

## Resource management for multimedia traffic over ATM broadband

satellite networks

Awadalla, Husam Osman

For additional information about this publication click this link. http://qmro.qmul.ac.uk/jspui/handle/123456789/3802

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk



# RESOURCE MANAGEMENT FOR MULTIMEDIA TRAFFIC OVER ATM BROADBAND SATELLITE NETWORKS

by

Husam Osman Awadalla

## SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Department of Electronic Engineering Queen Mary and Westfield College, University of London

March 2000

To my wife Romessa and children Ahmed, Aisha and Zeinab

## ABSTRACT

The explosion of the Internet and the convergence of voice, video and data services is redefining the way we communicate. Broadband networks with ATM as the backbone technology are being deployed at an accelerating rate to cope with the exponentially growing bandwidth demands. Network operators push satellite ever deeper into the hybrid fibre/satellite architectures to interconnect broadband islands and extend ATM capabilities and multimedia services to distant user communities who may be accessing the satellite networks through fixed and portable terminals. However, ATM over satellite comes with a host of problems that needs to be addressed to be able to provide seamless communications. Resource management is a major problem that may face the satellite network operator because of the latency and the limited bandwidth resource that characterise satellite communications.

Acknowledging this problem, different researchers have proposed new MAC protocols and up-link assignment schemes to enhance the bandwidth resource utilisation and achieve QoS provision. This thesis complements the already proposed work on MAC protocols by proposing different approaches to the problem of resource management to achieve enhanced resource utilisation and better QoS provision. Two novel scheduling strategies are proposed, evaluated and compared to a Weighted Round Robin (WRR) scheduling scheme. The main idea is to allow the lower-class connections to intermittently make use of network resources allocated to higher-class connections without causing an excessive deterioration of the higher-class connections perceived QoS. The scheduling strategies were both found to achieve a drastic reduction of the endto-end delay experienced by lower-class connections.

An innovative resource allocation scheme was also proposed and evaluated. The scheme relies on the prediction of traffic arrival, where the bandwidth is granted to a connection according to the received predicted traffic. The scheme was found to outperform classical

bandwidth-on-demand in the utilisation of both the ground and space segment resources. Different prediction techniques were investigated to identify a suitable technique in terms of prediction efficiency and implementation simplicity for incorporation as the core technique for the proposed scheme. This exercise has provided an added contribution from the thesis, which was the proposition of Exponential Weighted Moving Average (EWMA) prediction for dynamic resource allocation.

# **TABLE OF CONTENTS**

ABSTRACT	3
TABLE OF CONTENTS	5
LIST OF TABLES	9
LIST OF FIGURES	10
ACKNOWLEDGEMENTS	14
GLOSSARY	15
1. INTRODUCTION	18
1.1 Overview	18
1.2 Objectives of the thesis	20
1.3 Contributions of the thesis	21
1.4 Organisation of the thesis	22
2. ATM BROADBAND SATELLITE NETWORKS	25
2.1 Broadband networks	25
2.1.1 Multimedia services	25
2.2 B-ISDN and ATM	26
2.2.1 ATM network performance objectives	26
2.3 Broadband satellite networks: Benefits	28
2.4 Broadband satellite networks: Challenges	29
2.4.1 Error Control	29
2.4.2 Multiple Access (MAC) protocols	31
2.4.3 Resource Management	33
2.4.3.1 Uplink access schemes	34
2.4.3.2 Scheduling schemes	38
2.4.4 Traffic management procedures	41
2.4.4.1 Reactive congestion control schemes	42
2.4.4.2 Preventive congestion control schemes	42
2.4.4.3 Predictive congestion control schemes	43

	2.5 ATM performance objectives for satellite systems	44
	2.6 Non-GSO satellite networks	45
	2.7 Summary	46
3.	RESOURCE ALLOCATION IN BROADBAND SATELLITE	
	NETWORKS	47
	3.1 Overview.	47
	3.2 The broadband satellite network	48
	3.3 The Resource Management (RM) simulation tool	48
	3.3.1 The RM simulator architecture	49
	3.3.1.1 Requests Generation Process	51
	3.3.1.2 Traffic Scheduling Process	51
	3.3.1.3 Permits Generation Process	52
	3.3.1.4 Data Transmission Process	52
	3.3.2 Implemented traffic models	53
	3.3.3 Performance parameters employed	56
	3.4 Performance evaluation of Hung's implementation	57
	3.5 Validation of the RM simulator	62
	3.6 Summary	66
4.	NEW STRATEGIES FOR SCHEDULING BEST EFFORT TRAFFIC IN	
	ATM BASED SATELLITE NETWORKS	67
	4.1 Motivation for improving the delay performance of best effort	
	traffic	67
	4.2 Intra and inter-scheduling	68
	4.3 Classical approaches for reducing best effort delay	69
	4.3.1 WFQ approach	69
	4.3.2 Unoccupied slots approach	70
	4.4 The proposed scheduling strategies	71
	4.4.1 The ground segment implementation	71
	4.4.2 The space segment implementation	73
	4.4.2.1 Alternation of the scheduling priorities in the	
	PA strategy	73

	4.4.2.2 Alternation of the ownership of the fixed slots in the
	QFS strategy
	4.5 Comparisons of the PA and QFS strategies
	4.6 Summary
5.	SIMULATION STUDY OF THE 'PA' AND 'QFS' STRATEGIES
	5.1 Introduction
	5.2 Limitations of WFQ scheduling
	5.3 Simulation results for the PA strategy
	5.4 Simulation results for the QFS strategy
	5.5 Summary
6.	A PREDICTIVE RESOURCE ALLOCATION SCHEME FOR ATM
	SATELLITE NETWORKS
	6.1 Introduction
	6.2 The functional architecture of the predictive scheme
	6.3 Linear Predictor
	6.3.1 Adaptive least mean square error linear predictor
	6.4 Artificial Neural Network (ANN) predictor
	6.4.1 Fundamentals of Neural networks
	6.5 Exponential Weighted Moving Average (EWMA) predictor
	6.6 Impact of the predictive implementation on the ground segment
	6.7 Modifications to the RM scheduler simulation tool
	6.8 Summary
7.	SIMULATION STUDY OF THE PREDICTIVE RESOURCE
	ALLOCATION SCHEME
	7.1 Introduction
	7.2 Prediction of MPEG traffic using the three prediction
	techniques
	7.3 Simulation and results
	7.3.1 End-to-End delay and delay variation results (receiving Earth
	Terminal)

7.3.2 Ce	ell discard at the receiving Earth Terminal	121
7.3.3 Bu	affer space requirement at the sending Earth Terminal	126
7.4 The target scheme for the predictive resource allocation scheme		
7.5 Position of slots in the TDMA frame		
7.6 Summary 13		
8. CONCLUSIONS AND FURTHER WORK 1		
APPENDIX A	EXACT FLUID-FLOW ANALYSIS	138
APPENDIX B	THE BACKPROPAGATION LEARNING ALGORITHM	142
APPENDIX C	PUBLICATIONS	146
APPENDIX D	REFERENCES	148

## **LIST OF FIGURES**

Figure 2.1 : Format of up-link MF-TDMA frames
Figure 2.2 : Transmission of a cell using variable-rate demand assignment
Figure 2.3 : Distribution of the broadband satellite network functions
between the ground and space sections
Figure 2.4 : Resource management different functions. 46
Figure 3.1 : The broadband satellite network
Figure 3.2 : Two stage network model (adapted from [HUNG2])
Figure 3.3 : Connections generation modelling
Figure 3.4 : Signalling procedure for ATM cells transmission
Figure 3.5 : 2-state MMDP (ON/OFF) model.53
Figure 3.6 : Voice sources modeled as a superposition of identical ON/OFF
Sources
Figure 3.7 : MPEG Group of Pictures (GOP) pattern.55
Figure 3.8 : Mean delays of the data connections
Figure 3.9 : Mean delays of the voice and video connections
Figure 3.10 : Mean buffer sizes for the data connections
Figure 3.11 : Mean buffer sizes for the voice and video connections
Figure 3.12 : Cell loss experienced by the video connection
Figure 3.13 : Two-point delay distribution for the video connection
Figure 3.14 : Mean TDMA frame utilisation.62
Figure 3.15 : Observed queuing behavior
Figure 3.16 : Requests and Permits propagation within the simulated
network
Figure 3.17a : Cell loss with an on/off voice source.65
Figure 3.17b : Cell loss with an on/off data source.65
Figure 4.1 : Intra and inter-scheduling in the satellite network68
Figure 4.2 : The probabilistic alternation process action in the PA and QFS

## LIST OF TABLES

Table 2.1 : ITU's ATM QoS class definition and network performance	
Parameters [ITU4]	27
Table 2.2 : Assignment schemes for ATM traffic classes	36
Table 2.3 : ATM performance objectives for satellites (Class-1 services)	44
Table 3.1 : Simulation parameters for the satellite network	57
Table 7.1 : Statistic of the encoded sequences used in the simulation study	119

strategies
Figure 4.3 : Ground component implementation of the PA and QFS
strategies
Figure 4.4 : The on-board scheduler realisation of the PA strategy
Figure 4.5 : The on-board scheduler realisation of the QFS strategy
Figure 4.6 : Ground component implementation of the QFS strategy (no increase
in down-link signalling)
Figure 5.1 : Data1 mean delay (WFQ)
Figure 5.2 : Data2 mean delay (WFQ)
Figure 5.3 : Data1 mean buffer space requirement (WFQ)
Figure 5.4 : Data2 mean buffer space requirement (WFQ)
Figure 5.5 : Data mean delay (PA)
Figure 5.6 : Data mean buffer space requirement (PA)
Figure 5.7 : Video mean delay (PA)
Figure 5.8 : Video mean buffer space requirement (PA)
Figure 5.9 : Video cell loss (PA)
Figure 5.10 : Video cell delay variation (PA)
Figure 5.11 : Frame utilisation (PA)
Figure 5.12 : Data mean delay (QFS)
Figure 5.13 : Data mean buffer space requirement (QFS)
Figure 5.14 : Video mean delay (QFS)
Figure 5.15 : Video mean buffer space requirement (QFS)
Figure 5.16 : Video cell loss (QFS)
Figure 5.17 : Video cell delay variation (QFS)
Figure 5.18 : Frame utilisation (QFS)
Figure 5.19 : Data mean delay 'two different Z values' (QFS)
Figure 5.20 : Data delay and Video QoS with the PA and QFS strategies
Figure 6.1 : Autocorrelation function of the frame sizes of the Jurassic Park movie
Sequence
Figure 6.2 : Autocorrelation function of the GoP sizes of the Jurassic
Park movie sequence

Figure 6.3 : Future TDMA frames appearing during the same time period as that
of a predicted MPEG frame
Figure 6.4 : A predictive request being transmitted to the satellite
Figure 6.5 : Position of reserved slots in the TDMA frames
Figure 6.6 : Non adaptive Linear predictor
Figure 6.7 : Adaptive Linear predictor
Figure 6.8 : Linear prediction schemes for terrestrial and satellite environments
Figure 6.9 : Model of a neuron processing unit
Figure 6.10 : A sigmoid (hyperbolic tangent) activation function
Figure 6.11 : An Architectural graph of a multi-layer feedforward ANN
Figure 6.12 : Prediction of the MPEG trace using Neural Networks
Figure 6.13 : Optimisation of the discount parameters
Figure 6.14 : A conceptual design for the Earth terminal when an EWMA
predictive resource allocation scheme is implemented
Figure 6.15 : Requests generation for the real time video connections
Figure 7.1a : Prediction of a scene from a Formula One race event using the
Linear, EWMA, and Neural predictors (start of trace)
Figure 7.1b : Prediction of a scene from a Formula One race event using the
Linear, EWMA, and Neural predictors (middle of trace)
Figure 7.1c : Prediction of a scene from a Formula One race event using the
Linear, EWMA, and Neural predictors (end of trace)
Figure 7.2a : Comparison of the three prediction techniques (Formula One)
Figure 7.2b : Comparison of the three prediction techniques (Talk Show)
Figure 7.3 : EWMA prediction with and without the trend term
Figure 7.4 : Mean End-to-End delay performance (Terminator trace)
Figure 7.5 : Two-point delay distributions (Terminator trace)
Figure 7.6a : Cell discard rates (Terminator)
Figure 7.6b : Cell discard rate (Star Wars)
Figure 7.6c : Cell discard rate (Jurassic Park)
Figure 7.6d : Cell discard rate (Soccer)

Figure 7.6e : Cell discard rate (Formula One)	124
Figure 7.6f : Cell discard rate (News and advertisements)	125
Figure 7.6g : Cell discard rate (Talk show)	125
Figure 7.7a : Maximum buffer size (Terminator)	127
Figure 7.7b : Maximum buffer size (Star Wars)	127
Figure 7.7c : Maximum buffer size (Jurassic Park)	128
Figure 7.7d : Maximum buffer size (Soccer)	128
Figure 7.7e : Maximum buffer size (Formula One)	129
Figure 7.7f : Maximum buffer size (News and advertisements)	129
Figure 7.7g : Maximum buffer size (Talk show)	130
Figure 7.8a : Cell discard rates achieved when the slots are reserved at the	
start and end of the TDMA frame (Terminator)	131
Figure 7.8b : Maximum buffer size achieved when the slots are reserved at	
the start and end of the TDMA frame (Terminator)	132
Figure A1 : Burst scale congestion occurring as a result of arrival rate exceeding	
the service rate	138
Figure B1 : An architectural graph of the backpropagation algorithm (adapted	
from [HAYK2] and [FAN]	143

## ACKNOWLEDGEMENT

Many people have contributed to this dissertation in various ways. There is only space to acknowledge a few. This work would not have been possible without the encouragement, support, and wisdom of my supervisor, Professor Laurie Cuthbert. Special thanks to Dr. John Schormans for the countless technical discussions and comments.

I wish to thank all the past and present researchers in the Telecommunication Research Group, especially my good friends Eliane Bodenese, Felicia Holness, Arif Al-Hammadi, Robert Stewart, Veselin Rakocevic, Dr. Zhiyuan Lou, Dr. Tijana Timotijevic and Dr. Babul Miah.

## GLOSSARY

ABR	Available Bit Rate
ANN	Artificial Neural Network
AR	Autoregressive
ARMA	Autoregressive Moving Average
ARQ	Automatic Repeat Request
ATM	Asynchronous Transfer mode
BAF	Broadband Access Facilities project
BER	Bit Error Rate
B-frames	Bi-directional predicted frames
B-ISDN	Broadband Integrated Services Digital Network
BP	Backpropagation
CAC	Connection Admission Control
CBR	Constant Bit Rate
CC	Congestion Control
CCDF	Complimentary Cumulative Density Function
CDMA	Code Division Multiple Access
CDV	Cell Delay Variation
CER	Cell Error Ratio
CFDAMA	Combined Free/Demand Assignment Multiple Access
CFRA	Combined/Fixed Reservation Assignment
CLP	Cell Loss Priority
CLR	Cell Loss Ratio
CMR	Cell Misinsertion Rate
CTD	Cell Transfer Delay
DA	Demand Assignment
DAMA	Demand Assignment multiple Access
DFQ	Delayed Frame Queuing
ET	Earth Terminal
EWMA	Exponential Weighted Moving Average

FD/CDMA	Frequency/Code Division Multiple Access
FDMA	Frequency Division Multiple Access
FIFO	First In First Out
FODA-TDMA	First In First Out Time Division Multiple Access
GMDP	General Modulated Deterministic Process
GoP	Group of Pictures
GSO	Geostationary Satellite Orbit
HEC	Header Error Check
HRR	Hierarchical Round Robin
IBP	Interrupted Bernoulli Process
I-frames	Intra-frames
IP	Internet Protocol
IPP	Interrupted Poisson Process
ISDN	Integrated Services Digital Network
ISO	International Standards Organisation
ITU	International Telecommunications Union
ITU-T	ITU Telecommunication Standardization Sector
JPEG	Joint Picture Expert Group
kbit/s	Kilo bits per second
LEO	Low Earth Orbit
LMS	Least Mean Square
MAC	Multiple Access Protocol
Mbit/s	Mega bits per second
MCR	Mean Cell Rate
MCTD	Mean Cell Transfer Delay
MEO	Medium Earth Orbit
MF-TDMA	Multi-Frequency Time Division Multiple Access
MMDP	Markov Modulated Deterministic Process
MP	Measurement Point
MPEG	Moving Picture Expert Group
NLMS	Normalised Least Mean Square
NN	Neural Network
NRTT	Non-Real Time Traffic

nrt-VBR	real-time Variable Bit Rate
OAM	Operation and Management
OBP	Onboard Processor
OPNET	Optimised Network Engineering Tools
PA	Priority Alternation
PCR	Peak Cell Rate
PDH	Plesiochronous Digital Hierarchy
P-frames	Predicted frames
PLCP	Physical Layer Convergence Protocol
QFS	Quasi-fixed Slots
QoS	Quality of Service
RF	Radio Frequency
RM	Resource Management
RTD	Return Trip Delay
RTT	Real Time Traffic
rt-VBR	real-time Variable Bit Rate
S-ATM	Satellite ATM cell
SDH	Synchronous Digital Hierarchy
SECBR	Severely Errored Cell Block Ratio
SNR	Signal to Noise Ratio
TD/CDMA	Time/Code Division Multiple Access
TDMA	Time Division Multiple Access
TDMA	Time Division Multiple Access
UBR	Unspecified Bit Rate
VAD	Voice Activity Detector
VBR	Variable Bit Rate
VC	Virtual Channel
VCI	Virtual Channel Identifier
VP	Virtual Path
VPI	Virtual Path Identifier
WFQ	Weighted Fair Queuing
WRR	Weighted Round Robin

## **CHAPTER 1**

## **INTRODUCTION**

#### 1.1 Overview

The interest in the Asynchronous Transfer Mode 'ATM' transmission over satellite has grown considerably because of the ubiquitous feature of communicating through satellites, which allows users to connect to terrestrial networks from any location on the globe. Through satellites, network boundaries are extended enabling the delivery of ATM capabilities and multimedia services to a large population of users [AWAD1]. Multimedia services are mostly carried over ATM and significantly over IP in the future. However, backbone IP will be largely carried over ATM, hence the emphasis in this thesis is on ATM based satellite networks.

A number of multimedia satellite systems have already been proposed such as the Astrolink, EuroSkyWay, SkyBridge, and Cyberstar systems [ASTRO], [EUROW], [SKYBR], [CYBER], [LOSQU], [MERTZ]. Such systems will use transport architectures based on ATM [ORS1], [MERTZ].

Several researchers have targeted the problem of Quality of Service (QoS) provision for multimedia traffic over ATM based broadband satellite networks [ORS2]. The work reported in the literature proposes a number of methods for assigning the bandwidth resource in the satellite network to connections of different classes. Although the proposed assignment schemes can achieve QoS provision to the higherclass<sup>1</sup> connections, they can exacerbate the delay experienced by lower-class connections (such as best effort). This is because in those schemes, lower-class connections receive low priority in the assignment process of the already scarce satellite bandwidth resource.

<sup>&</sup>lt;sup>1</sup> The classifications of 'higher' and 'lower' suggests that higher-class connections receive high priority when allocating the network resources whereas lower-class connections receive low priority.

Best effort traffic such as email and web browsing may be a substantial component of the multimedia traffic over the satellite link. Thus from the point of view of the satellite network operator, it is vital to manage the bandwidth resource efficiently, not only to provide QoS to higher-class connections but also to minimise the delay of lower-class connections to maximise revenue.

In a satellite environment the resource management entity may face difficulty in achieving the delay and cell loss requirements for higher-class connections. The problems being:

- A satellite link is characterised by long propagation delay especially when Geostationary (GSO) satellites are used.
- The bit rates of satellite links are currently typically a few Mbit/s (2 ~ 34Mbit/s).

The above two problems can also impinge on the size of buffers at the earth terminal. For example, with a permit-based MAC protocol, arriving cells will need to wait for a long period of time (equal to the propagation delay plus the processing delay) until a permit is received from the satellite to transmit the cells. In the case of a highly loaded satellite network, the permits may allow for only a relatively small number of cells to be transmitted, resulting in a growing queue.

On-board processing (OPB) satellite technology, coupled with the use of efficient adaptive multiple access (MAC) protocols, can achieve reasonable QoS provision and differentiation for the higher-class connections, as well as saving in the buffering requirements at the earth terminals from which the connections originate. This has been reported in [HUANG], [HUNG1], [HUNG2], [NGOC1], [NGOC2], [NGUYE], [ORS3] [ORS4], [VIDAL], [ZEIN1] and [ZEIN2]. However, to achieve tight QoS requirements of higher-class connections, the resource management entity must increase the number of fixed slots in the TDMA frame (assuming a TDMA based satellite system is used) allocated to that connection, as has been reported in Hung *et al* work [HUNG1], [HUNG2]. An increase in the number of fixed slots leads to a situation where there are less slots (bandwidth) available for other connections,

especially the lower-class connections. Fixed slots may be wasted in some of the frames as a connection might not have enough ATM cells to send. Because of the latency problem, it is impossible to signal back to the satellite the status of the slots fast enough to reallocate them to other connections.

Most of the work reported so far does not address the problem of high delays for lower-class traffic in the broadband satellite networks. Also, according to the author's knowledge, most of the work that does exist has focused on the development of MAC protocols and assignment schemes for dynamic resource allocation for QoS provision for higher-class connections and none considered the implementation of prediction techniques for bandwidth requests for the resource allocation exercise. Having a predictive allocation scheme coexisting with the current or proposed MAC protocols can achieve better utilisation of the bandwidth resource and better QoS provision than is possible with having a classical bandwidth-on-demand scheme.

