. It is then assumed that the CC protocol would override the MAC assignments by restricting the terminals in the use of their assigned capacity. Clearly this approach would lead to inefficiency, as the unused assignments would be lost. The next two approaches offer remedies to this situation.

CC modulates MAC requests: Essentially, the congestion information is used to modulate terminal requests. The amount of capacity a terminal can request is dictated by the congestion of the downlink to which the traffic is destined. As a result, a terminal will always have the ability to use its assigned capacity, which in turn will allow maximum utilization of the uplink capacity. Two schemes that fall into this category have appeared in the literature: one based on ERICA (Explicit Rate Indication for Congestion Avoidance) and the second based on BRCA (Broadcast Rate Congestion Avoidance) [75][76]. The former is a modification of an algorithm proposed by the ATM forum, while the latter is a proprietary technique developed by EMS Technologies.

CC controls MAC assignments: This last scheme is the most complicated. The basic concept is to have the uplink scheduler (which implements the MAC protocol) use the congestion control information when making its assignments. A terminal's assigned capacity is then based on its requests and on the status of the downlink queue where traffic is destined. The major advantage of this scheme compared to the second approach is that the time to react to congestion is halved, so that congestion can be cleared much more expediently. In other words, this scheme aims to integrate the congestion control and MAC protocol. A variation of this idea, known as Weighted Fair BoD (WFBOD) is proposed in [77]. The work focuses on the entire resource management issue (at the terminals, on the uplink, on the downlink, and at the OBP switch downlink queues). The solutions are quite elegant, but the optimization to determine the assignments can be quite complex. Most of this optimization is performed at the scheduler.

Owing to the poor performance of scheme 1, it is not considered a viable approach for a multimedia satellite network. In this chapter, we compare one of the techniques in the literature (the one based on ERICA) with 2 techniques that fall into the 3rd category. Of the two proposed techniques, one relies on MF-TDMA and CFDAMA as an accessing scheme, while the other relies on MC-TDMA with random slot selection. In comparison to the schemes proposed in [77], our proposed solutions are rather simple but do require iterative processing. Note that all schemes can be classified under the general category of preventative congestion control. As will be shown, the main advantage of the proposed schemes is that they react quicker to changing queue levels.

4.3 Proposed Solutions

4.3.1 MF-TDMA with CFDAMA

The basic protocol has been described in Section 3.3. In this section, the required changes are highlighted in order to deal with the possible congestion at the output of the on-board switch. The main idea is to assign channels to earth stations in such a way that they do not lead to possible congestion. To accomplish this, the scheduler needs to have an idea of future traffic arrival patterns to the switch. Rather than use a form of prediction to determine future arrival pattern, we propose to use the assignment information as a measure of the arrival in one RTT. If one ignores the free assignment capacity, then all assigned channels should be used by earth stations, and this traffic will arrive at the earth station in one RTT. Figure 4-2 provides an example of this for a single spot beam and a RTT equal to 11 frames. If there is an overload condition at one of the satellite downlink queues during frame $i+23$, this will be as a result of traffic channels assigned in frame i . Say the request information sent from earth stations includes the downlink to which the traffic is destined. So for example, earth station r in uplink spot beam s , would make a request for 3 channels that would be destined to downlink spot beam t (r,s,t). During frame i , the scheduler would perform its scheduling algorithm by looking over the requests (r,s,t), and by making

assignments for each terminal r . After the assignment process, the scheduler has an assignment matrix, listing channels assigned per uplink spot beam and downlink spot beam (a_{ij})(see Table 4-1). Based on the column totals, the scheduler can already estimate the arrivals to each of its downlinks. If the scheduler sees that a downlink will be overloaded and hence cause a downlink queue buildup, it can use this information to revise its assignment. The algorithm is explained in detail in the next section.

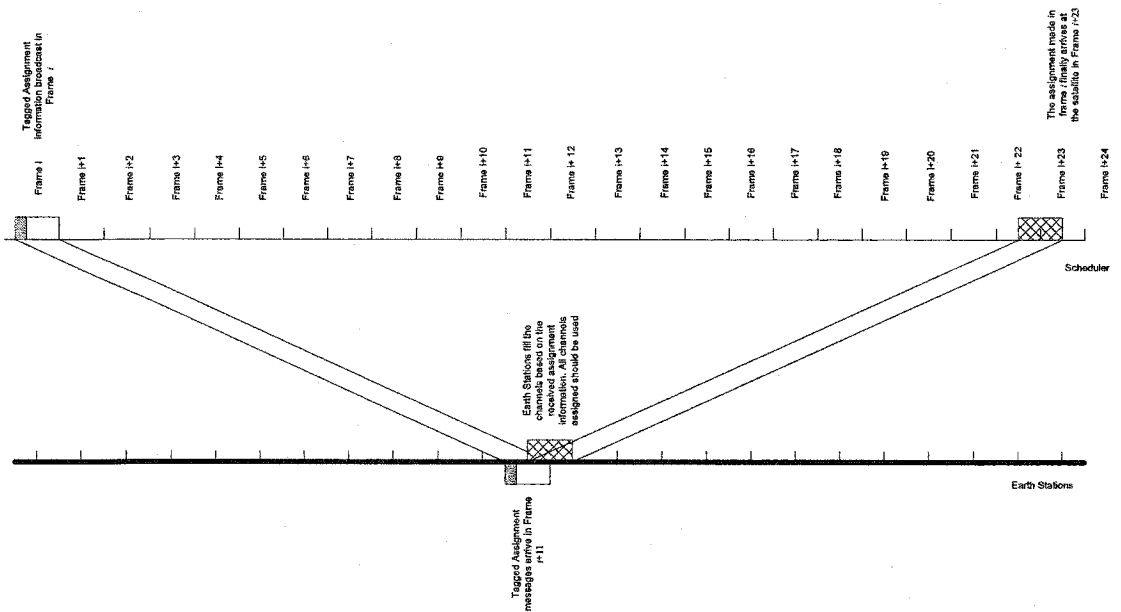


Figure 4-2: Arrivals to Satellite in Response to Reservations

Table 4-1: Assignment Matrix (Information per uplink/downlink beam)

Downlink \ Uplink	Spot Beam 1	Spot Beam 2	Spot Beam 3	Spot Beam 4	Total
Spot Beam 1	a_{11}	a_{12}	a_{13}	a_{14}	$a_{11} + a_{12} + a_{13} + a_{14}$
Spot Beam 2	a_{21}	a_{22}	a_{23}	a_{24}	$a_{21} + a_{22} + a_{23} + a_{24}$
Spot Beam 3	a_{31}	a_{32}	a_{33}	a_{34}	$a_{31} + a_{32} + a_{33} + a_{34}$
Spot Beam 4	a_{41}	a_{42}	a_{43}	a_{44}	$a_{41} + a_{42} + a_{43} + a_{44}$
Total	$a_{11} + a_{21} + a_{31} + a_{41}$	$a_{12} + a_{22} + a_{32} + a_{42}$	$a_{13} + a_{23} + a_{33} + a_{43}$	$a_{14} + a_{24} + a_{34} + a_{44}$	

4.3.1.1 Algorithm Details

The flowcharts shown in Figure 4-3 and Figure 4-4 describe the algorithm operation at both the earth station and the scheduler. At the earth station, the only major difference between a BP and OBP terminal is that in the latter, the packets must be differentiated by type (or capacity class VBDC, RBDC, CRA), and also by the packet's downlink beam destination. As a result, the earth station has a set of queues for each (type, downlink beam) pair. This is specifically shown in Figure 4-3. As can be seen, requests are made for each earth station by type and downlink beam.

At the scheduler, the algorithm is a small modification to the simple scheduler used in the BP system. The main idea is to perform the assignment process as usual (that is per uplink beam), and then create an assignment matrix (as in Table 4-1). Based on this matrix, assignments can be shuffled out of those downlinks that will be overloaded and into those that can handle the extra load. Care must be taken to be fair to all earth stations in all beams. The matrix must be updated after each assignment shuffle. The details are shown in Figure 4-4. The lightly shaded blocks denote the procedures that may lead to fairness concerns. For this work, we tried to achieve fairness by using a round-robin approach. To do this, the scheduler keeps track of the last station in each uplink to lose a shuffled channel, and the last station in each (uplink,downlink) to receive a shuffled channel. These are then searched in sequence as the shuffling takes place.

