Abstract

Variable Bit Rate (VBR) MPEG traffic is expected to be one of the widely used traffic sources for high speed networks, along with voice and data traffic. Encoded video traffic is a correlated and a bursty traffic with high value of peak to mean ratio (burstiness). For efficient traffic management in a high speed network, it is important to know the basic characteristics of such encoded video traffic.

This thesis deals with three main issues regarding the management of VBR MPEG traffic over ATM networks: Characterisation, Modelling and Multiplexing of VBR MPEG traffic. The statistical characteristics of MPEG traffic have been analysed. Our study is based on long and various video streams obtained from real *Movies*, *News* and *Sports* events. The work explores three statistical measures which are the main characteristics of MPEG streams, namely: Distribution, Autocorrelation Function (ACF) and Scene changes. Based on the statistical analysis, two simple Markovian based models are introduced: the histogram based model and the Detailed Markov Chain. Despite the simplicity of the models, it is possible to show some improvement in the statistical behaviour. The scene change measure is analysed in more detail because the scene changes are important reason for the fluctuations in the overall bit rate within the encoded video stream. This thesis presents two methods and algorithms to identify scene changes within the MPEG stream. Then, the scene change identification technique as well as the characteristics of scene changes are used to propose a construction of an MPEG source model. The model captures accurately the statistical behaviour of the actual sequence at different time scales.

For ATM networks, the challenge of QoS guarantees is to allocate an effective bandwidth for each video connection and a tradeoff should be achieved between improving the network utilisation and providing QoS guarantees. An allocation bandwidth approach based on a deterministic model for multiplexed VBR MPEG streams is presented. An arrangement for the multiplexed VBR MPEG streams is then presented. Furthermore, the impact of such arrangements on the allocated bandwidth is shown. Finally, the impact of the stream activity (amount of scene changes within MPEG stream) on the allocated bandwidth and the network multiplexing gain is explored.

TO

my parents and my wife.

Acknowledgements

Most of all I'd like to thank my supervisor Dr. Mourad Kara for his invaluable advise and help throughout my research and the preparation of this thesis. This work would never have been finished without his patience and support.

This thesis is dedicated to my parents. This thesis is a result of their efforts towards providing me with a good education and a loving home. I thank them from the bottom of my heart. I am especially thankful to my wife, Khulood, for her help and understanding through difficult times. I am deeply thankful for her patience over the last years. I am also deeply grateful to all my family members for their love and support.

I would like to thank the member of ATM-MM group for all their support and friendship throughout my studies at Leeds University: Rik Wade and Suhaidi Hassan. Special thanks to Dr. Mohammed Rahin and Mr. Karim Djemame for their advice and support.

I had the opportunity to make a lot of friends. It has been a great pleasure to meet such wonderful people and to have spent an enjoyable time with them.

I am grateful to Saudi Arabia government for funding my studying in Leeds.

Finally, I would like to thank the Almighty Allah and praise Him for numerous blessings during my study.

Declarations

Some parts of the work presented in this thesis have been published in the following articles:

M. Kara and A. Mashat, "Statistical Analysis and Modelling of MPEG Sources for Workload Characterisation of Distributed Multimedia Application", *Proceedings of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, D. Kouvatsos editor, ISBN 0 9524027 3 4, pp. 16.1-16.10*, July 1997.

A. Mashat and K. Jambi, "Distributed Multimedia: A Research Point of View", *Proceedings of Multimedia Symposium: Tools, Applications and Trands,B.Hassan editor, pp. 106-116*, December 5-6, 1995.

A. Mashat and M. Kara, "Scene Change Scale: An Indicator of Multiplexing MPEG Streams for The Assessment of QoS Guarantees Over ATM Networks", *Proceedings of 6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, D. Kouvatsos editor, ISBN 0 9524027 8 5, pp. 101.1-101.10* , July 1998.

A. Mashat and M. Kara, "The Impact of Synchronising MPEG Streams on Bansdwidth Allocation", *Proceedings of the fourteenth UK Computer & Telecommunications Performance Engineering Workshop, UKPEW'98. pp 102-117, editors R. Pooley & N. Thomas, ISBN 09524027 2* , July 1998.

A. Mashat and M. Kara, "Performance Evaluation of Scene Based Model for VBR MPEG Traffic", *To appear in the Proceedings of Performance99, Istanbul, Turkey* , August 1999.

Contents

List of Figures

List of Tables

Glossary

AAL	ATM Adaptation Layer
ABR	Available Bit Rate
ACF	AutoCorrelation Function
AR	Autoregressive
ATM	Asynchronous Transfer Mode
CAC	Call Admission Control
CBR	Constant Bit Rate
CCITT	Consultative Committee on International Telegraphy and Telecommunications
CLR	Cell Loss Ratio
CoV	Coefficient of Variation
CTD	Cell Transfer Delay
DCT	Discrete Cosine Transform
DMC	Detailed Markov Chain
GOP	Group of Picture
GTES	Generalisation of TES
FIFO	First In First Out
IQR	Interquartile Range
ITU	International Telecommunications Union
ITU-T	Telecommunication Standard Sector of ITU
JPEG	Joint Picture Experts Group
LRD	Long Range Dependency

MC	Markov Chain
MG	Multiplexing Gain
MMPP	Markov Modulated Poisson Process
MPEG	Motion Picture Experts Group
PMR	Peak to Mean Ratio
QoS	Quality of Service
Q-Q	Quantile-Quantile
SCS	Scene Change Scale
SRD	Short Range Dependency
TES	Transform Expand Sample
UBR	Unspecified Bit Rate
VBR	Variable Bit Rate
VoD	Video on Demand

Chapter 1

Introduction

1

1.1 Research Motivation

Asynchronous Transfer Mode (ATM) is an emerging standard for broadband networks which allows a wide range of traffic types to be multiplexed in a single physical network. The traffic types can range from real-time video to best-effort data. One of the most important benefits of ATM technology is its ability to provide Quality of Service (QoS) guarantees for applications. A QoS guarantee can define the form of bounds for end to end delay and data loss rate. ATM networks provide several classes of services to satisfy the QoS needs of various applications. Each class provides different QoS guarantees, based on application requirements. However, these requirements are quite difficult to achieve, largely due to an inherent tradeoff [Ryu96]. For instance, the network can always provide desirable QoS by allocating abundant network resources (based on a peak allocation) at the expense of low network utilisation. But, this is not always desirable, especially in the case of a burstiness traffic source such as video traffic.

In order to overcome and resolve this problem, essential characteristics of a traffic source must be extracted. Hence, for an efficient traffic management in a high speed network, it is important to have a working knowledge of the basic characteristics of multimedia traffic. This information can be used either to study the network utilisation, or to develop appropriate control schemes for handling multimedia traffic [Venturin95]. In order to achieve that, a real traffic

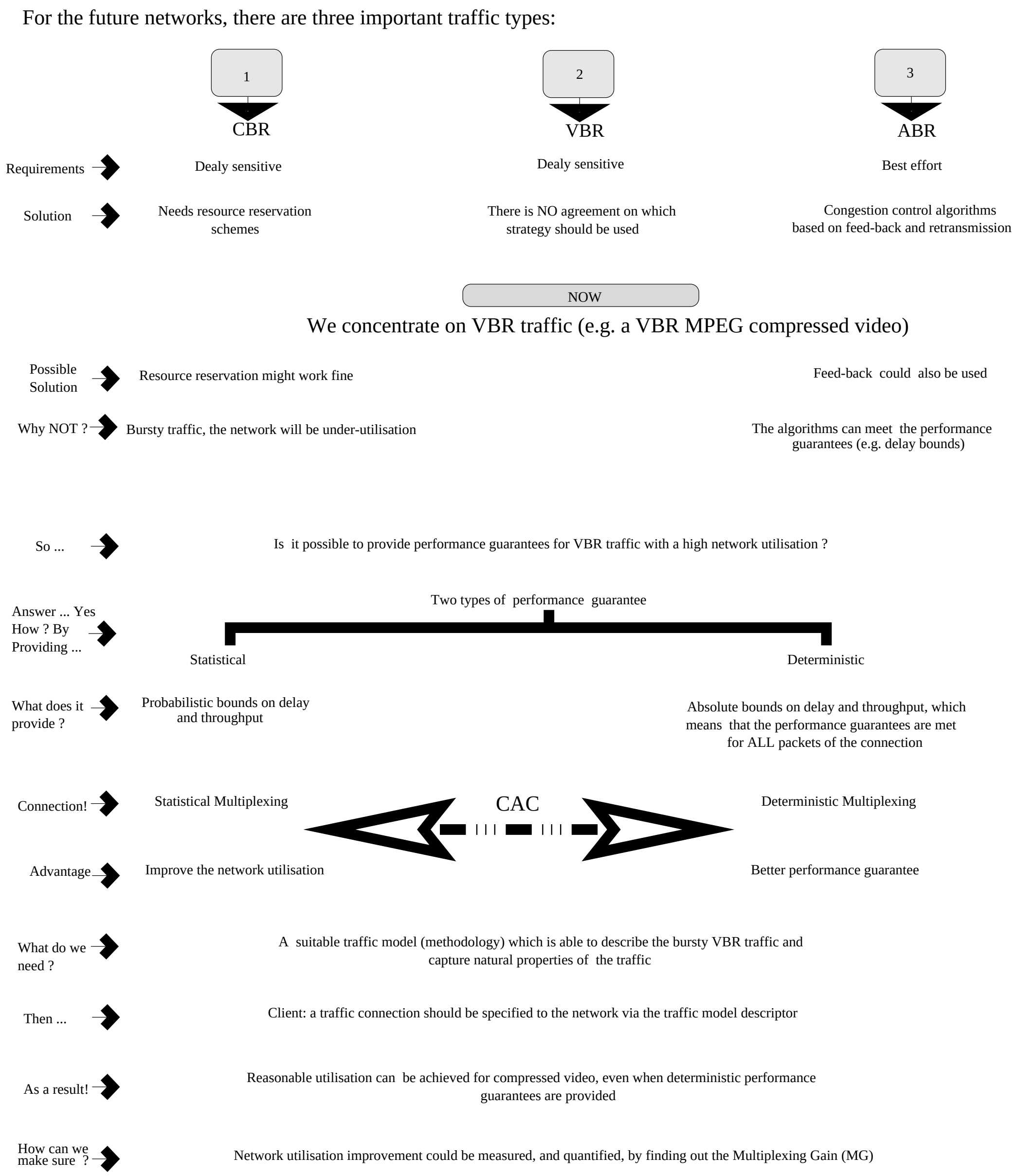


Figure 1.1: An Overview of the Problem

the QoS of VBR video traffic due to the compression technique which removes the redundancy from the video images. One reason for cell losses are buffer over-flows at a switching node [Onvural95]. The second requirement, the Cell Transfer Delay (CTD), can be observed in various ways, including coding, packetisation, propagation, transmission, switching, queueing and reassembly delay. The CTD requirement can be described as a set of delay constraints (or bounds). The delay constraints vary, according to video services. For instance, interactive services (such as video conferencing) require a short delay, while the delay bound is less important in the case of distribution services (such as the video on demand system).

In order to provide a guaranteed QoS for VBR video traffic, it should be managed efficiently and carefully. Managing VBR video traffic can be very difficult, due to the statistical properties[1] of the video stream, which are dependent on the coding scheme and the content of the video sequence [Kara97]. In addition, compressed video traffic (especially for multiple frame coding, e.g. MPEG) exhibits complex patterns which vary from one stream to another.

To evaluate the performance of ATM networks, and to provide a good guide for the design of various network control schemes, such as traffic management algorithms, it is important to have a good knowledge of the traffic source behaviour. In order to achieve this objective, direct observations of real traffic sources could be performed, or mathematical models could be constructed [Ni96]. However, the first option introduces several difficulties in formulating the real trace (which is obtained from a real traffic measurement) and applying a relevant analysis to it. On the other hand, a good mathematical model can be employed to characterise the real traffic source precisely, and to accordingly, produce efficient analysis techniques. If the model is able to predict the specific behaviour of a stochastic system accurately, then the model tends to be a good one. A model can possibly predict one behaviour accurately, while predicting another inaccurately. For instance, it is possible for a model to predict cell losses accurately, but can not predict the cell delay. Thus, one of the most important criteria for a model selection is based on the desired system metrics. Traffic models can be classified using various criteria: simplicity (number of parameters), modelling level (time scales) or according to the

[1] In this thesis, the terms “properties” and “characteristics” are used interchangeably.

stochastic process.

The source model[2] can not only be used for the traffic characterisation, but also to generate synthetic traffic, which is similar in behaviour to real traffic. There are many advantages to be achieved from generating a synthetic video traffic, especially in performance studies. Performance studies can not be carried out without providing actual video traces. Furthermore, a stochastic model encompasses many realisations (or sample paths) which represent 'structurally' similar, but not identical, streams [Krunz96]. Therefore, generated streams are ideal for statistical multiplexing studies.

Consequently, the challenge is to (1) introduce a traffic characterisation and modelling considering the important/essential characteristics of VBR MPEG traffic such as scene changes, (2) build a traffic model which captures the MPEG traffic behaviour in terms of statistical and queueing performance. Next section presents an overview of previous studies into the modelling of VBR video traffic.

1.2 An Overview of Video Source Models

Classical models, based on a *Poisson* arrival process, are not adequate for video traffic. The *Poisson* process assumes that the arrivals are independent, whereas for encoded video they are not [Izquierdo96]. Hence, new models to describe the encoded video traffic are needed. There are many traffic models which have been previously proposed for VBR video traffic, based on the traffic characterisation, e.g. see [Conti96] [Heyman92] [Rose95b] and [Venturin95]. Since the VBR of a video traffic depends largely on the compression scheme, most traffic models do not characterise the scene changes within the video stream. Moreover, some models were specifically developed for only low-activity video streams, and thus, were not appropriate to other streams with different levels of activity [Heymen92] [Li95] [Frost94].

One of the earlier models for VBR video traffic appeared in [Maglaris88]. A video source is described as a first order autoregressive process AR(1) with both a Gaussian and an exponential autocorrelation function (ACF). The main advantage of that model is simplicity. However, it is unable to capture accurately the statistical behaviour of the video sources. Another, more sophis-

[2] In this thesis, the terms "source model" and "traffic model" are used interchangeably.

ticated, model based on autoregressive moving average (ARMA) process was proposed in [Gruenefelder91]. The model showed an improvement in capturing the main statistical behaviour of the video source. However, a very short video sequence (a few seconds) has been used for the traffic characterisation.

Heyman et. al. observed that the number of ATM cells within an encoded video frame can be modelled along as Gamma distribution [Heyman92]. They also suggested that a multi-state Markov chain can be used for video traffic modelling in order to obtain more accurate results. However, their model was based only on video teleconferencing sequences, and is not, therefore, suited to more general video sources.

In [Skelly93], a histogram-based model was introduced to approximate the arrival rate of variable bit rate video traffic. The model was used to predict the queueing performance of a multiplexed stream at an ATM buffer such as the buffer occupancy distribution and cell loss rates. The video model is modelled as a Markov modulated Poisson process in order to investigate the approximation of the queueing performance. The model has been tested to approximate the bit rate of different video sequences (NTSC and MPEG) at frame-by-frame level. Experimental work has shown that the model approximates accurately the behaviour of video traffic in an ATM multiplexer. It has been also shown that the distribution and the presence of strong correlation are very important for capturing the queueing behaviour of video traffic. However, the sequences used in the simulation model were short (about 10 seconds), and hence do not give enough information of the long range correlation feature of video traffic.