## **1.2** Objectives of the thesis

To compete and win in a competitive environment, satellite network operators must not only offer revenue-generating enhanced services ahead of the competition but also be able to do this efficiently and guarantee QoS requirements. In order to provide the satellite network operator with a degree of differentiation to achieve improved return on investment and maximise revenue, the objectives of the thesis are:

- To increase the number of best effort or lower-class connections in the satellite network by achieving improved response times of end-to-end communication for this type of connections.
- To mitigate the impact of the latency problem on the response times of end-to-end communications of higher-class connections, specifically the real time video connections, with the aim of providing enhanced QoS (cell delay, cell loss, etc.) than is possibly achieved by using classical bandwidth-on-demand schemes.

## **1.3** Contributions of the thesis

This thesis proposes a novel approach to the problem of resource management for ATM based satellite networks that:

- 1. allows substantial reduction in the end-to-end delay experienced by lower-class traffic and,
- 2. achieves guarantees of tight QoS requirements of higher-class traffic.

The proposed resource management efficiently utilises both the space and ground segment resources. The two distinct areas of resource management that are targeted are scheduling and uplink bandwidth resource allocation. The main contributions of the thesis are:

- Two scheduling strategies for class-2<sup>2</sup>, 3 and 4 connections called the Priority Alternation (PA) and Quasi Fixed Slots (QFS) have been proposed and evaluated. It has been shown that by implementing either of the two strategies in a broadband satellite network, the delay performance of best effort (class-4) connections dramatically improves. The PA strategy allows the intermittent promotion of best effort connections priorities during the scheduling process. Similarly the QFS strategy allows the intermittent ownership of fixed slots (*TDMA frames slots*) already assigned to higher-class connections. Slight<sup>3</sup> degradation of the QoS of higher-class connections was observed as a result of implementing either of the strategies. However, the degradation is tolerable, and in the case where the QFS strategy is implemented, the degradation can be controlled.
- A predictive resource allocation scheme for the real time video connections (class-1) was proposed and evaluated. The connections were assumed to be MPEG coded. The scheme predicts the bandwidth requirement in future TDMA frames i.e. the required number of slots in the TDMA frames. This is achieved by predicting the sizes of the MPEG frames and mapping the predicted sizes to the

<sup>&</sup>lt;sup>2</sup> Higher-class connections are classes 1, 2 and 3 with class-1 being the most stringent class in terms of QoS requirements. The lower-class connections are class-4.

<sup>&</sup>lt;sup>3</sup> No appreciable change of the QoS from the user's point of view.

future TDMA frames as bandwidth requirements. Three prediction techniques were considered: Linear, Exponential Weighted Moving Average, and Neural prediction. The proposed predictive resource allocation scheme can achieve very tight QoS requirements of the video class-1 connections, utilising less bandwidth as well as less Earth Terminal buffer space.

• These predictive techniques vary in the degree of efficiency in terms of the ratio of the actual size of an MPEG frame to the prediction error. The Linear predictor that was adopted by Adas [ADAS] for MPEG traffic prediction was worse than the proposed EWMA predictor. The Neural predictor achieved the best results. However, although the EWMA predictor achieved slightly less efficient results than the Neural one, it is less computationally complex and more feasible to implement in a satellite network than both the Linear and the Neural predictors. The proposed EWMA predictor can thus be utilised in terrestrial networks and replace the Linear predictor in dynamic bandwidth allocation schemes such as the predictive scheme proposed in [ADAS].

### **1.4 Organisation of the thesis**

Following this introductory chapter, which summarises the motivation, objectives and contributions of the research, chapter 2 presents an overview to ATM based broadband satellite networks. First, broadband networks and multimedia services are briefly discussed. An introduction to ATM as the transfer mode for B-ISDN then follows. Here the ATM network performance objectives as laid down by the ITU are presented. The advantages of using ATM over satellite are discussed. The challenges are emphasised and a review is provided on error control, access protocols, resource management and access schemes, scheduling schemes and traffic management procedures. Having provided information on the above, the ATM Forum ATM performance objectives for satellite systems are discussed.

In order to develop a resource management scheme for the satellite network that is capable of guaranteeing high QoS requirements for the delay sensitive connections (such as real time video), and also increase the number of best effort connections using the satellite network, it is necessary to understand the limitations of the current proposed resource management schemes. One current proposed implementation is that proposed by Hung *et al* of the University of Waterloo, Canada. Chapter 3 therefore focuses on the performance of Hung's implementation. A resource management tool was built using the OPNET platform, which is then used to simulate Hung's implementation. The design of the simulation tool is discussed in this chapter as well as the traffic models implemented. The performance parameters used to evaluate the performance of Hung's implementation are summarised, and the results of the evaluation exercise are presented. There is also a section on the validation of the simulation tool.

Chapter 4 explains the proposed scheduling strategies (PA and QFS) for class-2, 3, and 4 connections. The motivation for proposing the two strategies is discussed and the hierarchical nature of scheduling in satellite networks is presented. After a review of weighted fair queuing (WFQ) and its inefficiency in reducing delays of best effort connections in the satellite network, a detailed explanation of the ground and space segment implementations for both of the proposed strategies is provided. Chapter 5 presents the simulation results of a WFQ scheduling scenario in a highly loaded satellite network. Simulation results for both the PA and QFS strategies are presented and a comparison between the two proposed strategies is given.

After explaining how the delay performance of the best effort traffic can be improved using the proposed scheduling strategies, chapters 6 and 7 concentrate on the improvement of QoS deliverance for the real time video (class-1) traffic. Chapter 6 explains the proposed predictive resource allocation scheme for ATM satellite networks. First the functional architecture of the predictive scheme is presented. The functional architecture basically explains how the future bandwidth requirements are generated and transmitted to the satellite i.e. how the prediction of the MPEG traffic is achieved and how the predicted MPEG frames are mapped onto the future TDMA frames. This is then followed by a discussion of three predictive techniques (Linear, EWMA and Neural). The impact of the predictive scheme on the design of the Earth Terminal is explained. Chapter 7 provides an evaluative overview of the predictive resource allocation scheme. Prediction results from the three predictive techniques are presented followed by a presentation of performance results of the predictive resource allocation scheme. MPEG traces with different properties are used to emphasis the superiority of the proposed predictive scheme over classical bandwidth-on-demand schemes.

Finally, Chapter 8 presents the conclusions of this thesis and suggests further research work.

## **CHAPTER 2**

## ATM BROADBAND SATELLITE NETWORKS

### 2.1 Broadband networks

ITU-T Recommendation I.113 [ITU1] defines the term "*broadband*" as "a service or system requiring transmission channels capable of supporting rates that are greater than the primary access rate".

The two main categories for broadband services according to ITU-T Recommendation I.211 [ITU2], are interactive services and distribution services. Interactive services can be further categorised into [ITU2] [SUN1] [ONVUR] :

- Message services; such as video and data mail.
- Conversational services; such as video telephony, video/audio information transmission, video conferencing, high-speed digital information, file and document transfer.
- Retrieval services; such as document and data, video and high resolution image.

Distribution services: such services can be without user presentation control (such as TV and services with user presentation control (e.g. Teletext).

#### 2.1.1 Multimedia services

The road map from narrowband networks to broadband is being driven by the anticipated demand for multimedia services. Onvural [ONVUR] refers to the term *"multimedia"* as: "the representation, storage, retrieval and distribution of machine-processable information expressed in multiple media, such as text, voice, video, graphics, image, audio, and video".

The two main features of multimedia traffic, from the network point of view, are that it imposes strict real time performance requirements and generates large amounts of bit streams, often with high bit rates.

## 2.2 B-ISDN and ATM

ATM has been conceived as a multi-service and cell-based technology ideal for supporting a wide variety of traffic types and access methods. ATM is promoted as the network convergence layer integrating legacy services and allowing fine-tuned allocation of network capacity to the different traffic types without the inefficiencies associated with traditional overlay network architectures.

Moreover, ATM's (QoS) tools enable sophisticated quality control beyond simple bandwidth measures. It is thus not surprising that ATM has been recognised by the ITU as the "target multiplexing and switching principle" for B-ISDN [ITU3].

ATM is a connection-oriented technique. In ATM, the information to be transferred is packed into fixed-size cells. An ATM cell has a 48 octet information field and a 5 octet header. The header field carries the information that pertains to the ATM layer functionality. Two important fields in the header are the Virtual Path Identifier (VPI) and the Virtual Channel Identifier (VCI) that constitute the routing information.

#### 2.2.1 ATM network performance objectives

The end-to-end ATM Layer network performance parameters and objectives for B-ISDN are defined in ITU-T Rec. I.356 [ITU4]. The performance parameters are the Cell Error Ratio (CER), Cell Loss Ratio (CLR), Cell Misinsertion Rate (CMR), Severely Errored Cell Block Ratio (SECBR), Cell Transfer Delay (CTD), Cell Delay Variation (CDV) and Mean Cell Transfer Delay (MCTD).

To accommodate the characteristics and the requirements of various traffic types, [ITU4] defines various Classes of Service:

• Class-1 (stringent Class) is a delay sensitive class and it is intended to support Constant Bit Rate (CBR) and real-time Variable Bit Rate (rt-VBR) services such as telephony and videoconference.

- Class-2 (Tolerant Class) is a delay tolerant class and supports Available Bit Rate (ABR) and non real-time Variable Bit Rate (nrt-VBR) such as video and data.
- Class-3 (Bi-Level-Class) supports VBR and ABR services such as high-speed data.
- The last class, Class-4 (Unspecified Class), supports Unspecified Bit Rate (UBR) services such as file transfers and email.

Table 2.1 Provides the ATM Layer performance objectives for the various service Classes. These objectives are not complete and may be revised in the future based on real operational experience.

	CTD	2-pt.	CLR0+1	CLR0	CER	CMR	SECBR
		CDV					
<u>Default</u>	no	no	no	no	4*10-6	1/day	10-4
<b>Objectives</b>	default	default	default				
<b>QoS Classes</b>							
Class 1	400 ms	3 ms	3*10-7	none	default	default	default
(stringent class)							
Class 2	U	U	10-5	none	default	default	default
(tolerant class)							
Class 3	U	U	U	10-5	default	default	default
(Bi-level class)							
Class 4	U	U	U	U	U	U	U
(Unspecified)							

# Table 2.1ITU's ATM QoS class definition and network performanceParameters [ITU 4]

In Table 2.1, the "U" above denotes an unspecified value. CLR0 represents the number of corresponding lost cells outcomes to the ratio of the number of cells transmitted with the Cell Loss Priority bit set to zero (CLP=0).

CLR1 represents the number of corresponding lost cells outcomes to the ratio of the number of cells transmitted with the Cell Loss Priority bit set to one (CLP=1). CLR0+1 represents the number of corresponding lost cells outcome to the total number of cells transmitted.

## 2.3 Broadband satellite networks: benefits

Satellite transmission systems are increasingly gaining recognition for their benefits and advantages that will play a significant role in establishing the global information infrastructure. Broadband satellite networks can be vital components of multimedia communication. Some of the benefits of satellite networks are listed below:

- Provision of broadband services over a wide geographical area, including remote, rural, urban and inaccessible areas.
- Simplicity of network expansions: new users can easily be added to the system by simply installing the broadband stations at customer premises.
- Flexible bandwidth-on-demand capabilities: this significant strength of satellite communications ideally matches the main characteristics of ATM networks, which provide bandwidth-on-demand and multimedia services.
- Flexibility in terms of network configuration and capacity allocation to different sites which use broadband networks in various geographical areas.
- Broadcast and multipoint-to-multipoint capabilities as well as fast network set-up, that can be useful in deploying multipoint-to-multipoint capabilities in broadband networks.
- Provision of alternative channels for connections where the bandwidth demands and traffic characteristics are unpredictable.
- Back up for optical fibre broadband networks: failure of fibre links, or problems of network congestion, can be overcome easily by routing traffic through a satellite channel on a demand basis.

## 2.4 Broadband satellite networks: challenges

Broadband satellite networks are fundamentally different form terrestrial networks. There are a number of challenges that must be resolved before broadband satellite networks can operate in full service. These challenges can be summarised as follows:

- Choice of efficient error control mechanisms to minimise the impact of burst errors on user applications as well as on the cell transport methods.
- Choice of flexible and complexity free MAC protocols and CAC algorithms to maximise the space and ground segments resources utilisation.
- Provision of a robust and highly efficient resource management entity capable of providing QoS guarantees to the different type of connections.
- Enhancement of terrestrial traffic and congestion control procedures to cope with the limited bandwidth and high delays (≈ 250ms for GSO) inherent in satellite communications.

The next subsections will explore in more detail the above challenges and the relevant research work carried out to address them.

#### 2.4.1 Error Control

Problems regarding reliable data transmission via satellite links are very important because they can affect user applications significantly. A characteristic of satellite links is the burst error phenomenon.

Burst errors in the satellite environment are caused by the variations in the satellite link attenuation and the use of convolutional coding to compensate for channel noise. ATM is well defined for a fibre optic transmission link with random error characteristics rather than error burstiness. Since the ATM Header Error Check (HEC) is only able to correct single-bit errors, burst errors in the ATM header cannot be corrected. Therefore, there might be a significant increase in ATM cell discard probability, which is defined as the ratio of the number of ATM cells that are discarded due to uncorrectable errors to the total number of cells received. A number of experiments have been undertaken to quantify the effect of burst errors on ATM transmission over satellite links. These experiments concentrated on the impact on user applications [KALTE], [IVANC] as well as on the cell transport mechanisms (SDH, PDH and PCLP) [CHITR1], [AKYIL].

Solutions to combat burst errors for reliable data transmission over ATM satellite links have been proposed by many workers in the field. They have identified four possible solutions to the problem:

- Employment of interleaving mechanisms. It has been shown that the ATM cell discard probability and the probability of undetected errors can be reduced significantly if interleaving is performed on the ATM cell header and/or payload [CHITR1], [FARSE], [LUNSF], [FAIR1], [FAIR2], [FAIR3] and [LIM]. The interleaving can be inter-cell or intra-cell interleaving and in a bit or byte basis.
- Implementation of error recovery algorithms. These are based on the automatic repeat request (ARQ) techniques such as selective-repeat ARQ and are only useful for loss-sensitive, delay-insensitive services.
- Implementation of efficient coding schemes. Previously, many satellite modems employed mainly convolutional codes with Viterbi decoding to achieve a typical BER ranging from 10<sup>-3</sup> to 10<sup>-5</sup> [MARSH]. However, this is regarded as unreliable for satellite ATM networks, which should support loss-sensitive ATM traffic. Currently, BER rates of 10<sup>-10</sup> are achievable through employment of an efficient coding scheme called concatenated coding [LUNSF], [TESFA]. The key feature of concatenated coding is that an outer code is added before the convolutional encoder, which means that the convolutional inner code is concatenated with the outer code before transmission. By using these combined codes, satellite links would show very long intervals between errors. The success of the concatenated coding depends on the choice of the outer code.
- Improving the ATM header error check (HEC). The suitability of the standard cell Header Error Control (HEC) for the transmission over satellite links has been

investigated in the STRATOSPHERIC [STRAT] project and a modification was suggested to the format of the cell to be transmitted over a coded satellite link. The modification allows the HEC code to correct 3 bits error in the header and to detect multiple errors. With such modification, the HEC requires 2 more octets than the traditional standard HEC.

Error correction is not considered further in this thesis; it is assumed that to the ATM layer the satellite link is to all intents and purposes almost error free (BER >  $10^{-10}$ ).

#### 2.4.2 Multiple Access (MAC) protocols

The are six generic ways of sharing the communication resources amongst users in satellite systems [PEYRA]:

- 1. Frequency Division Multiple Access, FDMA, in which sub-bands of frequency are specified.
- 2. Time Division Multiple Access, TDMA, in which periodic time slots are identified.
- 3. Code Division Multiple Access, CDMA, in which specified number of a set of orthogonal spread spectrum codes are allocated.
- 4. Space Division Multiple Access, in which spot beam antennas are used to separate radio signals.
- 5. Polarisation Division Multiple Access, in which orthogonal polarisation is used to separate signals, allowing reuse of the same frequency band.
- 6. Multiple Random Access, in which packets are transmitted by each user without any restriction on the time of transmission. Examples of such MAC protocols are the ALOHA and slotted ALOHA protocols.

Several types of multiple access can be combined [MARAL] such as Frequency/Code Division Multiple Access (FD/CDMA), Time/Code Division Multiple Access (TD/CDMA) and Frequency/Time Division Multiple Access (FD/TDMA). The latter is some times referred to as Multi-Frequency Time Division Multiple Access (MF-TDMA). The above combined-access schemes are better suited for ATM technology, especially MF-TDMA, which is considered to be the main candidate for future ATM satellite networks because it allows the possibility of "on-demand" allocation of bandwidth [HUNG1] and, therefore, takes advantage of the flexibility and statistical multiplexing capabilities of ATM. MF-TDMA is a variant of TDMA. In conventional TDMA, only one frequency is used; all earth terminals transmit and receive on a single frequency, whatever the destination of the bursts. Therefore, it does not provide power efficiency, and the satellite link speed is limited. To solve this inefficiency, MF-TDMA was proposed. MF-TDMA uses a much lower transmission rate enabling better multimedia communication than is possible with TDMA with reduction of the satellite antenna sizes, satellite and Earth Terminal transmission powers, Earth Terminal size, and increase of satellite network bandwidth. Each Earth Terminal may transmit on any one frequency at a given time. If the ATM cell payload capacity on each frequency is B Mbit/s and the number of frequencies is N, the overall payload capacity is N x B Mbit/s [AKYIL]. Figure 2.1 shows the MF-TDMA frame format.

Other variants of the generic access methods, incorporate a specific resource assignment scheme or combination of assignment schemes.

Although MF-TDMA can provide improved satellite network bandwidth, it is not widely employed in current satellite systems but it is assumed for all new multimedia satellite systems. Hence, TDMA is chosen as the multiple access scheme throughout this work due to its widespread use and high throughput when compared with other access schemes especially FDMA. The details of the MAC protocol are not needed for the work described in this thesis; it is assumed that there is a suitable scheme in operation.



Figure 2.1 Format of uplink MF-TDMA frames

#### 2.4.3 Resource Management

Sun *et al* [SUN1] have suggested that, for an effective implementation of resource management (RM), the allocation of the satellite link bandwidth should be mapped into the Virtual Path (VP) architecture in the ATM networks and that each connection mapped into the Virtual Channel (VC) architecture.

The RM entity has two main functions, namely resource allocation and flow control. Some of the sub-functions of the RM entity can span one or more other entities. For instance the uplink access (assignment) and the scheduling sub-functions, which are sub-functions of the resource allocation, are also functions of the MAC protocol.

A number of researchers (such as Ananasso *et al* [ANANA] ), deal with the uplink assignment problem in the context of the MAC protocol. Hence, they regard the uplink assignment as being solely a function of the MAC protocol. However, in this thesis, the assignment will be treated in the context of the resource management, since the resource management is the central theme of the thesis and not the MAC protocol.

The work described in this thesis reflects the fact that the resource management is dependent on both the uplink access scheme and the scheduling scheme.

#### 2.4.3.1 Uplink access schemes

According to Hung *et al.* [HUNG2], there are five specific uplink access (assignment) schemes to support the connections:

- Random access,
- Fixed assignment,
- Fixed-rate demand assignment,
- Variable-rate demand assignment and
- Free assignment.

In this section the overall functionality of each of the above schemes will be discussed and a mapping of the schemes to the different ATM service classes will be presented, as well as examples on how MAC protocols utilise the schemes to service the different types of traffic.

With random access, connections from different terminals may broadcast cells simultaneously; thus results in possible collisions. Collisions corrupt the data and hence retransmissions become necessary for reliable data communication. A retransmission represents an extra delay. Random access schemes cannot provide guarantees on delay and cell loss. Hence as Ors *et al* [ORS1] indicated, the use of random access is discouraged for delay sensitive applications but can be used with best effort applications where performance guarantees are not an issue.

With fixed assignment, a terminal's connection is permanently assigned a constant number of slots per frame (or some multiple number of frames) for the lifetime of the terminal or for the lifetime of the connection. The latter assignment scheme is referred to as "fixed-rate demand" assignment. Fixed assignment schemes can achieve provision of performance guarantees, even when the performance requirements are very stringent, but at the expense of frame utilisation. The frame utilisation suffers because when the connection is idle, the slots assigned to the connection are not utilised (wasted). The terminal-to-terminal delay when using this access scheme is the propagation delay (250 ms) plus the processing and queuing delays at the terminals and the satellite.

With the variable-rate demand assignment scheme, the slot(s) assignment depends on whether or not there are cells awaiting service at the connection's terminal queue. This works as follows: when a cell arrives at the terminal queue, signalling messages are sent to the satellite notifying it of the arrival. When the satellite receives this information, it dynamically assigns slot(s) to the connection. Variable-rate demand assignment is a guaranteed assignment scheme in that slots are assigned based on previously allocated resources (bandwidth), which is available for use whenever it is needed. It avoids collisions and efficiently uses the uplink capacity because the satellite is aware of the needs of the source and it responds to the need by assigning slots on a frame-by-frame basis.

If the connection does not need a slot that has been allocated to it during connection establishment, the satellite may assign the slot to others. The drawback of this scheme is the 250 ms delay from when the signalling information is sent to when the satellite's response is known (i.e. assignment) at the earth terminal. The timing diagram showing the delays associated with variable-rate DA is given in figure 2.2



Figure 2.2 Transmission of a cell using variable-rate demand assignment (adapted from [HUNG2])
Chitre *et al* [CHITR2] refers to the variable-rate demand assignment as dynamic bandwidth assignment. They specify three levels of fairness in assigning bandwidth to each earth terminal for the COMSAT Linkway 2000 satellite network [CHITR2]:

- 1. Outgoing fairness that ensures that all VCs originating from a particular terminal get a fair share of the bandwidth assigned to the terminal.
- 2. Incoming fairness that ensures that all VCs terminating at a particular terminal get a fair share of the down-link bandwidth assigned to that terminal.
- 3. System fairness that ensures that all VCs in the entire network get a fair share of the total system capacity in a fair manner.