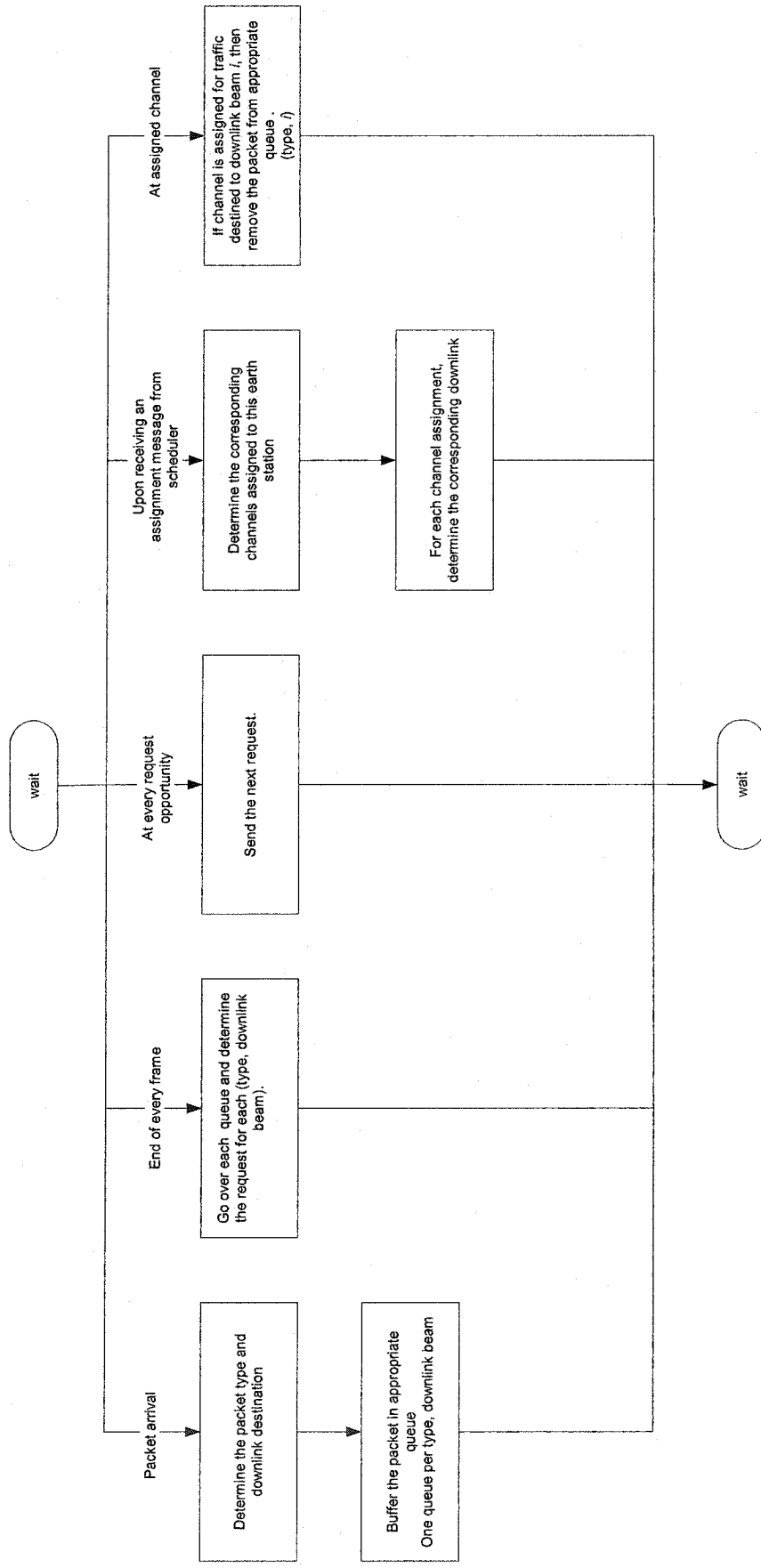


Figure 4-3: Procedure at Earth Stations

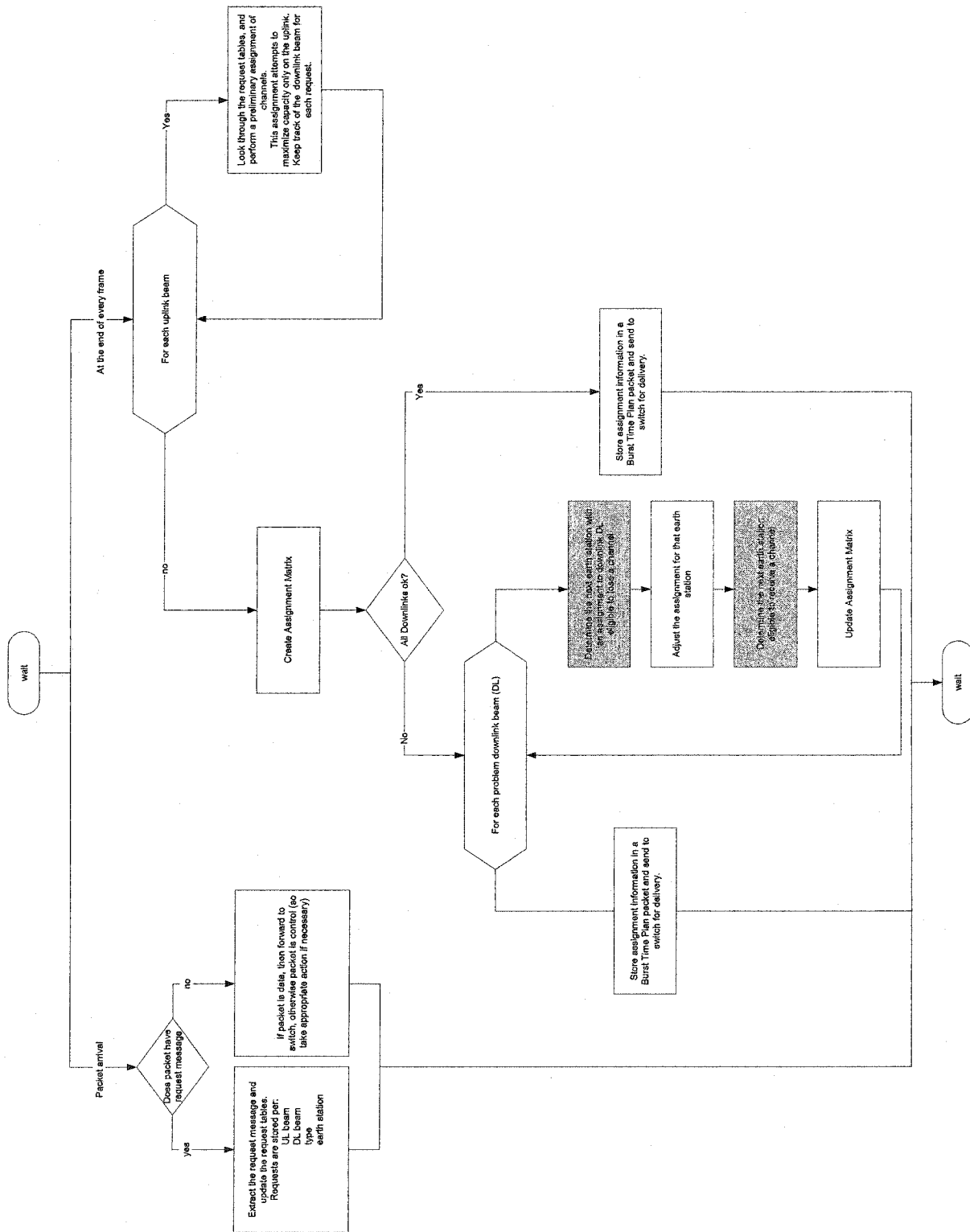


Figure 4-4: Procedure at Scheduler

4.3.1.2 Performance Results

In order to gauge the performance of the protocol, an end-to-end analysis would be required. This would involve the uplink access and the scheduling, the on-board switching, and the on-board queuing. As a result of the number of variables to consider, it would be difficult to determine a suitable analytical model of the network. An attempt was made in [79] to provide an end-to-end performance evaluation, but in order to carry out the analysis, the authors amalgamated all earth stations in a spot beam into one “super” station. Although the analysis provided a method to determine results for end-to-end delay, the procedure did not take into account the dynamics of the multiple access protocol at the earth station level. As a result of these limitations, performance results for our system will only be obtained through simulation.

In order to provide a comparison with the ERICA algorithm proposed in [76], we will assume an identical simulation set-up. The ERICA algorithm uses explicit control messages to control the rate of channel requests from an earth station. The control messages are regularly transmitted by the earth station, multiplexed within the information stream. At the scheduler, the control messages are updated with congestion information. This congestion information is in the form of the fair share of the resources that can be allocated to a particular earth station connection. The scheduler then sends the updated control message back on the downlink, where it is interpreted by the earth station that originally sent it. The earth station uses the congestion information contained within the message to modulate its capacity requests. As a result, capacity requests are based on congestion information that was updated by the scheduler $\frac{1}{2}$ a round trip time earlier. As a first step, the authors of [76] wanted to load the network in such a way as to guarantee congestion on one or more of the downlinks. This was achieved by having heterogeneous earth stations and by defining a loading scenario.

In a typical situation, one can argue that on average the traffic admitted into the satellite system would not overload any of the uplinks or downlinks. However, the traffic to be carried by the satellite network is bursty in nature and exhibits a high degree of variability. In addition, it is subject to multiplexing on-board the satellite. Situations can therefore arise when some of the downlinks are overloaded.

In our simulations, in order to trigger the congestion control mechanisms, we deliberately created congestion situations on some downlinks. The model used in simulations assumes a four beam network (beam 0 to 3). The loading on all uplinks was considered the same (and equal to 90%), but on downlinks the total traffic was distributed in an unbalanced way: 120%, 100%, 80% and 60% on downlinks 0 to 3, respectively. This means that downlink port 0 is overloaded (by 20%), downlink port 1 is barely overloaded, and downlink ports 2 and 3 are underloaded. This loading scenario is depicted in Figure 4-5. In this figure, the left side denotes the transmitting earth stations, while the right side denotes the receiving earth stations. There are 32 active earth stations per beam, and these are arranged in four groups per port (shown by the colored circles). Stations in a group transmit to the same downlink beam. The links between groups show the actual traffic distribution to a destination beam. As an example, for the loading scenario depicted, group 3 stations in beam 0 transmit to group 3 stations in beam 2. In addition, the number beside each of the groups denotes the load generated by the station terminals of that group. All stations in one group have been assigned the same load, but the load is different for stations in different groups (stations are heterogeneous).

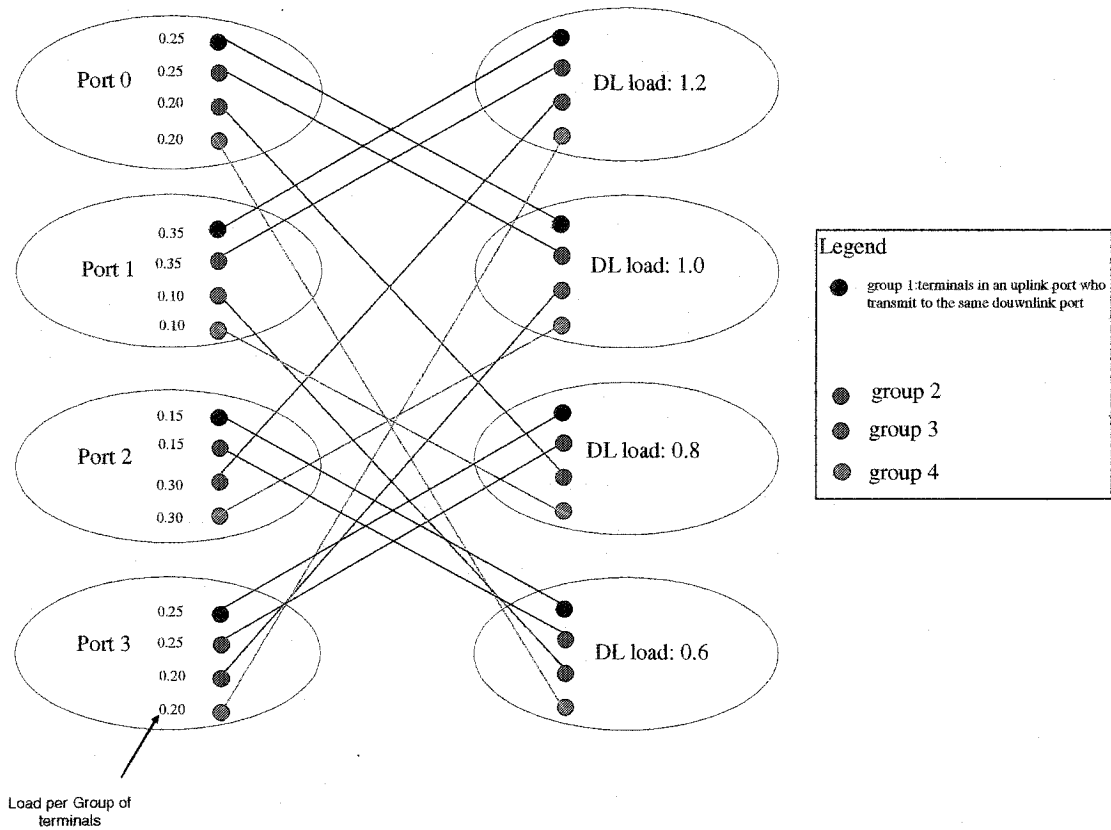


Figure 4-5: Loading Scenario for 4-Beam Network

Note:

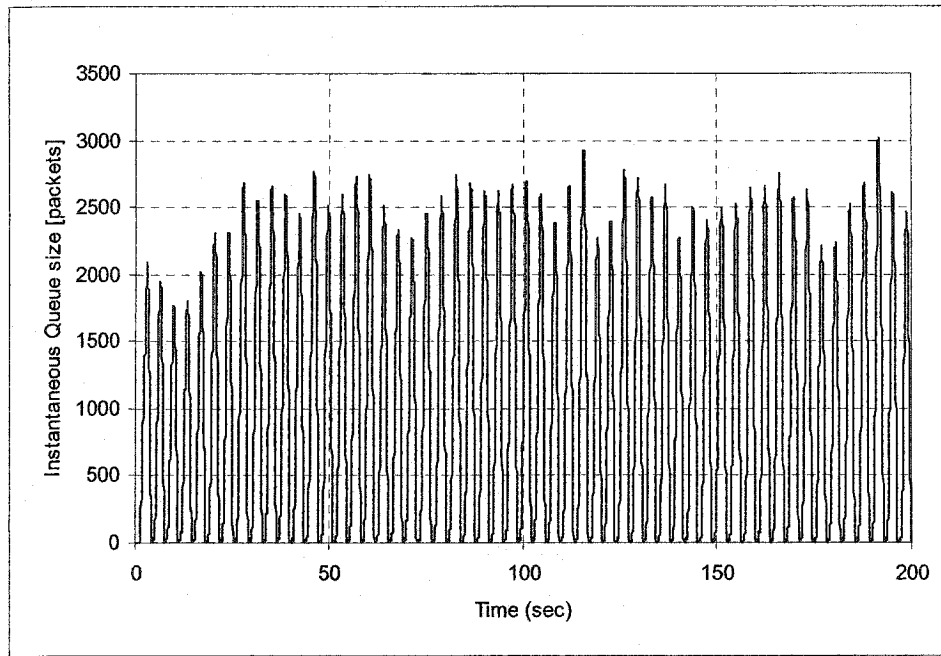
- each uplink beam is loaded to 90% (this means that stations within that beam generate (on average) 90% of the total uplink beam capacity)
- Each beam has 4 groups of earth stations
- Terminals in one group send traffic to the same downlink beam
- Load per group varies as indicated in the figure
- All stations in one group generate the same load

To study the effect on congestion control, we only loaded the network with jitter-tolerant traffic. Real-time traffic has priority both at the earth stations and at the satellite downlink queues. If capacity is not available in either the uplink or downlink, the real-time traffic will be discarded – it will not directly lead to congestion. It can however indirectly lead to congestion. In fact, the

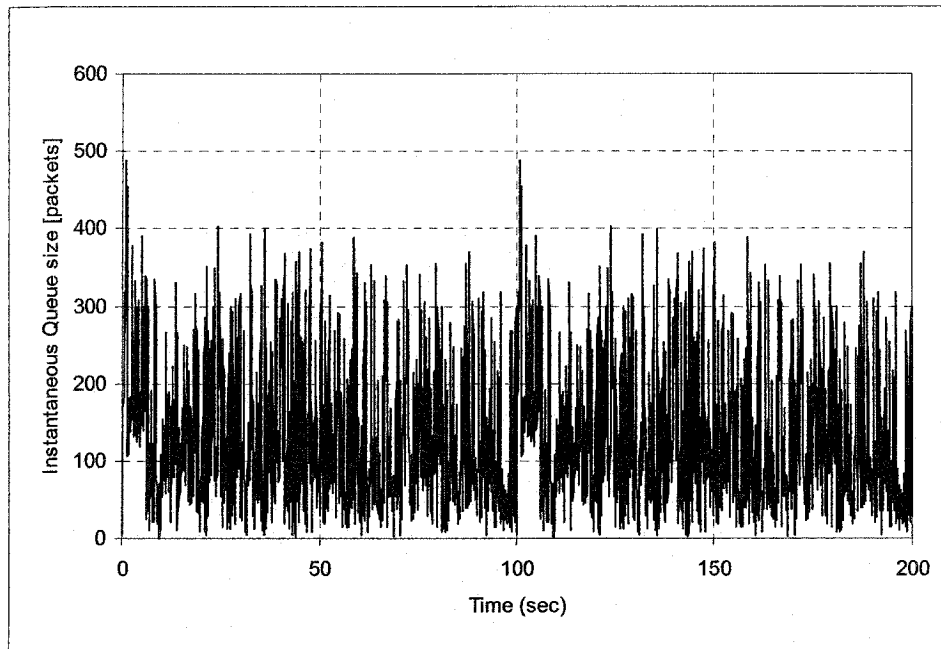
real-time traffic acts as background traffic and will consume capacity that would otherwise have been available to the jitter-tolerant traffic.

In all simulations, we assumed a capacity of 128 channels per uplink and downlink. Note that this is a 32 times reduction in capacity as compared to the simulation cases presented in Chapter 3. This was necessary in order to allow reasonable simulation run times. On the uplink the channels were divided between two carriers (there were 64 time-slots per carrier). With every channel supporting an ATM-type cell (48 byte payload), this corresponded to a maximum station burst-rate of 1.024 Mbps.

The results for the ERICA scheme and the proposed scheme (will be referred to as CC-MAC in figures) are shown in Figure 4-6 to Figure 4-8. Figure 4-6 (a and b) show the results for the instantaneous on-board queue size at the downlink queue for beam 0. Remember that the loading scenario continually tries to drive this queue further and further into congestion and as a result we expect the queue size to grow very large.

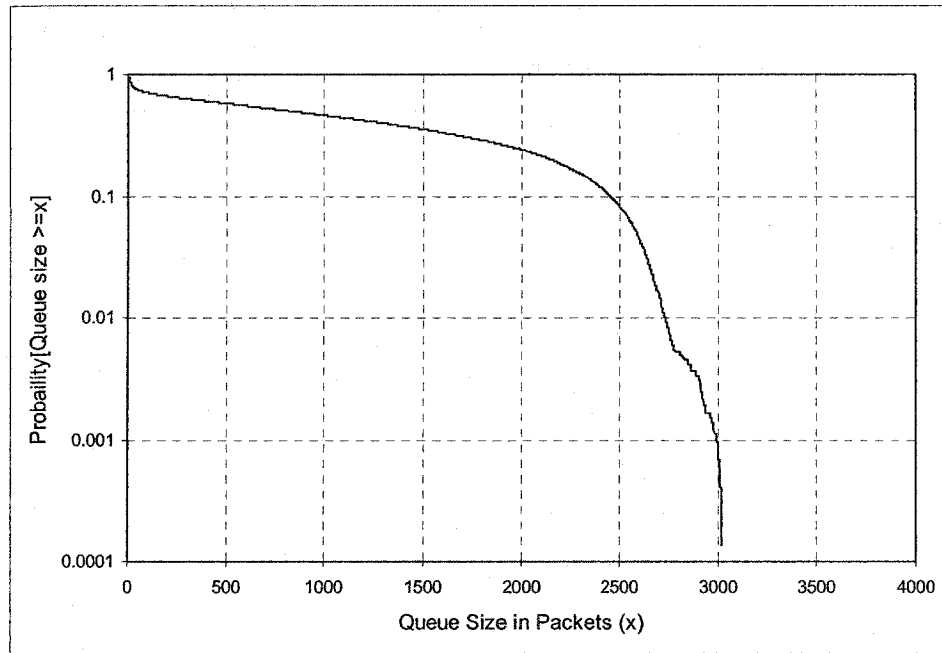


a) Instantaneous Queue: Using ERICA

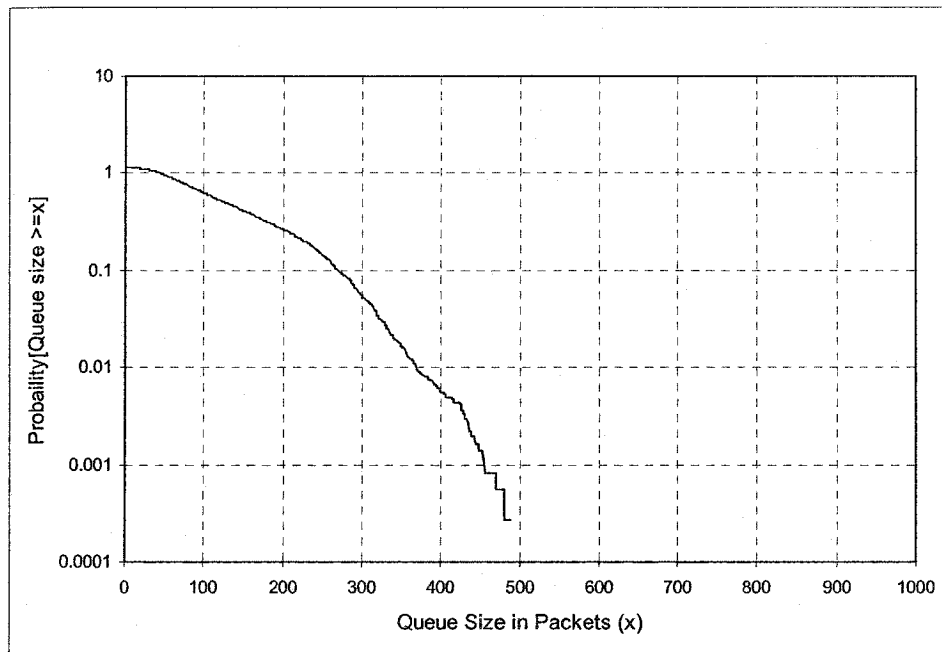


b) Instantaneous Queue Size: Using CC-MAC

Figure 4-6: Instantaneous DL Queue Size Comparison: DL Beam 0



a) Queue Survivor Function: Using ERICA



b) Queue Survivor Function: Using CC-MAC

Figure 4-7: Queue Survivor Function Comparison: DL Beam 0

Figure 4-7 shows the same information, but in terms of queue survivor function – that is, the probability that the queue size exceeds some size (X). In some papers, it is argued that the cell loss probability (CLP) for a fixed buffer of size X , can be estimated from the $\Pr[\text{Queue Size} > X]$. The sets of figures clearly show that both techniques can be used to limit on-board congestion. However, the CC-MAC protocol does a much better job – in fact it reacts faster to possible congestion, and keeps the queue sizes much lower. Note that for the ERICA protocol, the on-board queue size seems to oscillate – this is characteristic of the protocol, since it relies on congestion notification packets to transfer back and forth between the network. These packets cannot be too frequent, because they will consume the available uplink capacity. In addition, if they are too few, then the control will be sluggish (as is evidenced from the oscillatory behavior shown). Recently a technique known as BRCA has been proposed. Simulation results for this scheme show that its performance is many orders of magnitude better than ERICA. Whether this protocol is better than CC-MAC is an open question, and requires further study.

One last point should be made about the simulations. Clearly those stations trying to send traffic to downlink beam 0 will be at a definite disadvantage. Since the port is continually overloaded, the CC-MAC algorithm will restrict channels assigned to these stations. Since the traffic is still being generated, one would expect that the terminal queues would tend to grow indefinitely. Figure 4-10 shows the simulation results for a tagged earth station that generates jitter tolerant traffic to downlink beam 0. As expected, the queue size grows without bound.

Although the algorithm is iterative, the simulation settled to the final capacity solution after a few iterations. However in a practical implementation, the iterative process may take a long time, and it may be necessary to limit the number of assignments that are shuffled. If not all assignments leading to congestion are shuffled, this will obviously impact the performance of the algorithm. Future work on this protocol should definitely consider the impact of this limit.