Krunz et. al developed an MPEG model in which the number of cells per frame was determined by a Lognormal distribution [Krunz95]. This model was tested by conducting several simulation experiments. The simulation results showed that the model could not capture the multiplexing performance accurately especially at larger buffer sizes.

Transform Expand Sample (TES) processes are designed to match both the fitted distribution and the autocorrelation function of the original traffic. Melamed et. al. used TES processes to come up with a model for the number of bits per group of blocks for an H.261 video encoder [Melamed92]. They observed that there is a periodical component within the bit rate at the group of blocks level. They extracted these components, and then applied the TES processes on these extracted data sets. Their simulation results showed that the throughputs curve of the TES model performed slightly better than the AR model for system loads higher than 0.6 (with the same amount of QoS matrices). The model has been also used for a performance evaluation study of an integrated network [Melamed94] .

Most of the models mentioned previously did not handle video sequences with interpolated frames (such as 'B' frame in MPEG). Following the standardisation of the MPEG coding scheme, the characterisation of MPEG traffic began to be investigated. Reininger et. al. presented a composite model for MPEG traffic with three random processes associated with each frame type: I, P and B [Reininger94]. The model was based on the prediction of the B and P frames from the I frame within the same group of picture (frame). The model was mainly proposed in order to determine the performance of a multiplexer fed by a number of VBR MPEG sources in terms of multiplexing gains and cell losses. They have found that the model captured the deterministic periodical behaviour of MPEG sequences.

Generally, an MPEG video sequence contains *scenes, groups of pictures, frames and slices,* each of them corresponding to a different time scale. Most modelling studies have mainly focused on the statistical behaviour of MPEG traffic as a sequence of frames (at a frame level) without considering the impact of scene changes within the video stream; these are believed to introduce a large variation in the overall bit rate [Lazar93]. In addition, most of the proposed models were based on the characterisation of the three frame types of the MPEG stream (i.e. I, P and B). The challenge, therefore, is to introduce a model which captures the behaviour not only of one time scale, but several ones [Lazar93].

Due to the way MPEG is designed, MPEG video traffic introduces a great impact on the CLR when multiple video streams are multiplexing [Mashat98a]. The scene changes within the video stream are some of the most important issues which affect the statistical behaviour of MPEG traffic. This can be observed via the overall bit rate of an MPEG stream. Therefore, scene changes should be incorporated for the characterisation, and modelling, of MPEG traffic. Various models have been proposed for VBR MPEG traffic [Stamoulis94] [Heyman92] [Frost94], but only a few incorporate scene changes [Krunz96] [Lazar93] [Heyman96] [Rose95b]. In addition, some models are designed to adjust only, as a general rule, the main statistical parameters such as mean, variance and some initial lags of the autocorrelation function (short range dependency). However, these models tend to underestimate the losses ratio and delay bounds when they are used in the queueing performance. This is due to the fact that these models neglect the long range dependency which a video traffic exhibits, since the main reason of this feature is the different level of activities (scene changes).

Lazar et. al. developed a source model for an encoder sequence based on a Generalisation of TES process (GTES) [Lazar93]. The video sequence was modelled as a collection of subsequences. Each subsequence represented a scene within the video sequence. The model used two of the traffic layers: slice and frame. However, it considered only real time video sources and was applied to assess the impact of real time video sources on scheduling algorithms. In fact, the raw data which they used for the modelling was an encoded sequence using a coding algorithm similar, but not identical to JPEG. However, MPEG exposes different traffic characteristics. But these authors believed that the model could be applied at the slice level of the MPEG sequence.

In [Heyman96], a method was used to identify scene changes within several video sequences which have been coded using DCT (i.e. one frame coding). The work was an extension to their previous studies on modelling video conference traffic [Heyman92]. The video sequence was divided into several scenes. The number of frames within each scene was modelled using different distributions depending on the video sequence. Some sequences employed Weibull distribution, some used Gamma, while for others no simple model could be fitted. The video sequence was then modelled, primarily based on autoregessive models, in order to capture the autocorrelation features of the video sequence (using autocorrelation coefficient for intra-scene frames). The model used in Heyman's simulation studies was designed to predict the cell losses. Even though, the model was accurate for some sequences, but it overestimated the cell losses of others.

Rose studied the MPEG sequence as a group of scene classes [Rose95a]. A simple Markov chain model was built to model the amount of variation (in a group of frames) which could be tolerated for one scene. Each scene class was simply presented as a Markov chain state. Therefore, in this case, the number of states equaled the number of scene classes. However, this modelling approach leads inextricably to very large matrices.

In [Krunz96], the authors also used a method for identifying scene changes within an MPEG video sequence. However, the identification method employed in this study was based on analysing the changes only in consecutive 'I' frames, while the 'B' and 'P' were ignored. The size of 'I' frame was modelled by the sum of two random components: a related scene which reflected the overall level of scene activity, and an AR(2) component which accounted for the fluctuations of 'I' frames within a scene. The size of 'P' and 'B' frames were modelled using two random processes. The traffic model was tested to fit four empirical video sequences, and provided a good prediction of the queueing performance. However, the model was only tested on a limited number of video sequences, and did not cover a wide variety of video activities.

A new model for the simulation of MPEG video traffic was presented in [Reyes97]. The model was implemented on neural networks to adjust the autocorrelation and probability distribution functions of a given video traffic. The model uses a neural network to learn the conditioned histogram of the given traffic. By using neural networks, it is possible to benefit from their capacities for working in real time and interpolating unknown functions. These interpolations avoid the need of searching in transition matrices and reduce the amount of stored information. However, the model captures the autocorrelation function of the given video traffic for only first lags.

1.3 Thesis Components

This thesis is looking at the characterisation of MPEG traffic, focusing on different levels of activity. The major components of this thesis are as follows: in the field of traffic modelling, we define the statistical characteristics of VBR MPEG patterns in order to highlight the most important features of VBR MPEG traffic. Two models for VBR MPEG traffic are presented, describing the statistical behaviour of the traffic. We then extend our analysis of these two simple models in order to generate synthetic VBR MPEG traffic.

Following the importance of scene changes within an MPEG stream, an enhanced analysis on scene changes is conducted. In order to characterise VBR MPEG traffic based on the impact of scene changes, we present a methodology for classifying MPEG streams according to the amount of activities within each stream. We analyse variations in the bit rate of the MPEG stream, followed by an extensive scene change characterisation for MPEG video traffic. Two simple methods or techniques are offered in order to identify the scene changes within an MPEG video stream.

For the purpose of our analysis, we have employed real empirical data sets for various MPEG sequences to define and study the statistical characteristics. Based on the amount of scene changes, a *Scene Change Scale (SCS)* will be presented to exhibit the amount of activity within the MPEG stream. Furthermore, the impact of scene changes on QoS requirements will be explored. Our primary measure of interest is the CLR at an ATM multiplexer, because the amount of scene changes affects cell losses when multiple MPEG streams are multiplexing at an ATM multiplexer. Consequently, a novel scene-based model for MPEG video traffic is constructed, based on the scene change characterisation. The proposed model can be used to generate synthetic MPEG streams to be used in many performance studies, including buffer dimensioning and bandwidth allocation at video servers and network nodes.

The most important reason behind the use of VBR is the opportunity for increased multiplexing gains [Izquierdo96]. However, VBR multiplexing may cause data loss. This thesis deals with another issue regarding QoS guarantees by presenting a deterministic model for VBR MPEG traffic. By using this model, we analyse the multiplexing process of multiple MPEG streams. Next, we introduce an approach for stream alignment, before entering the network, to improve the network utilisation. Based on the classification results derived from the scene changes analysis, we will study the impact of the scene changes on the multiplexing gains for aligned streams, thereby showing the level of improvement in the network utilisation. Furthermore, based on the analysis of the stochastic behaviour of the statistical multiplexing for VBR traffic at the ATM switch, we present some simulation results showing the amount of multiplexing gained when the First In First Out (FIFO) discipline is implemented. We will be in a position to provide an answer to the question “Is it possible to improve the network utilisation with deterministic guarantees of QoS?”. We will also study the impact of the MPEG stream alignment on the network utilisation improvement in terms of calculating the allocated bandwidth for each multiplexed stream. Based on the scene change analysis, we will also study the impact of scene changes on the bandwidth allocation process.

1.4 Overview of the Thesis

In this thesis, three related problems, in the area of efficient management of VBR MPEG video traffic, namely characterisation, modelling and multiplexing of VBR MPEG traffic, are addressed. Most of the terminology used in this thesis is specific to ATM networks. The remainder of the thesis is organised as follows:

In chapter 2, a background study of the issues pertinent to the area of research will be presented. In addition, we provide a context for the problems we are exploring in this thesis. First, we cover several issues related to the ATM framework, including the main services provided by ATM. We then present several issues related to ATM networking, including the challenges of proving network support for QoS guarantees. A brief overview of various multimedia applications, including data types and their requirements are presented, followed by an overview of multimedia traffic modelling. Several traffic models, which have been proposed in the literature for the characterisation of multimedia traffic, are discussed. Some aspects dealing with the traffic management in ATM networks are addressed. Then, some important concepts of ATM switching and multiplexing processes will be given. We also present a tool for the performance evaluation of multimedia systems: an application level traffic generator.

Chapter 3 presents the statistical characterisation process of MPEG video traffic. We start with a general overview of the MPEG standard. We then present an analysis of various MPEG sequences, including a study of the statistical characteristics of VBR video traffic. We initially investigate some empirical data sets of various traced VBR MPEG streams. We then define the statistical characteristics of VBR MPEG patterns. In so doing, two main statistical features of an MPEG sequence will be explored: namely Distribution and Correlation. We also discuss the burstiness measurement for MPEG traffic including a presentation of various ways of analysing the burstiness within an MPEG sequence.

In Chapter 4, we present a methodology for classifying MPEG streams according to the amount of activity within each stream. We do so by analysing the variations in the bit rate of the MPEG stream. Then we present two methods in order to identify the scene changes within the MPEG stream. Based on the classification process of MPEG streams, we introduce a Scene

Change Scale (*SCS*) factor exhibiting the amount of activity within the MPEG stream. We perform several experiments on an ATM multiplexer to show the impact of scene changes on the performance of the multiplexer in relation to the results of *the SCS*.

We analyse the statistical modelling of VBR MPEG traffic in chapter 5. First, we present two simple and efficient models and try to describe the statistical behaviour of MPEG sequence. The traffic modelling approach is based on the Markov Chain methodology. Our analysis is subsequently extended to the design and implementation of a generator for synthetic traffic. In order to improve the statistical behaviour of the model, a novel scene-based model is proposed considering scene change characteristics.

In Chapter 6, we present a deterministic model for VBR MPEG traffic. Based on this model, we analyse the multiplexing processes of multiple MPEG streams. An approach for the stream alignment is introduced to achieve a better improvement on the network utilisation. Based on the stream classification results derived in Chapter 4, we study the impact of the scene changes on the multiplexing gain for aligned streams. We also perform several experiments on an ATM multiplexer to show the impact of multiple MPEG streams multiplexing.

In Chapter 7, we discuss and evaluate the results obtained from the studies in this thesis. A summary and a conclusion are provided in Chapter 8, including the future work.

Chapter 2

Multimedia Applications Over ATM Networks

In this chapter, we provide an overview of multimedia applications over ATM networks including their performance requirements. We begin with a short introduction to the ATM network, and proceed to discuss the basic concepts to ATM technology. We then provide a brief discussion of the various service classes in ATM networks for supporting various traffic types, presenting several issues regarding multimedia applications over ATM networks, including a brief description of their network requirements. For efficient management of multimedia traffic, a short introduction to various traffic models which are used to describe the behaviour of multimedia traffic will be provided. Finally, we introduce several issues relating to the traffic management algorithms and protocols, including the performance evaluation process using a realistic workload generation.

2.1 Overview of ATM Networks

ATM uses a multiplexing and switching technique, which in turn, uses a short and fixed size packet called a cell. It also called 'cell relay'. The ATM cell is 53 bytes long. Each cell contains two parts: header (5 bytes long) and payload (48 bytes long). The header contains information about the cell's route, such as virtual channel and path of the cell. The 48 bytes payload contains the used data, which is formatted in one of the adaptation layer formats. Each cell is tagged with a virtual channel identifier. Cells which belong to the same virtual channel will not important appear at periodic intervals. It is 'asynchronous' because different hosts connected to the same ATM line can transmit cells at different rates and place them into transfer media whenever they want. At an interval of time, the ATM line could include different cells belonging to different hosts. Therefore, several different media could be sent at the same time along the same ATM line without delay.

ATM is connection based. This means that, when an application wants to communicate with another application, it must request that a connection be made. The advantage of it being connection based is speed. Once the connection is built, no other routing is necessary.

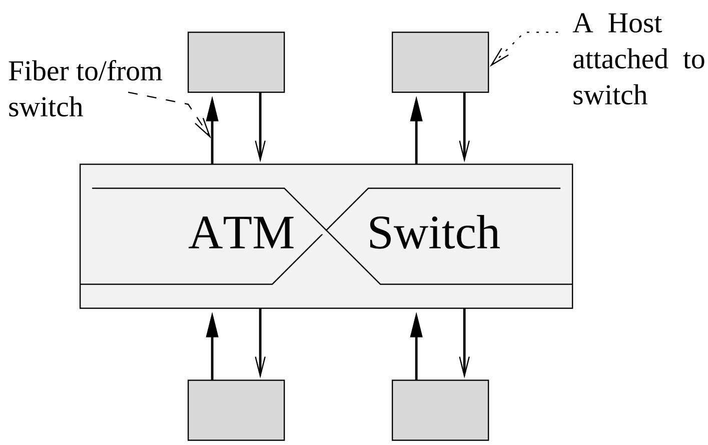


Figure 2.1: ATM Network Connection

To form an ATM network, more than one host is connected through the ATM line (fiber optic), using a special processor called an ATM switch. The ATM switch is designed to transfer data at extremely high speed. It is composed of a number of input and output ports, and the network which connects these two ports is called the *switch fabric*. Cells arrive at the input ports, pass through the fabric and exit through the output ports. A number of different fabric types exist, including simple crossbars and batcher banyan networks (section 2.9 presents an overview of ATM switch types). Figure 2.1 illustrates the connection between a computer (host) and an ATM switch. Because an ATM switch has finite capacity, multiple switches can be interconnected in order to form a larger ATM network.

The connection between the host and the ATM switch will be through User to Network Interface (UNI), where the connection between two ATM switches could be through either UNI or Network to Network Interface (NNI). Figure 2.2 illustrates the connections between multiple switches.

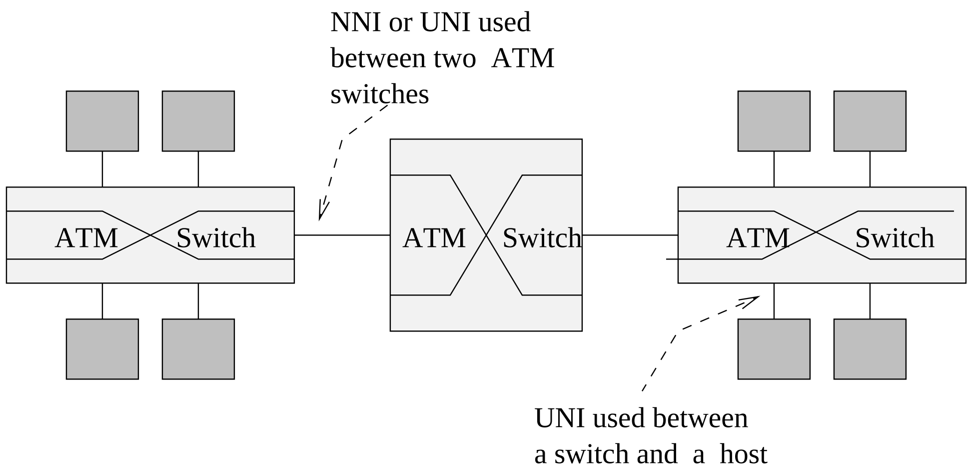


Figure 2.2: The Connections Between Multiple Switches.