Free assignment is concerned with the remaining slots in a frame, which have not been assigned by the fixed or demand schemes. From the satellite network's point of view, these remaining slots are considered to be the spare uplink capacity that can be freely assigned to connections in order to increase overall throughput, to relieve congestion at the ground terminal queues, or to reduce the terminal-to-terminal delay. Free assignment is the main assignment scheme for best effort connections. Table 2.2, shows the rate based and free assignment schemes that are possible for implementation with the different ATM traffic classes.

	Fixed-rate	Variable-rate	Free
	demand Assignment	demand Assignment	Assignment
CBR	$\checkmark$	-	-
rt-VBR	$\checkmark$	$\checkmark$	$\checkmark$
nrt-VBR	$\checkmark$	$\checkmark$	$\checkmark$
ABR	$\checkmark$	-	$\checkmark$
UBR	-	-	$\checkmark$

#### Table 2.2 Assignment schemes for ATM traffic classes

Different MAC protocol implementations lead to variation of the basic schemes. For instance, the FIFO Ordered Demand Assignment-TDMA (FODA-TDMA) [CELA1] uses fixed-rate demand assignment for stream traffic and variable-rate demand assignment for datagram traffic. This system uses three queues with different priorities at each earth terminal. The stream traffic FIFO queue has highest priority, followed by interactive data traffic FIFO queue. Bulk data traffic has the lowest priority since it is not delay sensitive.

Another protocol that has received considerable attention in the literature is the Combined/Fixed Reservation Assignment (CFRA) protocol [ZEIN1], [ZEIN2], [CELA1], [CELA2] and [ORS1]. Here a fixed-rate demand assignment for a minimum bandwidth is made at the beginning of a burst. If a burst is longer than a certain number of cells, extra capacity is requested using variable-rate demand. Hence this protocol tries to improve the performance of the anticipated reservation protocol by distinguishing between short and long bursts. Other protocols that make use of the assignment schemes are the Combined Free/Demand Assignment Multiple Access (CFDAMA) [NGOC1] and the Movable Boundary random/DAMA Access [NGUY1] where the frame is divided into three areas; a reservation area, a slotted Aloha channel and a DAMA channel. The boundaries between the three areas can be changed to cater for increases in the different traffic loads.

Recently, ORS *et al* [ORS1] and [ORS2] proposed a protocol that is based on the above mentioned schemes. The Random-Reservation Adaptive Assignment Protocol features a prioritised queuing of requests on-board the satellite and the use of random access for the UBR service class.

Following the same approach as that of Hung *et al* [HUNG1] and [HUNG2], the simulated ATM satellite network in this work will use a TDMA based access scheme in conjunction with the rate based and free assignment schemes. This allows for performance comparison between Hung's bandwidth on demand scheme and the author's proposed scheduling strategies and resource allocation schemes.

#### 2.4.3.2 Scheduling schemes

As mentioned earlier, the scheduling schemes are sub-functions of the resource allocation function. A scheduling scheme for multimedia satellite communications must satisfy two main criteria:

- It should be simple to implement and flexible enough to provide for a wide range of bandwidth allocations.
- It should avoid computational complexity, especially if the scheduler is housed on-board the satellite. Schedulers that use a numerical service deadline (a time stamp) for each cell and use that to determine which cell to serve, usually require massive computation complexity and storage requirements [GARRE].

The literature search revealed three research projects that targeted scheduling for broadband satellite networks. This is the ongoing work at the University of Waterloo in collaboration with Spar Aerospace Limited of Canada [HUNG1], [HUNG2] and the work reported by J.Huang *et al* [HUANG].

The researchers at the University of Waterloo proposed two scheduling implementations for uplink access. These are the Weighted Round Robin (WRR) and the Hierarchical Round Robin (HRR).

Under WRR, connections are given access opportunities in a round robin fashion. Each round robin cycle is called a WRR frame, which can be equal in length to the MF-TDMA frame (assuming such a TDMA based access scheme is used). In each frame, a connection can be assigned up to its allotted number of slots; the next connection is assigned consecutive slots immediately afterwards.

If a connection does not require any of its allotted number of slots, those slots are unassigned: such a scheduler is *non-work-conserving* or *idle*. The total number of assignments made in a WRR frame is the total of fixed and variable assignments. A free assignment algorithm is invoked after the WRR scheduler has completed an MF-

TDMA frame in order to use the unassigned slots to determine the free assignments for all connections.

The authors argue that the WRR can cause greater queuing delay jitter at the earth terminal when the WRR frame is large. The Hierarchical Round Robin (HRR) was proposed with the intention to reduce the delay jitter. HRR is a more flexible kind of round robin scheduling. HRR attempts to allocate bandwidth more "evenly" over a frame than WRR, and thereby reduce the delay jitter at the terminal queues caused by cell clumping. HRR is hierarchical in the sense that there are  $L \ge 1$  *levels* of HRR frames. A connection is allotted slots in one or more levels so that the cells are dispersed along the frame and not concentrated in one part of the frame.

Another purpose of the multiple levels is to provide different bandwidth granularities. The higher levels are supposed to offer a greater amount of bandwidth but have coarser granularity, and the opposite is true for the lower levels.

The authors provide delay bounds (maximum queuing delay) and also OBP buffer sizing results. The authors argue that the results for the OBP buffer size are sufficient to ensure that there is no overflow. However, the results for the delay bounds and OBP buffer size assumes only fixed rate allocation. Also the authors do not show through simulation how the clumping of cell is avoided and thus how the delay jitter is reduced.

J.Huang *et al* [Huang] advocates a scheduling strategy based on a Tuneable Leaky Bucket scheme. At a given time, voice, video and data calls from all earth terminals are admitted on the basis that the average traffic rate does not exceed the uplink capacity. The OBP then sets the token rates and bucket depths corresponding to the average and peak rates of the traffic type (e.g. voice, video, data) for each active earth terminal.

In each MF-TDMA frame, the OBP generates tokens into the buckets according to the set rates, and allocates to each active earth terminal a number of time frequency slots equivalent to the instantaneous capacity requested by the earth-station, or the number of tokens remaining in the bucket, whichever is smaller.

To maintain the cell loss probability below an acceptable threshold, the token rates and bucket depths are adaptively adjusted such that the flow of aggregate traffic to a downlink beam is temporarily reduced when the queue level becomes high according to the prediction made in the network controller.

This is the only scheme that utilises a scheduling strategy to regulate the traffic arriving at the down-link buffers. However, nothing is mentioned in Huang's work on how the fairness in allocations is changed or how the uplink frame utilisation and throughput are affected, when controlling the inbound traffic to the downlink buffers. Pocher and Leung [POCHE] recently proposed a scheduling architecture for the real-time component of the multimedia traffic in ATM satellite systems. Their proposed scheduling architecture can be viewed as distributed between the uplink, the downlink and the satellite queues. The underlying strategy for the architecture is a Delayed Frame Queuing protocol (DFQ) that defines the service priorities, deadlines and eligibility times of queued cells.

At the uplink queue, the authors propose a work-conserving DFQ that allows high throughput of real-time traffic and optimises the successful completion of contractual delay bounds guarantees. At the downlink, a non-conserving DFQ is proposed that allows CDV bounds to be guaranteed to end-user applications, and CDVT bounds to be guaranteed at the wireline ATM interface. At the satellite queues where there are several queue architectures, the researchers suggest that the simplest option that requires the least buffer capacity is a work-conserving DFQ.

Scheduling strategies for broadband satellite communications have not been extensively explored. The main reasons are that the satellite and the terminal are both limited in resources and computational capabilities and these limit the prospects of deployment of terrestrial based scheduling strategies in the satellite networks. In the work described in this thesis, an on-board scheduler will be considered but scheduling strategies that require huge computational complexity especially in the satellite will be avoided.

#### 2.4.4 Traffic management procedures

Congestion problems may occur when the demand for resources on-board the satellite exceeds its capacity. Congestion can rapidly increase the delay and severely limit the advantages obtained by dynamic resource sharing.

Traffic control functions that manage and control traffic to avoid congestion in ATM networks have been specified by ITU-T [ITU2] and the ATM Forum [ATMF1]. Two specific mechanisms are expected: proactive and reactive congestion control. The former allows the deletion of cells in a congested node (Selective Cell Discard), whereas the latter notifies the end users of congestion inside the network such that they should reduce their data rate (Explicit Congestion Indication).

In Selective Cell Discard, ATM cells with the CLP bit set are discarded, hence reducing the congestion (provided that there are sufficient cells of this type). However, though this method can be efficient from a control standpoint, higher-layer retransmission may occur, possibly leading to an increase in the total number of ATM cells being transmitted.

Explicit Congestion Indication is incorporated with a feedback mechanism and thus seems to be inappropriate for effective congestion control in satellite ATM environment because of the propagation delay.

When designing congestion control schemes for satellite systems, attention should be paid to the following unique characteristics of satellite communications:

- The long propagation delays between the earth terminals and the satellite, mandating predictive congestion control schemes.
- On-board storage is very expensive, i.e. congestion control schemes implemented on-board the satellite should require minimal buffering capabilities.
- On-board processing should be minimised (processing power and power consumption limitations), i.e. it is advisable to avoid complex and computational intensive procedures executed on-board.

There are in the literature a number of congestion control schemes relevant to broadband satellite systems. These schemes are discussed in the next three subsections.

#### 2.4.4.1 Reactive congestion control schemes

The scheme proposed by A.Baiocchi *et al.* [BAIOC] employs a reactive control mechanism, a so called rate-based congestion control, where the overall rate of the ABR cells sent to every down-link is controlled using a feedback mechanism so that the buffers are not saturated. In this scheme, the closer the ABR traffic rate comes to the available serving rate at the down-link buffer, the faster the decrease in the ABR rate, even if sudden downfalls of the rate cause loss of efficiency.

As mentioned earlier, a feedback mechanism is inappropriate with long propagation delay. However, in Baiocchi's scheme the allowed maximum value of the ABR rate is very much less than the service rate, so that the propagation delay does not impinge on the performance of the rate adaptation algorithm. Also the interval between two successive rate changes (increase or decrease) encompasses a whole round trip propagation delay.

Recently, Joo *et al*, of the University of Surrey [JOO] proposed a "buffer threshold" reactive congestion control scheme. In this scheme, the difference between a defined threshold and the average on-board buffer occupancy, which is determined over a control time, is used to regulate the rates of the incoming ABR traffic. Like the former scheme, the decrease in rate is performed rapidly according to an exponential control function. This scheme was analysed through simulation and was found to be robust and stable.

#### 2.4.4.2 Preventive congestion control schemes

P.Chu *et al* [CHU] proposed a preventive scheme that uses a global feedback signal to regulate the packet arrival rate of earth terminals. The satellite continuously broadcasts the state of its output buffer. When an earth terminal detects congestion in the satellite switch, it either reduces its arrival rate by discarding packets, or starts tagging excess packets as low priority. These low priority packets will be discarded on-board if congestion actually occurs.

However, the discarding scheme needs a lot of discarding at the earth terminal to achieve low on-board packet loss probability. Jain *et al* [JAIN] commenting on this scheme suggested that the tagging scheme could be a better alternative since it tolerates more uncertainty caused by the long propagation delay.

T.ORS *et al.* [ORS1] of the University of Surrey proposed a reactive/preventive congestion control policy that uses the Usage Parameter Control (UPC) function (for CBR and ABR traffic) and a dual Leaky Bucket (LB) configuration (for VBR traffic) at the ground segment. The congestion control policy uses at the space segment a Multiple Shared Buffer Scheduling (MSBS) scheme that allows cells with similar delay but different cell loss requirements to be placed in the same buffer with different loss priorities. Simulation results showed that the proposed congestion control policy is very effective for traffic management.

#### 2.4.4.3 Predictive congestion control schemes

The scheme proposed by Jang *et al.* [JANG] requires the estimation of the on-off sources characteristics (mean-on and mean-off durations) using a Maximum Likelihood estimation method. These estimated values are used to predict the transient cell loss probability at each downlink. When the QoS requirements are not met, the congestion control scheme determines the control parameters for source traffic shaping or controls the total number of connections in the system.

Le-Ngoc *et al* [LE-NG2] proposed the use of Neural Networks in a packet switch OBP satellite system to estimate traffic intensity in downlink queues and to predict traffic load status. Two neural networks are used. The first one estimates the traffic intensity from the number of packets arrived in a frame; the other one calculates congestion probability in the next two round trip delay periods.

Most of the congestion control schemes proposed for broadband satellite networks target ABR traffic. Also, how such schemes impact the resource allocation is not clear.

In the work described in this thesis, the congestion control function of the resource management entity is not considered.

# 2.5 ATM performance objectives for satellite systems

ATM Forum contribution number 98-0828 [ATMF2] proposes that the numerical values of ATM performance parameters for satellite systems can be derived by applying the allocations given in Table 2.1 to the performance objectives given in [ITU4]. As an illustration, the ATM performance objectives for a satellite link used in the international portion that provides Class-1 service and does not contain switching or cross-connect functions [ATMF2] is shown in table 2.3. To achieve such stringent performance objectives, especially the CTD objective, it is necessary to place the delay sensitive functions of the satellite system on board the satellite. Hung *et al* [HUNG2] suggested that the scheduler should preferably be placed onboard the satellite, as it is a delay-sensitive function. Other functions to be placed onboard, according to them, are switching, queuing, and flow control.

Spar Space Systems [GILDE] have been researching the feasibility of an on-board ATM switch and have proposed designs based on well-known terrestrial fabrics that meet mission requirements while attempting to minimise on-board costs. The author of this thesis recommends that the other sub-function of the resource allocation, the uplink assignment sub-function, should also be housed on board the satellite because it is delay sensitive.

Performance	ITU objective	ITU objective
parameters	end-to-end	satellite
CLR	3*E-7	7.5*E <sup>-8</sup>
CER	4*E <sup>-6</sup>	1.4*E <sup>-6</sup>
SECBR	1*E <sup>-4</sup>	3.0*E <sup>-5</sup>
СТД	400 ms	320 ms (max)
CDV	3 ms	Negligible
CMR	1/day	for further study

# Table 2.3ATM performance objectives for satellites (Class-1 services)[ATMF2]

Delay tolerant functions such as billing, Operation and Management (OAM) and the Connection Admission Control (CAC) functions can be placed in the ground segment. Figure 2.3 shows the distribution of the various satellite network's functions between the space and ground segments



Figure 2.3 Distribution of the broadband satellite network functions between the ground and space sections.

#### 2.6 Non-GSO satellite networks

ATM over satellite is achievable as has been successfully demonstrated in the CATALYST project [SUN2], where 10 to 150 Mbit/s ATM connections were used over a GSO satellite link for various broadband applications, including multimedia communications. ATM over Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) satellites is gaining momentum as LEO/MEO satellites achieve much smaller delays than GSO satellites as well as lower terminal power requirements [EVAN1], [EVAN2]. However, LEO/MEO satellites present many additional challenges such as cell sequence integrity due to the inter and intra-satellite handovers, and inter-satellite links. The dynamic network topologies of such orbits require complicated resource management, CAC, MAC and routing functions. Non-GSO networks are not considered in this thesis as they are out of the scope of the work described.

#### 2.7 Summary

This chapter presented a brief introduction to ATM technology and ATM based broadband satellite networks. The benefits and challenges of such a network were highlighted along with an outline of the different entities that it is composed of, such as the MAC protocol and resource management. The resource management entity is the targeted entity in this thesis, in particular the uplink resource allocation function, which has two distinct sub-functions: the uplink assignment and the scheduling as shown in figure 2.4. The rest of the thesis will concentrate on these two functions.



Figure 2.4 Resource management different functions

GSO satellites are chosen for this work because of their popularity, large coverage and fast deployment.

# **CHAPTER 3**

# RESOURCE ALLOCATION IN BROADBAND SATELLITE NETWORKS

#### 3.1 Overview

QoS provision for the different components of multimedia traffic in a satellite network can be accomplished through careful design of the MAC protocol coupled with an efficient resource allocation policy. The resource allocation should use robust bandwidth assignment and scheduling schemes.

Three uplink assignment schemes (fixed-rate demand, variable-rate demand and free assignment) have been discussed in the previous chapter. Several scheduling schemes have also been identified: WRR, HRR and the DFQ based scheduling schemes.

The combination of the above uplink assignment schemes and a WRR on-board scheduling scheme will be referred to as Hung's implementation through out this thesis. The HRR scheduler will not be considered as it can only exist with MF-TDMA based satellite networks as well as being essentially a WRR scheduling scheme. Similarly, the DFQ based scheduler will not be included as it was designed specifically for real-time traffic on both the uplink and downlink and does not cater for best effort traffic.

In this chapter a TDMA based satellite network loaded with delay sensitive and best effort connections will be considered. The performance of multimedia traffic under Hung's implementation will be evaluated through simulation. Prior to Hung's implementation evaluation, the simulation tool built by the author will be discussed and validated. A study of the performance of a classical resource management such as Hung's is important for comparison with the proposed resource management that will be discussed in the next chapters.

## 3.2 The broadband satellite network

Currently, most satellites act as transparent repeaters, re-transmitting the RF signal after amplifying it. To be able to efficiently provide more advanced B-ISDN communications, satellites should ideally possess on-board processing (OBP) and switching capabilities [GILDE]. The broadband satellite communication system that is considered in this work uses an on-board packet switch in conjunction with an on-board scheduler to provide broadband services to a large population of Earth terminals. The Earth terminals assumed in this work can be fixed or portable, with bursty traffic and peak bit rates (up to 8Mbit/s for portables) as shown in figure 1. TDMA is used for the uplink and TDM is used for the downlink.



Figure 3.1 The broadband satellite network

## 3.3 The Resource Management (RM) simulation tool

The performance of the resource management function in a network loaded with heterogeneous connections with different QoS requirements can be difficult to analyse with mathematical modelling. This is because the formulation of tractable mathematical models for such complex networks would entail many restrictive assumptions [ORS1] that can lead to oversimplification of the problem resulting in misleading results. The other two options are either to use direct measurements or

simulation. The author adopted the latter option, as there are currently no available satellite ATM networks with on-board processing. Hence, to qualitatively evaluate the performance of multimedia traffic under Hung's implementation, a simulation tool was developed. The RM simulator was built using a commercially available network simulation tool OPNET [OPNET]. Using OPNET, rapid development of simulators incorporating ATM and satellite models is possible. Effort can be concentrated on the modelling of the scenarios rather than the development of the underlying simulator infrastructure. The simulation level adopted for the tool was *cell level* as the aim for the simulation experiments was to monitor the QoS parameters (CTD, CLP etc.).

OPNET is a discrete event simulator that allows the user to specify the system through a graphical interface in terms of incrementally decreasing levels of abstraction from the network level down to the individual process level. The top level (the network level) is where the topology of the simulated network model is defined; the interconnection of networks (e.g. ATM links, satellite links.) is also defined at this level. At the second level (the node level) the elements that make up the network nodes and their interconnections within the networks are defined; these elements include queues, processes, sources, receivers and transmitters. The functionality of each process or queue element is defined in terms of a finite state diagram and the transition between states; this is the third level. The fourth and final level is where the processing in each of the states in the finite state diagram is defined in C code.

#### 3.3.1 The RM simulator architecture

The satellite network viewed from any connection can be considered to be two tandem output buffer switches [HUNG2] and hence can be modelled as a two-stage model as shown in figure 3.2. As the main interest is in the uplink resource management; the assignment function, scheduling function and signalling are modelled for the uplink stage. For the downlink stage, only the signalling part is modelled. At the network level, all earth terminals are grouped in one module, because mobility is not a concern as the satellite is assumed to be geostationary. Also, this approach allows the modelling of a large number of connections without the need of assigning a module to each terminal. The other module is the satellite where an OBP is assumed to exist. At the node level, the earth terminal module is implemented using a connection generating process and a global queue.



Figure 3.2 Two stage network model (adapted from [HUNG2])

The satellite module is implemented using a process node. At the process level, the generating process can be viewed as a parent process capable of invoking child processes where each child process is an individual connection of a particular nature (voice, video, data) as shown in figure 3.3. The global queue consists of a number of subqueues equal to the total number of connections.



Figure 3.3 Connections generation modelling

Each connection is assigned a subqueue to emulate an earth terminal buffer. The TDMA frame is partitioned into a signalling and information (data) areas where the signalling slots are further divided into mini-slots. Each connection has a dedicated mini-slot.

#### 3.3.1.1 Requests Generation Process

Using their dedicated mini-slots, connections signal their requests for bandwidth (outof-band signalling). A request is basically the buffer occupancy (subqueue occupancy), which can be determined using Hung's recursion [HUNG2] as shown below:

$$occupancy(n) = [occupancy(n - 1) + a(n) - o^{fx}(n) - o^{v}(n) - o^{fr}(n)]^{+} (equ 3-1)$$

where:

a(n) is the number of cell arrivals to the terminal queue of a particular connection at the end of the nth TDMA frame.  $o^{fx}(n)$  is the number of the TDMA frame slots assigned by fixed-rate demand.  $o^{v}(n)$  is the number of the TDMA frame slots assigned by variable-rate demand.  $o^{fr}(n)$  is the number of the TDMA frame slots assigned by free assignment. The assignments are received by the earth terminal during the *n*th frame. Here  $x^+ = \max(x, 0)$ .

#### 3.3.1.2 Traffic Scheduling Process

The satellite receives the requests from all connections and groups them according to their QoS requirements. This is achieved through the VPI and VCI information of each connection. WRR scheduling is implemented, whereby the group of connections with stringent QoS requirements receives the highest priority in the scheduling process. Within each group, connections are allocated part of the bandwidth (frame's slots) in a round robin manner. In this work, there are three groups: voice, video and data (the voice group has the highest priority).