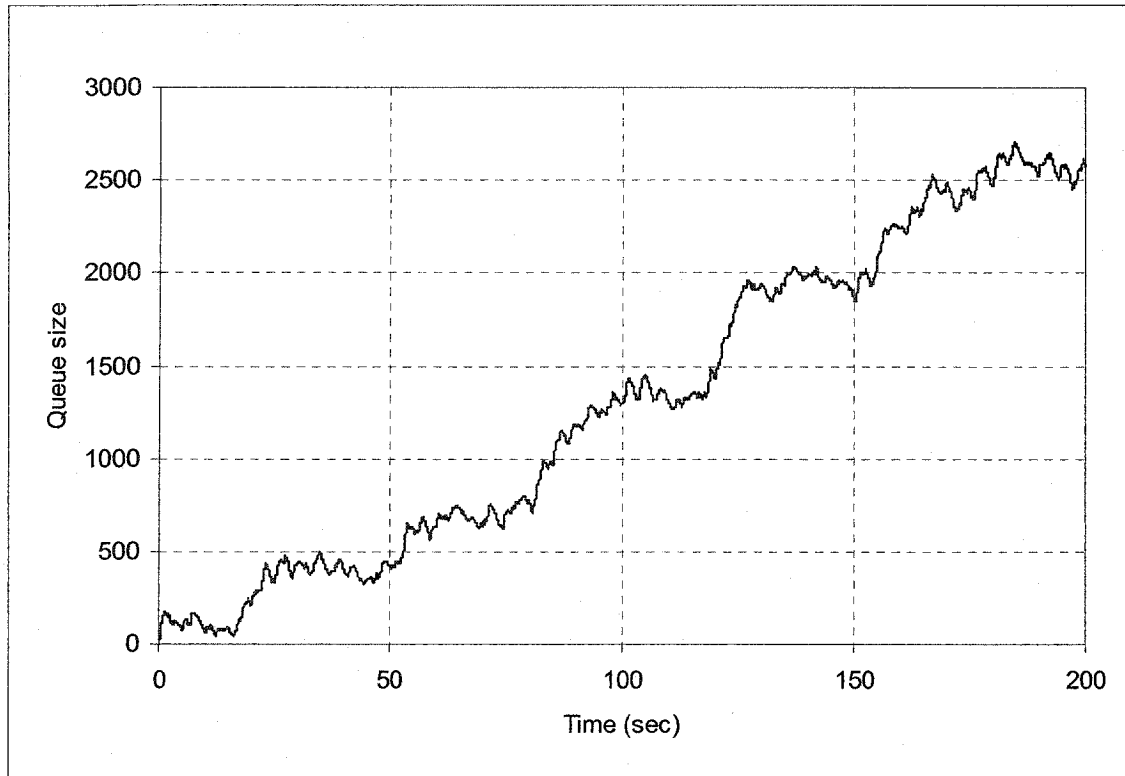


Figure 4-8: Instantaneous Earth Station Queue Size (for earth station sending to DL Beam 0)

4.3.2 MC-TDMA with Random Slot Selection

In Chapters 2 & 3, it was argued that CDMA has such inherent advantages, that it should not be overlooked as a multiple access scheme – even for a multimedia satellite network. Performance results show that assigning a spreading code to each earth station and allowing the earth station to select the timeslot in which to transmit can lead to an optimum delay, but can suffer from soft blocking. In order to compensate for the soft blocking, we proposed a very simple solution that depended on a flow control parameter (p_{fc}). The main idea was to trade the jitter-tolerant delay for soft blocking – basically delaying users at the earth stations when soft blocking is a problem. In Section 3.4, it was contended that the parameter p_{fc} could be set up based on the number of active calls and their traffic profile. In an OBP system with multiple spot beams, we would also like to control the congestion at the downlink queues. Since there are no reservations, the

approach described in the previous section would not be useful. However, close examination will reveal that the congestion problem is very similar to the soft blocking problem. In fact, congestion occurs as a result of the jitter-tolerant traffic being multiplexed by the on-board switch. If congestion is imminent, we could prevent it by controlling the *flow* of packet transmissions from the earth stations. So the flow control would be based on two factors:

1. the traffic from earth stations for a particular (uplink,downlink) pair
2. imminent congestion status

In the following, the algorithm details are highlighted, but no performance results are provided. As CDMA systems are interference limited, a true end-to-end system evaluation requires that system imperfections be considered on the uplink. This requires that a link budget be established, as well as consideration for adjacent beam interference, imperfect power control, and the propagation weather conditions.

4.3.2.1 Algorithm Details

The basic algorithm is shown in Figure 4-9 and Figure 4-10. The scheduler on board the satellite is aware of the number of admitted calls from all earth stations. It uses this information to make a Load Matrix for each uplink/downlink pair. This matrix is very similar to the Assignment Matrix discussed in section 4.3.1, but rather than having assignment information, it stores the load information from an uplink to a particular downlink¹¹ (See Table 4-2). ρ_{ij} shows the load from uplink beam i to downlink beam j . The scheduler uses this information to determine the flow control parameter for each uplink (i) and downlink (j) pair $p_{fc_u}(i), p_{fc_d}(j)$. A lower value of $p_{fc_u}(i)$ implies that the earth stations in uplink i are pressured to refrain from transmitting jitter-tolerant traffic, while $p_{fc_d}(j)$ implies that the earth stations with traffic to downlink beam j are pressured to refrain from transmitting jitter-tolerant traffic. In addition, to this mechanism,

¹¹ We assume that the scheduler is aware of this information as a result of a call admission algorithm.

we propose that the scheduler also monitor the status of each downlink queue. So if downlink queue k becomes congested, the scheduler can reduce $p_{fc_d}(j)$ accordingly. All flow control probabilities can be regularly broadcast on the downlink.

Table 4-2: Load Matrix (Information per uplink/downlink beam)

Downlink \ Uplink	Spot Beam 1	Spot Beam 2	Spot Beam 3	Spot Beam 4	Total
Spot Beam 1	ρ_{11}	ρ_{12}	ρ_{13}	ρ_{14}	$\rho_{11} + \rho_{12} + \rho_{13} + \rho_{14}$
Spot Beam 2	ρ_{21}	ρ_{22}	ρ_{23}	ρ_{24}	$\rho_{21} + \rho_{22} + \rho_{23} + \rho_{24}$
Spot Beam 3	ρ_{31}	ρ_{32}	ρ_{33}	ρ_{34}	$\rho_{31} + \rho_{32} + \rho_{33} + \rho_{34}$
Spot Beam 4	ρ_{41}	ρ_{42}	ρ_{43}	ρ_{44}	$\rho_{41} + \rho_{42} + \rho_{43} + \rho_{44}$
Total	$\rho_{11} + \rho_{21} + \rho_{31} + \rho_{41}$	$\rho_{12} + \rho_{22} + \rho_{32} + \rho_{42}$	$\rho_{13} + \rho_{23} + \rho_{33} + \rho_{43}$	$\rho_{14} + \rho_{24} + \rho_{34} + \rho_{44}$	

At the earth stations, the generated traffic is queued according to type. We again assume the capacity types defined by the DVB-RCS standard – (namely CRA, RBDC, VBDC). At the start of every frame, the earth station randomly selects the time slots it will use for its real-time traffic. All other slots are available for jitter-tolerant traffic. For each of these, the earth station makes a pair of Bernoulli trials. The first determines if the uplink can accept the packet. If successful, a second Bernoulli trial is made to see whether the downlink beam can support the packet. If this is also successful then the packet is transmitted. If $p_{fc_u}(i)$ and $p_{fc_d}(j)$ are chosen properly, then both soft blocking and congestion can be controlled.