An application request for a connection must specify the quality of service (QoS) that is required so that the connection which is made will meet the application's requirements. There are many common metrics which are used to define QoS, including *Cell Loss Ratio (CLR), Latency Traffic Type (Class), Maximum Burst Rate and Sustained Burst Rate.* This thesis will focus mainly on the *Cell Loss Ratio* (*CLR*). CLR can be defined as the number of cells which are dropped (due to buffer overflow in switches), before reaching their destination, divided by the total number of transmitted cells.

2.1.1 ATM Advantages

ATM networks have many advantages over existing networking technologies. The following lists only the more significant ones:

- ATM provides a high bandwidth through end-to-end switching topology. Furthermore, an ATM switch can support greater switching capacity than classical packet switches, and can support more applications simultaneously,
- ATM networks also support quality of service guarantees to applications. This can be achieved using various resource reservation algorithms and mechanisms,

- Various types of applications, with very diverse traffic characteristics, can be supported by using the switching and multiplexing methodologies of ATM. This feature makes ATM more flexible in terms of dividing the bandwidth into a number of virtual connections arbitrarily,

- Because ATM can support all types of application, a single networking platform can be built over ATM in order to handle all the different types of application. Therefore, an ATM network can be used as an integrated platform to manage and deliver not only one type of application, but multiple types; and

- The statistical multiplexing in ATM provides a significant gain when bursty applications are multiplexed. By doing so, the available bandwidth can be used efficiently in most cases.

2.1.2 ATM Adaptation Layer (AAL)

ATM adaptation layer (AAL) is a part of the ATM standard. AAL is directly above the ATM layer to provide end-to-end service. The role of AAL is to support different classes of application in the ATM network. It also has a number of important functions that satisfy diverse application requirements. In other words, it refines the QoS offered by the ATM layer, and offers some limited flow control of the data. It may, therefore, enhance the service provided by the ATM layer (a layer below the AAL layer). Thus, a computer (or user application) interacts with ATM through an ATM adaptation layer, and is, therefore, responsible for making the network behaviour transparent to the application. AAL has the ability to detect and correct errors, such as lost or corrupted cells [Comer95]. Information received by AAL from a higher layer is segmented or packetised into ATM cells. Cells received by AAL from the ATM layer are reassembled to form back the information.

There are different types of adaptation layer protocols, as different applications require different services and different QoS. Therefore, when a channel is established, adaptation layer protocol must be specified and agreed between the two hosts.

All AALs provide Segmentation and Reassemble (SAR) functions, which split up user data into cells and deliver them to the ATM layer, and reassemble them into user data at the receiving end to form back the user data. AAL can be described as two part or two sub-layer. The first sub-layer is SAR, while the second provides a management of data flow from, and into, SAR. Four types of AAL have been recommended by the Telecommunication standard sector of International Telecommunications Union (ITU-U), namely AAL1, AAL2, AAL3/4 and AAL5. The following provides a short description of these:

- **AAL1:** This type is for applications which require information transmitted with a constant bit rate, as well as, strict timing control (such as real time voice). AAL1 can also indicate lost or corrupt information, which is not recoverable by the AAL itself.

- **AAL2:** It offers information transmission with a variable bit rate. It also provides a timing information transmission between the source and the destination.

- **AAL3/4:** This type is a combination of AAL3 and AAL4. It provides a service for data applications in connection, or connectionless, modes. It has been recommended by ITU-T for the transfer of information with loss-sensitive property, but not for delay.

- **AAL5:** This type is a simplified version of AAL3/4. It has been specified by the ATM Forum to offer a service with less overhead and better error detection. The ATM Forum also specified AAL5 for signalling the UNI and NNI in the Broadband Integrated Service Digital Network (B-ISDN).

2.2 Classification of Services in ATM Networks

For future broadband networks, a large number of services must be provided to support a variety of applications with a wide range of QoS requirements. In general, these applications can be classified into real-time and non real-time. The real-time application requires a bound on delivery delays of each transmitting packet, while the non real-time application can usually be serviced in the 'best-effort' mode, where the available network capacity is divided among the applications sharing it. These applications are not usually sensitive to delay.

Many different services classes have been defined by ATM Forum to meet the various application requirements. For real-time services these are Constant

Bit Rate (CBR) or Variable Bit Rate (VBR), for non real-time applications: Available Bit Rate (ABR) and Unspecified Bit Rate (UBR).

CBR: With CBR service, a constant bandwidth is reserved for each connection throughout its duration. CBR service generates traffic at a constant rate, and can be described by its peak rate. The burstiness of a CBR traffic source is equal to *one*, as all periods are transmitted at the peak rate, making the network easy to manage. However, this is an inefficient use of the bandwidth resource (see Figure 2.3). Since the amount of traffic generated by most applications varies over time, it is possible to reserve less bandwidth in the network than the peak rate.

VBR: VBR service has been introduced to provide an efficient use of the bandwidth resource and to achieve high resource utilisation (see Figure 2.3). Most multimedia applications are assumed to be VBR sources. VBR traffic can be described as a bursty source. The peak-to-average (burstiness) of a VBR source is often much greater than one. This service can be classified into two classes: real-time VBR (rt-VBR) and no-real-time VBR (nrt-VBR). The rt-VBR service is almost identical to the CBR service, that it is for VBR instead of CBR applications. Real time streaming applications which send at variable bit rates can use this service. On other hand, the nrt-VBR service provides bandwidth guarantee at a peak rate, but it provides no guarantee in delay bounds.

UBR: This service was proposed to support non-real time applications which only need best-effort service. Therefore, it does not offer any service guarantees. Thus, it has minimum priority among all the other classes. The problem with the UBR is that there are no cell loss ratio guarantees for these applications, while many of the non-real time applications expect a packet loss rate similar to existing Local Area Networks (LANs). Thus, this is one of the key motivations behind the next service (ABR service).

ABR: This service is an improvement of the UBR service, reducing cell loss ratio and providing a more efficient use of the available network resources. ABR service is intended for best-effort applications requiring a guaranteed minimum rate, and uses a rate-based feedback approach to control congestion. This service attempts to dynamically share the available bandwidth among all ABR connections in a fair manner. The user connection therefore, may send at the peak rate when the network has a low load level. This will help to increase the network efficiency.

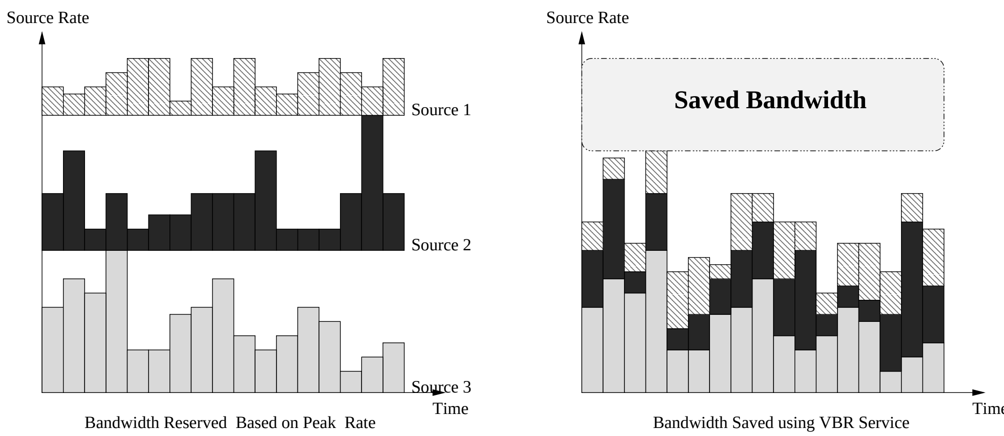


Figure 2.3: Bandwidth Usage, employing CBR and VBR services.

2.3 QoS Support in ATM Networks

Quality of service is the most often used in network terminology, it is to be used as a resource management tool within a computer network. The main goal of QoS is to provide an infrastructure facilitating negotiation between the client and the network for an acceptable connection within the capabilities of a given network system.

Support for QoS guarantees traffic transmission over ATM networks and is crucial to the success of many multimedia applications, including, for example, video on demand. One of the most important problems facing the network service provider is that of providing QoS guarantees to users, while maintaining the network utilisation as efficiently as possible. The network can be managed efficiently by performing a number of control techniques, including routing, bandwidth allocation, call admission control (CAC) and scheduling functions. Selecting appropriate paths for an incoming connection will improve the network utilisation.

CAC controls the acceptance of connections into the network. A good CAC algorithm should result in the maximum possible utilisation of resources.

Scheduling policies manage the priority of the serving connection's cells, based on traffic class. For instance, a connection with stringent QoS requirements will be served first over other connections, so that the QoS requirements can be met.

Some of the problems in providing QoS guarantees, along with high resource utilisation in an ATM network with multiple connections, are as follows [Rampal95]:

- Traffic burstiness: In the case of transmitting a bursty traffic, high efficiency can be obtained mainly by statistically sharing out the network resources. By doing so, the network utilisation is improved, but this can lead to difficulties in characterising the performance of each multiplexed connection;

- Statistical resource sharing techniques rely on accurate source traffic models. However, most of the multimedia sources are difficult to model. Therefore, resource allocation techniques within the network may result in performance levels different from those calculated using the source model at connection request stage; and

- For each incoming connection, the admission control function has to be done on the fly. The use of exhaustive numerical analysis is thus not feasible.

2.4 Multimedia Application

Multimedia can be considered as a communication tool that can be used to communicate almost anything. For instance, the first application of multimedia was electronic games. In fact, multimedia applications can be used by different sectors, such as education, training, industry, entertainment and business [Bunzel94]. By incorporating animation and sound to an application, the message, regardless of the content, has a far better chance of reaching its users. As an example in the business sector, multimedia can serve a variety of needs. It can be used in presentations, point-of-sale, video conference and training. A list of multimedia applications has been classified by ITU recommendations [ITU93]. These applications introduce great requirements for the communication network system. Table 2.1 presents some of the requirements for multimedia applications [Nahrstedt95]. Some of these requirements can not be met for several reasons. According to [Herrtwich91], the requirements of multimedia applications can not be met for three main reasons :

- System resources do not meet the needs of multimedia applications efficiently: In the past, high-quality audio and video could not be handled at all by standardised system resources. However, with the new technology, this has changed. Now, high performance resources meet the requirements of multimedia applications. This is caused by having fast processors (such as RISC), data compression schemes, optical disks and fiber networks. Resources are already handling both audio and video, but they need to be managed correctly to use them in a multimedia environment. For instance, video needs a high bandwidth since video frames must be displayed at a rate of 30 frames per second. Therefore, video frames need to be transmitted in a compressed form. This implies that, if the performance of resources is high, the use of the resources has to be regulated to meet the needs of multimedia applications. Without such a regulation, no guarantee on the behaviour of a given system can be provided.

- The system resources are not scheduled carefully: Careful scheduling in a multimedia system means the ability to multiplex a resource so that the throughput and delay requirements of the multimedia applications are met. The traditional scheduling techniques are not, in this sense, carefully scheduled. Therefore, a multimedia application which uses more than one resource at the same time needs a special scheduling technique to support the throughput and the delay requirements.

- There is no reservation of the resource capacity according to individual needs: The best scheduling method is useless if the system resources exceed its capacity. This means that resource reservation should be used as a mechanism to control resource access. Therefore, if an application makes a reservation, it will ensure that part of the resource capacity will always be available. Also, the reservation will be denied if the application does not need the reserved resources. By applying this mechanism, conflicting applications will be avoided, nor will they be disturbed by each other.

Application	Data Type	Bit Rate	Delay (End-To-End)	Loss
Telephone/CD audio	Audio	16-128 Kbps	0-150 ms	10^{-2}
MPEG,H.261 or NTSC and PAL TV	Video	1.86-20 Mbps	$\approx$ 250 ms	$10^{-2} - 10^{-11}$
File Transfer	Data	0.2-10 Mbps	$\approx$ 1 sec	10^{-11}

Table 2.1: Requirements of Multimedia Applications

However, reservation will cause some problems. First, reservation requires the periodic workload to be foreseeable. Second, it should be known how accurate the forecast can be. Third, it is not always possible to calculate or monitor the duration of a task execution. A solution to this problem is possible by reserving for the worst case, which is based on the assumed maximum workload that may never occur in practice. Moreover, reservation could be optimised to get a better resource utilisation, and avoid the rejection of a reservation as much as possible. Also, the resource manager should have the ability to recover any error, or conflict, caused by the reservation mechanism.

As a result, it should be taken into consideration that, in building a multimedia platform, performance alone will not be a solution. This implies that resource usage should be regulated by careful scheduling using, reservation mechanisms [Mashat95].

2.5 Standards Supporting Multimedia Applications

It is widely predicted that multimedia applications, such as video conference, will become widespread in the future. Without standards, products from different vendors will not be able to inter-operate, and therefore participate, in the same system. The following are some organisations which help to produce and adopt a standardisation for multimedia data and application to be provided to the vendors and system's developer.

- **ATM Forum:** The ATM Forum is a non-profit organisation, formed in 1991. The objective of ATM Forum is to accelerate the use of ATM products and services through a rapid convergence of inter-operability specifications. Another of its objectives is promoting industry co-operation and awareness.

- **IMTC:** The IMTC stands for the International Multimedia Teleconferencing Consortium, Inc. It is also a non-profit organisation. The fundamental goal of IMTC is to bring all organisations involved in the development of multimedia teleconferencing products and services together to help create and promote the adaptation of the required standards. The IMTC is trying to make consistent standards, such as the ITU T.120 and H.320 suites. In addition, IMTC helps to educate the end user to use open standards. This is done through public statements and publications.

- **ITU-T:** The ITU-T is the Telecommunication standard sector of ITU (International Telecommunications Union). ITU was formed in 1947, under United Nation (UN). It has two committees which have a direct impact on multimedia data transfer; the Consultative Committee on International Radio (CCIR) and the Consultative Committee on International Telegraphy and Telecommunications (CCITT). The second committee was renamed by ITU-T. This organisation is responsible for the creation of several multimedia standards, such as H.261 (for video conferencing). ITU provides some recommendations to help the cooperative work to adopt consistent and reliable standards. These recommendations can be employed by a developer to create a system which could be widespread in the future.

- **MPEG:** The Moving Picture Experts Group was established to develop a common format for coding and storing digital video and associated audio information. MPEG is a working group of ISO/IEC which is in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination (more details could be found in the next chapter).

2.6 Multimedia Traffic Types

In this section, we present the main classifications of multimedia traffic: Data, Audio and Video. Each traffic class could be characterised according to traffic behaviour and QoS requirements. Audio and video represent real-time traffic, while data represents non-real time traffic. Each traffic class differs from all the others in its characteristics. We will now briefly itemise briefly these classes, and their main characteristics.

2.6.1 Data Traffic

This type of traffic can be produced by computer-oriented services, such as file transfer and terminal emulation. This includes interactive data traffic (for example telnet) and bulk data traffic (for example FTP service). Traffic behaviour varies from one type to another. For instance, the pattern of interactive traffic is extremely unconnected, with silent intervals, while bulk traffic is a bursty traffic with large burst lengths [Liu92]. Data traffic is very sensitive to cell loss, therefore it demands a very small cell loss rate, but delay requirements are not strict (in most cases). However, for some real time applications (such as visualisation), data traffic could be sensitive to both data loss and delay.