An important factor in the scheduling process is the number of data slots ( $o^{frame}$ ) in the TDMA frame. For a delay sensitive connection the allocation should be such that  $0 \le o^{fx} + o^v + o^{fr} \le o^{frame}$  and for a best effort connection  $0 \le o^{fr} \le o^{frame}$ .

Moreover the number of slots allocated to any connection must be less than or equal to the available pool of data slots (*oavailable*). This pool shrinks in size as connections are allocated slots during the scheduling process period.

#### 3.3.1.3 Permits Generation Process

Once a scheduling process is executed, the allocations are broadcast to all terminals through the downlink channel. This is achieved by delivering a packet<sup>1</sup> to the Earth terminal node after a delay of 0.125 seconds to account for the propagation delay. The packet specifies the number of slots allocated to each connection in the TDMA frame that will appear after 0.125 seconds.

#### 3.3.1.4 Data Transmission Process

The receipt of a packet at the earth terminal node from the satellite invokes the transmission of the ATM cells at the terminal queues. All subqueues will be emptied according to the number of slots allocated to each connection. The author assumes that a slot is sized to accommodate one ATM cell so that the number of slots allocated is equal to the number of cells to be transmitted. Figure 3.4 presents the modelled signalling procedure.



#### Figure 3.4 Signalling procedure for ATM cells transmission

<sup>&</sup>lt;sup>1</sup> The packet has the same size as that of an ATM cell but not the format.

#### **3.3.2** Implemented traffic models

Multimedia traffic is either real time traffic (RTT) such as CBR voice, VBR voice and VBR video, or non-real time traffic (NRTT) such as data, which is usually transmitted in ABR/UBR cells. Several general source traffic models such as the ON/OFF model and the General Modulated Deterministic Process (GMDP) model are available for the multimedia source description. The ON/OFF model, which is a special case of the GMDP and which is referred to as a two state Markov Modulated Deterministic Process (MMDP), is able to describe most of the multimedia traffic adequately [KOWTH].

In the modelled satellite network, VBR voice streams are considered rather than CBR sources as it is assumed that an earth terminal will possess a sensitive Voice Activity Detector (VAD) which means ATM cells are generated and delivered only during the ON state. An ON/OFF model alternates between active and silent periods as shown in figure 3.5 where *a* and *b* are the transition rates between the ON and OFF states. During the activity period (sojourn time), ATM cells are transmitted with constant interarrival time *T*, where T = 1/PCR. The duration of the activity and silent period are exponentially distributed with mean 1/a and 1/b respectively.



Figure 3.5 2-state MMDP (ON/OFF) model

To reduce the simulation time, the voice sources were modelled by one continuoustime birth-death process [WEIN] which can be viewed as the superposition of multiple identical ON/OFF sources as shown in figure 3.6. The simulation time is reduced because for N sources 2N states are required for the ON-OFF model and N states for the birth-death model [ORS3].



Figure 3.6 Voice sources modelled as a superposition of identical ON/OFF sources

The data streams are modelled using the Interrupted Bernoulli Process (IBP) ON/OFF processes as in [KOWTH]. The IBP is a discrete version of the Interrupted Poisson Process (IPP) where the time is slotted and where a slot in an active state can probabilistically contain a cell [WU]. The active and silent periods are geometrically distributed.

VBR video traffic is expected to be one of the major traffic types that need to be supported by broadband satellite networks. Most of the video encoding in broadband networks will be done using the MPEG standard (ISO Moving Picture Expert Group) [PILIP]. An MPEG stream contains deterministic periodic sequences of frames referred to as a Group of Pictures (GoP) [ROSE1].

Frames can be encoded in three modes: intra-frames (I-frames), forward predicted frames (P-frames) and bi-directional predicted frames (B-frames) as shown in figure 3.7.



Figure 3.7 MPEG Group of Pictures (GOP) pattern

An I-frame is encoded as a single image, with no reference to any past or future frames. The encoding scheme used is similar to JPEG compression for still images. A P-frame is encoded relative to the past reference frame, the future frame (which is the closest following I- or P-frame), or both. The encoding algorithm for B-frames is similar to P-frames, except that the motion vectors may refer to areas in the future reference frames.

In this work, a number of empirical MPEG-1 video traces were used from movies such as "Jurassic Park", and sport events such as the world cup soccer final 1994, as well as other video traces. The traces vary in the degree of GoP-by-GoP correlation and frame-by-frame correlation as well as the scene activity. Most of the traces are each about half an hour long and are available from the University of Wuerzberg internet site [ROSE2]. The sequence of MPEG I, P and B frames that was used is IBBPBBPBBPBB.......(12 frames/GoP) and there are 40,000 frames in total per trace (each frame time is 40 ms). A two hours long MPEG trace of the movie "Star wars"

was also used. The trace is available from the Bellcore internet site [BELLC] and like the University of Wuerzberg traces, has an IBBPBBPBBPBB sequence.

#### 3.3.3 Performance parameters employed

The performance of the simulated satellite network was evaluated using five performance parameters:

- Cell loss ratio (CLR): the ratio of the total number of lost cells to total transmitted cells from one sender to a receiver. Cells are classified as lost if they arrive after a maximum transfer delay (max CTD).
- Mean end to end delay, which is equal to:
   Earth terminal queuing delay + propagation delay + on-board processing delay
- Mean buffer occupancy: the main performance parameter for the ground segment resources (required buffering at the earth terminal).
- Mean frame utilisation: a measure of the percentage of occupied slots in all the TDMA frames.
- ➢ Delay variation: In this work the author follows the approach adopted in the Broadband Access Facilities (BAF) project [BAF] where the distribution of the 2-point Cell Delay Variation (CDV) is used as a measure of the delay variation. The 2-point CDV parameter describes the variability in the pattern of cell arrival events at the output of a connection portion (*the receiver terminal; which can be viewed as measurement point MP*<sub>1</sub>) with reference to the pattern of corresponding events at the input to the portion (*the sender terminal; or measurement point MP*<sub>2</sub>). According to ITU I.356 [ITU4], the 2-point CDV (v<sub>k</sub>) for cell k between *MP*<sub>1</sub> and *MP*<sub>2</sub> is the difference between the absolute cell transfer delay (d<sub>1,2</sub>) between those MPs i.e. v<sub>k</sub> = x<sub>k</sub> d<sub>1,2</sub>. In all simulations, the value d<sub>1,2</sub> was chosen to be equal to the propagation delay between the receiving and sending terminals (RTD ≈ 250ms).

# 3.4 Performance evaluation of Hung's implementation

To realise Hung's implementation, delay sensitive connections should receive the highest priority in the scheduling process. However, if voice and video receive equal priority, voice connections tend to suffer because the video connections tend to usurp most of the available capacity as video usually has high peak cell rates. Hence, in the simulation, voice connections received the highest priority, then video connections and finally data connections.

The satellite network was loaded with five voice sources, one video source and two data connections. The parameters are shown in Table 3.1.

Uplink capacity	2 Mbit/s
TDMA frame duration	10 ms
Signalling slots per frame	3 slots
Mini-slots per signalling slot	10 mini-slots
RTD	250 ms
Voice -	
mean ON duration	0.352 s
mean OFF duration	0.650 s
PCR	150 cells/s
Data -	
mean ON duration	0.5 s
mean OFF duration	1.9 s
PCR	1000 cells/s
Video ( MPEG trace from the	
movie Star Wars [BELLC] ) -	
PCR	11579 cells/s
MCR	975 cells/s

#### Table 3.1 Simulation parameters for the satellite network

The parameters for the voice ON/OFF sources are widely used in the literature while those of the data were chosen to reflect the behaviour of data sources that are bursty with high PCR and long mean OFF durations.

The number of sources (voice, video, data) was chosen such that the totality of the traffic in terms of the aggregated PCR is higher than the link rate. In such a scenario, when the load is high the data connections will suffer, as most of the bandwidth resource will be allocated to voice and video connections. The degradation in the end-to-end delay performance of the data connections can then be observed.

The performance of the satellite network is observed as the number of fixed slots allocated to the video connection is increased. A parameter that serves as a measure of the fixed allocation is defined as in equation 3-2:

#### $p = fixed \ slots \ allocation \ rate / MCR$ (equ 3-2)

In the equation (p) is defined as the "relative fixed allocation rate" which is equal to the ratio of the fixed allocation rate to the mean cell rate (MCR).

Figures 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, and 3.14 provide results for Hung's implementation when applied to the considered satellite network. The voice connections, which have the most stringent QoS, are not affected, as they receive the highest priority in the WRR scheduling process. The mean delay suffered by both the data connections increases with increasing values of (p) (Figure 3.8). The mean delays are in terms of seconds for high values of (p). This is also true for the mean buffer size for both data connections (Figure 3.10) where the required buffer space increases four fold as the amount of fixed bandwidth allocated to the video connection is increased to almost three times its MCR.

The video connection achieves better mean delay, buffer space utilisation, cell loss and delay variation performance (Figures 3.9, 3.11, 3.12, and 3.13) when increasing the fixed allocation i.e. increasing (p). However, increasing (p) not only degrades the performance of the best effort connections (as this starves them of slots) but moreover impacts the utilisation of the frame as seen in figure 3.14, where the mean frame utilisation drops to very low values. The explanation is that, during some periods the video connection might have few cells or might have no cells awaiting transmission in the buffer so that some or all of the video fixed slots allocated to it in the TDMA frame are wasted.











Figure 3.10 Mean buffer sizes for the data connections



Figure 3.11 Mean buffer sizes for the voice and video connections



Figure 3.12 Cell loss experienced by the video connection



Figure 3.13 Two-point delay distribution for the video connection



Figure 3.14 Mean TDMA frame utilisation

#### 3.5 Validation of the RM simulator

The primary objective of the simulator validation activity is to demonstrate that the simulator is able to correctly model the behaviour of the chosen network scenarios. To achieve this objective, the individual models within the simulator must operate correctly and the various models that comprise the simulator must inter-work correctly. In order to test the basic functionality of each of the simulator's process models, the packet tracing functions of the OPNET simulation tool were utilised. These generate time stamped console messages that display the contents of a packet received by a process model. They enable the propagation of the following messages through the simulated satellite network to be monitored:

- Signalling-
  - Connections requests from earth terminals (sub-queues) to satellite
  - Permits (slots allocation information) from satellite to earth terminals
- Cell level traffic sources creation-
  - Enqueing and dequeing of cells at earth terminals
  - Destroying of cell traffic at receiving terminals

As a means of secondary validation, the queuing behaviour was observed. Figure 3.15 shows the terminal queue for the "Star Wars" trace (used to simulate real time video) when the uplink rate was  $8^2$  Mbit/s and the TDMA frame duration was 40ms. The large link rate of 8Mbit/s ensured that there was always bandwidth available for the terminal requests. The queue size was probed every 400ms and the simulation lasted for 37 seconds. The variation in the queue size demonstrates that the enqueing and dequeing operation was taking place.



Figure 3.15 Observed queuing behaviour

To ensure the proper interworking between the process models of the simulator, namely the Requests Generation Process, Traffic Scheduling Process, Permits Generation Process and Data Transmission Process models, the requests from the earth terminal were probed at 400ms intervals for the first four seconds, during the same simulation as above. The received requests at the satellite and received permits from the satellite to the earth terminal where also probed. Figure 3.16 clearly shows that the time between the sending of a request by the earth terminal and receipt of the request by the satellite is equal to the propagation delay (125ms). Also the time duration from when the satellite sends a permit to the Earth Terminal and the receipt

 $<sup>^2</sup>$  For the sake of validation of the RM simulator, the link rate and frame size had to be increased beyond the values previously used for Hung's implementation study. This meant that there was no bandwidth constraint such that the bandwidth requested should equal to the bandwidth granted.

of the permit by the Earth Terminal is again equal to the propagation delay. This indicates that the timing relationships are functioning properly in the RM simulator. Furthermore the 3 elements in the triplet (sent request, received request and received permit) all have the same value which demonstrates that the processes are behaving as expected when the scenario was such that there was bandwidth available for all requests.



Figure 3.16 Requests and Permits propagation within the simulated network

To validate the sources feeding into the simulated network, a simple experiment was performed whereby the basic ON/OFF source which was used for the voice and data connections was fed to a simple FIFO queue. The queue size was fixed (2 cells) and the server rate was varied. The cell loss probability (CLP) was observed and was then compared to the analytical results for the same queue. The analysis was based on the Exact fluid-flow analysis [SCHO1], [SCHO2], and [SCHO3]. The derivation for the formula used for the CLP is shown in Appendix A.

The simulation experiment and analysis were performed for a typical ON/OFF voice source with peak bit rate of 64 kbit/s (150 cells/s), mean on duration of 0.356s and mean off duration of 0.650s. The server rate was varied in 1 cell/s increments from 120 to 150 cells/s. A bursty data source was also considered with a peak bit rate of

425 kbit/s (1000 cells/s), mean on duration of 0.5s and mean off duration of 1.9s. The server rate was varied from 600 to 1000 cells/s. Figures 3.17a and 3.17b show the experimental results closely match the analytical results, thus indicating that the ON/OFF OPNET model is a true and accurate representation of the ON/OFF source.



Figure 3.17a Cell loss with an ON/OFF voice source



Figure 3.17b Cell loss with an ON/OFF data source

The slight mismatch between the simulation and analytical results in figures 3.17a and 3.17b is due to the fact that in the simulation of the ON/OFF source the number of ATM cells generated during the ON period is rounded down to an integer value. This meant that for the case of the voice source that had a low cell rate, the cell loss was noticeably less than the theoretical value especially at high service rate.

#### 3.6 Summary

This chapter discussed a generic resource allocation scheme for ATM satellite networks (Hung's implementation). A detailed overview was presented of a resource management (RM) simulator developed by the author for the performance evaluation of uplink assignment and scheduling schemes. Using the simulator tool, the performance of Hung's implementation was analysed. It has been shown that by adopting Hung's implementation in the satellite network, QoS requirements can be guaranteed for the delay sensitive connections. However, this is at the expense of best effort connections which experience huge delays and increased buffer space requirement. Also some of the fixed slots allocated to the higher-class connections might be wasted in a number of frames, resulting in degraded frame utilisation.

The next chapters present improved WRR scheduling strategies that enhance the delay performance of best effort traffic and a novel uplink resource allocation scheme that is capable of delivering stringent QoS guarantees to real time video traffic with lower utilisation of both space and ground segment resources.

# **CHAPTER 4**

# NEW STRATEGIES FOR SCHEDULING BEST EFFORT TRAFFIC IN ATM BASED SATELLITE NETWORKS

# 4.1 Motivation for improving the delay performance of best effort traffic

Looking at the Internet, it is clear that best effort services, such as file transfer, remote terminal access, simple electronic mail and web browsing make up the vast majority of today's traffic. In fact, the Internet Protocol (IP) suite was built around the best effort service model [DERMO]. However this situation will change as new multimedia services appear and because some components of a multimedia service will require some QoS guarantees.

It is anticipated that best effort traffic will continue to constitute a substantial proportion of the overall traffic [HUTER], despite the predicted growth in real-time video and voice services.

In ATM networks, best effort traffic carried using the UBR traffic class is categorised as "class-4". In ATM satellite networks, proposed scheduling schemes (such as WRR) are class-based schemes i.e. the granting of a high priority status in the scheduling process depends on the class of the connection. Moreover, the uplink assignment function allocates a fixed amount of the bandwidth to connections belonging to class-1, class-2 and class-3. A good example of a resource allocation scheme that comprises a class-based scheduler and an uplink assignment function, is Hung's implementation that was discussed in the previous chapter.

The combination of the uplink assignment function and the class-based scheduler can lead to large delays for best effort traffic.

The purpose of this chapter is to describe in detail the author's proposed strategies to reduce the effects of the satellite resource allocation schemes on the best effort traffic and hence considerably reduce the end-to-end delay of best effort connections. This will improve the perceived QoS seen by the customer and hence can act as a differentiator between operators.

## 4.2 Intra and inter-scheduling

Scheduling in the satellite network can be thought of as being hierarchical in nature, consisting of two distinct stages [AWAD2]:

- Inter-scheduling stage: this is a global operation, where the on-board scheduler distributes the satellite capacity according to the totality of the bandwidth requests it receives.
- Intra-scheduling: this is a local operation, where the total bandwidth allocated by the on-board satellite to the Earth Terminal is redistributed by the terminal to the connections originating from it.





The WRR scheduling function in Hung's implementation is an inter-scheduling function. The proposed strategies explained later can be employed at the intra or inter-scheduling stages. However, because in this work comparison with only Hung's implementation is considered, the strategies will be employed at the inter-scheduling stage.

# 4.3 Classical approaches for reducing best effort delay

To reduce the mean delay of best effort traffic in the satellite network, two obvious approaches can be considered.

- Implement a Weighted Fair Queuing (WFQ) function for the best effort connections.
- Utilise the unoccupied slots of the higher-class connections for the best effort connections.

# 4.3.1 WFQ approach

WFQ is a term often used in the literature on ATM to describe a behaviour of a scheduler or queuing mechanism that attempts to ensure relatively fair distribution of the bandwidth to all connections within its domain. The WFQ approach that is widely referred to in the literature [GALL1] and [GALL2] is to regulate the emission of cells by computing the virtual finishing time (service tag) of each cell arriving at a queuing system and serving the cells accordingly. This approach has a significant implementation problem in the form of high complexity in that computation [GARRE]. Such complexity may be too high for the inter-scheduling level of the satellite system, where the on-board scheduler has to process the requests from best effort connections in the satellite network. Even assuming that each terminal computes the virtual finishing times of each of the arriving cells and signals this information to the satellite, the limited computational power of the OBP may not be adequate to cope with the task of inspecting each virtual finishing time.

A simple and straightforward approach for WFQ is that mentioned by Handel *et al* [HANDE] where each connection is allocated part of the bandwidth proportional to a relative value or weight as shown in equation 4-1.

 $proportionate-bandwidth [i] = \frac{total-bandwidth * weight [i]}{summation of weights of all active connections} \quad (equ 4-1)$ 

where  $1 \le i \le N$  and N is the total number of best effort connections in the network. This equation applies when all active best effort connections use their proportionate amount. When they use less than that, the unused amount is redistributed to the remaining active connections using the same equation.

Although this reduces to WRR scheduling within the best effort class of connections, it should not be confused with Hung's implementation of WRR. Hung does not associate weights to connections or classes of connections but uses the term 'weight' to explain that higher-class connections receive high priority in the scheduling process e.g. all class-1 connections will receive the highest priority in every scheduling instant followed by class-2 connections and so on.

Using WFQ scheduling in the intra or inter stages for best effort connections may reduce the delay experienced by the best effort connections with the highest weights, but those with the lower weights will suffer more.

#### 4.3.2 Unoccupied slots approach

Any connection may at some time have few or no ATM cells to transmit resulting in some or all of its allocated slots being wasted. Such slots may be reallocated to other connections. This approach can only be employed at the intra-scheduling stage where each terminal can check its buffer to determine how many cells from the higher-class connections are awaiting transmission. If there are none, or there are fewer than the allocated slots, the unused slots are then reallocted to the best effort connections.

This approach can have a local impact on the mean delay of best effort connections i.e. only the best effort connections originating from an individual terminal will benefit from this approach. If the load in the satellite network is such that some terminals have a high proportion of best effort connections while others have a high proportion of higher-class connections, the intra-scheduling stage is of little use. At the inter-scheduling stage it may prove difficult if not impossible to reallocate any unused slot, because of the latency problem.

#### 4.4 The proposed scheduling strategies

The author proposes two new strategies that are capable of drastically reducing the end-to-end delay of best effort connections with minimal effect on higher-class connections. One strategy is based on the alternation of the priorities of connections in the scheduling process while the other is based on the alternation of the ownership of some of the fixed slots allocated to the higher-class connections (between these connections and the best effort connections). The conceptual and performance aspects of both strategies were published in [AWAD3], [AWAD6] and [AWAD8]. The former strategy will be called the Priority Alternation (PA) strategy and the latter will be referred to as the Quasi-Fixed Slots (QFS) strategy.

The alternations of scheduling priorities and fixed slots are performed in an intermittent and controlled manner.

Both strategies have the same algorithmic component that resides at the ground segment (Earth Terminals). The algorithm controls the alternation processes of the scheduling priorities in the PA strategy and the fixed slots ownership in the QFS strategy.

#### 4.4.1 The ground segment implementation

At the connection set-up phase, the author proposes that users requiring services of class-3 and class-2 would be presented with the option of having two QoS states. The first state is the desired cell transfer delay  $(CTD^{des})$  and the second state is the maximum tolerable cell transfer delay  $(CTD^{max})$ . The closer  $(CTD^{max})$  is to  $(CTD^{des})$ , the higher the charge the user incurs.

During the connection lifetime, and at equally spaced update periods, the sending Earth Terminal determines the mean delay. The mean delay is equal to the mean
waiting time at the Earth Terminal buffer plus the RTD. The update period should be long enough (more than the Round Trip Delay: RTD  $\sim 250$  ms) to allow for the strategy to have an impact on the mean delay of the best effort connections and to avoid instability in the system.

The probabilistic algorithm resides at each Earth Terminal and indicates whether or not a delay sensitive connection can tolerate a demotion of its scheduling priority or relinquishing of fixed slots. The probability of demotion of priority or relinquishing of slots increases as the measured mean delay (m) approaches ( $CTD^{des}$ ). This scheme can be represented by the following conceptual diagram (figure 4.2):



Figure 4.2 The probabilistic alternation process action in the PA and QFS strategies

The above scheme can be realised with a linear relationship as shown below:

$$Prob = \begin{cases} 1 , m > CTD^{max} \\ \frac{m - CTD^{des}}{CTD^{max} - CTD^{des}} \\ 0 , otherwise \end{cases}, m > CTD^{max} (equ 4 - 2)$$

A random number between 0 and 1 is generated and if it is found to be greater than (*Prob*), a status bit in a signalling cell (assuming out of band signalling is used) is set to one, otherwise it is set to zero. The signalling cell may have the same size as an ATM cell, but not the format, and may be used in the satellite network for operation and management (OAM) aspects, congestion control, error control etc. The following flowchart diagram (figure 4.3) further explains the ground segment implementation:





#### 4.4.2 The space segment implementation

The next two subsections will explain how the scheduler performs the alternation of priorities in the PA and fixed slots in the QFS strategies.