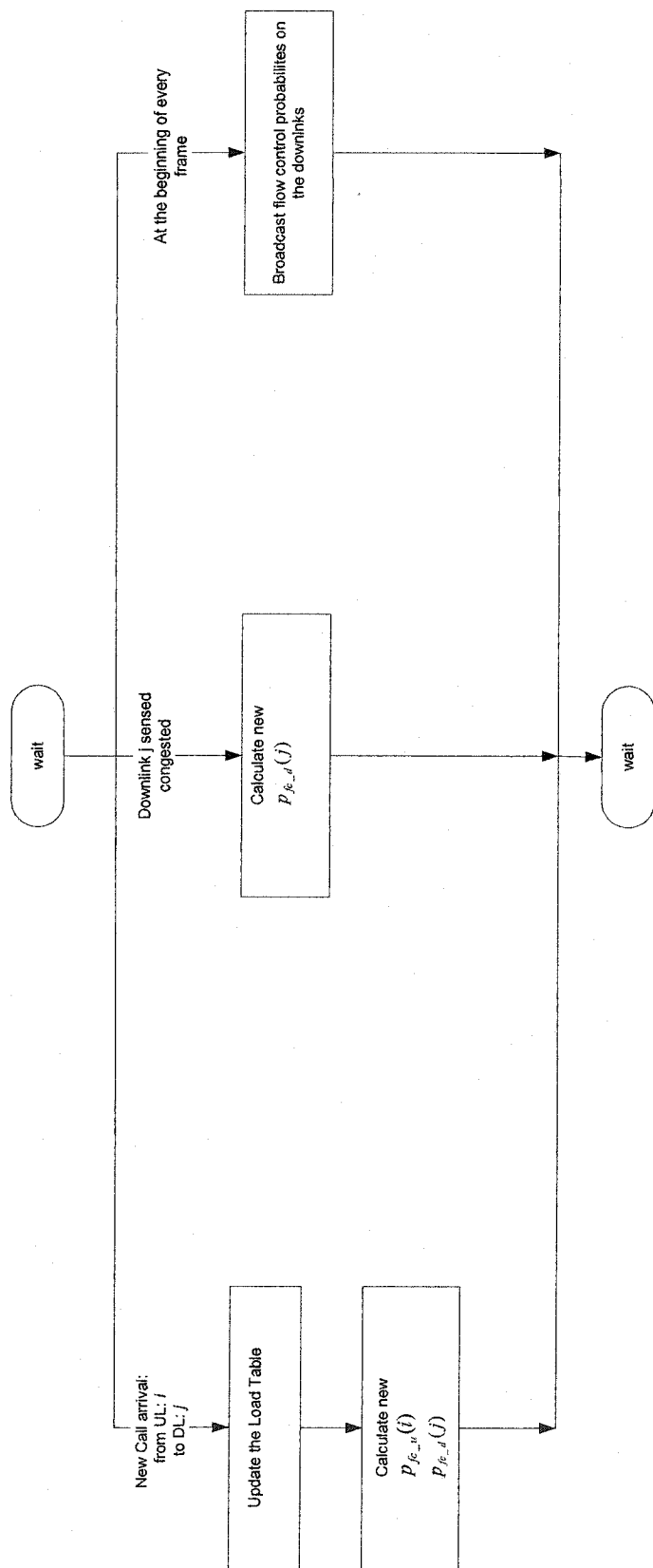


Figure 4-9: Procedure at Scheduler

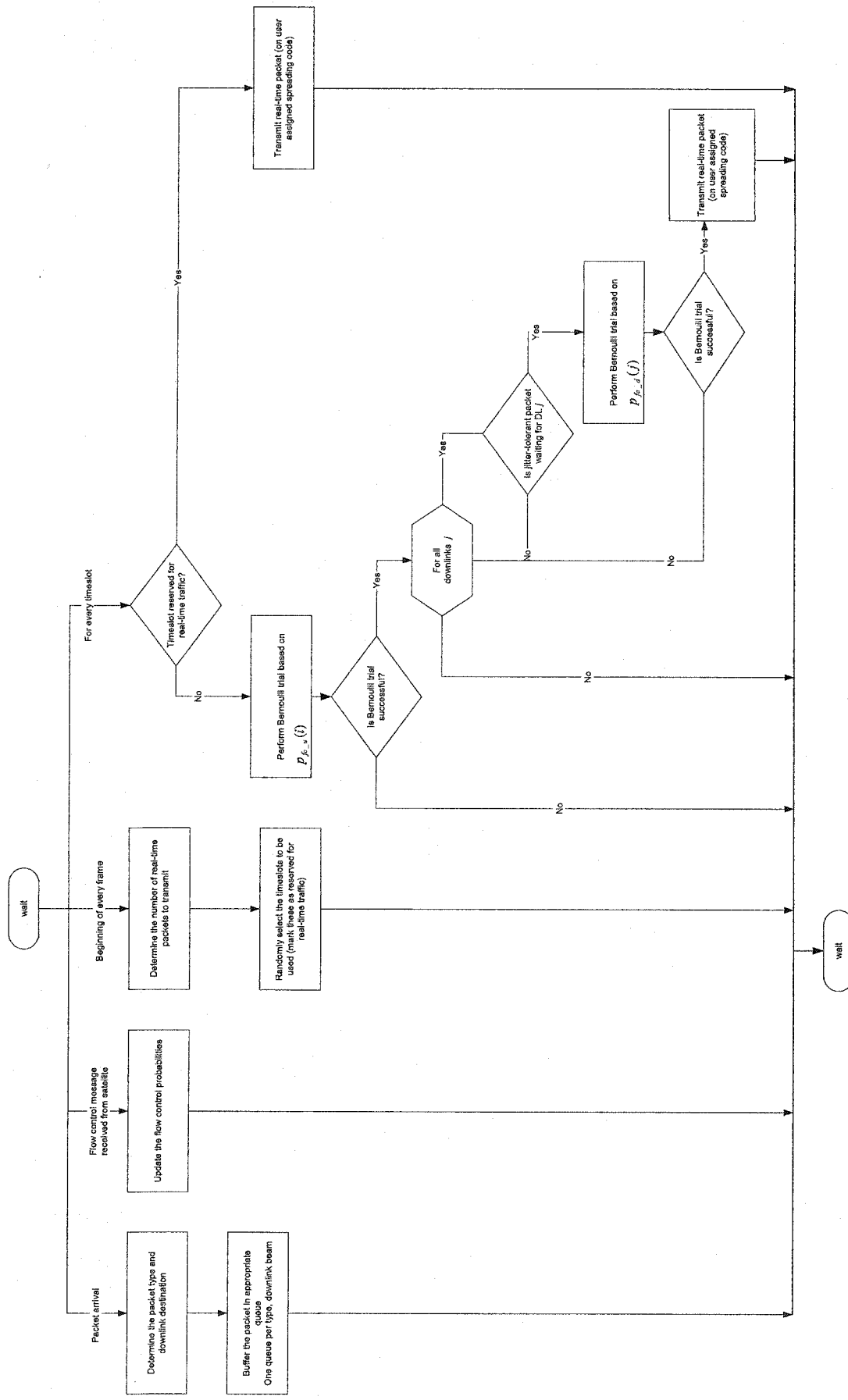


Figure 4-10: Procedure at Earth Stations

Chapter 5

Conclusions and Suggestions for Further Research

In this thesis work, we have provided solutions for a multimedia satellite network, capable of integrating with the QoS frameworks used in today's terrestrial networks. The main focus was on multiple access protocols and their performance under multimedia traffic (both real-time voice and video, and jitter-tolerant data).

Two multibeam satellite network configurations were considered: bent pipe and on-board processing, and two media access protocols were proposed for both. In the bent-pipe system, the first protocol is a very straightforward implementation of the protocol that is being suggested by DVB-RCS. Our main contribution was a performance analysis of the protocol in terms of jitter-tolerant traffic delay and real-time loss probability. For the former, our solution strategy was based on the Matrix Geometric method. In order to make the problem tractable, we modeled all sources as MMPPs. Simulation results have shown that this is acceptable under the conditions tested. An assumption was also made regarding the service mechanism at high loads. This tended to overestimate the jitter-tolerant traffic delay. For the latter, we used the uniformization technique to determine the distribution of real-time arrivals, as this allowed calculating the real-time loss probability. The second protocol was based on MC-TDMA channelization. The technique performed extremely well under all load conditions, but did suffer from soft blocking. We proposed a distributed mechanism to trade jitter-tolerant traffic delay for soft-blocking. Simulation results verified that the technique provided the necessary control.

The two media access schemes were then adapted to fit an on-board processing environment. The changes were necessary to deal with possible congestion at the downlink queues resulting from the multiplexing of traffic at the output of the baseband switch. For the CFDAMA scheme, the

scheduler uses the calculated channel assignments as an indication of future traffic arrival to the satellite. It then shuffles assignments that would lead to congestion on one of the downlink queues. The algorithm was compared against an ERICA based technique and showed a marked improvement in queue size. The algorithm effectively prevents congestion, as it only assigns channels to those earth stations that send traffic to uncongested downlinks. A modification was also suggested for the MC-TDMA media access control algorithm. The algorithm is based on the same principle used to limit soft blocking. The main idea is to only use those timeslots that have passed through two levels of flow control – one for the uplink to limit soft blocking and the second for the downlink to limit congestion.

In addition, the OPNET simulation model developed can be considered an important contribution for future network performance studies. The model is very flexible in terms network architecture (number of earth stations, location of scheduler, uplink and downlink frame formats, etc) and in the makeup of each earth station (traffic types, traffic models, etc.).

5.1 Suggestions for Further Research

The overall thesis scope was quite complex, and as a result, numerous areas still require further investigation. Namely:

- In the analysis of the CFDAMA system, an assumption was made as to the distribution of the service mechanism to the jitter-tolerant traffic queue. The assumption overestimates the delay at high loads. Research to determine the distribution would be useful and may be applicable to a host of other uplink access problems.
- The jitter-tolerant traffic delay in an MC-TDMA system remains an open issue. This is complicated by the need to consider the physical layer imperfections.
- The use of variable spreading as an option to reduce soft blocking. Since CDMA-based systems are not bandwidth-limited, it is possible to vary the coding gain for certain traffic types to counter soft blocking. The same argument can be extended as a means to control possible congestion of downlink queues.

- On the simulation side, research is required on
 - The impact of making requests every N frames (superframe) as opposed to every frame. In a real network, the number of earth stations is expected to be very large and it may not be possible to guarantee that an earth station receives a request opportunity every frame.
 - The impact of heterogeneous terminals on system performance.
 - The performance of the proposed 2-level flow control MC-TDMA system. This would require inclusion of physical layer imperfections in the simulation model.
 - The impact of capping the amount of assignment shuffles permitted per frame on the performance of the CC-MAC.

Appendix A

Review of Channelization Techniques

As mentioned in Chapter 1, the multiple access problem can be divided into methods of channelization and methods of channel access. The former refers to the technique by which the available resources are divided into channels, each capable of supporting a packet. There are numerous combinations of channelization and channel access that have been proposed in the literature. When selecting the most appropriate of these, it is important to remember the criteria or figures of merit, which will be used in their evaluation. These criteria go hand in hand with the proposed network environment/architecture. Basically, for a multiple access technique to be deemed acceptable, it must be capable of meeting the following four criteria:

- Must support various traffic types with vastly different quality of service requirements. Typically, these techniques should provide low average packet delay and minimal real-time loss, all while maximizing the channel utility. This implies that the performance of these schemes should be measured with traffic models that are commensurate with the traffic type. Unfortunately, the performance of the majority of MAC protocols is based on Poisson type traffic, which is not a valid model for the multimedia traffic considered.
- The technique should function in a bent-pipe as well as OBP GEO satellite environment. This implies a large propagation delay and precludes the use of multiple access schemes that require terminals to have full knowledge of the network state.
- Should be able of servicing relatively high load earth stations, with information rates ranging from 384 Kbps to 2048 Kbps. The number of active earth stations per spot-beam should be fewer than 100.
- For the GEO system, the MAC should employ a frame type structure with fixed size cells, in order to reduce the on-board switch complexity.

A.1 Channelization Techniques

The methods of channelization are classified in terms of the resource axis (of Figure A-1) that is shared. Time division and frequency division (TD/FD) lead to the well-known techniques of time division multiple access (TDMA) and frequency division multiple access (FDMA), respectively. In the former, the time axis is divided into a number of time slots, each capable of carrying a single packet. The channel uses the full available bandwidth, and is transmitted at maximum power (the PSD is fully used). In contrast, for FDMA, the frequency axis is segmented into sub-bands, with the channel occupying the entire time and PSD axis. Interestingly, the time and frequency axis can be considered dual sides of a single limitation [85]. With all other things equal (modulation, coding, pulse shaping, etc) higher bit rates require a proportionally larger bandwidth. Notice that a common feature of both TDMA and FDMA is that channels use the entire power spectral density band. Spread spectrum techniques, on the other hand, are a family of channelization approaches that attempt to share this band. This sharing is achieved via spreading codes – hence the name code division multiple access (CDMA)¹². The channel uses the entire frequency and time axis, but only a small portion of the PSD axis. The actual spreading is accomplished by using pseudorandom codes (Gold, Kasami, etc.). At the receiver, the channels can be isolated as long as the crosscorrelation between the spreading codes of the channels, is low. For a review of CDMA and other spread spectrum techniques, the reader is referred to the texts by Dixon [86] and Ziemer and Peterson [87]. These three basic methods of division can be combined into a number of hybrid techniques. These include multi-code TDMA (MC-TDMA), multi-frequency TDMA (MF-TDMA), multi-code FDMA (MC-FDMA), and multi-code multi-

¹² Strictly speaking, this refers to Direct-Sequence CDMA (DS-CDMA), which uses the spreading codes to spread the transmitted bandwidth. There is also a frequency hopped version of CDMA, where the spreading code is not used for bandwidth expansion, but rather to hop the carrier between a number of frequency sub-bands. In this thesis, only DS-CDMA has been considered, and the term CDMA is used exclusively to refer to this type.

frequency TDMA (MC-MF-TDMA). These are briefly described below, and depicted in Figure A-1.

MC-TDMA: in this scheme, channels correspond to specific spreading codes and time-slots.

MF-TDMA: Channels correspond to unique frequency-time slots.

MC-FDMA: This is a combination of FDMA and CDMA. The available bandwidth is first divided into sub-bands, and each channel is assigned a unique spreading code to fill this band. Note that the same spreading codes can be assigned to channels in adjacent bands, as the frequency division maintains the orthogonality.