2.6.2 Audio Traffic

This type of traffic includes any traffic that carries sound and voice information. It differs from data traffic in its QoS requirements. Delay is the most critical performance requirement (see Table 2.1), and the cell loss is also important with an acceptable cell loss probability 10^{-2}. However, the behaviour and properties of audio traffic depends on the adapted encoding scheme [Stamoulis94]. For example, audio traffic can be transferred over the network in various compressed modes such as Digital Speech Interpolation (DSI). This type of traffic requires a low bit rate compared to video sources. It also produces short burst lengths [Liu92].

2.6.3 Video Traffic

Video traffic can be divided into two main groups: image retrieval and real-time video. Each group is associated with certain QoS requirements. A typical image retrieval application requires the transmission of a succession of images at irregular intervals, while the bit rate of the video traffic varies. There are two main factors that have influence on the behaviour of video traffic: QoS requirements and the encoding schemes [Stamoulis94]. Video traffic is delay sensitive. The limit of real time services delay is 250 ms [Yousef97]. In addition, it can be cell loss sensitive, with an acceptable cell loss probability of 10^{-2} *to* 10^{-11}.

2.7 Multimedia Traffic Modelling

The traffic characterisation is used to develop a model which captures the main features of the traffic. There are many traffic models that have been proposed, starting from a basic model to more complex ones [Izquierdo96], [Conti96], [Heyman92], [Doulamis96], [Rose95b], [Habib92], [Daigle86] and [Frost94]. The more significant characteristics of most of the traffic models presented in the literature are *mean rate*, *peak rate* and *burst length*. However, in the case of encoded video traffic, there are more traffic characteristics that need to be explored, due to the way video traffic is encoded. Most of the traffic models are based on a stochastic process. The stochastic process can be classified into three main types:

- **Independent:** A traffic source whose autocorrelation function decays to zero at lag one (such as the Poisson process);
- **Short-Range dependent (SRD):** A traffic model tends to be described as short range dependent if its autocorrelation function decays to negative-exponentially fast; and
- **Long-Range dependent (LRD):** If the autocorrelation function does not decay exponentially, then the traffic can be described as long range dependent.

In most cases the validity or 'goodness' of a model is determined by comparing the simulation results using the empirical data as the source, and results from using the model [Izquierdo96]. In this thesis, we will follow this methodology to validate our models.

2.7.1 The Importance of Traffic Characterisation

ATM networks are expected to support different types of applications, utilising a wide range of characteristics. Unfortunately, there is no satisfactory agreement on the characteristics of the various types of multimedia applications in accurate manner. The degree of understanding traffic characteristics for different types of applications varies widely. For instance, the characteristics of audio sources have been studied for several decades and are reasonably well understood, while VBR video sources still remain largely incomprehensible, due to their unpredictable features.

For efficient traffic management in a high speed network, it is important to know the basic characteristics of multimedia traffic. This information can be used to study the network utilisation [Venturin95]. In addition, it can be used to develop appropriate control schemes for handling multimedia traffic. In order to achieve that, a traffic source model should be developed, based on measurements of the existing multimedia applications. [Rose95a] presents three main reasons why models for video traffic should be developed:

- The statistical properties of video traffic have a remarkable impact on the network performance. By extracting these properties, we will be able to decide which property is the cause of the performance problems;

- The computational complexity of simulations, especially long simulation runs, can be reduced using traffic models and standard analytical tools such as discrete time analysis; and

- A traffic model can be used to determine the traffic descriptors, which are necessary at the connection phase.

2.7.2 Traffic Sources

There are many traffic models that have been proposed [Izquierdo96], [Heyman92], [Doulamis96], [Rose95b], [Habib92], [Daigle86], [Frost94]. Some of these are more appropriate than others for a given type of traffic. Generally, the vast majority of multimedia traffic models are based on a stochastic process and most of these models use Markov Chain (MC) method because of its ease of use to characterise the alternating arrival process [Habib92]. The simplest and the most commonly used traffic model is the simple Poisson model which assumes that the arrival process is a Poisson process. However, for a high speed network, the traffic is more bursty than in a Poisson process [Liu92] [Paxson95]. Complex traffic models are useful only when their parameters can be estimated accurately.

In some models, it is possible to achieve different classes of traffic characterisation by varying the model parameters, even when the model is simple (for instance On/Off source model). An overview of these models can be found in [Frost94], [Rose95a] and [Izquierdo96]. Table 2.2 shows some possible source models for data, audio and video traffic. Some of these models will be presented briefly.

Traffic Type	Possible Traffic Source Model
Data	Bernoulli and Poisson Process
	Compound Poisson Process
	Train Model (Idle/Active)
	Persistent Model (with a maximum permitted rate)
Audio	ON/OFF
	Two Markovian States Model
Video	Autoregressive Model (AR)
	Detailed Markov Chain Model
	Markov Modulated Poisson Process
	Self Similar (Fractal Process)

Table 2.2: Some Traffic Models

2.7.2.1 Poisson Process

The Poisson process is the oldest type of traffic model, dating back to the advent of telephony. The model can be characterised by a random arrivals in an interval of time t while the interarrival time $\{A_i\}$ is a negative exponentially distributed with a rate of λ; $P\{A_i \leq t\} = 1 - exp(-\lambda t)$ (see Figure 2.4). The arrivals are independent from each other, meaning the past has no effect on the future arrivals (memory-less).

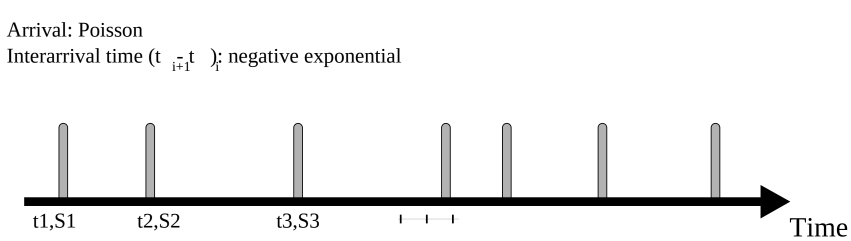


Figure 2.4: Poisson Arrivals Model

2.7.2.2 Interrupted Poisson Process

The Interrupted Poisson Process model (IPP) is a commonly used traffic model which can be used to model bursty traffic. This model is characterised by two Markovian states: State 1 (Active) and State 2 (Idle), each being associated with a bit rate λ and 0 respectively and the Sojourn time (i.e. holding time) for each state is exponentially distributed (see Figure 2.5). These states alternate continuously. During the Active time, the interarrival times are exponentially distributed (Cells arrive in a Poisson manner). This model can be described by the parameters t_{state1}, $B_{state1}\ and\ \lambda$ as follows:

- The average time (duration) for the State 1 (Active) period (t_{state1}): $t_{state1} = \frac{1}{\alpha}$;
- The average time (duration) for the State 2 (Idel) period (B_{state2}): $B_{state2} = \frac{1}{\beta}$; and
- Generation rate for Active state (λ).

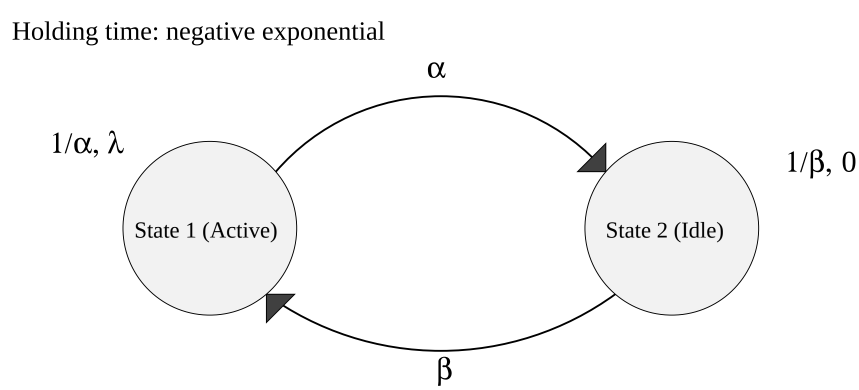


Figure 2.5: Interrupted Poisson Process

2.7.2.3 On/Off

This model is widely used to model many B-ISDN services (such as VBR). In addition, it can be used to model various classes of traffic with various degrees of burstiness [Stamoulis94]. It can be described as a two-state Markovian-based model, which alternates continuously between *Active* and *Idle*. Each state presents a period of time, namely *On* and *Off*. The traffic will be generated during the *On* periods, while no traffic is generated during the *Off* periods (see Figure 2.6). In most cases, each period takes either exponential or geometric random variables (depending upon the choice of the time axis as either being continuous or discrete), with a mean $\frac{1}{\alpha}$ for the *On* period and $\frac{1}{\beta}$ for the *Off* period. During the On period, the interarrival time (T) is constant. The two periods are independent of each other. Thus, the On/Off model cannot be used to model the overall correlation of the traffic. However, it has been stated that *on/off* model can be used to reflect some of the characteristics of a superposed VBR video traffic [Helvik95]. In other hand, this model is unable to reflect the periodicities feature of video traffic due to coding output pattern.

On/off model can be described by the parameters P, t_{On} , *and* m as follows:

- Peak arrival rate (P): $P = \frac{1}{T}$;
- The average of the *On* period (t_{On}): $t_{On} = \alpha^{-1}$; and
- The fraction of time in which the system is in the *On* period (m): $m = \frac{\alpha^{-1}}{\alpha^{-1}+\beta^{-1}}$.

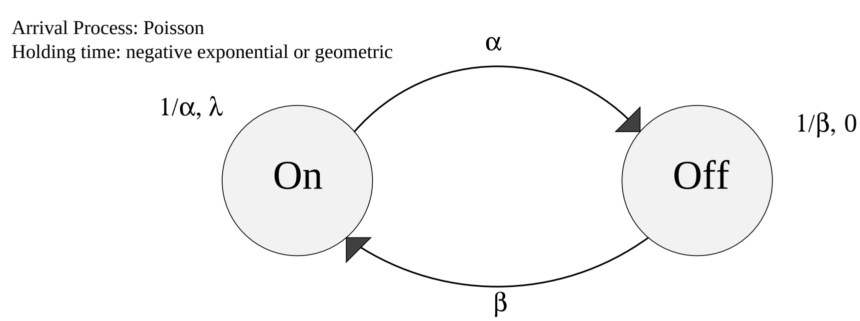


Figure 2.6: On - Off Model

2.7.2.4 MMPP

Markov Modulated Poisson Process (MMPP) is one of the most powerful models and has been used in many modelling research studies such as in [Frost94] and [Izquierdo96]. MMPP is essentially an m-state continuous-time Markov chain (see Figure 2.7). Each state i is associated with a bit rate λ_i and a mean holding time $1/r_i$. The arrival process in each state is Poisson, and the holding times are exponentially distributed. In fact the IPP model is a special case of the two-state MMPP model wherein State 1 represents *Active* with bit rate of λ and State 2 represents *Idle* with bit rate of 0. The two-state MMPP model has been used to approximate the superposition of packetised voice sources, together with data traffic arrival [Heyman92]. However, MMPP can only be used to model a short-term correlation [Liu92].

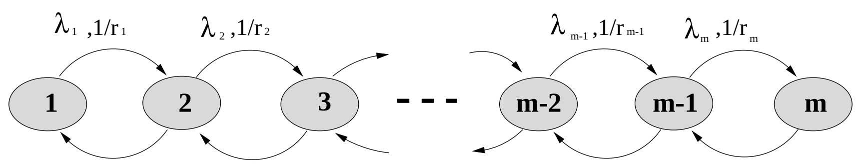


Figure 2.7: Markov Modulated Poisson Process

2.7.2.5 Deterministic

This is a very simple model based on generating traffic with a constant inter-arrival period. Therefore, it has the ability to describe the constant bit rate (CBR) service of the ATM network. This model can show better performance reflections on the lower level of an ATM source [Helvik95].

2.7.2.6 Fluid Flow

The traffic source is viewed as a stream of fluid which is characterised by a flow rate. This model always sends traffic with the maximum permitted rate. This source model imposes heavy constraints on the network, therefore it is appropriate for testing the fairness and the throughput of the traffic service [Liu92]. The model eliminates statistical delays, which could be caused by random traffic generators. Thus, it would be possible to achieve deterministic and reproducible simulation results.

2.7.3 Selection Criteria

In order to select the appropriate source model for an ATM network, a set of selection criteria could be considered [Stamoulis94], and include:

- **Actual Source Approximation:** A model to be selected should approximate to the actual source. It is very important for the source model to capture the main statistical characteristics of the actual source which would influence the performance of an ATM network when fed by such a source.

- **Simplicity:** The model should be kept as simple as much as possible. This could be achieved by using only a few parameters to describe the source model. In addition, if the model employs Markov Chain method, the number of states should be kept small.

- **General Source:** It is preferred to construct a general source by which one will be able to cover and fit a wide range of traffic types. For instance, by varying the parameter values of the model, it is possible to capture a different type of traffic. However, if one requires a general source, and a simple model at the same time, then conflict will arise. Thus, a tradeoff should be achieved between the simplicity and the generalisation property.

- **Ease of Implementation:** The source model should be easy to implement in the case of simulation experiments. For instance, the model may be incorporated into a simulation-based experiment. Thus, it has to be implemented easily as a traffic source generator.

- **Accuracy:** The model should give accurate results when performance measures are considered (such as delay and cell loss probability), corresponding to the actual source .

2.8 Traffic Management in ATM Networks

Traffic management mechanisms are important processes in an ATM network to protect the network against traffic congestion. In general, traffic management can be considered at two different levels namely: connection level and cell level. At the first level, topmost level, a process is performed in order to give a decision to allow a new connection to be admitted to the network or not. If the connection is allowed to enter the network then a bandwidth is allocated to the connection based on the QoS requirements of the connection (it can be described as a contract between the user and the network). At cell level, the policing mechanisms is performed to ensure that the contract between the user and the network is not violated (such as leaky bucket algorithm).

Several proposals for resource management algorithms are currently emerging, or under development, to provide guaranteed performance communication (the guarantee being either statistical or deterministic) [Gallasi90] [Belhaj97] [Pancha93]. For instance, bandwidth allocation is one task of these algorithms which determines the amount of bandwidth required by the connection to provide the required QoS. In the following sections, we present a brief definition for some traffic management algorithms.

2.8.1 Bandwidth Allocation

In ATM, bandwidth allocation deals with the amount of bandwidth required by a connection for the network to provide the required QoS. Mainly, there are two approaches for bandwidth allocation, namely deterministic multiplexing and statistical multiplexing. Peak bandwidth is allocated for deterministic multiplexing. This approach can cause large amounts of bandwidth to be wasted for a bursty connection. On the other hand, with multiplexing gains can be achieved by employing the statistical multiplexing approach.

The allocated bandwidth for multiplexed connections is less than the sum of their peak rates. Therefore, the statistical multiplexing allows more connections to be multiplexed in the network than deterministic multiplexing, thereby allowing better network utilisation. However, QoS guarantees are satisfied in the case of deterministic multiplexing, while they can be only statistically guaranteed in the case of statistical multiplexing.

2.8.2 Call Admission Control (CAC)

One way of preventing the network congestion is to perform a process called Call Admission Control (CAC). A new user connection with QoS requirements should go through the CAC procedure to decide whether to accept the new connection or not (i.e. reject it). If the new connection is accepted, a bandwidth will be allocated for this connection. Thus, CAC determines the amount of bandwidth required by a connection for the network, providing the required QoS. In fact, bandwidth allocation works as apart of the call admission control algorithm. Because the allocated bandwidth should be done on the fly, the algorithm should be kept simple to meet the real time requirement.