## 4.4.2.1 Alternation of the scheduling priorities in the PA strategy

The on-board scheduler checks the status bit of the received signalling cell belonging to a higher-class connection. If the bit is set to one, the scheduler will swap the priority of the higher-class connection with that of a best effort connection as shown in figure 4.4. All or some of the best effort connections will be promoted in the scheduling process, one connection at each scheduling instant. The scheduling instants are spaced out by the TDMA frame duration. The promotion continues for a period of time equal to the update period. The choice of the best effort connection to be promoted depends on how large is its demand for bandwidth or on the scheduling function e.g. round robin or WFQ.



Figure 4.4 The on-board scheduler realisation of the PA strategy

#### 4.4.2.2 Alternation of the ownership of the fixed slots in the QFS strategy

As in the former strategy, the on-board scheduler checks the status bit of the received signalling cell. If the bit is set to one, a slot or a number of slots that were previously (*at call-setup*) allocated to the higher-class connection to use as fixed slots, are considered free slots.

The best effort connections will use these slots starting from the future frame appearing after RTD/2 seconds (*time from when the assignment is made at the* 

satellite to the time when it is received at the Earth Terminal) and in all future frames within an update period.

During these future frames, the best effort connections will possess ownership of the fixed slots according to their demand for bandwidth or according to the scheduling function implemented as shown in figure 4.5.



Figure 4.5 The on-board scheduler realisation of the QFS strategy

### 4.5 Comparisons of the PA and QFS strategies

Both strategies can achieve reduction in the mean delays of the best effort connections but only the QFS can adaptively increase or decrease the amount of reduction. This is achieved by adjusting the number of the fixed slots from a connection that is to be included in the ownership alternation exercise. In the QFS strategy, if the number of best effort connections in the network increases above a certain threshold the number of fixed slots per TDMA frame of a connection to alternate ownership 'S' can be adjusted by varying 'q' in equation 4-3 shown below:

$$S = \left[ \begin{array}{c} q * \lambda * f \end{array} \right] \quad , \ 0 \le q < 1 \qquad (equ \ 4-3)$$

Where ' $\lambda$ ' is the fixed slots allocation rate and 'f' is the TDMA frame size in seconds.

Obviously both strategies have an impact on the signalling load at the uplink side due to the need to signal to the satellite the status of the higher-class connections. However, not all of the higher-class connections will opt for less than the achievable stringent requirements. Hence the signalling load is a function of the number of the higher-class connections employing any of the two strategies as well as the duration of the up-date period. The shorter the update period the higher the signalling load. There is also an impact on the MAC protocol where a bit or a number of bits must be provided in the signalling cell in the case of out of band signalling, or in the satellite ATM (S-ATM) cell in the case of in-band signalling. The QFS strategy can cause an increase in the signalling load at the downlink side as the OBP needs to signal to the terminal to let it know if the higher-class connection can use the fixed slots or not. To avoid this increase in the downlink signalling load, the Earth Terminal can periodically abstain from using some of the fixed slots whenever a signalling cell with a set status bit is transmitted to the satellite. The duration of abstention starts after a period of time equal to the RTD and continues for a period of time equal to the update interval. This will require precise timing capabilities at the terminal. Figure 4.5 for the QFS strategy can be modified as shown in figure 4.5.



# Figure 4.6 Ground component implementation of the QFS strategy (no increase in downlink signalling)

# 4.6 Summary

In this chapter the PA and QFS strategies to reduce the mean delay of best effort traffic were detailed. The following are the key features of the proposed strategies:

- Non-adhoc scheduling approach the strategies can be implemented in conjunction with any class-based scheduling scheme.
- Minimum impact on the design of the OBP most of the complexity that is added to the satellite network is confined to the ground segment.
- Efficiency reductions in the end-to-end delays endured by the best effort connections, is achieved with minimal degradation of class-2 and 3 connections and with increased frame utilisation performance.
- Charging applicability the proposed strategies can also form part of the pricing/charging architecture of the satellite network operator to maximise the revenue generation.

The next chapter will concentrate on the performance aspects of the proposed strategies.

# SIMULATION STUDY OF THE 'PA' AND 'QFS' STRATEGIES

# 5.1 Introduction

In the previous chapter, two new strategies (PA and QFS) for scheduling best effort traffic in an ATM satellite environment were proposed. The concept of allowing higher-class connections to opt at call set-up for two QoS states was described. Having identified the ground and space segment implementations for both strategies it is necessary to investigate using simulations [LAW], [PITTS] the enhancement achieved in the delay performance of best effort traffic and also the impairment caused to the QoS of the higher-class (class-2 and 3) traffic as a result of employing either of the two strategies in the satellite network.

The load scenario, the TDMA frame and the uplink capacity were the same in all simulations as those listed in Table 3.1 of chapter 3. Larger frame sizes and greater capacities will increase the number of events that the simulator has to handle resulting in excessively long periods to achieve steady state. Also the number of connections (especially the higher-class connections) should be large enough such that best effort connections do not get allocated enough slots in every frame. Such a situation provides an optimum environment to test the proposed strategies. However, large numbers of connections will increase the number of events and hence a compromise between the number of connections and the level of mean delay experienced by the best effort traffic is essential for the study.

The limitation of incorporating a WFQ function for the best effort traffic will be highlighted through simulations. An ATM satellite network that uses either the PA or QFS strategies was simulated and the results compared with the results obtained from Hung's implementation in chapter 3. The comparisons allow the evaluation of the delay performance of the lower-class traffic and the QoS degradation of the higher-

class traffic. All the simulations were carried out using the RM simulator discussed in chapter 3.

# 5.2 Limitations of WFQ scheduling

To demonstrate the limitations of the WFQ scheduling, each of the two data connections is allocated part of the bandwidth proportional to a relative value, or weight, as in equation 4-1. In the first simulation experiment, the data1 connection was allocated 75% of the total bandwidth while in the second experiment it was allocated 85%. As expected, it can be seen in figure 5.1 that the mean delay of data1 connection drops with increasing weight. However, the mean delay of data2 connection increases to very high values as shown in figure 5.2 due to it being denied large part of the bandwidth. The buffer size required for the data1 connection also decreases (figure 5.3) with increasing weight but that of data2 increases twofold when data1 weight was 0.75 and threefold when it was 0.85 (figure 5.4).

The experiments show that allocating part of the bandwidth to a connection when the bandwidth resource is already scarce, such as in the case of a satellite network with high load, may cause large increases in the mean delay and buffering requirement for the other connections.



Figure 5.1 Data1 mean delay (WFQ)



Figure 5.3 Data1 mean buffer space requirement (WFQ)



Figure 5.4 Data2 mean buffer space requirement (WFQ)

# 5.3 Simulation results for the PA strategy

For performance evaluation under various QoS states scenarios and also for comparison purposes, a parameter that relates  $CTD^{des}$  and  $CTD^{max}$  is introduced. This is the "QoS states ratio" denoted by '**Z**' as shown in equation 5-1:

$$\mathbf{Z} = (CTD^{max} - CTD^{des}) / CTD^{des}$$
equ 5-1

The PA strategy was implemented in the simulated satellite network and Z was kept constant throughout all the simulation runs. Three experiments were performed each, with a different Z value (0.01, 0.02, 0.05). Figures 5.5 and 5.6 show noticeable improvements in the delay performance and buffer space requirement for the data connections over Hung's classical scheme. The degradation in the delay performance, buffer space requirement and cell loss of the video connection (figures 5.7, 5.8 and 5.9) is slight, even when Z is as high as 0.05. This is also true for the cell delay variation (figure 5.10) experienced by the video connection.

One observation is that increasing Z does not yield much improvement in the delay performance of the data connection and has little effect on the QoS of the video connections. This is evident from figure 5.11 where the frame utilisation shows

minimal increase as Z is increased from 0.01 to 0.05, indicating that only few empty slots were occupied by cells from the data connections. This suggests that the PA strategy has a limit beyond which it becomes ineffective in trying to reduce the delay of the data traffic.

Moreover, renegotiating lower<sup>1</sup> *CTD<sup>max</sup>* values will not achieve noticeable improvement in the QoS of the video traffic. Hence, though the PA strategy achieves impressive reduction in the end-to-end delay experienced by the best effort traffic it is a rigid strategy as the degradation in the QoS performance of the higher-class traffic can not be controlled. The rigidity of the PA strategy was the main motive for the author proposing the QFS strategy.



<sup>&</sup>lt;sup>1</sup> If  $CTD^{max}$  equals  $CTD^{des}$  (i.e. **Z** approaches zero) there will be an abrupt transition to the delay and buffer space values of Hung's implementation. This is because the PA strategy is a two level strategy where as Hung's implementation is a one state. This also applies to the QFS strategy as can be seen in figures 5.12, 5.13, 5.19 and 5.20. In the QFS strategy an abrupt transition can occur if either **Z** or *q* approaches zero.



Figure 5.6 Data mean buffer space requirement (PA)





Figure 5.8 Video mean buffer space requirement (PA)





Figure 5.10 Video cell delay variation (PA)



Figure 5.11 Frame utilisation (PA)

# 5.4 Simulation results for the QFS strategy

The satellite network was simulated with the QFS scheduling strategy and 'S' (*the number of fixed slots per TDMA frame of a video connection to alternate ownership*), was varied by changing the value of 'q' in equation 4-3 of chapter 4. The QoS states parameter Z was again kept constant.

The result from the simulation study shows an improved data delay performance and data buffer space requirement as q is increased (figures 5.12 and 5.13). This is also evident from the increase in the frame utilisation in figure 5.18, which demonstrates that the data cells occupied some of the fixed slots. The degradation in the QoS of the video connection (delay, buffer space requirement, cell loss and cell delay variation) increased with increasing values of q (figures 5.14, 5.15,5.16 and 5.17). This is an expected outcome as the video connection is intermittently starved of its fixed slots.

An increase of q beyond 0.3 for this specific load scenario does not substantially improve the data performance but degrades the video QoS even further. Hence it can be concluded that q should be small. The choice of q may prove difficult, as different load scenarios will have different maximum q values. A sensible approach would be for the satellite network operator to specify a default value for q (at call set-up) and allow the higher-class connections to renegotiate for lower values if the QoS becomes intolerable.

Figure 5.19 shows the data delay performance when q was fixed at 0.05 and Z was changed to 0.01 in one simulation and to 0.1 in a second simulation. The improvement was not substantial and hence it can be concluded that, similar to the PA strategy, renegotiating a different  $CTD^{max}$  value will not have a noticeable improvement on the data delay performance or the higher-class QoS.









Figure 5.15 Video mean buffer space requirement (QFS)





Figure 5.17 Video cell delay variation (QFS)





Figure 5.20 shows the data delay against the perceived QoS (video cell loss) when z = 0.05 and p = 2.7. The data delay achieved by the PA strategy was 0.58 seconds but the

cell loss was around 5\*10<sup>-3</sup>. Although the data delay achieved by the QFS does not approach that achieved by the PA strategy, different video QoS values are achievable.



Figure 5.20 Data delay and Video QoS with the PA and QFS strategies

# 5.5 Summary

This chapter presented the results of the simulated satellite network when the PA and QFS scheduling strategies were implemented. Although the QFS strategy cannot achieve the same level of data delay reduction as the PA, it allows greater flexibility in controlling the degradation in QoS suffered by the higher-class connections as a result of including the strategies in the satellite network as part of a flexible charging scheme. It should be noted that if the traffic load is such that the number of class-2 and class-3 connections is not large or if there is a high proportion of class-1 connections, the PA and QFS strategies may be ineffective in reducing the response time of end-to-end communications of class-4 connections. A complimentary approach may be to enhance the resource allocation for the class-1 connections to achieve a saving in the satellite network resources thus realising more resources to allocate to the best effort connections.

# **CHAPTER 6**

# A PREDICTIVE RESOURCE ALLOCATION SCHEME FOR ATM SATELLITE NETWORKS

# 6.1 Introduction

In chapter 3 the effect on best effort traffic due to an increase in the number of allocated fixed slots to the high-class connections in the satellite network, was demonstrated. In chapters 4 and 5 the PA and QFS scheduling strategies were presented and studied. The PA and QFS strategies are effective when there is a large number of class-2 and class-3 connections opting for two QoS states. However there are two cases when both strategies may be of little use in improving the delay performance of best effort traffic in the satellite network:

- Not enough class-2 and/or class-3 connections opting for two states.
- A high proportion of class-1 connections with stringent QoS requirements.

In this chapter a different approach for improving the delay performance of best effort traffic will be presented. The approach adopted is to efficiently allocate the bandwidth resource to the real time connections (class-1), specifically the video connections, through predicting the required number of slots in future frames rather than signalling to the satellite the immediate requirement for slots. As has been shown in [AWAD4], [AWAD5], and [AWAD7], this new approach achieves a saving in the number of fixed slots, as fixed slots are only required to compensate for the prediction errors. Less fixed slots results in more bandwidth to all connections, in particular to best effort traffic.

All simulations were carried out using the RM simulator that was developed by the author for scheduling ATM traffic over a satellite link. The simulator required a

number of modifications to include prediction algorithms and process the predicted data.

The functional architecture of the predictive resource allocation scheme will be discussed. This will then shed light on how the bandwidth requirements are predicted. Having addressed this crucial issue, three techniques for prediction will be explained, and finally the impact of the proposed scheme on the ground segment of the satellite network will be outlined as well as the modifications required to the RM scheduler simulation tool.

# 6.2 The functional architecture of the predictive scheme

In chapter 3 the main property of MPEG video, the periodicity of deterministic frames, was highlighted. This property, plus the fact that there is correlation between the frames and between the GoPs [ROSE1] [PILIP] [CHAND] and [CHIOT], allows for the prediction of future frames. The autocorrelation function p(i) is given by equation eq. 6-1 [ROSE1]:

$$p(i) = E[(x(n) - \mu) \cdot (x(n + i) - \mu)] / \delta^2$$
 (equ 6-1)

where the frame size process is described by X(n),  $\mu$  is a mean, and  $\delta^2$  is a variance of the frame size process.

The frame-by-frame correlation and the long range correlation between the GoPs (GoP-by-GoP correlation) for an MPEG trace are shown in figures 6.1 and 6.2. From figure 6.1 it can be seen that the trend of autocorrelation function generally follows the GoP pattern. The highest peaks are due to the correlation between the 'I' frames, the next highest peaks are due to the correlation between the 'P' frames and the troughs are due to the correlation between the 'B' frames.

Hence, if the bit size of a frame from a certain class (I, B or P) is known it may be possible to predict the sizes of the next frames of the same class (especially the I frames). Figure 6.2 suggests that predicting frames within the same GoP or within the next few GoP gives better results compared to predicting frames in distant GoPs.



Figure 6.1 Autocorrelation function of the frame sizes of the Jurassic Park movie sequence



Having predicted the bit size of an MPEG frame, it was then possible to estimate the number of ATM cells arriving during the time period of all the TDMA frames, and parts of TDMA frames, encompassed by the MPEG frame as shown in figure 6.3



# Figure 6.3 Future TDMA frames appearing during the same time period as that of a predicted MPEG frame

It is assumed that the TDMA frame duration is less than that of the MPEG frame. This is a valid assumption since a long TDMA frame duration exacerbates the end to end delay and requires a higher storage capacity in the transmitting and receiving earth terminal buffer memories [MARAL]. Another important assumption is that the MPEG video traffic is deterministically smoothed over the MPEG frame period, i.e. cells emitted during a frame period are evenly spaced [ADAS], [BAIOC].

The predicted value 'E' of the integer number of cells that will arrive as a result of the predicted MPEG frame during any TDMA frame that will exist between the times (t + j) and (t + j + M) can be determined as shown in equation 6-2:

$$E = \left[ F(t+j) * \Phi * T/M \right] / 384 , \ 0 < \Phi \le 1$$
 (equ 6-2)

Where F(.) is the predicted size of a future MPEG frame, j is the number of seconds after which the future MPEG frame starts from the time the prediction is made, T is the TDMA frame duration, M is the MPEG frame duration and  $\Phi$  determines how much of the TDMA frame is encompassed by the MPEG frame.

Figure 6.4 shows the satellite receiving a request containing a prediction of the number of cells arriving at the ET's buffer instead of a request containing the current number of cells awaiting transmission (i.e. buffer occupancy).



Figure 6.4 A predictive request being transmitted to the satellite

To allow for possible under-prediction, a number of fixed slots ' $\phi$  ' in every TDMA frame (or multiple of frames) are allocated to the real time video connection. Hence (*assuming each slot accommodates one cell*) the earth terminal signals a total demand of:

 $D = E + \phi \tag{equ 6-3}$ 

The fixed slots ' $\phi$  ' is determined by the level of required QoS and the connection would be allocated the fixed slots according to the declared traffic parameters (MCR, PCR etc.). This is similar to Hung's implementation, but the number of fixed slots required to achieve a certain QoS level is drastically less than in Hung's implementation. For example when the allocation rate was 4 slots per second (see figure 7.6a) the predictive resource allocation scheme achieved a cell discard rate of approximately 10<sup>-4</sup> whereas Hung's classical scheme achieved a mere 5\*10<sup>-2</sup>. The

reduction in slots leads to fewer resources being required, which is a major contribution of this thesis.

The functional architecture for the predictive resource allocation scheme takes into account the position of the slots in the TDMA frame. If slots are reserved for the predicted cells earlier than the actual arrival of the cells, some of the cells will miss their reserved slots and thus may have to be transmitted in the next frame. This scenario is presented in figure 6.5 where the upper part of the figure shows a TDMA frame with slots reserved at the start of the frame (*the blue coloured slots*). The last two arriving cells will have to be transmitted in the next frame, possibly resulting in the cells themselves or future cells being delayed by more than the CTD. The lower part of figure 6.5 shows a TDMA frame with reserved slots at the end of the frame. Here all the cells (*shown in red colour*) are transmitted during the current frame.



Figure 6.5 Position of reserved slots in the TDMA frames

Cells delayed by more than the maximum CTD may be lost and hence the cell loss in the case of the predictive resource allocation scheme is not only due to the prediction error but also due to the position of the slots reserved for the predicted cells in the TDMA frame. The next chapter will provide results for cell loss comparison of two cases where the slots are reserved at the start of the frame and where they are reserved at the end of the frame. It should be noted that the start and end positions are extremes (i.e. minimum and maximum) and that from the satellite network point of view the end position exacerbates the end-to-end delay and a compromised position would be the middle of the frame. Also, to reduce the cell delay variation and avoid cell clumping the allocated slots should not be close to each other.

## 6.3 Linear Predictor

Predicting future MPEG frames using an adaptive least mean square (LMS) error linear predictor, was first proposed by A. Adas [ADAS]. The adaptive technique achieves small prediction errors that are almost white noise, enabling the use of small buffer sizes and resulting in high utilisation and small delays [ADAS] in ATM networks. The adaptive technique does not require any prior knowledge of the statistics, nor assumes stationarity.

#### 6.3.1 Adaptive least mean square error linear predictor

The *k-step* linear predictor predicts x(n + k) using a linear combination of the current value and previous values of x(n) [HAYES], [MARKE]and [ADAS]. The  $p^{th}$  order linear predictor can be expressed as in equation 6-4 shown below [ADAS]:

$$\hat{x}(n+k) = \sum_{l=0}^{p-1} w(l) x(n-l)$$
 (equ 6 - 4)

where w(l), for l = 0, 1, ..., p-1 are the linear prediction filter coefficients as shown in figure 6.6



Figure 6.6 Non adaptive Linear predictor

The error *e*(*n*) in prediction can be written as [ADAS]:

$$e(n) = x(n + k) - W^{T} \mathbf{X}(n)$$
(equ 6-5)  
where,  
$$W = [w(0), w(1), \dots, w(p-1)]^{T}$$
$$\mathbf{X}(n) = [x(n), x(n + 1), \dots, x(n - p + 1)]^{T}$$
$$e(n) = x(n + k) - x(n + k)$$

The optimal linear predictor in the mean square sense is one that minimises the mean square error  $\xi$  [ADAS], where:

$$\xi = E\{e^2(n)\}$$

To find the vector W that minimises  $\xi$ , the gradient is set to zero [ADAS]:

$$\nabla \xi = \nabla E\{e^2(n)\}$$
$$= -2 E\{e(n)\mathbf{x}(n)\} = 0$$

Substituting the value for e(n) from 6-5:

$$\nabla \boldsymbol{\xi} = -2 E\{(\boldsymbol{x}(n+k) - \boldsymbol{\mathcal{W}}^T \boldsymbol{\mathcal{X}}(n)) \boldsymbol{\mathcal{X}}(n)\} = 0$$

Taking expectations, the above equation can be written in matrix form [ADAS] as:

$$\mathbf{R}_{\mathbf{X}} \ \mathcal{W}^T = \mathbf{P} \ (\mathbf{k}) \tag{equ 6-6}$$

where an element  $r_{\mathbf{X}}(k)$  of the matrix **P** (k) is defined as  $E\{x(n + k)x(n)\}$ . Equation 6-6 is referred to as the Wiener-Hopf equation for linear prediction [BERAN], which requires a knowledge of the auto-correlation of  $\mathbf{X}(n)$  and the assumption of wide sense stationarity (*i.e. mean, variance and auto-covariance of*  $\mathbf{X}(n)$  *do not change with time*) to be able to solve it [ADAS]. Following the work of A. Adas [ADAS], the LMS adaptive approach is used in this work as an on-line algorithm for predicting the frame sizes. The operation of an adaptive linear predictor is shown in figure 6.7. The prediction coefficients W(n) are time varying. The errors,  $\{e(n)\}$ , are fed back and used to adapt the filter coefficients in order to decrease the mean square error.