MC-MF-TDMA: This is a hybrid combination, which is based on a division of all three axis shown in Figure A-1.

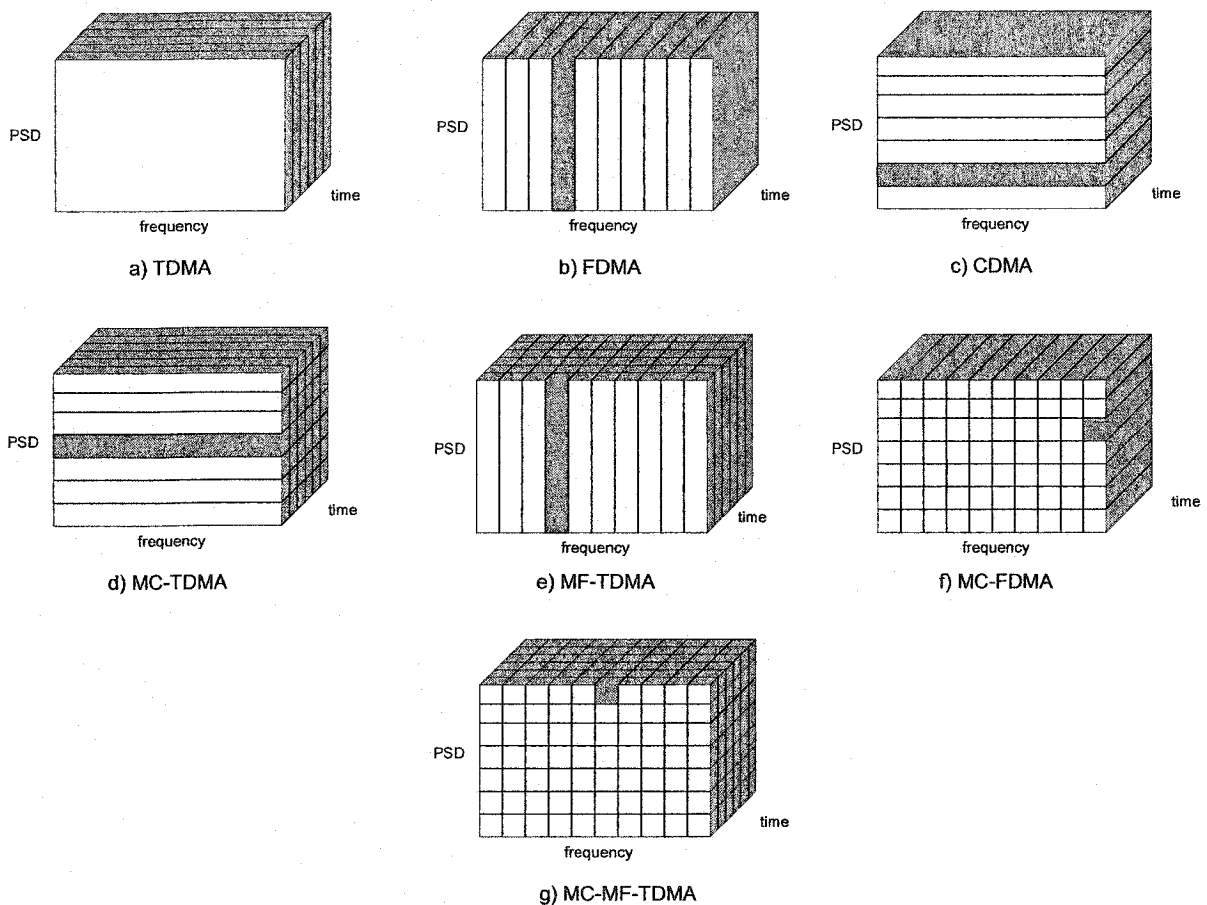


Figure A-1: Channelization Techniques

A.1.1 Radio Capacity

The basic channelization techniques can be compared in terms of radio capacity and information capacity. The former refers to the number of users that can be supported in a given bandwidth. This capacity generally depends on the user station bit rate, the transmission modulation and coding technique employed by these users, and the available bandwidth. The bit rates are normally taken to be constant, equal to R_b , the modulation as BPSK, the coding rate as r , and the available bandwidth as B_w . The details of the analysis are shown in [88]. The capacity (M) for five of the channelization techniques are shown below. The results for the other two techniques attain a similar form [88].

$$M_{FDMA} = \frac{B_w r}{KR_b} \frac{1}{1 + G_{FDMA}} \quad (A-1)$$

$$M_{TDMA} = \frac{B_w r}{KR_b} \frac{1}{1 + G_{TDMA}} \quad (A-2)$$

$$M_{CDMA} = \frac{1}{1 + 6\alpha} \left[1 - \frac{3}{2} \frac{1}{(E_b / N_0)} \frac{B_w}{R_b} + \frac{3}{2} \frac{1}{(E_b / N_0)_{\min}} \frac{B_w}{R_b} \right] \quad (A-3)$$

$$M_{MF-TDMA} = \frac{B_w r}{KR_b} \frac{1}{1 + G_{TDMA}} \frac{1}{1 + G_{TDMA}} \quad (A-4)$$

$$M_{MC-TDMA} = \frac{1}{1 + 6\alpha} \left[N_t + \left(\frac{-3}{2} \frac{1}{(E_b / N_0)} \frac{B_w}{R_b} + \frac{3}{2} \frac{1}{(E_b / N_0)_{\min}} \frac{B_w}{R_b} \right) \frac{1}{(1 + G_{TDMA})} \right] \quad (A-5)$$

where G_{FDMA} and G_{TDMA} are guard bands associated with the respective techniques. These guard bands are used to relax network synchronization requirements for TDMA, and to reduce the filtering requirements for FDMA. The remaining variables are defined below:

K = the frequency reuse factor

= the number of spot-beams over which the frequency is reused (this is equivalent to the cluster size in cellular networks);

α = attenuation provided by the satellite antenna to earth stations in adjacent beams;

N_t = is the number of time slots in the hybrid TDMA frame;

(E_b/N_0) = signal bit energy-to-single sideband thermal noise PSD, referred to as SNR;

$(E_b/N_0)_{min}$ = minimum SNR required to achieve a given performance (it takes into account all forms of interference).

In the above, it is assumed that the CDMA systems employ a frequency reuse factor of 1 ($K = 1$).

This is possible as users in adjacent beams are given their own spreading codes, in which case their transmissions only act to increase the multiple access interference.

The following interesting observations can be made by examining equations (A-1) to (A-5).

1. Channelization techniques that employ time division and/or frequency division exclusively are bandwidth limited.
2. These same techniques are strongly influenced by the frequency reuse factor (K). The smallest K that ensures adjacent beams will have non-overlapping frequencies is four (See Figure A-2).
3. The performance of the spread spectrum systems is interference limited, and depends on α , the processing gain P_G ($P_G = B_W/R_b$), and the required $(E_b/N_0)_{min}$ to achieve a certain performance.

4. The radio capacity of spread spectrum systems is not directly affected by the coding rate (r). Consequently, very powerful forward error correction (FEC) codes can be used (low r), reducing the required $(E_b/N_0)_{min}$, and increasing M [89].

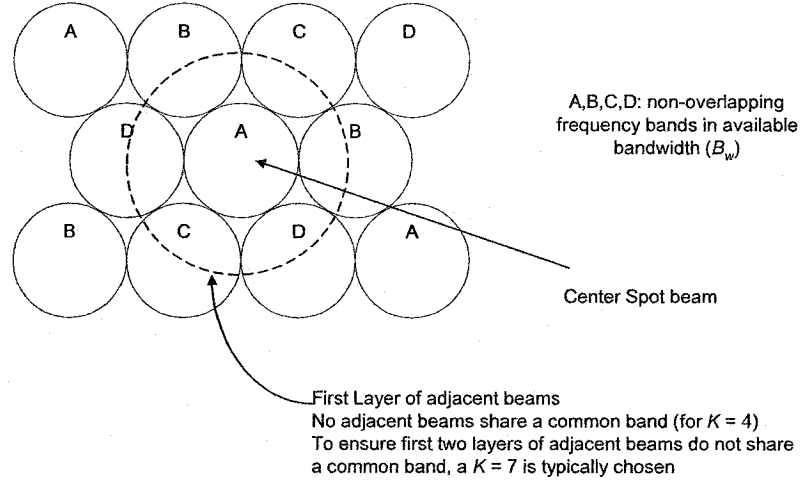


Figure A-2: Frequency Reuse for a Multibeam Environment

Figure A-3, shows the radio capacity of FDMA, TDMA, and CDMA, for a bandwidth of $B_w = 120$ MHz, and a fixed user bit rate $R_b = 384$ Kbps. The TD/FD techniques employ a coding rate of $r = 1/2$, no guard bands (i.e., $G_{FDMA} = G_{TDMA} = 0$), and a reuse factor of $K = 4$. The results for CDMA also assume no guard bands, but as a frequency reuse factor of 1 was selected, we considered the impact of adjacent beam interference (ABI). Based on the specifications given in [90], the adjacent beam attenuation was taken to be 15.5 dB ($\alpha = 10^{-1.55}$). Both cases assumed an E_b/N_0 of 10 dB. The results are plotted vs. $(E_b/N_0)_{min}$. Since the TD/FD techniques are bandwidth limited, they have a single operating point, at $(E_b/N_0)_{min} = E_b/N_0 = 10$ dB (shown by the * on the graph). The capacity of CDMA, on the other hand, can be increased if the requirement for $(E_b/N_0)_{min}$ is relaxed (by using a low rate FEC code, for instance). In practice, in order to ensure proper functioning of the on-board synchronization circuitry and the like, there is a minimum $(E_b/N_0)_{min}$ which can not be exceeded.

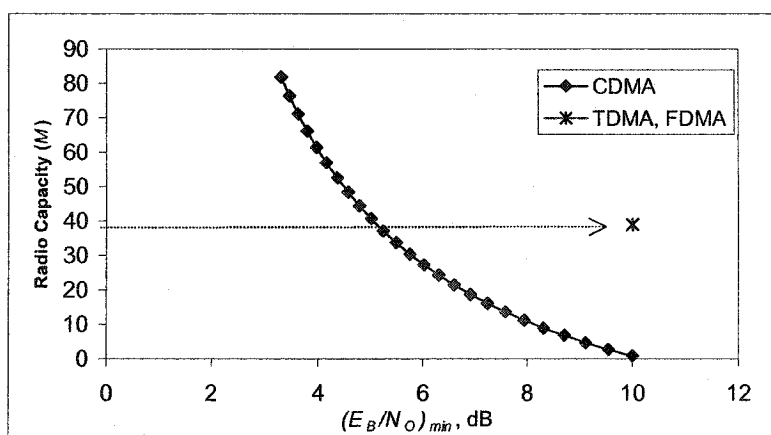


Figure A-3: Radio Capacity Comparison (TDMA, FDMA,
CDMA)

The graph clearly shows that under certain conditions, the capacity of CDMA exceeds that of TDMA and FDMA, which agrees with some of the findings in the available literature (albeit for different circumstances). See for instance [91],[92], and [93].