Some of the CAC algorithms require a specific traffic model and some require only the traffic parameters such as peak rate and average rate. There are many ways to classify the CAC schemes (algorithms). For instance, [Perros96] classified the CAC schemes according to the principle that was used to develop the schemes. Determining and calculating the required bandwidth for each connection is one of the most important aspects to quantify any CAC algorithm.

2.8.3 Effective Bandwidth

Statistical multiplexing improves network utilisation by allowing bursty sources to share bandwidth on demand and allocate a bandwidth for each source (connection). The allocated bandwidth should be less than the source peak rate. In order to take advantages of statistical multiplexing, the network should be able to approximate and determine the minimum required bandwidth for each source as a function of QoS, buffer size (at the multiplexer) and the traffic parameters. This bandwidth is commonly known as the effective bandwidth.

Therefore, the effective bandwidth allocation (equivalent capacity) can be defined as the service rate with corresponding QoS requirements (such as cell loss probability and minimum delay). The allocated effective bandwidth is computed to be close to the long range average (*mean*) rate and far from the *peak* rate.

2.9 ATM Switching and Multiplexing

In an ATM network, cells have to be merged from different sources and routed to different destinations via switch paths. In this way, the cells will share the transmission links for part of their journey. In fact, an input to an ATM switch within a switching element could be an output from another multiplexer. The process of multiplexing and switching cells involves temporary storage of cells in a finite sized buffer and the arrival cells form a queue in order to be served. Therefore, the main tasks for the switch and the multiplexer are to provide a temporary storage for the arrival cells, and then route them to the correct outgoing port while maintaining their QoS requirements. Another advantage of the multiplexing process is to enable a large number of sources to share network resources, such as the buffer and link capacities.

An ATM switch can be either blocking or nonblocking switch. To understand nonblocking switching, one needs to understand the role of blocking in the switch. Block occurs in a switch when a cell can not immediately access an idle outgoing port to which it would have access. This problem may occur because a buffer in the succeeding stage is full or because a cell at the head of the port of a queue can not be immediately switched. However, in nonblocking switch, cell can always immediately access a desired idle outgoing port.

There are many ways to arrange a switch to provide temporary storage [Perros96], depending on where buffers can be placed in the switch: Input buffering, Output buffering and Cross-point matrix. In this thesis, we assume that a non-blocking ATM switch is a multiplexer with an output buffering (see Figure 2.8) whereby multiple sources have been multiplexed into a buffer with one outgoing port (link). Then, the call admission control should be applied through each outgoing port of the switch.

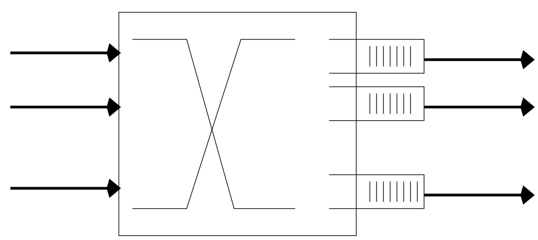


Figure 2.8: Outgoing Non-Blocking ATM Switch

Each outgoing port, and its buffer, could be represented as a queueing process. This type of process is known as an ATM Multiplexer [Perros96]. We usually consider a model where a number of sources emit their traffic streams directly into the multiplexer which has one output port. This is an idealisation, because in reality most source streams are multiplexed into a smaller number of trunks when they enter a switch. It is obvious that this makes no significant difference to the results [Roberts91].

Any queueing process could be described as arrival customers, service time by the server, number of service channels and the buffer capacity. The arrival customers can be specified as an input to the buffer with an average number of arrivals per unit of time, or they could just as easily be described by the average time between the arrivals. The arrival average time could be either constant (deterministic) or variable (stochastic). The service time can also be described by a service rate (or link speed).

Through this thesis, we will study a case in which the arrival customers are VBR MPEG streams and there is only one service channel to serve the customers, while the system capacity is the waiting space (see Figure 2.9). The waiting space could be finite or infinite, but in a real system the capacity must be finite. If the system capacity has been exceeded, then any incoming arrivals will be lost. Furthermore, larger buffer sizes will increase the waiting time for arrivals to be served. Therefore, QoS guarantees should be satisfied before establishing any new connection.

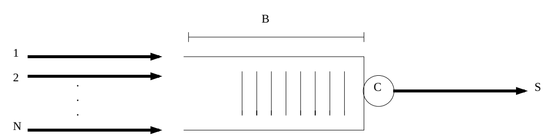


Figure 2.9: Buffering at Outgoing Port

2.10 Performance Evaluation of Multimedia Systems

Several multimedia applications are emerging, due largely to recent advances in fiber optics and hardware technology. An important issue in the successful development and delivery of these future applications is a system platform supporting the applications QoS requirements. Several proposals for resource scheduling algorithms and protocols are currently emerging, or under development to provide guaranteed performance communication support for these platforms (the guarantee being statistical or deterministic).

An important factor in the performance evaluation process of these resource scheduling algorithms and protocols is the input workload selected. The workload is made up of traffic flows generated by sources of different media types (audio, video, text, etc.) with different characteristics determining their behaviour and QoS requirements. Choosing the right workload for the performance evaluation of algorithms and protocols is crucial as different workloads will lead to selecting different algorithms and protocols (as well as their internal parameters). Most performance experiments tend to select traffic sources without considering profiles generated by applications, hence obtaining results from system-level sources [Kara95]. While the best algorithms and protocols would have been selected for the particular workload selected, it remains to be seen how these will perform under the workload generated by real applications.

In this thesis, we propose source models reflecting the behaviour of a real MPEG video traffic. We also provide steps required to integrate the MPEG source onto the workload model in order to generate a synthetic traffic emulating a realistic MPEG workload.

2.10.1 Workload Generation

Any study of high speed networks requires a workload to test the performance of designs based on a particular traffic model [Liu92]. There are many traffic models which approximate the traffic characteristics are used in these performance studies [Schuler96] [Celandroni97]. Thus, a workload can be presented as generated traffic with specific characteristics. Generally, two different approaches to generate traffic for measurement and simulation may be identified as:

- **Replay of saved traffic:** This is a storage based generation, where a pre-recorded or predefined traffic sequence is reproduced during the measurement. This approach is suitable for initial functional testing because of its determinism and simplicity. However, this approach is limited due to the availability of reasonable memory. Therefore, this approach is unsuitable for validation.

- **Stochastic based:** This approach can be regarded as a *'Black Box'* approach. The traffic is generated according to the class of the stochastic process (e.g. renewal) or source (e.g. on-off). The traffic is generated based on the process or source parameters. By using this approach, it is possible to generate a long traffic stream with various behaviours. However, the quality of the generated traffic depends on the how well the selected traffic parameters are.

A number of models have been proposed as approximations of individual and/or aggregate traffic sources in a a high speed network. However, it is argued that complex traffic models are useful only when their parameters can be estimated accurately. In this thesis, we will focus to generate a synthetic VBR video traffic based on the approximation of MPEG traffic behaviour.

2.10.2 An Application Level Traffic Generator

A more appropriate design for a realistic workload model is one that emulates multimedia applications. An example of such a tool has been under development at the University of Leeds, called an application level traffic generator [Kara95]. This traffic generator distinguishes itself from other traffic generators because it aims to capture the behaviour of these applications in its traffic generation. Two key design issues addressed in the traffic generator are calibration and validation. The calibration process determines the profile and patterns of the different media types required by an application. Validation is the process of comparing the traffic generated by scenarios from the traffic generator with those of real applications. An important part of this application-level traffic generator is the availability of multimedia traffic sources. The traffic generator should provide a wide variety of multimedia applications, including video, audio and data traffic. The architecture of the workload model in a greater details can be found in [Kara95] and [Kara97]. However, the design objectives of the traffic characterisation and generation architecture are:

- **Application-oriented traffic generation:** The aim of this architecture is to provide a realistic set of scenarios to the algorithm and protocol designer from which they can test their new designs.

- **Calibration:** The traffic pattern generated by applications needs to be characterised as a set of services with temporal relationships. This objective is to ensure that the architecture features a calibration process that is in place to extract the essential features of an application.

- **Validation:** This is the process of verifying that the traffic generated by the workload software reflects realistically what the real application is emulating. A level of confidence as well as an interval of confidence are usually provided with the validation process. The level of confidence, placed upon this tool as a reliable instrument to drive experiments rests in the validation objective.

- **To provide an interface for algorithms and protocols:** This is to ensure that the interface between the architecture and the algorithms developers is (1) independent of any specific transport provider, (2) conforms to the OSI RM, (3) widely implemented across different platforms.

The traffic generator operates by specifying :

- The configuration of the network (i.e. the participating nodes),
- A profile of the scenarios, and
- A set of libraries of the algorithms and protocols to be tested.

A profile file contains the specification of the experiments, and acts as a binding for the scenario, configuration, and algorithms and protocols required, as well as experimental settings for repetitive experiments.

Chapter 3

Characterisation of VBR MPEG Video Traffic

3.1 Introduction

Variable rate video traffic requires a careful treatment by the network. For instance, a sufficient bandwidth (with a little wasted bandwidth) should be provided, allowing a minor error for video traffic transmission. In the case of MPEG traffic, cell losses are crucial because most of the original video redundancy has been removed by the MPEG data compression process. Therefore, as stated previously, the knowledge of traffic characterisation is an important issue in ATM networks especially for efficient traffic controls. Some measures are necessary for characterising the burstiness of an encoded video source. For example, if the number of multiplexed sources are large and the traffic intensity is low, then the performance of ATM networks depends largely on the distribution information (such as average, deviation-to-average ratio and peak-to-mean ratio). On the other hand, if the traffic intensity is high, other information must be considered (such as autocorrelation and coefficient of variation) [Nomura89].

In the case of a compressed video traffic, three main measures can be introduced, namely: Distribution, Autocorrelation and Coefficient of variation. There are some other measures which characterise the overall variation (or unsteadiness) of the encoded traffic, such as the duration of the peak and the scene change durations. These measures provide a greater number of characteristics of the traffic over a period of time. This chapter deals with the main statistical information of VBR MPEG traffic including, distribution and correlation information, while the next chapter will examine the other traffic measures.

3.2 MPEG Overview

Nowadays, video has become an increasingly important component of multimedia communications because of increasing user demand for video and rapid advances in coding algorithms. The focus of this thesis is on a particular coding algorithm which has recently received a great deal of attention, namely the Motion Picture Experts Group (MPEG) standard. In 1988 MPEG was founded under ISO/SC2, with a charter to standardise video coding algorithms aimed for digital storage media having bit rates at up to about 1.5 Mbits/s.

MPEG is an example of variable bit rate video traffic. Generally speaking, video sequences contain a significant amount of statistical and subjective redundancy within, and between, frames. The ultimate goal of the video source coding is the bit-rate reduction for the storage and transmission. This is done by exploring both statistical and subjective redundancies, and to encode a 'minimum set' of information using entropy coding techniques. This usually results in a compression of the coded video data when compared to the original source data. The performance of video compression techniques depends on the amount of redundancy contained in the image data, as well as, on the actual compression techniques used for coding. With practical coding schemes, a trade-off between the coding performance (high compression with sufficient quality) and the implementation complexity is targeted [Sikora98].

The MPEG digital video coding techniques are statistical in nature. Usually, video sequences contain statistical redundancies in both temporal and spatial directions. MPEG compression techniques rely upon a basic statistical property, namely inter-pel (or inter-pixel) correlation, including the simple correlated translatory motion between consecutive frames. Thus, it is assumed that the magnitude of a particular picture pel can be predicted from nearby pels within the same frame (using Intra-frame coding techniques) or from pels of a nearby frame (using Inter-frame techniques). Consequently, during scene changes of a video sequence, it is clear that the temporal correlation between pels in nearby frames is small (or even not exist), and the video scene assembles accordingly a collection of uncorrelated still pictures. In this case Intra-frame coding techniques are appropriate to explore spatial correlation in order to achieve efficient data compression. The MPEG compression algorithms employ Discrete Cosine Transform (DCT) coding techniques on image blocks of 8×8 pels to efficiently match spatial correlations between nearby pels within the same picture. However, if the correlation between pels in nearby frames is high, e.g. in cases where two consecutive frames have similar or identical content, it is desirable to use Inter-frame DPCM coding techniques employing temporal prediction. In order to achieve high data compression (hybrid DPCM/DCT coding of video), a combination of both temporal motion compensated prediction followed by a transform coding of the remaining spatial information is used . The basic units that the MPEG algorithm uses as follows (see Figure 3.1):

- **Block:** A block is the smallest coding unit in the MPEG algorithms. It is made up of 8×8 pels and it is the basic unit in the intraframe DCT coded frames.
- **Macroblock:** A macroblock consists of a 16×16 pel segment.
- **Slice:** It is a horizontal strip within a frame, and is the main processing unit in MPEG. The coding of a slice is done independently from its adjacent slices.
- **Picture:** A picture is a single frame in a video sequence.
- **Group of Pictures:** The Group Of Pictures (GOP) is a small sequence of a deterministic pattern of pictures.
- **Sequence:** A sequence contains a series of pictures (or GOPs).

3.2.1 MPEG Coding

The MPEG coding algorithm was developed initially to store a compressed video on a digital-storage media [Pancha93]. MPEG is a flexible coding scheme which makes this type of coding widely available, and the most frequently used standard for video encoding [Bunzel94]. A variety of video applications (including video conferencing) use the MPEG coding scheme for reducing the

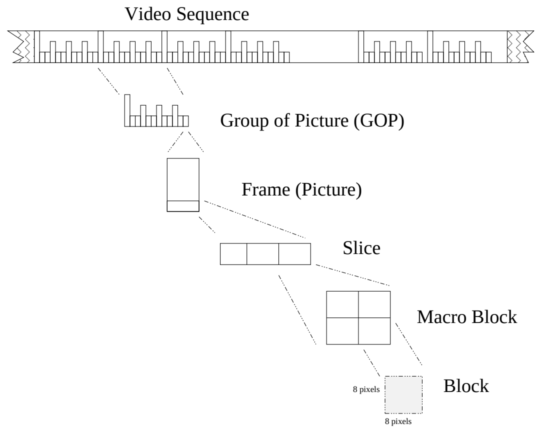


Figure 3.1: The Basic Units in MPEG

required bandwidth [Rose95a]. There are two main types of MPEG coding schemes for video: MPEG-I and MPEG-II.

- **MPEG-I:** The first Draft International Standard (DIS) released by the committee, ISO 11172 (MPEG-I), was drafted in 1991, and finally issued in 1992. MPEG-I was intended to be generic (although the initial target applications were constrained to digital storage media). The standard is independent of a particular application and therefore is mainly described as a toolbox. The user may decide which tools to select to suit the particular applications envisaged. This implies that only the coding syntax is defined, and that the decoding scheme is standardised. MPEG-I defines a hybrid DCT/DPCM coding scheme, with motion compensation similar to the H.261 coding standards. Further refinements in prediction and subsequent processing were introduced to provide a level of functionality required for random access in digital storage media.

- **MPEG-II:** Studies on MPEG-II started in 1990 with an initial aim of issuing a standard for the coding of TV-pictures with CCIR Rec. 601 resolution at data rates below 10 Mbit/s. In 1992 the scope of MPEG-II was enlarged to suit the coding of High Definition Television (HDTV). The DIS for MPEG-II video was issued in early 1994. The video coding scheme used in MPEG-II is again generic and similar to that of MPEG-I, however, with further refinements and special consideration of interlaced sources. Furthermore, many functionalities, such as 'scalability' were introduced. In order to keep the implementation complexity low for products not requiring the full video input formats supported by the standard (e.g. SIF to HDTV resolutions), so called 'Profiles', describing functionalities, and 'Levels', describing resolutions, were introduced to provide separate MPEG-II conformance levels.