Figure 6.7 Adaptive Linear predictor

The algorithm for prediction is the same as that of A. Adas [ADAS], which is to start with an initial coefficient W(0) and then compute for each new data point  $\nabla \xi$ . As the statistics are not known and may change with time, the expectation is replaced with an estimate and thus  $\nabla \xi$  can be thought of as the one point sample average  $e(n)\mathbf{X}(n)$ . Taking a step size  $0.5\mu$  in the negative direction to point to the bottom of the error surface, the LMS filter coefficients are the updated using equation 6-8 as shown below [ADAS], [HAYK1]:

$$W(n+1) = W(n) - 0.5\mu\nabla\xi \qquad (equ 6-7)$$
$$= W(n) + \mu e(n) \mathbf{X}(n) \qquad (equ 6-8)$$

As has been shown in [HAYK1], the LMS will converge in the mean if  $\mathbf{x}(n)$  is stationary and if  $0 < 1/\mu < 2/\lambda_{max}$ , where  $\lambda_{max}$  is the maximum eigenvalue of  $\mathbf{R}_{\mathbf{x}}$ .

The normalised LMS (NLMS) is used as it is less sensitive than the LMS to the step size  $\mu$  [ADAS]. The NLMS will converge in the mean [HAYK1] if  $0 < \mu < 2$ . Since at time *n* the value of x(n + k) is not available to compute e(n), e(n - k) is used instead.

The *k*-step linear predictor update equation (equ 6-9) becomes:

$$W(n+1) = W(n) + \mu e(n-1) \mathbf{X}(n-1) / \|\mathbf{X}(n-1)\|^2$$
(equ 6-9)  
where,  $\|\mathbf{X}(n-1)\|^2 = \mathbf{X}(n)^T \mathbf{X}(n)$ 

Adas's scheme predicts the next immediate frames of the same class i.e. the MPEG sequence is fragmented into three subsequences (I, P, and B). The predicted frames appear at 40ms ahead of time for a B frame to 480ms for an I frame (*assuming a GOP of 12 frames and a MPEG frame duration of 40ms*). In satellite communications this will not be possible because of the latency problem. Prediction of MPEG frames should be for frames appearing at or more than 250ms ahead in time.

In this work the number of subsequences chosen ensures that the above criteria is met. The choice of the number of subsequences ' $\varsigma$ ' for the 1-step linear predictor is as follows:

$$\varsigma = h * G \tag{equ 6-10}$$

here 'h' is the minimum number of GoPs required such that  $(\varsigma * M) \ge 250$ ms, and G is the number of frames in a GoP. M is the MPEG frame duration.

The MPEG traces used in this work have all the same number of frames per GoP (12 frames) and each frame duration is 40ms .  $\varsigma$  is thus capped at 12 in the simulation work that will be discussed in the next chapter. Figure 6.8, shows Adas's linear prediction implementation, and the author's proposed implementation for a satellite network environment. The author's scheme requires twelve linear predictors (*for the MPEG traces used*) that can increase the complexity at the ET. Hence, there is a need to investigate other prediction techniques. In [AWAD9] a linear predictor with twelve predictors was compared to the EWMA scheme of section 6.5 and was found to be

better in terms of prediction efficiency suggesting that for terrestrial networks with high propagation delays a linear predictive scheme may not be a good option.



Figure 6.8 Linear prediction schemes for terrestrial and satellite environments

# 6.4 Artificial Neural Network (ANN) predictor

Artificial Neural networks (ANN's) have been extensively used in studies of ATM networks especially in the areas of congestion and flow control, [LIU1], [LIU2], [LIU3], [CHANG], [CHONG], [BALES] and [HIRAM]. In [MARS] and [LIU1], ANN's were tested on Autoregressive (AR) Markov video models and in [LIU2], [LIU3], [CHANG], [MOH] and [CHONG] they were tested on MPEG traces.

The MPEG traces can be characterised by a general autoregressive moving average (ARMA) process [CHANG] that is essentially a non-linear and non-stationary process. An ANN captures the non-linearities of a process as it has generalisation capability that makes it flexible and robust when faced with new and/or noisy data

patterns. Hence it is better suited than the linear predictor (explained in the previous section) in the adaptive prediction of a nonstationary time series [YU].

## 6.4.1 Fundamentals of Neural networks

The architecture of an ANN resembles in some respects that of a human brain, which was the original inspiration for them [MARS]. The various individual building blocks (neurons) of the brain are relatively simple units that decide on the basis of incoming signals whether or not to fire off an impulse. Whereas this mechanism appears to be relatively simple, the brain system as a whole is incredibly complex, due to the enormous number of neurons, the number of interconnections between them and the fact that all the neurons function autonoumsly and in parallel. This great number of connections is essential, since the learning process in the brain depends on the growth of new connections or the breaking of existing ones.

Many mathematical models for the (human) brain have been developed. Although they may differ considerably from one another in detail, they have some similar characteristics of information processing. These models were then used to implement ANNs.



The basic unit of a neural network is the neuron processing unit shown in figure 6.9.

#### Figure 6.9 Model of a neuron processing unit

An ANN is composed of large numbers of neurons that are interconnected in a certain topology. Each neuron accepts a number of input signals (from other neurons)  $x_1$ ,  $x_2$  ...... $x_p$  and has one output signal  $y_k$  that can be input to other neurons. There is a set of "synapses", each of which is characterised by a weight of its own. Specifically, a signal  $x_j$  at the input of synapse j connected to neuron k is multiplied by the synaptic weight  $w_{kj}$ . An adder sums those weighed inputs and the result is called the linear combiner ( $u_k$  in figure 6.9).

An activation function  $\varphi$  defines the output of a neuron in terms of the activity level at its input. The activation function that is commonly used in ANN's is the sigmoid function [TARAS] that is a differentiable squashing function, usually a hyperbolic tangent function (figure 6.10).



Figure 6.10 A sigmoid (hyperbolic tangent) activation function

The model of a neuron shown in figure 6.8 also includes an externally applied threshold  $\theta_k$  that has the effect of lowering the net input of the activation function. This model was first proposed by Mc Cellot and Pitts in 1943 [McCAL].

The manner in which the neurons of a neural network are interconnected, i.e. the network structure, can be divided into four classes [FAN]: single-layer feedforward network, multi-layer feedforward network, recurrent network and lattice structure. The multi-layer feedforward network model (figure 6.11) is commonly used because of its simplicity and because it has the ability to learn dynamic system characteristics through nonlinear mappings [HAYK2].

Multi-layer feedforward networks consists of an input layer, output layer and a number of in between layers referred to as hidden layers. The networks are trained in a supervised manner with a highly popular algorithm known as the BackPropagation (BP) algorithm that is based on the error-correction learning rule. It may be viewed as a Least Mean Square (LMS) algorithm. The details of the BP algorithm are given in Appendix A.



Figure 6.11 An Architectural graph of a multi-layer feedforward ANN

Once the training is completed, an ANN can be computationally inexpensive even if it continues to adapt on-line.

As stated in the previous section (section 6.3.1), the earth terminal must be able to predict at least 250 ms ahead to account for the round trip delay. A feedforward three layered (30-15-10) ANN was trained on each trace. Fifteen successive MPEG frames

were used as inputs to the NN to predict the twenty third successive frame which starts after ( [MPEG frame size<sup>1</sup>] \* 7 ) 280 ms from the end of the fifteenth frame, as shown in figure 6.12. The extra 30 ms ensures that processing and queuing delays are catered for.

As there are no rules for choosing the number of inputs as well as the number of hidden layers and neurons, a number of experiments were required to achieve a near optimum configuration for the ANN. It was found that the more inputs the ANN has the better the prediction. However, it was also noticed that the improvement is not substantial for more than ten inputs. In fact with some of the tested traces, when the number of layers of the ANN was increased beyond three layers, the learning worsened. Also the more inputs and hidden layers, the greater the complexity that makes the implementation of the ANN at the earth terminal unattractive.



Figure 6.12 Prediction of the MPEG trace using Neural Networks

A training file of 4000 samples (1/9 of the whole MPEG trace) was used to train the ANN (*this was carried out for each MPEG trace*). All inputs were scaled so that they

<sup>&</sup>lt;sup>1</sup> The MPEG frame size of all the traces used in the thesis was 40 ms

lay in the range (0-1). Following the same approach as that of Moh *et al* [MOH], the scaling was achieved by dividing all samples of each training file by the size of the largest frame in the respective trace.

### 6.5 Exponential Weighted Moving Average (EWMA) predictor

In the Linear and Neural Network predictors, the MPEG traffic stream (trace) was treated as a time series. The Exponential Weighted Moving Average (EWMA) analysis explained below was also used to treat the MPEG trace as a time series; however, the MPEG trace was assumed to be generated by an additive model of the form:

$$X(t) = T(t) + S(t) + R(t)$$
  $t = 1, 2, \dots$  (equ 6-11)

Here T(t) is the trend, S(t) the seasonal term, and R(t) is the irregular or random term.

Each MPEG frame of a GOP pattern can be regarded as a unique seasonal term such that there are twelve seasonal effects  $S_1, S_2, \ldots, S_{12}$ . Each seasonal effect has a period of *s* unit times; that is, it repeats after *s* time periods:

$$S(t+s) = S(t) \qquad \forall t \qquad (equ \ 6-12)$$

The random term is set to zero as in [JANAC], the *j*th seasonal is denoted as S(j) and the period is taken as *s*. Following [JANAC], three smoothed series can be written as:

$$M(t) = a [X(t) - S(t - s)] + (1 - a) [M(t - 1) + T(t - 1)] \quad |a| < 1 \quad (equ \ 6-13)$$

$$T(t) = c [M(t) - M(t - 1)] + (1 - c) T(t - 1) \qquad |c| < 1 \quad (equ \ 6-14)$$

$$S(t) = d [X(t) - M(t)] + (1 - d) S(t - s) \qquad |d| < 1 \quad (equ \ 6-15)$$

M(t) is a local approximation and *a*, *c* and *d* are the discounting parameters. The prediction P(t + k) of X(t + k) made at time *t* is:
$$P(t+k) = M(t) + kT(t) + S(t+k-s)$$
 (equ 6-16)

This simple model for prediction has been used by a number of researchers, especially for forecasting economical time series where considerable success for prediction was reported [GRANG]. Equations 6-13, 6-14, and 6-15 can be reduced to two equations if the trend is neglected (i.e. setting T(t) = 0). The two equations to consider are equation 6-15 and equation 6-17 shown below:

$$M(t) = a [X(t) - S(t - s)] + (1 - a) M(t - 1) \qquad |a| < 1 \qquad (equ \ 6-17)$$

The prediction achieved by using a model without the trend term does not differ greatly from that with the trend term included. This will be demonstrated in the next chapter.

The choice of the discounting parameters a and d is not altogether obvious. It was observed that, at least for the MPEG traces used in this work, large values of a and small values of d make the smooth series react rather quickly to changes. The author takes the view that a sensible strategy is to initialise the smoothed series with the first known data value and then to minimise:

$$\sum_{i=1}^{N} \frac{(X(t) - P(t-k))_{i}^{2}}{N}$$
 (equ 6-18)

where N is the number of predicted MPEG frames.

This is really a gradient descent technique to find the least mean square error LMS and is not a difficult numerical problem [AWAD9].

Optimisation of the a and d discounting parameters was achieved for all the traces used, and for example the optimum values for one movie trace proved to be 0.560996 and 0.200000, and for a Formula One racing trace were 0.713994 and 0.120000.

Figure 6.13 represents a plot of the normalised LMS against the discount parameters. The mean square error values are normalised by dividing the values by the maximum computed mean square error.



Figure 6.13 Optimisation of the discount parameters

The EWMA predictor proved to be more efficient than the Linear predictor advocated by Adas [ADAS] for dynamic resource allocation schemes in terrestrial networks as has been shown in [AWAD9].

#### 6.6 Impact of the predictive implementation on the ground segment

Implementing a predictive resource allocation scheme for the real time video connections at the ground segment (Earth terminals) entails modifications to the terminal design. Instead of generating the bandwidth demands of all connections at the terminal using Hung's recursion (equation 3-1), the demands of the real time video connections are generated using the predicted MPEG frames values as was shown in figure 6.4. Hence for example, in the case of the EWMA scheme, the bandwidth generation process will have two flows as inputs (figure 6.14):

- 1. A continuous flow indicating the buffer occupancy of cells form all connections excluding the real time video sources.
- 2. A flow that is short lived (at connection set-up) from the real time video sources.

This second flow (*originating from the rt-video sources*) contains information about the MPEG properties of the video streams, e.g. the frames duration and GoP pattern. These properties are required by the predictive function, which resides at the "Bandwidth requests generation" block, as shown in figure  $6.14^2$ .



Figure 6.14A conceptual design for the Earth terminal when an EWMApredictive resource allocation scheme is implemented

#### 6.7 Modifications to the RM scheduler simulation tool

The RM scheduler discussed in chapter 3 requires a modification to the Requests generation OPNET process. This is essential as the requests sent from the ETs to the

<sup>&</sup>lt;sup>2</sup> The value of 'k' that appears in the figure is equal to 125ms such that the predicted number of ATM cells is the number of cells expected to arrive during the TDMA frame or part of frame appearing after an RTD.

satellite would in the case of a real-time video, contain the predicted number of cells to arrive during future frames and not the buffer occupancy, as was the case in the PA and QFS scheduling strategies. The Requests generation process is modelled by two sub-processes (figure 6.15):

- 1. MPEG frames prediction sub-process, where cell sizes of future frames (*appearing at least 250ms ahead in time*) are predicted. This sub-process contains three functions that perform the prediction of the MPEG frames. A function is selected at the start of the simulation run:
  - Back Propagation function for neural prediction; using the BP algorithm in appendix B.
  - Linear prediction function; using equations 6.4 to 6.10.
  - EWMA function using equations 6.11 to 6.18.
- 2. TDMA frames prediction sub-process, where outputs from the previous subprocess are mapped into TDMA frames (as was shown in figure 6.3) to achieve a prediction of the number of cells arriving in future TDMA frames. These predicted values are then transmitted to the satellite as bandwidth requests.





For ease of implementation, the prediction functions were written (in 'C' code) and compiled outside the OPNET environment as external functions to the RM simulator.

#### 6.8 Summary

This chapter presented a new predictive resource allocation scheme for the real-time video connections (class-1 connections), that is capable of efficiently allocating the uplink bandwidth to these connections and thus allowing for more bandwidth for all other connections especially the low-class connections.

The MPEG coded traffic properties were discussed and the mapping of the MPEG frames to TDMA frames was explained.

Three predictive techniques for predicting the MPEG traffic were discussed. These were the linear, the neural, and the EWMA prediction techniques. The EWMA predictor is a simple and an efficient predictor that requires fewer computations than the linear predictor and does not need to be trained on the traffic as the neural predictor.

The next chapter will present the simulation results of the three prediction techniques using different MPEG traces.

### **CHAPTER 7**

## SIMULATION STUDY OF THE PREDICTIVE RESOURCE ALLOCATION SCHEME

#### 7.1 Introduction

Chapter 6 presented a predictive resource allocation scheme that has the potential to efficiently allocate the satellite bandwidth resource to the class-1 connections, specifically to the real time video connections, thus realising more bandwidth for other classes of connections such as the best effort. This chapter will concentrate on the simulation results of the three predictive techniques (Linear, Neural and EWMA). The aim is to identify the technique that not only achieves efficiency in QoS provision but also is the least complex and more practical for implementation in the ground segment. The latter aim is vital since the Earth Terminal may have limited power and memory resources. This is especially more likely to be the case for portable terminals.

First, the predicted MPEG traces achieved using the three techniques will be presented and compared to the original MPEG trace. Second, the simulation aspects for the broadband satellite network using each of the prediction techniques will be discussed and results presented for the cell loss at the receiving terminal and buffer space requirements at the sending terminal. These results are compared to those of the classical scheme. Third, some results will be shown to emphasise the importance of the position of the slots reserved for the predicted cells in the TDMA frame. Finally, the suitability of the EWMA as the target scheme for predictive resource allocation will be highlighted.

# 7.2 Prediction of MPEG traffic using the three prediction techniques

Three different sub-sequences (fifty frames each) from the MPEG trace of a Formula One race event [ROSE2] are shown in figures 7.1(a), 7.1(b) and 7.1(c). Also shown

are the predictions using the Linear, EWMA, and Neural predictors. The author chose this trace as it has very high scenic activity with sharp transitions from one scene to the next and thus it can serve as a challenging trace for prediction. Other similar traces were also tested. It can be seen that the three techniques achieve reasonable prediction when there are no rapid changes between the scenes. The traces shown in figure 7.1(a), 7.1(b), and 7.1(c) also serve as a validation for the external prediction functions which were included in the RM simulator as mentioned in chapter 6.

Although it is difficult to determine which technique yields the best result from these short sub-sequences of the trace, it is evident that the Neural predictor achieves the best overall prediction, especially of the 'I' frames. This can be an indication that the cell loss performance may be better when using a Neural predictor since the 'I' frames are characterised by large bit sizes. The Neural predictor has one shortcoming: at some times it over-predicts the 'B' and 'P' frames. The reason for this negative aspect and for its superior capabilities in predicting the 'I' frames is that the correlation between the 'I' frames is usually higher than that between the 'B' and that between the 'P' frames [ROSE1]. Also another reason might be that the training file presented to the ANN had large numbers of 'I' frames which meant that the learning by the ANN of the 'I' frames properties was enhanced. The Neural predictor overall performance is because of the abilities of the ANN to capture the non-linearities inherent in the trace, which is difficult to capture using the linear or the EWMA predictors. The EWMA prediction is noticeably better than that of the Linear prediction especially with high scenic activity periods as in figure 7.1(b) and 7.1(c). The enhanced performance of the EWMA predictor is because the discount parameters were optimised.

The Linear predictor often over-predicts the 'P' frames and under-predicts the 'B' frames. This is mainly due to the value of the step size ' $\mu$ ' in equation 6-8. Recall, from chapter 6 that according to [HAYK1] the NLMS will converge in the mean if:  $0 < \mu < 2$ . Accordingly, in this work the values of ' $\mu$ ' were chosen through trial and error such that ' $\mu$ ' is between these two numbers. It was noticed that any slight deviation from the values chosen would render the linear prediction unstable, with hugely over-predicted and under-predicted bit sizes. In the opinion of the author, the Linear predictor as implemented by [ADAS] and [HAYK1] needs to include an algorithm to adapt the values of ' $\mu$ '. However to add such an algorithm to the Linear predictor would increase the already large computational complexity of this predictor possibly making it difficult to incorporate in the Earth Terminal.



Figure 7.1a Prediction of a scene from a Formula One race event using the Linear, EWMA, and Neural predictors (start of trace)



Figure 7.1b Prediction of a scene from a Formula One race event using the Linear, EWMA, and Neural predictors (middle of trace)



Figure 7.1c Prediction of a scene from a Formula One race event using the Linear, EWMA, and Neural predictors (end of trace)

To determine which of the three techniques yields the best result it will be necessary to look at the errors (residuals) after prediction. The errors resemble white noise and thus a good approach for evaluating the accuracy of the prediction is to consider the signal-to-noise ratio as has been proposed by Adas [ADAS]. Figure 7.2(a), shows the inverse signal-to-noise ratio for the three predicted traces. The inverse signal-to-noise ratio can be evaluated as shown in equation 7.1.

$$SNR^{-1} = \frac{\sum \chi^{2}(n)}{\sum e^{2}(n)}$$
(equ 7-1)

The Neural predictor achieves the least  $SNR^{-1}$  and hence at least for this MPEG trace (Formula One) it can be considered the suitable technique to use. This in fact was true for all the traces considered in this work. For example, in contrast to the Formula One trace a talk show trace [ROSE2] with very few abrupt changes between scenes was used for testing the three techniques and the outcome confirmed that the Neural predictor produces better results ( figure 7.2(b) ).



Figure 7.2a Comparison of the three prediction techniques (Formula One)



Figure 7.2b Comparison of the three prediction techniques (Talk Show)

The EWMA prediction in the previous two examples was achieved with the trend term included as in equations 6-13 to 6-16. The improvement in prediction with the trend term is not that much from the prediction without the term included. This is clearly shown in figure 7.3 for the Formula One trace.



Figure 7.3 EWMA prediction with and without the trend term

#### 7.3 Simulation and results

In this work seven empirical video traces are used for the simulation studies. The traces were obtained from a public ftp site [ROSE2] and they represent a wide range of video applications. The sequences have different statistical properties, for example typical TV sequences like sports, news and music clips lead to MPEG sequences with high peak bit rates and high peak-to-mean ratios compared to movie sequences. These properties result from the rapid movements of a lot of small objects, which increase the amount of data necessary to encode the sequence. The reason for using a number of different traces is to demonstrate that a predictive resource allocation scheme for real-time video is viable whatever the MPEG trace statistical properties, and that it always achieves better results than what is achievable using the classical on-demand allocation schemes currently proposed for broadband satellite communications.

The Movies, TV sport events and other TV sequences were encoded at the University of Wurzburg in Germany using a UC Berkeley MPEG-1 software encoder [ROSE1]. The statistics of the encoded sequences are shown below in Table 7.1.