A.1.2 Information Capacity

The second measure of capacity, referred to as the information capacity, reflects the maximum number of bits that can be transmitted per use of a channel. This capacity depends on the noise in the system, and the transmitted signal energy. The results in this section are derived in [88], and follow closely those presented in Alencar and Blake [94]. Essentially, the channel is modeled as having a discrete input and a continuous output (X_k and Z_k , respectively). X_k denotes this binary input, and Z_k denotes the continuous output of a matched filter receiver, as shown in Figure A-4. The output of this channel (Z_k) depends on the thermal noise, the adjacent beam interference, and the multiple access interference (if any). The information capacity of the resulting channel is defined as the maximum mutual information [95], taken over the input probability distribution ($p_i = \Pr[X_i = k]$). That is,

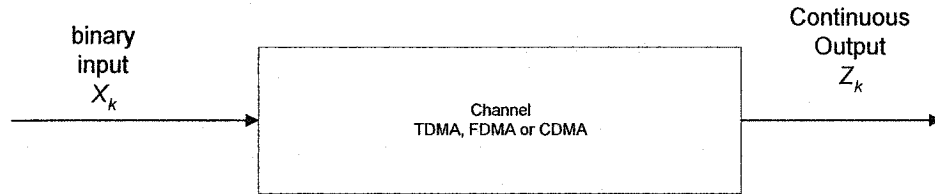


Figure A-4: Modeling of Multiple Access Communications System

C_k = information capacity

= maximum number of transmitted bits/channel use (for a single channel)

$$= \max_{p_i} I(X_k; Z_k) \quad (\text{A-6})$$

$$= \max_{p_i} (h(Z_k) - h(Z_k | X_k)) \quad (\text{A-7})$$

where $h(X)$ is the differential entropy of random variable X $\left(h(X) = - \int_{-\infty}^{\infty} f_X(x) \log(f_X(x)) dx \right)$.

The details of the derivation can be found in [88]. The final results are shown below for low and high signal-to-noise ratios.

$$C_k \cong \log_2(e) \left[\frac{\mu^2}{2\sigma^2} \right] \text{ for low SNR} \quad (\text{A-8})$$

$$C_k \cong 1 - \frac{2\pi}{(\pi^2 - 8) \ln(2)} \left\{ \sqrt{2\pi} Q \left[\frac{4}{\pi} \frac{\mu}{\sqrt{2}\sigma} \right] e^{((8-\pi^2)/\pi^2)\mu^2/2\sigma^2} - 4Q \left[\frac{\mu}{\sigma} \right] \right\} \text{ for high SNR} \quad (\text{A-9})$$

where $\mu = \sqrt{\frac{E_b}{2R_b}}$ and $\sigma^2 = \frac{E_b}{6R_b P_G} (M^{(int)} - 1) + \frac{N_0}{4R_b}$. $M^{(int)}$ denotes the number of interfering

users sharing a time-slot. For TDMA and FDMA, $M^{(int)}=1$ (only one user is active in a time slot at any given time). For a case with M channels, the total capacity would be equal to $C=MC_k$. The results for a case with $M=20$ and a processing gain of $P_G=128$, are shown in Figure A-5. Notice that the information capacity of CDMA is very close to that of TDMA and FDMA – in fact, they only differ by about 1 bit/transmission for an E_b/N_0 of 5dB. What is important to note is that although CDMA channels are not orthogonal, the capacity of these, is nonetheless close to that of the orthogonal channelization techniques of TDMA and FDMA. Furthermore, these results for CDMA were obtained assuming a non-cooperative type receiver. These receivers treat all channels independently, with each contributing MAI to each other. Performance is expected to be much better for cooperative type receivers [96].

In summary, the results for radio capacity and information capacity show that CDMA based channelization techniques should not be ignored as possible candidates for our proposed environment.

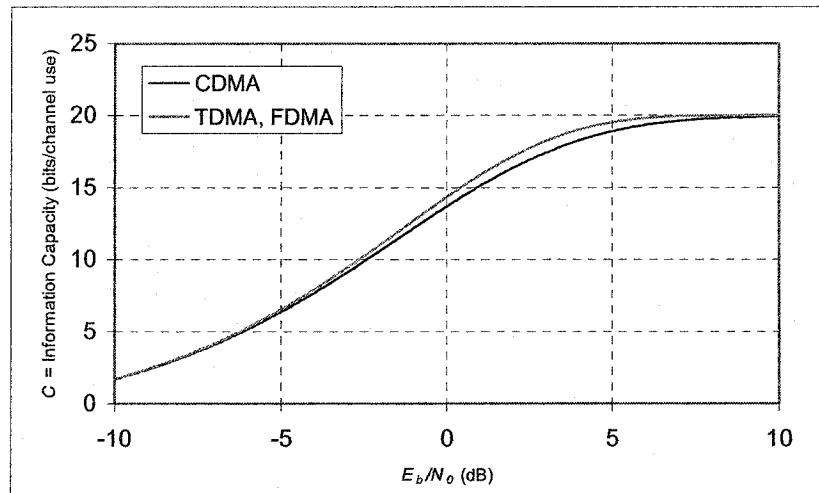


Figure A-5: Information Capacity Comparison (TDMA, FDMA, CDMA)

References

- [1] V. Labrador. (2004, February). Whither Broadband? What's in Store for 2004. *SATMAGAZINE.COM* [Online]. pp. 12-14. Available: <http://www.satmagazine.com/feb2004/feb2004.pdf>
- [2] Satlynx, "SATLYNX Building Your Interactive relationships," Available: <http://www.satlynx.com/>, Accessed: February 15, 2004.
- [3] W. Partridge, "Application for authority to launch and operate the Voicespan system," *filing with Federal Communications Commission*, AT&T Corporation, December 1993.
- [4] Teledesic, "Teledesic Internet in the Sky," Available: <http://www.teledesic.com/>, Accessed: February 15, 2004.
- [5] ASTROLINK, "Astrolink Powering Today's Communication revolution," Available: <http://www.astrolink.com/>, Accessed: February 15, 2004.
- [6] L. Wood, "Lloyd's Satellite Constellations - Overview," Available: <http://www.ee.surrey.ac.uk/Personal/L.Wood/constellations/overview.html>, Accessed: February 15, 2004.
- [7] G. Maral, *VSAT Networks*. Toronto: John Wiley and Sons, 1995.
- [8] D.J. Bem, T.W. Wieckowski, and R.J. Zielinski. (2000). Broadband Satellite Systems. *IEEE Communications: Surveys and Tutorials* [Online]. Vol. 3, No. 1, pp. 1-15. Available: <http://www.comsoc.org/livepubs/surveys/public/1q00issue/zielinski.html>
- [9] R. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues*. Boston: Artech House Inc, 1994.
- [10] European Space Agency, "ESA Telecommunications," Available: <http://telecom.esa.int/telecom>, Accessed: August 1, 2004.
- [11] W. Buerkle and M. Trefz, "On-Board Switching Architectures for Multimedia Satellite Systems," *Space Communications*, Vol. 17, No. 1-3, 2001, pp. 215-229.
- [12] R. Padovani, "Reverse Link Performance of IS-95 Based Cellular Systems," *IEEE Personal Communications*, Vol. 1, No. 3, Third Quarter 1994, pp. 28-34.
- [13] A.J. Viterbi, A.M. Viterbi, and E. Zehavi, "Performance of Power-Controlled Wideband Terrestrial Digital Communication," *IEEE Transactions on Communications*, Vol. 41, No. 4, April 1993, pp. 559-569.
- [14] *Digital Video Broadcasting: Interaction Channel for Satellite Distribution Systems*, European Telecommunications Standards Institute Standard ETSI EN 301 790 v1.3.1, March 2003.

- [15] *Data-Over-Cable Service Interface Specifications DOCSIS 2.0: Radio Frequency Interface Specification*, Cable Television Laboratories Inc Standard SP-RFiv2.0-I04-030730, July 2003.
- [16] A.L. Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*. New York: McGraw Hill Higher Education, 2004.
- [17] R. Braden, D. Clark, and S. Shenkar, "Integrated Services in the Internet Architecture: an Overview," IETF RFC 1633, June 1994.
- [18] S. Shenkar, C. Partridge, and R. Guerin, "Specification of Guaranteed Quality of Service," IETF RFC 2212, September 1997.
- [19] J. Wroclawski, "Specification of the Controlled-Load Network Element Service," IETF RFC 2211, September 1997.
- [20] R. Braden, L. Zhang, S. Herzog, and S. Jamin, "Resource ReserVation Protocol (RSVP) – version 1 functional specification," IETF RFC 2205, September 1997.
- [21] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," IETF RFC 2475, December 1998.
- [22] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," IETF RFC 2598, June 1999.
- [23] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB Group," IETF RFC 2597, June 1999.
- [24] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, "Requirements for Traffic Engineering Over MPLS," IETF RFC 2702, September 1999.
- [25] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," IETF RFC 3031, January 2001.
- [26] A. Hung, M. Montpetit, and G. Kesidis, "ATM via satellite: A framework and implementation," *Wireless Networks*, Vol. 4, Issue 2, February 1998, pp. 141-153.
- [27] A. Baiocchi, M. Listanti, N. Blefari-Melazzi, and C. Soprano, "An ATM-like System Architecture for Satellite Communications Including On-board Switching", *International Journal of Satellite Communications*, Vol. 14, No. 5, September 1996, pp. 389-412.
- [28] N. Iuoras, T. Le-Ngoc, M. Ashour, and T. Elshabrway, "An IP-based Satellite Communication System Architecture for Interactive Multimedia Services," *International Journal of Satellite Communications*, Vol. 21, No. 4-5, July-October 2003, pp. 401-426.
- [29] T. Ors and C. Rosenberg, "Providing IP QoS over GEO Satellite Systems using MPLS," *International Journal of Satellite Communications*, Vol. 19, No. 5, September/October 2001, pp. 443-461.