Coder	Rate	I-frame	P-frame	B-frame	Average
MPEG-I	1.15 Mbit/sec	150,000	50,000	20,000	38,000
MPEG-II	4.00 Mbit/sec	400,000	200,000	80,000	130,000

Table 3.1: Examples of Typical Frame Sizes (in bits) for MPEG-I and MPEG-II

Table 3.1 shows the typical frame sizes for both MPEG-I and MPEG-II [MPEG99]. It is important to mention that MPEG-II was built on the powerful video compression capabilities of MPEG-I standard. Therefore, MPEG-I and MPEG-II specifications are similar [Gringeri98]. Since both MPEG types use the same compression concept and our analysis is applicable to both types [Krunz97], therefore, we will use MPEG-I throughout the thesis. However, the most important differences between MPEG-I and MPEG-II can be addressed in the following points:

- MPEG-I is meant for progressive sequences, whereas MPEG-II is optimised for interlaced pictures, so that it can represent a progressive sequence;
- MPEG-II supports a higher bit rate than MPEG-I; and
- MPEG-II has more profiles and layers depending on the targeted application.

Two further types of MPEG are currently under development: MPEG-4, a standard for multimedia applications, and MPEG-7, a content representation standard for information search [ISO/IEC97]. A project for developing an MPEG-III was originally exist, intended for HDTV applications, but the project was cancelled when HDTV was added to the MPEG-II standard.

The basic scheme of MPEG coding is to predict motion from frame to frame in a temporal direction, and then to use DCTs to organise the redundancy in the spatial directions. Thus, MPEG coding is a combination of interframe and intraframe coding techniques. Considering the output of an MPEG-I encoder, the reduction can be achieved by producing three types of frames: I, P and B (see Figure 3.2):

- **I Frame (Intra frame):** I frames are simply frames coded as a still image. The coding of this type of frame does not need any reference to another frame. Temporal redundancy is not taken into account. An 'I' frame is always an access point in the video sequence.

- **P Frame (Predictive frame):** P frames are predicted from the most recently reconstructed I or P frame. This frame is coded using a motion compensated prediction mechanism which exploits both spatial and temporal redundancies.

- **B Frame (Bidirectional predictive):** B frames are predicted from the closest two I or P frames, one in the past and one in the future. Coding B frame achieves the highest possible compression ratios.

As a result, MPEG-I can be distinguished from other encoding schemes by bi-directional temporal prediction [Conti96]. Each of these frames uses a different coding algorithm. An MPEG encoder repeats these frames periodically. Each frame contains a two dimensional array of picture elements called pels. The output of the encoded stream (the sequence of decoded frames) contains a deterministic periodic sequence of frames such as [IBBPBBPBBPBB] which is called Group Of Pictures (GOP). The selection of the encoding sequence is a tradeoff between latency, compression and error propagation. The B and P frames are preferred to I frame in terms of reducing the overall data rate for compressed video stream. However, I frame is necessary because it can be used to terminate the propagation of error. This is due to the ability of decoding the I frame without a reference to any other frames. Thus, it is typical to limit the maximum length of the GOP.

Generally, an MPEG video stream can be classified into three main layers: scene layer (containing similar images), GOP layer (containing a deterministic periodic sequence of frames) and frame layer (with different types of frames). The duration of these layers varies from several seconds to tens of milliseconds. MPEG traffic can be characterised using different levels: macroblock, slide, frame, GOP, or even the entire MPEG stream. We will use the frame and GOP levels for our statistical analysis.

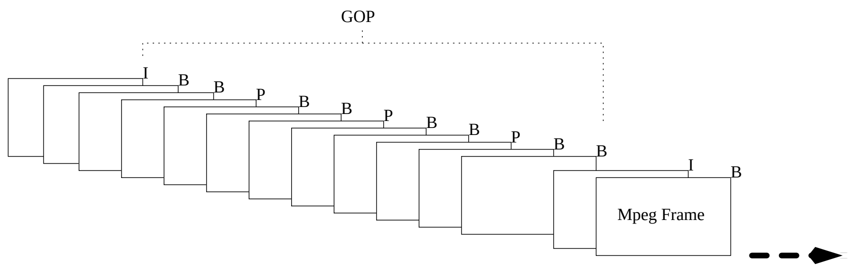


Figure 3.2: Encoded MPEG Video Sequence

3.3 VBR Codec Output

Generally, the overall bit rate of an encoded video stream depends on two resolutions: the temporal resolution (frame/sec) and the spatial resolution (pels/cm^2). The bit rate can vary from tens of kilobits to hundreds of megabits per second, based on the quality of the coded stream. Compression coding schemes (e.g. interframe coding) reduce the amount of data transmission which results in an essentially variable rate. The principle of interframe coding is to realise changes between successive frames with respect to a base picture (frame) predicted from previously received frames. Thus, the amount of data per frame can vary substantially, according to the degree of movement in the transmitted stream. Typically, characteristics of a video stream (sequence) based on the amount of data in an encoded frame can be sketched as follows (see Figure 3.3):

- According to the variation in the bit rate, sharp peaks can occur due to large scene changes. However, rate variations are relatively slight within the same scene.
- The shape of the bit rate distribution is typically bell shaped (or mound shaped). However, the skewness direction is based on the type of video stream (e.g. video conference, movie or sport).
- The autocorrelation function is another factor which can be used to characterise the dependencies feature of successive frames within the same video sequence.

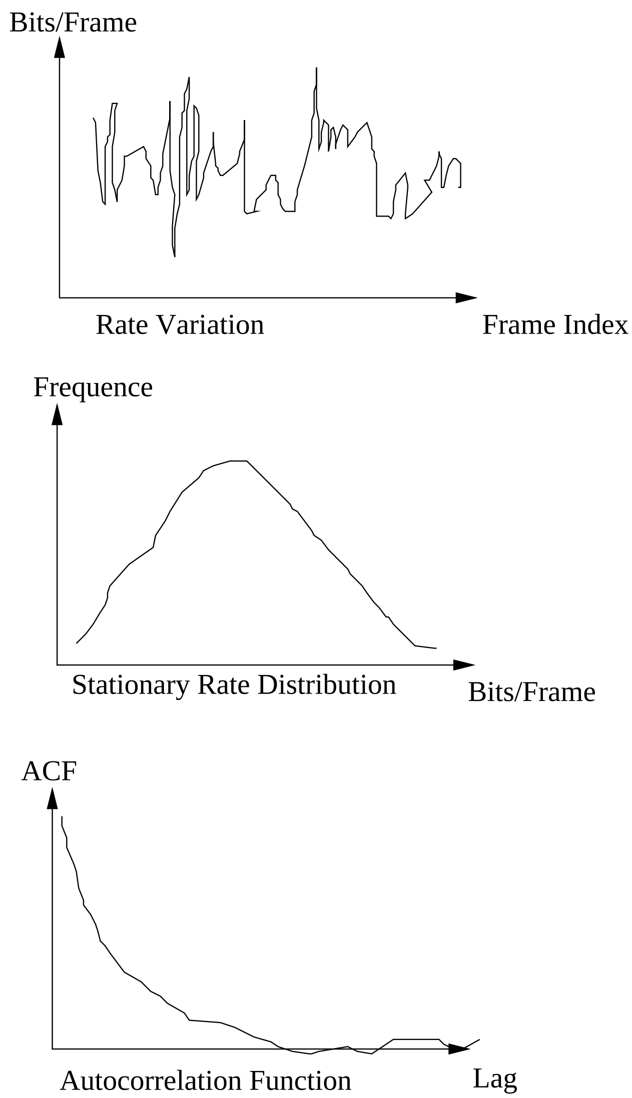


Figure 3.3: Characteristics of VBR video output

3.4 Statistical Analysis of an MPEG Pattern

In this section, we examine the characterisation of VBR MPEG streams in terms of their statistical behaviour by employing the three measures (distribution, autocorrelation function and scene changes). These measures comprise the focus of our statistical study because the distribution parameters are important to describe and understand the main features of MPEG traffic, while the correlations have an enormous impact on the queueing performance of a statistical multiplexer [Sriram86]. Furthermore, one of the major reasons for fluctuations in the overall bit rate is scene changes within the video stream [Lazar93]. The scene change will be discussed in greater detail in the next chapter while, in this chapter, we analyse MPEG behaviours, and examine some statistical characteristics of MPEG traffic using the first two measures (distribution and correlation).

It is difficult to characterise video traffic by using a short sequence of real data (of only a few seconds). For our statistical analysis, we use a long (about 30 minutes) sequence of real MPEG video which contains 40000 frames. Empirical data sets for MPEG video streams have been retrieved from the ftp site [Wurzburg95]. The video sequences have been encoded at the Institute of Computer science, University of Wurzburg. These sets represent frame size traces from MPEG-I encoded video sequences. Each frame consists blocks with an encoder input of 384x288 pels (Berkeley MPEG-encoder ver. 1.3 has been used with 12 bit colour information). The traced videos were captured in motion-JPEG format from VCR (VHS) with a captured rate between 19 to 25 fps. Table 3.2 summarises the encoding parameters. The traffic prameters (shown in the table) have been taken from [Wurzburg95].

Encoder	MPEG ver. 1.3
No. of Frames	40000 frames
Quantization Values	I=10, P=14 and B=18
Encoded Pattern	'IBBPBBPBBPBB'
GOP Size	12 frames
Encoder Input	384X288 pel
No. of Slices	1
Rate	19-25

Table 3.2: Parameters of the Encoded Sequence

In order to characterise the statistical behaviour of MPEG streams, we describe the number of cells per frame as a time series $\{X_i, i \geq 0\}$. In that way, the statistical parameters can be defined as follows:

$$\Delta = sup_{i \geq 0}\{X_i\}$$

$$\nabla = inf_{i \geq 0}\{X_i\}$$

$$\mu = E[X_i]$$

$$\sigma^2 = Var[X_i]$$

where Δ, ∇, μ, *and* σ are the peak rate , minimum rate, mean and standard deviation of $\{X\}$ respectively.

Table 3.3 presents the most important statistical parameters (frame-based) that we have obtained for some video sequence classes, including movies, sports events, TV shows and video conference. The table gives a general picture of the video sequence in terms of the variation and burstiness features. From this table, we can observe that some events (such as 'Race') lead to an MPEG sequence with a high peak. Moreover, some other events (such as 'Star Wars') have a high peak-to-mean ratio. The statistical properties of these video sequences are different, depending on the moving activities of the sequence. For example, the size of B frames in the sports sequences has a large amount of change (in some cases it has the same size as P frames (see Figure 3.4 (b)). This indicates a large number of movements in the input encoded sequence. As the result, the amount of activity within the video stream affects the frame sizes of the same GOP. Therefore, the amount of activities need to be considered at the statistical analysis process.

Generally speaking, it is possible to classify an MPEG sequence into three classifications, according to the amount of movements during the video sequence: namely High, Moderate and Low activity classes. The behaviour of the video sequences for the same class is almost the same, therefore 'Race', 'Dino' and 'video conference'[1] sequences have been chosen and they will be analysed in detail. The 'Race' sequence represents the High activity class, the 'Dino' sequence is representative of the Moderate activity class, while the 'video conference' represents the Low activity class.

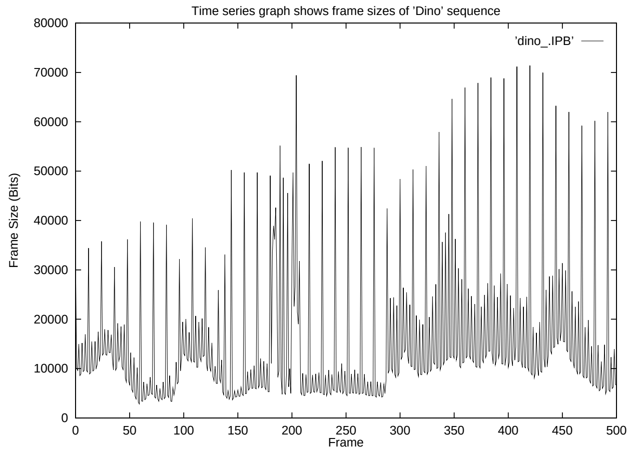
Video Sequence	Sequence Type	Mean (μ) Cell/frame	CoV($\frac{\sigma}{\mu}$)	Stdev(σ) Cell/frame	Peak Cell/frame	Peak/Mean
Dino	Jurassic Park Movie	35	1.13	39	312	9.14
StarWars	Star Wars Movie	25	1.38	34	325	13.4
Race	Formula 1 car race	80	0.69	55	527	6.58
News	TV News	40	1.27	51	495	12.36
Talk	TV Talk show	38	1.14	43	279	7.34
video Conference	Set-top Conference	16	1.93	30	121	7.66

Table 3.3: Simple Statistical Parameters for Some MPEG Sequences (Frame-Based)
Where : CoV is the Coefficient of Variation$\frac{\sigma}{\mu}$ And Stdev is Standard Deviation σ.

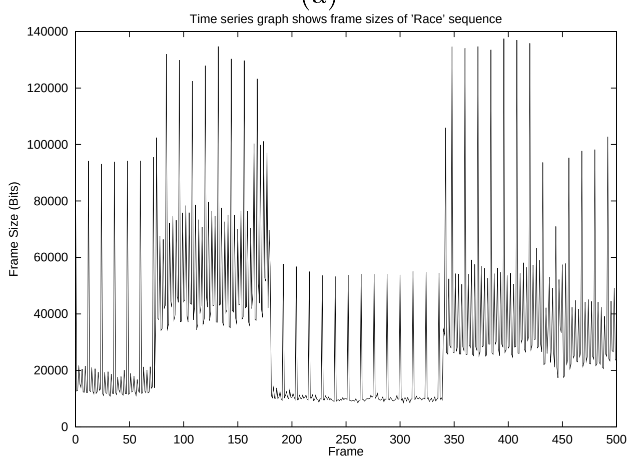
The size of GOP (summation of the frame sizes for every 12 consecutive frames) is influenced by the video activity within the same video sequence. For instance, if there is a lot of movement within a scene, the GOP size will be high; if there is minimum movement, the GOP size will be low. Figures 3.4 (a-c) show the time series plots for some video sequences compared at frame sizes. A high level of activity causes the sizes of both P and B frames to be enlarged. For instance, the frame size of type I is very large in the 'Race' sequence (Figure 3.4 (b)). In comparison, the B and P frames in the 'video conference' sequence are small compared to the I frames. This is because the amount of movement in the video sequence is not large. In other words, there is a small amount of scene changes.

The GOP plays the most important role concerning the autocorrelation effects of an MPEG video stream coded with different frame types, because it fixes the periodic nature of the stream. This unique property of an MPEG coded video prevents us from using video models which are based on statistical data from video sequences which have only one frame type, or ignore the GOP

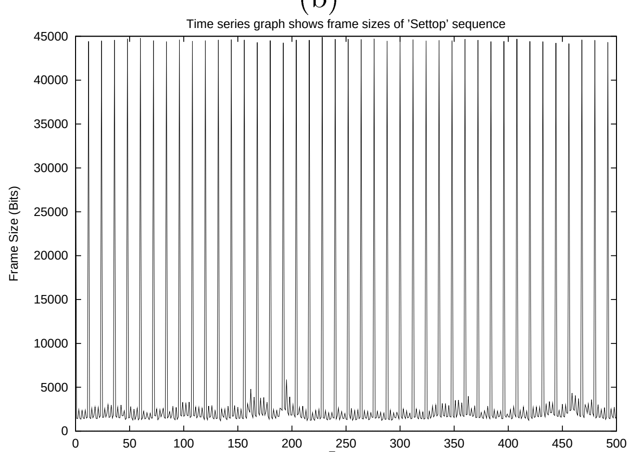
[1] In this thesis, 'video conference' and 'Settop' are referred to the same video stream



(a)



(b)



(c)

Figure 3.4: Time Series Plot

structure altogether. Thus, we focus on the GOP level for our characterisation process.