Sequence	Frames	Frames	<u>Frames</u>	Bit rate	Bit rate
	Mean	COV	Peak/mean	Mean	Peak
	[bits]			[Mbps]	[Mbps]
Jurassic	13,078	1.13	9.1	0.33	1.01
Park					
Terminator	10,904	0.93	7.3	0.27	0.74
Formula	30,749	0.69	6.6	0.77	3.24
One					
Talk show	14,537	1.14	7.3	0.36	1.00
Soccer	25,110	0.85	7.6	0.63	2.29
News &	15,358	1.27	12.4	0.38	2.23
Ads					
Star Wars	15,599	1.16	11.9	0.36	4.24

#### Table 7.1 Statistic of the encoded sequences used in the simulation study

As the emphasis in this simulation study is on the resource allocation as a whole (both fixed and on-demand) and not on the fixed allocation as was the case when evaluating the PA and QFS schemes, p of equation 3-2 will be modified to represent the relative slots allocation rate as shown below:

$$p = slots allocation rate / MCR$$
 (equ. 7-2)

The two performance measures that are considered in the simulation study are the cell discard probability at the receiving Earth Terminal and the maximum buffer size at the sending terminal. The maximum buffer size is used instead of the mean size as it was felt that the averaging process might mitigate the sharp increases in buffer size due to under prediction, a phenomenon that is especially true of Linear prediction. The cell discard performance gives good indication of the delay performance as the performance objective for cell discard is a cell being discarded if at the receiving end it is found to be delayed by more than the maximum tolerable delay. However, for the sake of completeness the delay performance for one trace will be shown.

#### 7.3.1 End-to-End delay and delay variation results (receiving Earth Terminal)

Figure 7.4 shows the mean end-to-end delay experienced by cells generated using the Terminator trace. The relative slots allocation rate p was fixed at 4.2 for the classical and predictive allocation schemes simulations. The three predictive schemes show better delay performance than the classical scheme. The linear predictive scheme is slightly worse than both the EWMA and the Neural schemes. However, not much can be deduced from this figure regarding the performance of the EWMA to that of the Neural.



Figure 7.4 Mean End-to-End delay performance (Terminator trace)

The delay variation gives more information on the performance of the different schemes as shown in figure 7.5 where it can be seen that the predictive schemes vastly outperform the classical scheme. The Neural scheme outperforms both the EWMA and Linear schemes. For example, the ratio of the number of cells delayed by more than 295 ms to the number of transmitted cells drops to  $2*10^{-5}$  in the Neural predictive case while it drops to  $2*10^{-4}$  in the EWMA and to  $8*10^{-4}$  in the Linear cases and drops to only  $9*10^{-2}$  in the classical case.



Figure 7.5 Two-point delay distributions (Terminator trace)

#### 7.3.2 Cell discard at the receiving Earth Terminal

In these simulations the performance objective for cell discard was: any cell received at the receiving terminal and found to be delayed by more than 280ms is to be discarded. The cell discard rate is defined as follows:

Cell discard rate =  $\frac{\text{total number of cells delayed by more than 280ms}}{\text{total number of received cells}}$ 

(equ. 7 - 3)

The relative slots allocation rate p was varied in all simulations by changing the value of  $\phi$  in equation 6-3. Figures 7.6(a) to 7.6(g) clearly prove the superiority of the predictive resource allocation schemes over that of the classical scheme in achieving low cell discard rates with lower allocation rates. The cell discard rate figures of the Movies traces (figures 7.6(a), 7.6(b) and 7.6(c)) clearly show that the Neural predictor achieves better results than both the EWMA and the Linear predictors especially at medium to high allocation rates. The TV sports events and the TV sequences (figures 7.6(d) to 7.6(g)) also show the Neural predictor outperforming the EWMA and the linear predictors. In figure 7.6(e) (where the Formula One trace was used) the graph of the linear predictor starts at an allocation rate value greater than 1 (approx. 1.33). This indicates that the Linear predictor over-predicts when there are very high scene changes and abrupt movements between changes (characteristics that are typical of such a racing event). In figure 7.6(g) of the Talk show the EWMA and Linear predictors achieves similar results to those of the Neural predictor. This is due to the fact that there are very few changes between the scenes, which means that the learning of the ANN of the variations in the bit sizes of the MPEG frames becomes more difficult to achieve.

Overall it can be deduced from the cell discard figures that the EWMA predictor results are comparable to those of the Neural predictor. This is especially evident from figures 7.6(c) to 7.6(g).



Figure 7.6a Cell discard rates (Terminator)





Figure 7.6c Cell discard rate (Jurassic Park)



Figure 7.6e Cell discard rate (Formula One)



Figure 7.6f Cell discard rate (News and advertisements)



Figure 7.6g Cell discard rate (Talk show)

#### 7.3.3 Buffer space requirement at the sending Earth Terminal

In these simulations the buffer occupancy at the sending Earth Terminal was monitored at each cell arrival event and at the end of the simulation the maximum buffer size detected was then recorded. This was performed for different values of p (the relative slots allocation rate) resulting in plots of the maximum buffer size against p. It can be seen from figures 7.7(a) to 7.7(g), that as in the previous case of cell discard simulations the predictive schemes in general outperform the classical scheme in buffer space requirement.

The Neural scheme achieves by far the best results. For example, in figure 7.7(a) for a relative allocation rate of 4.8 the Neural scheme required a maximum buffer space for just one cell, the EWMA scheme required space for 24 cells, the Linear scheme required space for 41 cells and the classical scheme required space for 78 cells. However, at low allocation rates the predictive schemes tend to cause an increase in buffer space requirement. This is true for all the three predictors but it is especially true for the Linear predictor which tends to "under-predict" the most. An under-prediction can cause a rapid increase in the number of cells awaiting transmission and hence resulting in an instantaneous increase of buffer space leading to large maximum values as has been shown in the figures. If terminal buffer space is a prime factor for the satellite network operator then the predictive schemes should be implemented at medium to high relative allocation rates.

It is also worth noting that in figure 7.7(e) of the Formula One trace, there was not much reduction of buffer space as a result of implementing the predictive schemes which emphasises that for very active traces prediction can be a difficult task with little impact on the use of the ground segment resources.

The EWMA predictor proved to be comparable to the Neural one in the utilisation of buffer space as can be seen in figures 7.7(e) and 7.7(f) when the allocation rate is high.







Figure 7.7b Maximum buffer size (Star Wars)



Figure 7.7c Maximum buffer size (Jurassic Park)



Figure 7.7d Maximum buffer size (Soccer)



Figure 7.7e Maximum buffer size (Formula One)



Figure 7.7f Maximum buffer size (News and advertisements)



**7.4** The target scheme for the predictive resource allocation scheme Efficient utilisation of both the ground and space segment resources of the broadband satellite network can be achieved through predicting the uplink bandwidth requirements of the real time video connections and allocating to each connection part of the bandwidth according to the predictions. The target scheme for prediction should be simple and avoid complexity if it is to be implemented at the ET's.

The Linear predictive scheme when compared to both the EWMA and Neural scheme can be seen to produce the worst results of cell loss and buffer space requirement due to the difficulty in the choice of the step size ' $\mu$ ' as has been discussed in section 7.2. Although the Neural scheme consistently achieves the best results it involves training the ANN on a portion of the traffic and also scaling the inputs to the ANN between one and zero. There is no rule of thumb for configuring the ANN in terms of the size and make up of the training set and in terms of the number of inputs, neurons and hidden layers. An ANN trained on one type of traffic will degrade prediction if a different type of traffic is presented to it. The scaling also presents another set of problems; for example, what value(s) should be used for scaling the inputs to the ANN and how should the value(s) be derived from the traffic characteristics (e.g. PCR) or from the MPEG stream properties (frame size). In this work the scaling is achieved by dividing the MPEG frame sizes by the largest frame size to ensure that no input value is greater than one. In practice it is difficult to have *a priori* knowledge of the maximum MPEG frame size.

The EWMA scheme achieves good results that are sometimes comparable to those of the Neural scheme. It requires less computations than the Linear prediction and does not require to be trained on the traffic or require that the inputs to be scaled as is in the Neural scheme. The EWMA scheme can thus be viewed as the target scheme for the predictive resource allocation for the broadband satellite network.

#### 7.5 Position of slots in the TDMA frame

Figures 7.8(a) and 7.8(b) demonstrate the effect of the position of the reserved slots in the TDMA frame. A worst scenario for the reserved (allocated) slots is to place them at the start of the TDMA frame. Here it can be seen that the cell loss and buffer size when slots are reserved at the start of the frame (*in blue*) increase to values that are much greater than those achieved by reserving slots at the end of the frame (*in red*).



Figure 7.8a Cell discard rates achieved when the slots are reserved at the start and end of the TDMA frame (Terminator)



Figure 7.8b Maximum buffer size achieved when the slots are reserved at the start and end of the TDMA frame (Terminator)

#### 7.6 Summary

This chapter presented simulation results for the proposed predictive resource allocation scheme using three different prediction techniques, namely the Linear, EWMA and Neural prediction techniques. The predictive resource allocation schemes achieved superior results to those achieve by the classical scheme in terms of the following QoS parameters:

- cell loss
- buffer space requirement
- delay (CTD and CDV)

Although the Neural scheme achieved the best results of the three predictive schemes, the EWMA scheme achieved comparable results with no need for training on the traffic or scaling the traffic. The EWMA scheme requires less computations than the Linear scheme and moreover can be easily optimised.

The results also clearly shows that the position of the slots allocated using the predictive schemes should ideally be as close as possible to the end of the TDMA

frame so as to ensure that most of the cells destined to arrive during the frame period would arrive before the start time of their allocated slots. However, if the CTD and CDV are taken into consideration, the reserved slots should be concentrated around the middle of the TDMA frame to reduce the end-to-end delay. To avoid cell clumping the reserved slots should not be close to each other.

## **CHAPTER 8**

## **CONCLUSIONS AND FURTHER WORK**

A broadband satellite network operator can achieve differentiation and boost revenues if the tight QoS requirements of users running mission critical applications (class-1 users) can be met with reduced allocated network resources. Furthermore delays experienced by best effort users should also be reduced to tolerable levels especially during times of heavy usage and high load. An efficient implementation of network resource management can allow the realisation of the above two quests. However in satellite communications the latency problem and the limited bandwidths pose difficulties in dimensioning the network resources between the users, and guaranteeing the QoS requirements of the users.

This thesis delivered a novel approach for the problem of resource management in broadband satellite networks by proposing two scheduling strategies for class-2, 3, and 4 connections and a predictive resource allocation scheme for class-1 connections.

Both scheduling strategies utilise an algorithm that takes into account the mean delays of class-2 and class-3 connections. In one strategy, the PA strategy, the algorithm initiates the intermittent demotion of the priorities of the higher-class connections and promotion of the priority of class-4 connections. In another strategy, the QFS strategy, the algorithm initiates the intermittent acquisition of fixed slots in the TDMA frame by class-4 connections. These fixed slots are owned by higher-class connections (2 and 3 but not 1). Both strategies have been evaluated through simulation and the results have shown drastic reduction of the end-to-end delay experienced by best effort connections as well as improvement in the TDMA frame utilisation. Another feature of using either of the strategies in the satellite network is that the strategies can be implemented in conjunction

with other class-based scheduling schemes. Also, most of the complexity is confined to the ground segment and minimum change is required to the design of the OBP.

Simulation results also showed that there was a slight degradation of the QoS of higherclass connections. However, the results revealed that though the QFS cannot achieve the level of best effort connections delay reduction as the PA strategy, it allows better flexibility in the control of the QoS degradation suffered by the higher-class connections.

MPEG coding is now the standard coding for video applications, and thus real time video over satellite links will be mostly MPEG coded. The proposed predictive resource allocation scheme utilises the inherent features of MPEG video: the frame-by-frame correlation and deterministic periodicity of frames, to predict the traffic arrival process. The predictive scheme aims to achieve provision of tight QoS guaranties to video connections (class-1 connections). The performance of the scheme was evaluated using simulation and the results obtained showed that the scheme outperforms the classical bandwidth-on-demand based schemes, such as Hung's scheme. In the simulated satellite network, the cell discard rate at the receiving terminals dropped significantly as a result of implementing the predictive scheme and the buffer space requirement at the sending terminals showed a noticeable drop. Low cell discard rates can be achieved with less allocated bandwidth than is possible with the classical schemes.

Three prediction techniques (Linear, EWMA and Neural) were used for the proposed functional architecture of the predictive scheme with the aim of identifying the technique that provides efficient prediction of the MPEG frames, and offers the least complexity, bearing in mind the limited resources (memory and power) of the Earth Terminal. The results obtained proved that the Neural technique provides the best prediction while the Linear gives the worst prediction. However, the Neural technique requires training on portions of the MPEG traffic streams, and scaling of the inputs to the Neural network. The EWMA technique achieves slightly less efficient prediction and does not require *a priori* knowledge of the traffic characteristics or learning as with the Neural technique and is less complex than the Linear technique. Hence an added contribution from this

work is the identification of the EWMA technique as a better candidate than the Linear technique for predictive based dynamic bandwidth allocation schemes as was previously proposed by Adas *et al* [ADAS].

The limitations in implementing the PA and QFS strategies manifest themselves in situations where the traffic mix in the satellite network is mainly of class-1 nature and/or of class-4 nature (*i.e. not enough of class-2 and class-3 connections*). Other technical limitations are the increase of the signalling load and impact on the MAC protocol because of the requirement of communicating the probabilistic information to the satellite. Earth Terminal power and memory constraints may prove to be a limitation when computing the end-to-end delay at the end of each update period.

The main significance of this work is the proposition of predictive resource allocation for real time video traffic in the satellite network. The predictive resource allocation scheme has been shown to achieve better QoS guarantees than the classical bandwidth on demand schemes that are currently used for broadband satellite communications. Of course, the main limitation is that the proposed predictive scheme can only be used with MPEG coded video. Other limitations are the increase in the signalling load since at call set-up the properties of the MPEG stream must be signalled to the satellite. Two separate bandwidth requests (Buffer occupancy and predicted cell arrivals) are generated by the terminal at the start of each TDMA frame. Hence, there is an impact on the MAC protocol since the MAC must provide slots in out of band signalling or piggyback the requests with the data in inband signalling. With fixed Earth Terminals, where a number of users access the satellite network, it may prove to be more efficient if each user sends his predictive information. The Earth Terminal would then transmit the predictive information as a generic request which means that no modification to the design of the terminal is required.

The results of this thesis have been disseminated in various workshops, conferences and journals. Appendix C shows a list of publications that have resulted from this research.

#### **Further work**

The objectives set out for this research have been achieved. However, more work is required to evaluate the impact of the proposed scheduling strategies on the congestion control function at the downlink buffers of the satellite. The proposed scheduling strategies will increase the rate of transmission of cells from best effort users to the satellite, which may flood the downlink buffers and thus accelerate the onset of congestion at the satellite.

Policing at the User Network Interface (UNI) was not considered in this work i.e. traffic from the sources accessing the satellite network was not regulated prior to entering the network. At certain instances during a connection lifetime the source transmission rate may be capped by the policing algorithm and therefore a situation may develop whereby the predicted number of arrivals is very much higher than the actual number of arrivals. This can increase the number of wasted slots in the TDMA frames. More research is required to evaluate the impact of traffic policing on the predictive resource allocation scheme.

It would also be interesting to evaluate the performance of the predictive scheme when a number of MPEG coded video traffic streams share the satellite capacity. In such a scenario the allocations will be constrained by the available bandwidth. One possible solution to increase the available bandwidth to the video connections (class-1) is to implement a PA or QFS scheduling strategy and intermittently allow both class-1 and class-4 connections to use some of the fixed slots of class-2 and class-3 connections.

Although this work has targeted ATM over satellite, the strategies and schemes proposed can be adapted to cater for IP over satellite. Frame units (*where the IP packets are segmented into fixed size units*) are sent to the satellite and TDMA slots would accommodate one or more of the frame units. However, the delay due to the segmentation and reassembling of the IP packets may be an issue.

#### **APPENDIX A**

#### **EXACT FLUID-FLOW ANALYSIS**

The analysis considers queuing at the burst level of a single on/off source feeding an ATM buffer. The key feature of this analysis is that it deals fundamentally with rates. When the source rate exceeds the queue's service rate, the queue begins to fill and ultimately, for a finite queue, it overflows. This is referred to as burst-scale congestion [ROBER] as shown in figure A1. In the literature, burst-scale congestion is analysed using fluid-flow techniques based on those of Anick et al [ANICK] and Tucker [TUCKE]



Figure A1 Burst scale congestion occurring as a result of arrival rate exceeding the service rate

Schormans and Pitts [SCHO1], [SCHO2], [SCHO3] use a rate-based model, analogous to standard fluid-flow analysis, but one which considers changes to the queue size in discrete steps rather than approximating them as continuously variable as in fluid-flow analysis. Consider figure A1 and define *R* to be the actual cell arrival rate, *C* the actual cell transmission rate, *E*[ON] the expected on time for the source, and *E*[OFF] the expected off time for the source. When the buffer is not full and the source is in the ON state, the queue will fill at a constant rate of (R - C), but if the buffer is full, then the cells will be lost at a rate of (R - C). Once the ON/OFF source

model has entered the OFF state, it will remain there for at least one time slot and once the source has entered the ON state, it will generate at least one 'excess-rate' arrival. Define 's' to be the probability that after each time slot in the OFF state the source remains in that state and define 'a' to be the probability that after each arrivals in the ON state the source generate another 'excess-rate' arrival. The ON/OFF periods are geometrically distributed [SCHO2] and hence according to [SCHO3]:

*E*[number of 'excess-rate' arrivals in an ON period] = 1/(1 - a)*E*[number of 'excess-rate' arrivals in an ON period] = 1/(1 - s)

Now

E[number of 'excess-rate' arrivals in an ON period] = E[ON time](R - C)

SO

a = 1 - 1/(E[ON time](R - C))

also

E[number of time slots in an OFF period] = E[OFF time]C

so

s = 1 - 1/(E[OFF time]C)

Define  $p[k] = p[an \text{ 'excess-rate' cell finds k cells in the queue on arrival] and N to be$ the buffer capacity,*CLP*to be the exact cell loss probability. For the level betweenstates N - 1 and N, equating probabilities gives:

a.p[N-1] = p[N](1 - a)

For the level between states N - 2 and N - 1, equating probabilities gives

$$a.p[N-2] = p[N](1 - a)s + p[N - 1](1 - a)s$$
$$= p[N - 1]s$$
$$p[N - 2] = p[N - 1](s/a) = p[N]s(1 - a)a^{2}$$

$$= p[N](s/a)^2(1 - a)/s$$

Similarly,

$$p[N-3] = p[N-2](s/a) = p[N](s/a)^3(1-a)/s$$
$$p[N-I] = p[N-I+1](s/a) = p[N](s/a)^i(1-a)/s$$

By definition,

$$\sum_{i=0}^{N} p[i] = p[N] + \sum_{i=1}^{N} p[N-1] = 1$$

so substituting for p[N-1] gives:

$$p[N] = \frac{1}{\sum_{i=1}^{N} \left( \left( \frac{s}{a} \right)^{i} \left( \frac{1-a}{s} \right) \right)}$$
(equ A - 1)

Equation A.1 can be rearranged as it is the sum of a geometric progression:

$$p[N] = \frac{1}{1 + \left(\left(\frac{s}{a}\right)^{N} - 1\right) \cdot \left(\frac{1 - a}{s - a}\right)}$$
(equ A - 2)

p[N] is the probability that an 'excess-rate' cell finds N cells in the queue on arrival and is therefore lost. Hence, the cell loss probability for 'excess-rate' cells is p[N]. However, the overall cell loss probability *CLP* needs to be relative to all the cells, i.e. the whole rate R, not just the excess rate R - C. Thus:

$$R.CLP = (R - C)p[N]$$

This yields:

$$CLP = \frac{R - C}{R\left[1 + \left(\left(\frac{s}{a}\right)^{N} - 1\right) \cdot \left(\frac{1 - a}{s - a}\right)\right]}$$
(equ A - 3)

Equation A.3 allows the network operator to estimate the cell loss for a given ON/OFF source and buffer size, alternatively it allows the operator to dimension the buffer sizes and/or capacities at the different nodes for a given CLP requirement.

## **APPENDIX B**

## THE BACKPROPAGATION LEARNING ALGORITHM

In backpropagation (BP) neural networks the differences (errors) between the desired outputs and the outputs form nodes in the output layers, are propagated to nodes in the previous layers so as to adapt the weights. The NN is trained by initially selecting small random weights and thresholds and then presenting all training data repeatedly. Weights are adjusted after every trial until weights converge and a cost function is reduced to an acceptable level. This cost function is equal to the mean square error and is minimised using a gradient search technique.

Figure B1 (adapted form [FAN] and [HAYK2]) shows an architectural graph for BP learning that incorporates both a forward and backward phases of the computations involved in the learning process. The upper part of figure B1 accounts for the forward phase while the lower part accounts for the backward phase. The notations used in the figure are as follows [HAYK2]:

- $X_p$  = inputs to the first layer (the input layer i.e. l = 0), where p is number of inputs
- $\mathbf{w}^l$  = synaptic weight vector of a neuron in layer l
- $v^l$  = vector of net internal activity levels of neurons in layer *l*
- $y^l$  = vector of function signals of neurons in layer l
- $\theta^l$  = vector of function signals of neurons in layer *l*
- $\delta^l$  = vector of local gradients of neurons in layer *l*
- **e** = error vector represented by e<sub>1</sub>, e<sub>2</sub>, ...., e<sub>q</sub> as elements, where q is the number of outputs



## Figure B1 An architectural graph of the backpropagation algorithm (adapted from [HAYK2] and [FAN]

[FAN] summarised the BP algorithm in five distinct phases. These were the Initialisation, Presentation of the training examples, Forward computation, Backward computation and Iteration phase:

- 1. *Initialisation*: Weights and threshold levels are set to uniformly distributed small random numbers.
- Presentation of training examples: An epoch of training examples is presented to the network where the training data set is {[x (n), d (n)]; n = 1,2,..., N} and where x (n) is the input and d (n) is the desired output. For each example in the set, perform the following sequence of forward and backward computations in steps 3 and 4, respectively.
3. *Forward computation*: Computation of the activation potentials and function signals of the network is performed by proceeding forward through the network, layer by layer. The net internal activity level  $v_i^l(n)$  for neuron *j* in layer *l* is:

$$V_{j}^{l}(n) = \sum_{i=0}^{p} W_{ji}^{l}(n) Y_{i}^{l-1}(n)$$
 (equ B-1)

where  $y_i^{l-1}(n)$  is the function signal of neuron *i* in the previous layer l-1 at iteration *n* and  $w_{ji}^{l}(n)$  is the weight of neuron *j* in layer *l* that is fed from neuron *i* in layer l-1. For i = 0,  $y_0^{l-1}(n) = -1$  and  $w_{j0}^{l}(n) = \theta_j^{l}(n)$ , where  $\theta_j^{l}(n)$  is the threshold applied to neuron *j* in layer *l*. As the sigmoidal nonlinearity of each neuron is defined by a logistic function, i.e.,  $\varphi(x) = \frac{1}{1 + \exp(-x)}$ , the function (output) signal of neuron *j* in layer *l* becomes:

$$y_{j}^{l}(n) = \frac{1}{1 + \exp(-v_{j}^{l}(n))}$$
 (equ B-2)

If neuron *j* is in the first hidden layer (i.e., l = 1),  $y_j^0(n) = x_j(n)$ , where  $x_j(n)$  is the *j*th element of the input vector **x** (n). If neuron *j* is in the output layer (i.e., l = L),  $y_j^L(n) = o_j(n)$ . The error signal is:  $e_j(n) = d_j(n) - o_j(n)$ , where  $d_j(n)$  is the *j*th element of the desired output vector **d** (n).