- [30] S. Kota and M. Marchese, "Quality of Service for Satellite IP Networks: a Survey," *International Journal of Satellite Communications*, Vol. 21, No. 4-5, July-October 2003, pp. 303-349.
- [31] R. Rom and M. Sidi, *Multiple Access Protocols*. New York: Springer-Verlang, 1990.
- [32] S. Tasaka, *Performance Analysis of Multiple Access Protocols*. London: MIT Press, 1986.
- [33] N. Abramson, "VSAT Data Networks," *Proceedings of the IEEE*, Vol. 78, No. 7, July 1990, pp. 1267-1274.
- [34] D. Rachaudhuri, "ALOHA with Multipacket Messages and ARQ-type Retransmission Protocols – Throughput Analysis," *IEEE Transactions on Communications*, Vol. 35, No. 7, July 1987, pp. 767-772.
- [35] L.G. Roberts, "ALOHA Packet System with and without Slots and Capture," *Computer Communications Review*, Vol. 5, No. 2, April 1975, pp. 28-42.
- [36] W. Yue, "The Effect of Capture on Performance of Multichannel Slotted ALOHA Systems," *IEEE Transactions on Communications*, Vol. 39, No. 6, June 1991, pp. 818-822.
- [37] J.I. Capetanakis, "Tree Algorithm for Packet Broadcasting Channels," *IEEE Transactions on Information Theory*, Vol. 25, No. 5, September 1979, pp. 505-515.
- [38] D. Makrakis and K.M.S. Murthy, "Spread Slotted ALOHA Techniques for Mobile and Personal Satellite Communication Systems," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 6, August 1992, pp. 985-1002.
- [39] D. Raychaudhuri, "Performance Analysis of Random Access Packet-Switched Code Division Multiple Access Systems," *IEEE Transactions on Communications*, Vol. 29, No. 6, June 1981, pp. 895-901.
- [40] R. Binder, "A Dynamic Packet-Switching System for a Satellite Broadcast Channels," *Proceedings ICC' 75*, pp. 410-514.
- [41] N. Calendroni and E Ferro, "The FODA-TDMA Satellite Access Scheme: Presentation, Study of the System, and Results," *IEEE Transactions on Communications*, Vol. 39, No. 12, December 1991, pp. 1823-1831.
- [42] I.M. Jacobs, R. Binder, and E.V. Hoverstein, "General Purpose Packet Satellite Networks," *Proceedings of the IEEE*, Vol. 66, No. 11, November 1978, pp. 1448-1467.
- [43] D.J. Goodman, et.al., "Packet Reservation Multiple Access for Local Wireless Communications," *IEEE Transactions on Communications*, Vol. 37, No. 8, August 1989, pp. 885-890.
- [44] T. Le-Ngoc, and Y-D. Yao, "CREIR: A Multiaccess Protocol for On-board Processing Satellite Systems," *Proceedings CCECE'91*, pp. 26.1.1-26.1.4.
- [45] A.K. Elhakeem, S. Bohm, M. Hachicha, T. Le-Ngoc, and H.T. Mouftah, "Analysis of a new Multiaccess Switching Technique for Multibeam Satellites in a Prioritized ISDN

- Environment," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 2, February 1992, pp. 378-390.
- [46] T. Ors, "Traffic and Congestion Control for ATM over Satellite to provide QoS," Ph.D. dissertation, School of Electronics and Physical Sciences, University of Surrey, Guildford, England, 1998.
 - [47] A. Iera, A. Molinaro, and S. Marano, "Call Admission Control and Resource Management Issues for Real-Time VBR Traffic in ATM-Satellite Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 11, November 2000, pp. 2393-2403.
 - [48] G. Acar and C. Rosenberg, "Algorithms to Compute Bandwidth-on-Demand Requests in a Satellite Access Unit," *Proceedings 5th Ka Band Utilization Conference*, Taormina, Italy, October 1999.
 - [49] J. Mohamed, "Combined Free/Demand Assignment Multiple Access Protocols for Packet Satellite Communications," M.A. Sc. Thesis, Department of Electrical and Computer Engineering, Concordia University, 1993.
 - [50] S. Krishnamuthy, "Combined Free Demand Assignment Multiple Access (CFDAMA) for Integrated Data Voice Satellite Communications," M.A. Sc. Thesis, Department of Electrical and Computer Engineering, Concordia University, 1994.
 - [51] E.A. Wibowo, A. Iuoras, P. Takats, J. Lambadaris, and M. Devetsikiotis, "Guaranteeing Quality of Service in Packet-Switched Satellites by Medium Access Control," *Proceedings Canadian Conference on Broadband Research (CCBR) 98*, Ottawa, Canada, June 1998.
 - [52] Z. Jiang, Y. Li and V.C.M. Leung, "A predictive demand assignment multiple access protocol for broadband satellite networks supporting internet applications", *Proceedings ICC'02*, New York, April 2002, pp. 2973-2977.
 - [53] C.L.I and K.K. Sabnani, "Variable Spreading Gain CDMA with Adaptive Power Control for True Packet Switching Wireless Networks," *Proceedings ICC'95*, pp. 725-730.
 - [54] C.L.I, G.P. Pollini and R.D. Gitlin, "Performance of Multi-Code CDMA Wireless Personal Communications Networks," *Proceedings VTC'95*, pp. 907-911.
 - [55] C. Baugh, "Europe: Leading the Pack for Satellite Standards," *Interspace*, April 2003, pp. 6-7.
 - [56] Wildblue Communications, "WildBlue Communications – Press Release, January 5, 2004," Available: <http://www.wildblue.com/>, Accessed: February 15, 2004.
 - [57] OPNET Technologies Inc., "OPNET Modeler – Accelerating Networking R&D", Available: <http://www.opnet.com/>, Accessed February 15, 2004.
 - [58] D. Anick, D. Mitra, and M.M. Sondhi, "Stochastic Theory of Data-Handling System with Multiple Sources," *The Bell System Technical Journal*, Vol. 61, No. 8, October 1982, pp. 1871-1894.

- [59] J.F. Hayes, *Modeling and Analysis of Computer Communications Networks*. New York: Plenum Press, 1986.
- [60] M.F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and their Application*. New York: Marcel Dekker, 1989.
- [61] D.M. Lucantoni, "New Results on the Single Server Queue with a Batch Markovian Arrival Process," *Stochastic Models*, Vol. 7, No. 1, 1991, pp. 1-46.
- [62] P.T. Brady, "A Model for Generating on-off Speech Patterns in two way Conversation," *Bell System Technical Journal*, September 1969, pp. 2445-2472.
- [63] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J.D. Robbins, "Performance Models of Statistical Multiplexing in Packet Video Communications," *IEEE Transactions on Communications*, Vol. 36, No. 7, July, 1988, pp. 834-843.
- [64] W.E. Leland, S.M. Taqqu, W. Willinger, and D.V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (extended version)," *IEEE/ACM Transactions on Networking*, Vol. 2, No. 1, February 1994, pp. 1-14.
- [65] H.J. Fowler and W.E. Leland, "Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 7, September 1991, pp. 1139-1149.
- [66] R. Di Girolamo and T. Le-Ngoc, "Performance of a Multimedia SATCOM System using MC-TDMA," Technical Report: Department of Electrical and Computer Engineering, Concordia University, 1996.
- [67] R. Di Girolamo and T. Le-Ngoc, "Performance of CFDMA in a Multimedia SATCOM System using MF-TDMA," Technical Report: Department of Electrical and Computer Engineering, Concordia University, 1996.
- [68] S.B. Kim, M.Y. Lee, and M.J. Kim, " Σ -Matching Technique for MMPP Modeling of Heterogeneous ON-OFF Sources," *Proceedings Globecom 94*, pp. 1090-1094.
- [69] R. O. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues*. Boston: Artech House, 1994.
- [70] H. Heffes and D.M. Lucantoni, "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, September 1986, pp. 856-868.
- [71] S.N. Subramanian and T. Le-Ngoc, "Traffic Modeling in a Multi-media Environment," *Proceedings CCECE'95*, pp. 828-841.
- [72] R. Gussella, "Characterizing the Variability of Arrival Processes with the Indexes of Dispersion," *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 2, February 1991, pp. 203-211.
- [73] C. Loo, "Statistical Models for Land Mobile and Fixed Satellite Communications at Ka Band," *Proceedings VTC'96*, pp. 1023-1027.

- [74] J. Gilderson and J. Cherkaoui, "Onboard Switching for ATM via Satellite," *IEEE Communications Magazine*, Vol.35 No.7, July 1997, pp. 66-70.
- [75] A. Iuoras, P. Takats, et. al. "Quality of Service-Oriented Protocols for Resource Management in Packet-Switched Satellites," *IEEE Journal of Satellite Communications*, Vol 17, No. 2-3, March- June 1999, pp. 129-141.
- [76] A. Iuoras, R. Di Girolamo, et. al. "An Efficient Protocol for Congestion Management in Satellite Networks," *Proceedings 5th Ka Band Utilization Conference*, Taormina, Italy, October 1999.
- [77] G. Acar, "End-to-End Resource Management in Geostationary Networks," Ph.D. Dissertation, The Imperial College of Science, Technology, and Medicine, University of London, 2001.
- [78] A.M. Law and W.D. Kelton, *Simulation Modeling and Analysis*. McGraw Hill: New York, 1991.
- [79] T.V.J. G. Babu, "Performance Analysis of Broadband Multimedia Networks," PhD Dissertation, Department of Electrical and Computer Engineering, Concordia University, 2001.
- [80] UMTS Forum, "UMTS Forum: UMTS Forum Home," Available: <http://www.ums-forum.org/servlet/dycon/ztumts/ums/Live/en/ums/Home>, Accessed: August 1, 2004.
- [81] J. N. Daigle and J. Langford, "Models for Analysis of Packet Voice Communications Systems," *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, September 1986, pp. 847-855.
- [82] J. N. Daigle, *Queueing Theory for Telecommunications*. Reading: Addison-Wesley Publishing Co., 1992.
- [83] M.F. Neuts, "The c-server Queue with Constant Service Times and a Versatile Markovian Arrival Process," in *Applied Probability – Computer Science: The Interface*, R.L. Disney and T.J. Ott, Ed. Boston: Birkhauser, 1982, pp. 31-70.
- [84] V. Ramaswami, "Stable Recursion for the steady state vector for Markov chains of M/G/1 type," *Stochastic Models*, Vol. 4, No. 2, 1988, pp.183-188.
- [85] P. Jung, P.W. Baier, and A. Steil, "Advantages of CDMA and Spread Spectrum Techniques over FDMA and TDMA in Cellular Mobile Radio Applications," *IEEE Transactions on Vehicular Technology*, Vol. 42, No. 3, August 1993, pp. 357-364.
- [86] R.C. Dixon, *Spread Spectrum Systems*. New York: John Wiley and Sons, 1984.
- [87] R.E. Ziemer and R.L. Peterson, *Digital Communications and Spread Spectrum Systems*. New York: Macmillan Publishing Company, 1991.
- [88] R. Di Girolamo, "Study on the Feasibility of CDMA in a Satellite Environment," Technical Report: Department of Electrical and Computer Engineering, Concordia University, 1995.

- [89] A. J. Viterbi, "Very Low Rate Convolutional Codes for Maximum Theoretical Performance of Spread-Spectrum Multiple Access Channels," *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 4, May 1990, pp. 641-649.
- [90] G. Epstein, et. al., "Application for authority to launch and operate the Spaceway system," *filing with Federal Communications Commission*, Hughes Communication Galaxy, December 1993.
- [91] P. Jung, P.W. Baier, and A. Steil, "Advantages of CDMA and Spread Spectrum Techniques over FDMA and TDMA in Cellular Mobile Radio Applications," *IEEE Transactions on Vehicular Technology*, Vol. 42, No. 3, August 1993, pp. 357-364.
- [92] K.G. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, and C. Wheatley, "On the Capacity of a Cellular CDMA System," *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, May 1991, pp. 303-312.
- [93] W.C.Y. Lee, "Overview of Cellular CDMA", *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, May 1991, pp. 291-302.
- [94] M.S. Alencar and I.F. Blake, "The Capacity for a Discrete-State Code Division Multiple-Access Channel," *IEEE Journal on Selected Areas in Communications*, Vol. 12, No. 5, June 1994, pp. 925-937.
- [95] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.
- [96] A. Duel-Hallen, J. Holtzman and Z. Zvonar, "Multiuser Detection for CDMA Systems," *IEEE Personal Communications*, Vol. 2, No. 2, April 1995, pp. 45-58.