Video Sequence	Sequence Type	Mean (μ) Cell/GOP	Stdev(σ) Cell/GOP	Peak Cell/GOP	CoV($\frac{\sigma}{\mu}$)	Peak/Mean
Asterix	Cartoon	698	325	2811	0.46	4.02
Atp	ATP Tennis Final	684	255	2052	0.37	2.99
Bond	Movie	759	284	2490	0.37	3.27
Dino	Jurassic Park Movie	408	164	1634	0.40	4.00
Fuss	Football Game	847	322	3336	0.38	3.93
Lambs	Movie	228	138	1203	0.60	5.26
Movie	Movie	446	224	1776	0.50	3.97
Mr Bean	TV Series	550	275	2248	0.49	4.08
MTV	Musical Program	769	358	3335	0.46	4.33
News	News	645	282	2472	0.43	3.82
Race	Formula 1 Race	961	362	3470	0.37	3.61
Sbowl	Sport	734	270	2213	0.36	3.01
Simpson	Cartoon	580	247	2182	0.42	3.75
Soccer	Soccer Game	784	376	3050	0.48	3.88
Starwar	Movie	290	167	1170	0.57	4.02
Talk	Talk Show Program	454	147	1225	0.32	2.69
Term	Movie	340	117	1061	0.34	3.11
Settop	Video Conference	187	32	372	0.17	1.98

Table 3.4: Simple Statistical Parameters for MPEG Sequences (GOP-Based) Where : CoV is the Coefficient of Variation$\frac{\sigma}{\mu}$ And Stdev is Standard Deviation σ.

Table 3.4 shows the most important statistical parameters of several MPEG sequences at GOP level. This table includes various types of MPEG sequences such as movies, sports and even news and 'video conference' events. This means that we cover wide range of MPEG sequences. What is obvious from the table is the great diversity of statistical parameters of the sequences. The sequences are similar in that they use the same coding parameters and picture format, and were retrieved in the same way from analog video tape. Thus, even though the streams contain the same amount of data before the coding took place, the outcome varies widely. From the table, it is possible to describe the general behaviour of an MPEG sequence in terms of its mean rate (Cell/GOP) and burstiness. From the table, we can also conclude that some sequence types (such as sport and movie events) lead to MPEG sequence with a high peak bit rate and a high peak-to-mean ration. However, even the statistical parameters of the sequence of the same type are not stable. This leads to difficulties in finding traffic classes for MPEG sequences. In the following sections, MPEG sequences can be described by two main factors: statistical distribution of each frame type, and the dependency between them.

3.4.1 Sequence Distribution

In this section, we discuss the statistical distribution of the empirical data sets in order to discover the best distribution. The statistical distribution parameters can be used in the process of approximating the frame and GOP sizes. Figure 3.5 shows the corresponding histogram for the empirical data set of 'Dino' sequence. The shape of the histograms suggest that the true underlying distribution is skewed to the right. Since the densities of the Gamma, Weibull and Lognormal distributions can all take on shapes similar to that of the histogram, we propose them as candidates for the desired distribution. In order to use these distributions in our analysis, we need first to determine the parameters of the three proposed distributions. It is important to note that both Gamma and Weibull distributions with their shape parameters have an appearance similar to the typical histograms of these two distributions while the Lognormal distribution always has this general shape [Law91]. We will now assess how well our three particular distributions represent the actual distribution of the MPEG empirical data sets.

There are many ways to test the 'Goodness' of the data. One way to test the Goodness is to fit the data histogram with the distribution curve. Previous studies have shown that the frame size of video sequences with various (vary from low to moderate) number of movements could be described by either a Gamma or a Lognormal distribution [Rose95a] [Hyman92]. In this section, we discuss the fitting of video sequences (at GOP level) to the three distributions. The study considers a wide range of MPEG sequences (with different amount of activities). The Gamma density function is given by:

$$f_x(x) = \frac{\lambda^{\alpha} x^{\alpha-1}}{?\,(\alpha)} e^{-\lambda x} \quad , \; x, \alpha, \lambda > 0$$

where x is the random variable (a GOP size in bits), $\alpha \, and \, \lambda$ are the shape

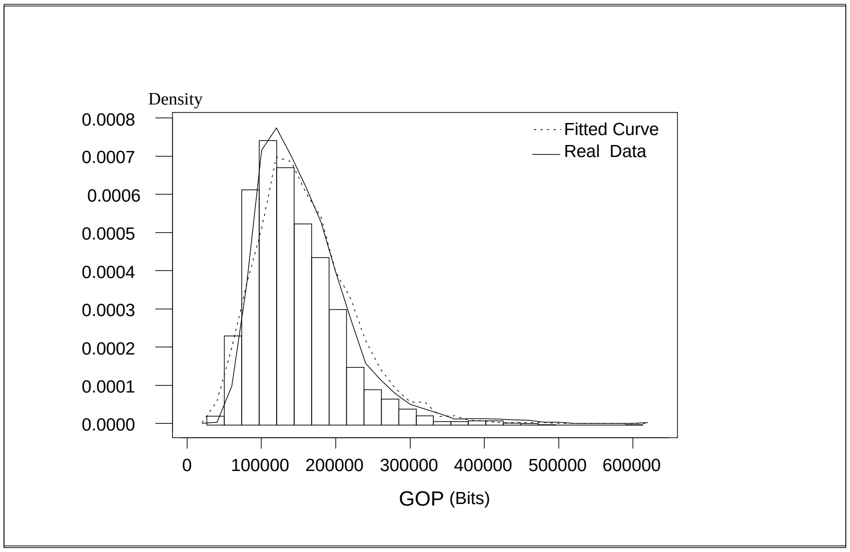


Figure 3.5: The Data Histograms and The Corresponding Gamma Curve

and scale parameters, and $?(\alpha)$ is the Gamma function and is defined as:

$$?(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

The mean and the variance are defined as:

$$E[x] = \alpha \cdot \lambda$$

and

$$Var[x] = \alpha \cdot \lambda^2$$

while the Lognormal density function is given by:

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{lnx-\mu}{\sigma})^2}$$

Then, the mean and the variance are defined as:

$$E[x] = e^{(\mu+\frac{1}{2}\sigma^2)}$$

and

$$Var[x] = w(w-1)e^{2\mu}$$

where

$$w = e^{\sigma^2}$$

The Weibull density function is defined as follows:

$$f(x) = \alpha\lambda^{-\alpha}x^{\alpha-1}e^{-(\frac{x}{\lambda})^{\alpha}} \quad , \quad x, \alpha \, and \, \lambda > 0$$

Accordingly, the mean and the variance are further defined as:

$$E[X] = \frac{\lambda}{\alpha} ? \left(\frac{1}{\alpha}\right)$$

and

$$Var(x) = \frac{\lambda^2}{\alpha} \left\{ 2? \left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \left[? \left(\frac{1}{\alpha}\right) \right]^2 \right\}$$

Figure 3.5 shows that there is an agreement between the data histogram and the curve of the corresponding Gamma distribution (for 'Dino' sequence). The data represents the GOP sizes. In [Rose95a], It has also been confirmed that frame size (for I, P and B) gives the same results to fit Gamma distribution. In many cases, comparing the histogram and the curve of the corresponding distribution is not enough. Thus, we need to use a more accurate method to fit the data distribution.

The fractile diagram method (Quantile-Quantile plot or Q-Q plot) is another way to find the closest distribution for an empirical data set (I, P and B frames or GOP size) [Doulamis96]. The method plots the quantiles of the data versus the quantiles of the fitted known distributions (more details on Q-Q plot can be found in [Law91]). The Q-Q plot can be used to amplify the differences which exist between the tails of the data and the fitted distribution. We have used a statistical software called MINITAB to analyse, calculate and plot the Q-Q plot [Minitab89]. Figure 3.6 shows an example of the MINITAB macro for plotting Q-Q. The following algorithm was employed for our calculations:

$\{X_i \, , \, N > i > 0\} \, contains \, an \, empirical \, data \, set$
$LET \;\; \alpha \, AND \, \lambda \; are \, the \, shape \, and \, scale \, parameters \, of \, the \, data \, set \, \{X_i\}$
$\{X_{sorted_1}\} = Sort \, \{X_i\}$
$Calculate \, the \, Empirical \, Cumulative \, Distribution \, Function:$

$(ECDF(\{X_i\}))\, of\, the\, data\, set$
$PLOT\ \sqrt[3]{\{X_{sorted_i}\}}\,,\ \sqrt[3]{ECDF(\{X_i\})}\,;\ N > i > 0$

In order to test the Q-Q plot on one of the frame types, Figure 3.7 (a) shows Q-Q plot for I frames of the 'Dino' sequence and fitting the Gamma distribution. The Q-Q plot shows that there is an agreement with the Gamma distribution because most of the data plot is 'nearly' linear shaped. However, Figure 3.7 (b) shows that GOP sizes for 'Dino' have a better agreement with the Gamma distribution than the 'video conference' sequence in Figure 3.7 (c). This is because some of the data do not fit Gamma distribution.

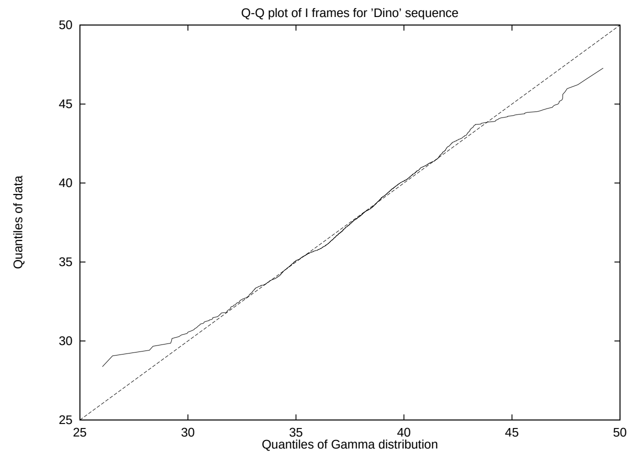
In order to examine the 'Goodness' of the data to the Lognormal distribution, we draw a normal probability plot for the $Log(X_i)$, where X_i for $i < N$ are the GOP sizes. The method is based on the fact that a variable X has a Lognormal distribution if $log(x)$ has a Normal distribution with a mean μ and a standard deviation σ [Monk91]. We have used again MINITAB to draw the normal probability plot for various MPEG sequences at the GOP level. The plot uses Anderson-Darling test for the normality test [Minitab89]. Figures 3.8 (a-d) show the normal probability plots of the $Log(GOP)$s for 'Dino', 'Race', 'Movie' and 'video conference' sequences. The vertical axis represents a probability scale while the horizontal axis represents a data scale. MINITAB fits and draws a least-squares line to the points (GOP) that estimate the cumulative distribution function for the population from which data are drawn. If the $Log(GOP)$ fits the Normal distribution then the data points should be over the line. We have observed from the Figures that the 'Race' sequence shows an agreement with the line while a better agreement can be observed in the case of the 'Dino' sequence. However, the 'Movie' sequence shows less agreement and the 'video conference' sequence shows a poor agreement with the line. Thus, generally speaking, Lognormal distribution can be used to describe GOP sizes of MPEG sequence.

Weibull distribution has been also suggested to describe video sequences [Heyman96]. We have used the Weibull plot to test the 'Goodness' of the data set. Similar to the normal probability plot, the vertical axis has a probability scale and the horizontal axis has a data scale. Figures 3.9 (a-c) show the Weibull probability plot for the 'Dino', 'Race' and 'video conference' sequence. From the Figures, we can observe that Weibull distribution exhibits a weak

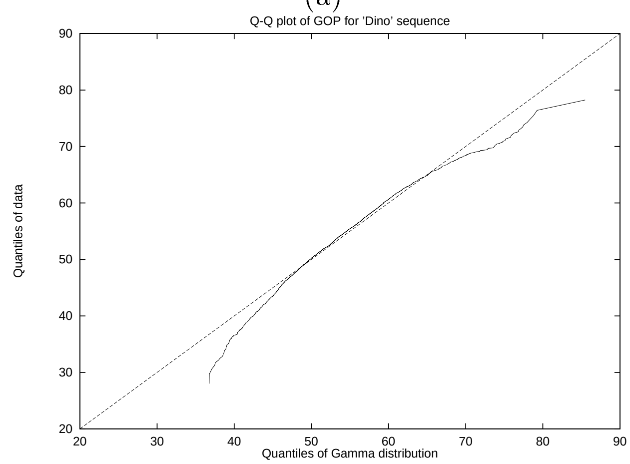
```
# Macro for Gamma Quantile-Quantile  Plot

    # This Macro is a modification ver.  of the original Macro which
  was written by Terry Ziemer, Minitab, Inc.
    # You will be prompted for the number of the column where your
  data
    # is stored.  The data will be sorted and an Empirical Cumulative
    # Distribution Function (ECDF) calculated.  Both the sorted data
  and
    # ECDF will be stored by the macro.  The sorted data will be
  stored in
    # column 99 and the ECDF in column 100.  In addition, the macro
  uses
    # columns 95-98.  If you don't have enough columns in your environment,
    # edit the macro and change the values of k1-k6.
    noecho
    oh 0
    let k1 = 95
    let k2 = 96
    let k3 = 97
    let k4 = 98
    let k5 = 99
    let k6 = 100
    note
    note Enter the number of the column which contains the input
  data
    note
    set 'terminal' c100;
    nobs 1.
    note
    copy c100 k7
    note Enter the Gamma Shape Parameter
    note
    set 'terminal' c100;
    nobs 1.
    copy c100 k8
    sort ck7 ck5
    let k10 = count(ck7)
    set c100
    1:k10
    let ck6 = (c100 - (3/8))/(k10 + (1/4))
    invcdf ck6 ck4;
    gamma k8 25291.689.
    let ck4 = ck4**(1/3)
    let ck3 = ck5**(1/3)
    plot ck3 ck4;
```

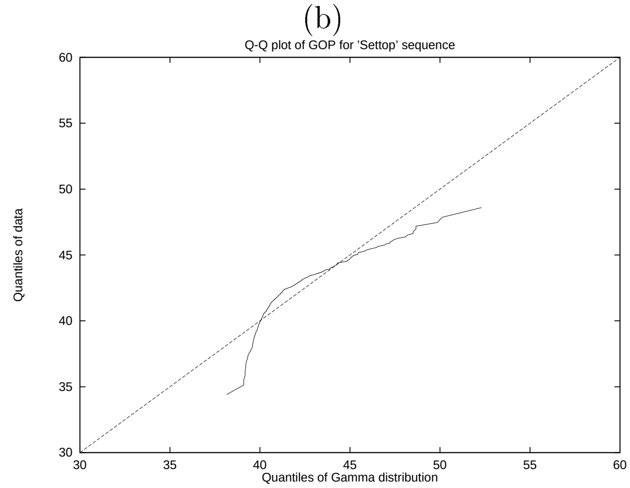
Figure 3.6: Q-Q Macro



(a)

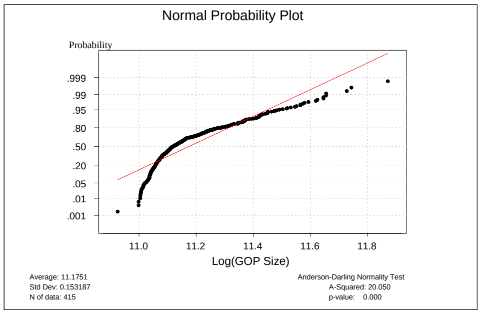


(b)