4. *Bacward computation*. Computation of the  $\delta$ 's (i.e., the local gradients) of the network is performed by proceeding backward, layer by layer:

$$\delta_{j}^{L}(n) = e_{j}^{L}(n) O_{j}(n) [1 - O_{j}(n)] \text{ for neuron } j \text{ in output layer } L \quad (equ B - 3)$$

$$\delta_{j}^{L}(n) = y_{j}^{l}(n)[1 - y_{j}^{l}(n)]\sum_{k} \delta_{k}^{l+1}(n) y_{kj}^{l+1}(n) \text{ for neuron } j \text{ in hidden layer } l$$
(equ B - 4)

The weights of the network in layer *l* are then adjusted by:

$${}^{l}_{W_{ji}}(n+1) = {}^{l}_{W_{ji}}(n) + \eta \, \delta^{l}_{j} \, y^{l-1}_{i}(n)$$
(equ B - 5)

where  $\eta$  is the learning rate parameter.

To achieve faster convergence a momentum term is added and weight changes are smoothed by:

$$W_{ji}^{l}(n+1) = W_{ji}^{l}(n) + \eta \delta_{j}^{l} Y_{i}^{l-1}(n) + \alpha [W_{ji}^{l}(n) - W_{ji}^{l}(n-1)]$$
(equ B - 6)

where  $\alpha$  is the momentum constant.

5. *Iteration*: Steps 2 to 4 are repeated for all training examples of different epochs with randomised examples until the free parameters of the network stabilize their values and the average squared error computed over the entire training set reaches a minimum small value.

# **APPENDIX C**

## **AUTHOR'S PUBLICATIONS**

#### Journal papers

- [AWAD4] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Framework for supporting real time VBR traffic over satellite networks, Electronics Letters, IEE, Vol. 35, No. 8, pp 616-617, 1999
- [AWAD6] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Improving the delay performance of 'best effort' traffic over ATM satellite networks, Electonics Letters, IEE, Vol. 35, No. 17, pp. 1430-1432, 1999
- [AWAD8] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Resource allocation strategy for ATM-based satellite systems, Electronics Letters, IEE, Vol. 36, No. 3, pp. 281-282, 2000

#### **Conference papers**

- [AWAD1] H. O. Awadalla, Personal communications in the context of future broadband satellite networks, Fourth communication networks symposium, Manchester, 1997
- [AWAD2] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, An algorithmic component of an intelligent Earth Terminal for ATM satellite networks, Sixth IFIP workshop on performance modelling and evaluation of ATM networks, Ilkley, 1998
- [AWAD3] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Efficient scheduling strategies for multimedia traffic in broadband satcom systems, 7<sup>th</sup> International conference on telecommunication systems, Nashville, 1999
- [AWAD5] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Predictive resource allocation for real time VBR traffic in ATM satellite networks, Third IFIP

workshop on the management and design of ATM networks, London, 1999

[AWAD7] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Predictive resource allocation for real time video traffic in broadband satellite networks, Fifth IFIP international conference on broadband communications (BC'99), Hong Kong, 1999.
 Also published in Broadband Communications: Convergence of Network

Technologies, edited by D. H. K. Tsang, and P. J. Kuhn, Kluwer Academic Publishers, pp. 509-519, 1999

[AWAD9] H. O. Awadalla, L. G. Cuthbert and J. A. Schormans, Achieving tight QoS guarantees for video traffic in packet-switched networks, accepted paper, UKTS, London, 2000

# **APPENDIX D**

### REFERENCES

- [ADAS] N. Adas, Supporting real time VBR using dynamic reservation based on linear prediction, INFOCOM'96, pages 3:1476-1483, San Francisco, 1996
- [AKYIL] I. F. Akyildiz and S.H. Jeong, Satellite ATM networks: A survey, IEEE Communications Magazine, pages: 35(7):30-34, 1997
- [ANANA] F.Ananasso and L.V. Giorgio, The role of VSAT systems in broadband networks and intelligent networks, International Journal of Satellite Communications, pages 13:11-29, 1994
- [ANICK] D. Anick, D. Mitra and M. M. Sondhi, Stochastic theory of a data handling system with multiple sources, BSTJ, Vol. 61, pages 968-981, 1982
- [ASTRO] http://www.astrolink.com
- [ATMF1] ATM Forum, ATM Traffic Management Specifications, version 4.0, ATM/Forum/95-0013R6, 1995
- [ATMF2] ATM Forum, A progress report on the standards development for satellite ATM networks, ATM Forum/98-0828, 1998
- [BAF] Broadband Access Facilities project, Report on Performance Studies BAFDeliverable 25, project number: RACE 2024, 1994
- [BAIOC] A. Baiocchi and M. Listanti, An ATM-like system architecture for satellite communications including on-board switching, International Journal of Satellite Communications, pages(14):384-412, 1996
- [BALES] F. Balestrieri, P. L. Panteli, V. Dionissopoulos, T. G. Clarkson, ATM connection admission control using pRAM based Artificial Neural Networks, Six IFIP workshop on performance modelling and evaluation of ATM networks, 1998
- [BELLC] Public ftp: thumper.bellcore.com, Directory:pub/vbr.video.trace
- [BERAN] J. Beran, R. Sherman, M.S. Taqqu and W. Willinger, Long-range dependence in variable-bit-rate video traffic, IEEE Transaction on Communications, pages: 43(2):1566-1579,1995
- [CELA1] N. Celandroni, E. Ferro and F. Potorti, Demand Assignment TDMA

satellite schemes: Distributed and centralised solutions, IEEE ICC, pages:2:911-915, 1996

- [CELA2] N. Celandroni, E. Ferro, F. Potorti, and G. Maral, Bursty data traffic via satellite: Performance comparison between two TDMA access schemes, European conference on satellite communications, pages 170-175, 1997
- [CHAND] K. Chandra and A.R. Reibman, Modelling one and two-layer variable bit rate video, IEEE/ACM Transactions on Networking, pages 7(30):398-413, 1999
- [CHIOT] T.Chiotis, C. Stathis and B. Maglaris, The impact of Self-similarity on statistical multiplexing of MPEG video data, Sixth IFIP workshop on performance modelling and evaluation of ATM networks, pages 14:1-10, Ilkley, 1998
- [CHITR1] D.M. Chitre, D.S. Gokhale, T. Henderson, J.L. Lunsford, and N. Mathews, Asynchronous Transfer Mode (ATM) operation via satellite: Issues, Challenges and Resolutions, International Journal of satellite Communications, pages 12:211-222, 1994
- [CHITR2] P. Chitre and F. Yegenoglu, Next-generation satellite networks: Architectures and implementations, IEEE Communications Magazine, pages 37(30:30-36, 1999
- [CHONG] S. Chong, S. Li and J. Ghosh, Dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM, Proceedings of IEEE INFOCOM'94, pages 81-90, 1994
- [CHU] P.P. Chu, W.D. Ivancic and H. Kim, On-board closed-loop congestion control for satellite based packet switching networks, Technical memorandum 106446 AIAA-94-1062, NASA Lewis research Center, Cleveland, Ohio, USA, 1994
- [CYBER] http://www.cyberstar.com
- [DERMO] T. MC Dermott and T. Van Landegem, New services mean new requirements for IP networks, Alcatel Telecommunications Review, 2<sup>nd</sup> quarter, pages 131-135, 1999
- [EUROW] http://www.euroskyway.alespazio.it
- [EVAN1] B. G. Evans, Satellite Communication Systems, 3<sup>rd</sup> edition, IEE, 1999
- [EVAN2] B.G. Evans and R. Tafazolli, Future multimedia communications via

satellite, IJSC, Vol. 14(6), pages 14(6): 467-474.

- [FAIR1] G. Fairhurst and H.S. Cheng, The effect of satellite errors on the performance of ATM, 3<sup>rd</sup> IFIP workshop on performance modelling and evaluation of ATM networks, 1995
- [FAIR2] G. Fairhurst and H.S. Cheng, Performance issues of interconnecting ATM islands, IEE ATM over satellite colloquium, 1996
- [FAIR3] G. Fairhurst and H.S. Cheng, Techniques for improving the performance of ATM over satellite and point-to-point RF links, 5<sup>th</sup> IFIP workshop in performance modelling and evaluation of ATM networks, Ilkley, UK, 1997
- [FAN] Z. Fan, Some aspects of traffic control and performance evaluation of ATM networks, Ph.D. thesis, School of Engineering, University of Durham, 1997
- [FARSE] J. Farserotu and A. Tu, Test analysis of low data rate ATM over SATCOM, IEEE MILCOM, pages 224-230, 1995
- [GALL1] R. G. Gallager and A. K. Parekh, A generalised processor sharing approach to flow control in integrated services network: the single-node case, IEEE/ACM Transactions on Networking, Vol.1, pages 3:344-357, 1993
- [GALL2] R. G. Gallager and A. K. Parekh, A generalised processor sharing approach to flow control in integrated services network – the multiple node case, IEEE INFOCOM'93, Vol.2, pages 521-530, 1993
- [GARRE] M. W. Garrett, A service architecture for ATM: From applications to scheduling, IEEE Network Magazine, pages 6-14, May/June 1996
- [GILDE] J. Gilerson and J. Cherkaoui, Onboard switching for ATM via satellite, IEEE Communications Magazine, pages 7:66-70, 1997
- [GRANG] C. Granger, P. Newbold, Forecasting economic time series, Academic Press, New York, 1977
- [HANDE] R. Handel, M. N. Huber, S. Schroder, ATM networks: Concepts, protocols and applications, 3<sup>rd</sup> edition, Addison-Wesley Publishing Company, 1998
- [HAYES] M.H Hayes, Statistical Signal Processing and Modelling, John Wiley and Sons, 1996
- [HAYK1] S. Haykin, Adaptive Filter Theory, Prentice Hall, 1991
- [HAYK2] S. Haykin, Neural Networks: A comprehensive foundation, Macmillan

College Publishing Company, 1994

- [HIRAM] A. Hiramatsu, ATM communications network control by neural networks, IEEE Trans. Neural Networks, pages 1:122-130, 1990
- [HUANG] J. Huang, T. L. Ngoc, and J. F.Hayes, Broadband satcom systems for multimedia systems, IEEE ICC, pages 906-910, 1990
- [HUNG1] A. Hung, M. J. Montpetit, G. Kesidis, and P. Takats, A framework for ATM via satellite, IEEE Globecom, pages 1:1020-1025,1996
- [HUNG2] A. Hung, M. J. Montpetit, and G. Kesidis, ATM via satellite: A framework and implementation, ACM-Baltzer Wireless Networks special issue on hybrid and satellite communications networks, pages: 4(2):141-153, 1998
- [HUTER] M. Huterer, J. Minnis, P. O'Connell and E. Traupman, Next generation networks: A cost effective combination of transport and switching layers, Alcatel Telecommunications Review, 2<sup>nd</sup> quarter, pages 109-117, 1999
- [ITU1] ITU-T Recommendation I.113, Vocabulary of terms for broadband aspects of ISDN, 1991
- [ITU2] ITU-T Recommendation I.121, Broadband aspects of ISDN, 1990
- [ITU3] ITU-T Recommendation I.150, B-ISDN ATM functional characteristics, 1995
- [ITU4] ITU-T Recommendation I.356, B-ISDN ATM layer cell transfer performance, 1996
- [IVANC] W.D. Ivancic, P. Chu, and D. Shyy, Technical issues regarding satellite packet switching, International Journal of Satellite Communications, pages 12:257-267, 1994
- [JAIN1] R. Jain, S. Fahmy, F. Lu, and S. Kalyanaraman, A survey of congestion control techniques and data link protocols in satellite networks, http://www.cis.ohio-state.edu/~jain/papers/sat surv.htm
- [JANAC] G. Janacek, L. Swift, Time Series: forecasting, simulation and applications, Ellis Horwood, Chichester, 1993
- [JANG] Y.M. Jang, A. Ganza, and F.Y. Hayes, predictive congestion control for broadband satellite systems, IEEE ICC, Dallas, 1996
- [JOO] J.K. Joo, ATM on-board satellite congestion control for multimedia services, MPhil. Thesis, University of Surrey, 1999
- [KALTE] T. Kaltenschnee and S. Ramseier, Impact of burst errors on ATM over satellite – Analysis and Experimental results, 10<sup>th</sup> IEE International

Conference on Digital satellite Communications, pages 236-243, 1995

- [KOWTH] S. Kowth and D. R. Vaman, A generalized ATM traffic model and its application in bandwidth allocation, IEEE ICC, pages 1009-1013, 1992
- [LAW] A. M. Law and W. D. Kelton, Simulation modelling and analysis, McGraw-Hill, 1982.
- [LIM] S.H. Lim, and D.M. An, Impact of cell unit interleaving on header error control performance in wireless ATM, IEEE GLOBECOM, 1996
- [LIU1] Y. Liu, Adaptive intelligent traffic management in ATM networks, Ph.D. thesis, University of Miami, 1995
- [LIU2] Y. Liu and C. Douligeris, Feedback congestion controleer for ATM networks using a Neural network predictor, SOUTHCON'95, New York, pages 398-402, 1995
- [LIU3] Y. Liu and C. Douligeris, Rate regulation with feedback controller in ATM network – A neural network approach, IEEE Journal on Selected Areas in Communications, Vol. 15, pages 2:200-208, 1997
- [LOSQU] G. Losquadro and M. Marinelli, The EuroSkyWay system for interactive multimedia and the relevant traffic management, Third Ka-Band utilization conference, pages 17-24, 1997
- [LUNSF] J. Lunsford, S. Narayanaswamy, D. Chitre, and M. Neilbert, Link enhancement for ATM over satellite links, IEE 10<sup>th</sup> International Conference on Digital Satellite Communications, pages 129-136, 1995
- [MARAL] G. Maral and M. Bousquet, Satellite Communications Systems: Systems, Techniques and Technology, 3<sup>rd</sup> Edition, John Wiley and Sons, Chichester, UK, 1998
- [MARKE] J.D. Markel and A.H. Gray, Linear Prediction of Speech, Springer-Verlag, New York, 1982
- [MARS] P. Mars, J. R. chen and R. Nambiar, Learning Algorithms: Theory and applications in signal processing, control and communications, CRC Press, 1996
- [MARSH] W.J. Marshal and A.Q. Je, Approaches and requirements for tactical Satcom ATM integration, Proceedings of MILCOM, pages 212-216, 1995
- [MERTZ] I. Mertzanis, G. Sfikas, R. Tafazolli, and B. Evans, Protocol architectures for satellite ATM broadband Networks, IEEE Communications Magazine,

pages 3:46-54, 1999

- [MOH] W. M. Moh, M. J. Chen, N. M. Chiu and C. D. Liao, Traffic prediction and dynamic bandwidth allocation over ATM: A neural network approach, Computer Communication, pages 563-571, 1995
- [NGOC1] T. L. Ngoc, and S. V. Krishnamurthy, Performance of Combined Free/Demand Assignment Multi-Access (CFDAMA) protocol with preassigned request slots in integrated voice/data satellite communications, IEEE ICC, pages 1572-1576, 1995
- [NGOC2] T. L. Ngoc, and S. V. Krishnamurthy, Performance of combined free/demand assignment multiple-access schemes in satellite communications, International Journal of Satellite Communications, pages 14:11-21, 1996
- [NGOC3] T. L. Ngoc, and H.R. Mehrvar, ANN approach for congestion control in packet switch OBP satellite, IEEE ICC, pages 810-814, 1995
- [NGUYE] T. Nguyen and T. Suda, Survey and evaluation of multiple access protocols in multimedia satellite networks, IEEE Southeast conference, pages 408-413, 1990
- [ONVUR] R.O. Onvural, Asynchronous Transfer Mode Networks, Performance Issues, 2<sup>nd</sup> Edition, Artech House, Norwood, MA, 1995
- [OPNET] OPNET Online Documentation; Version 3.0A, MIL3 Inc, Washington
- [ORS1] T. Ors, Traffic and congestion control for ATM over Satellite to provide QoS, Ph.D. Thesis, University of Surrey, 1998
- [ORS2] http://www.ee.surrey.ac.uk/Personal/T.Ors/atmsat/projects.html
- [ORS3] T. ORS, Z. SUN and S. P. W. Jones, ATM networking with VSATs, Fourth IFIP workshop on performance modelling and evaluation of ATM networks, pages 37:1-10, Ilkley, 1996
- [ORS4] T. ORS, and B. G. Evans, Performance Evaluation of a MAC protocol for ATM over satellite, COST 257 Management committee meeting TD(97)053, 1997
- [PEYRA] H. Peyravi, Multiple Access Control (MAC) Protocols for the Mars regional network: A survey and assessment, Report for NASA, Lewis Research Centre, Cleveland, Ohio, U.S.A, 1995

- [PILIP] S. Pilipovic, Z. Ilic, and M. Kos, The source model of VBR video based on modulated TES process, The International Society for Optical Engineering (SPIE) Conference on Performance and Control of Network Systems III, Boston, 1999
- [PITTS] J. M. Pitts, Cell-Rate simulation modelling of Asynchronous Transfer Mode telecommunications networks, Ph.D. Thesis, University of London, 1993
- [POCHE] H. L. Pocher and V.C.M. Leung, Real-time multimedia scheduling policies for end-to-end delay jitter and loss guarantees across ATM satellite systems, IEEE Journal on Selected Areas in Communications, pages(17)2:314-325, 1999
- [ROBER] J. W. Roberts, Variable-bit-rate traffic control in B-ISDN, IEEE Communications Magazine, Vol. 29, pages 9:50-56, 1991
- [ROSE1] O. Rose, Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM systems, Research Report Series, Report 101, Institute of Computer Science, University of Wuerzberg, Germany, 1995
- [ROSE2] ftp-info3.informatil.uni-wuerzburg.de, Directory: /pub/MPEG/traces
- [SCHO1] J. A. Schormans and J. Pitts, Introduction to ATM design and performance, John Wiley & Sons, 1996
- [SCHO2] J. A. Schormans, J. Pitts, K. Williams and L. G. Cuthbert, Equivalent capacity for on/off sources in ATM, Electronics Letters, Vol. 30, pages 21:1740-1741, 1994
- [SCHO3] J. A. Schormans, J. Pitts, and L. G. Cuthbert, Excat fluid-flow analysis of single on/off source feeding an ATM buffer, Electronics Letters, Vol. 30, pages 14:1116-1117, 1994
- [SKYBR] http://www.skybridgesatellite.com
- [STRAT] STRATOSPHERIC project, Cell discarding in satellite links, document number 2016/MET/WP1/DR/C/012/b1/A2, Issue 2, 1993
- [SUN1] Z. Sun, F.P. Coakley, B.G. Evans, Satellite ATM for broadband ISDN, telecommunication Systems, pages 4:119-131, 1995
- [SUN2] Z. Sun, T. Ors and B.G. Evans, Interconnection of broadband islands via satellite – Experiments on the RACE II CATALYST project, Transport protocols for High-Speed broadband networks workshop at GLOBECOM'96, London, 1996

- [TARAS] L. Tarassenko, A Guide to neural Computing Applications, John Wiley & Sons Inc., New York, 1998
- [TESFA] Y.A. Tesfai and S.G. Wilson, FEC schemes for ATM traffic over wireless links, http://watt.seas.virginia/papers/main\_milcom.ps
- [TUCKE] R. C. F. Tucker, Accurate method for analysis of a packet speech multiplexer with limited delay, IEEE Transactions, Vol. COM-36, pages 479-483, 1988
- [VIDAL] L. Vidaller, J. Aracil, A. Martinez, J. Perez and A. Ruiz, Network architectures and multiple access for ATM over satellite, Proceedings of the IEEE Gigabit Networking workshop, 1995
- [WEIN] C. J. Weinstein, Fractional speech loss and talker activity model for TASI and for packet switched speech, IEEE Transactions on Communications, pages 26:1253-1257, 1978
- [WU] X. Wu, S. Wu, H. Sun and L. Li, Dynamic slots allocation multiple access protocol for wireless ATM networks, IEEE ICC, pages 1561-1565, 1997
- [ZEIN1] T. Zein and G. Maral, A dynamic allocation protocol for a satellite network integrated with a B-ISDN, fourteenth AIAA Int'l. Communications Satellite Systems conference, pages 185-191, 1992
- [ZEIN2] T. Zein, G. Maral, T. Brefort and M. Tondriaux, Performance of Combined/Fixed Reservation Assignment (CFRA) scheme for aggregate traffic, COST 226 Integrated space/terrestrial networks final symposium, pages 183-199, Hungary, 1995