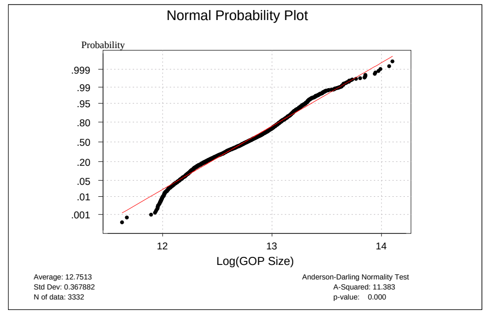


(c)

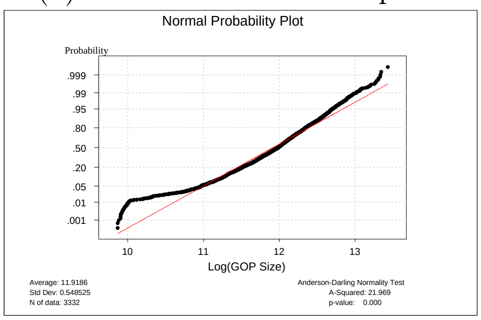
Figure 3.7: Q-Q Plots



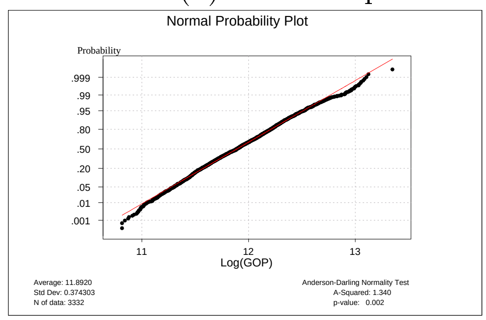
(a) 'video conference' Sequence



(b) 'Race' Sequence



(c) 'Movie' Sequence

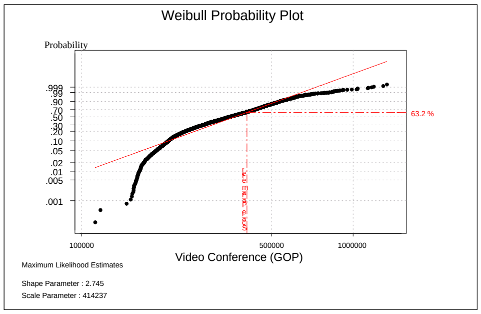


(d) 'Dino' Sequence

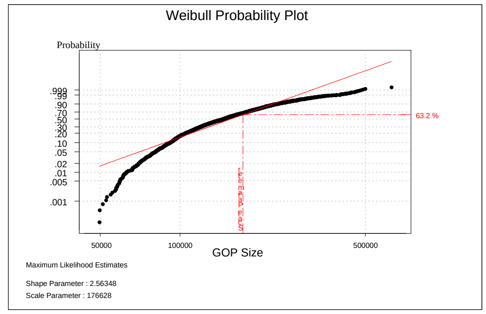
Figure 3.8: Examining Lognormal Distribution Using Normal Probability Plot

agreement for the GOP sizes of MPEG sequence. However, only the middle points (i.e. the GOPs which are close to the mean value) approximate the corresponding Weibull distribution in the case of the 'Dino' and 'Race' sequence, while the 'video conference' sequence does not approximate the corresponding Weibull distribution.

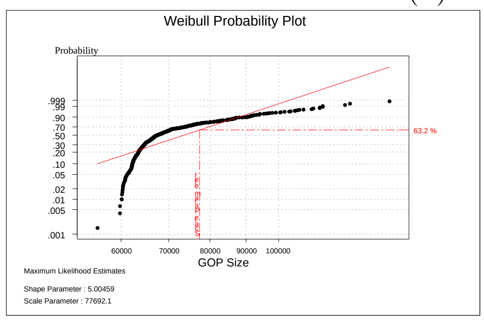
Consequently, there is not yet a good agreement for the most fitting distribution of all MPEG video traffic due to the different degrees of activity or the amount of movement within the video stream. Generally speaking, Gamma or Lognormal distributions can be used to approximate the frame sizes and GOP size of most of the types. In most cases, there is no large difference between Gamma and Lognormal distributions [Rose95a]. However, perfect agreement of the histogram and the approximation can not be achieved because the data set is finite. Heyman et al. suggest that the Gamma distribution is a good fit for a low activity video sequence (based on one frame coding scheme) and may not depend on the coding algorithm used [Hyman92].



(a) ‘Race’ sequence



(b) ‘Dino’ Sequence



(c) Video Conference

Figure 3.9: Weibull Probability Plot

3.4.2 Sequence Correlation (Dependency Feature)

In this section, we examine the dependency features of the MPEG sequence. This information comprises one of the most important features of an MPEG sequence. The analysis of any video sequence shows dependencies between the frames, and between the GOPs within the same video sequence. We have measured the correlation between the summation of each frame type for each GOP within the same MPEG sequence. Table 3.5 depicts the statistical results that we obtained. It shows that dependency exists among the three frame types, and therefore they should not be represented with three independent processes. It can be seen that there is a strong correlation between B and P frames while there is a weaker correlation between I and B, and , I and P. Another result that is shown in table 3.5 is that the correlation factor is negative due to the lack of activities within the ‘video conference’ stream.

Also, these dependencies can be measured by using the autocorrelation function (ACF). We compute the autocorrelation of the frame sizes $\{Z_i : i = 1, 2, 3, ...\}$. ACF can be given by the following equation:

Sequence	I and $\sum$P	I and $\sum$B	$\sum$B and $\sum$P
Dino	0.352	0.306	0.896
Video Conference (Settop)	-0.274	-0.278	-0.964

Table 3.5: Correlation Between I, P and B frames

$$r_k = \frac{\sum_{t=k+1}^{N}(Z_t - \overline{Z})(Z_{t-k} - \overline{Z})}{\sum_{t=1}^{N}(Z_t - \overline{Z})^2}$$

We plot the r_k against lag k:

$$\{k : k = 1...N,\ where\ N\ is\ the\ number\ of\ data\ frames\}$$

Figures 3.10 and 3.11 illustrate the correlation between I, P and B frames for the 'Dino' and 'video conference' sequences. It is clear that the I frames cause a large positive peak, followed by another smaller positive peak from the P frames, while B frames cause the negative and smallest peaks.

In Figures 3.10 and 3.11, the shape of the curve is the result of the periodic coding pattern (the pattern [IBBPBBPBBPBB] is repeated) and the different mean sizes of the frame types. The pattern between two I frame peaks is repeated with slow decaying behaviour. In addition, correlation exists between the GOPs within the same sequence. In the case of the 'Dino' sequence (see Figure 3.12), the GOP correlation curve shows a slow decay behaviour for larger lags, meaning that there is a long-range dependency (LRD) between the GOPs sequence [Izquierdo96]. In contrast, the GOP correlation curve for 'video conference' (Settop) and 'Race' sequences exhibit rapid decay behaviour for small lags. Therefore, they have a short-range dependency (SRD).

Another parameter that can be used to approximate the long range dependencies is the Hurst exponents (H) for the video sequence [Rose95a]. The value of the H parameter gives an indicator of the dependencies power. The value of the Hurst parameter is ranged from 0.5 to 1. If the sequence does not have long-range dependencies, then the H parameter will be 0.5 (as in Poisson process). In contrast, a larger value for the H parameter means a greater amount of movement in the video sequence. Table 3.6 indicates the H parameter for some video sequences. For instance, 'video conference' own low value for the H parameter, meaning that there are no long range dependencies. It is interesting to note that the 'Race' sequence has high H value, while the

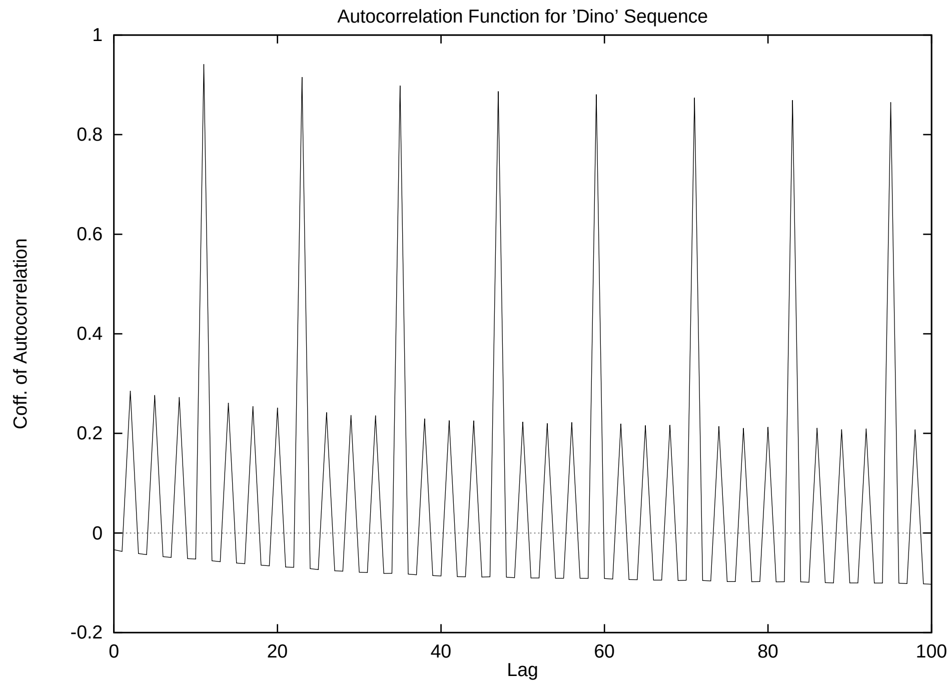


Figure 3.10: Autocorrelation Function for 'Dino' at Frame Level

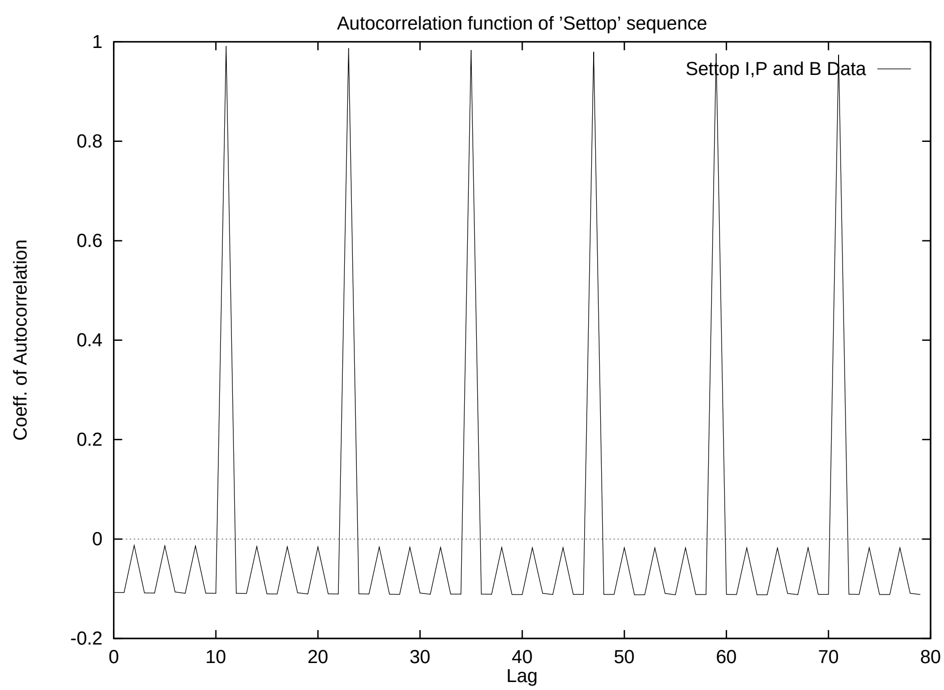


Figure 3.11: Autocorrelation Function for 'Settop' at Frame Level

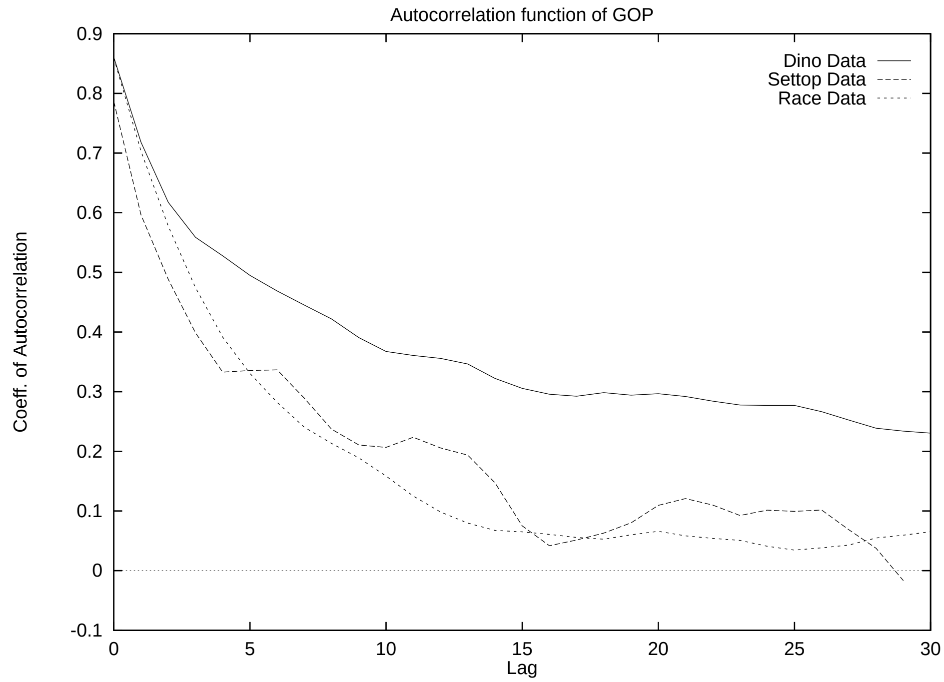


Figure 3.12: Autocorrelation Function for Several Sequence at GOP Level

autocorrelation function of its GOP decays rapidly in small lags (see Figure 3.12). There are many ways of estimating the H parameter, and most of them do not give the same value [Izquierdo96]. Thus, the H parameter is beyond the scope of this current studies. However, it can be used here just as an indicator to show the level of dependencies in the sequence.

Video Sequence	Sequence Type	Hurst Exponent (H)
Dino	Movie	0.88
Talk_1	TV talk show	0.87
News	TV news	0.79
Starwars	Movie	0.74
Race	TV sports event	0.99
Settop	Video conferencing	0.53

Table 3.6: Hurst Parameter (from [Rose95a])

3.5 Burstiness Measurement for MPEG Traffic

Burstiness is an important measurement for traffic characterisation. It plays a critical role in determining the network performance [Onvural94]. For instance, burstiness affects the queueing behaviour at an ATM multiplexer. However, there is not a single and widely-accepted notion of burstiness of video traffic [Molnar97]. In fact, burstiness is one of the connection parameter that a user might be expected to declare in order to provide the network management with information which can help to achieve an efficient network control as well as high resource utilisation. In other words, burstiness is one of the most effective factors in using the bandwidth of a network efficiently with the VBR service. In addition, exploring the burstiness of the traffic provides a better understanding of the correlation characteristic for the video traffic. Moreover, the burstiness information helps us to locate a suitable traffic model. For instance, the traditional traffic models (such as Poisson) cannot be used to model bursty traffic because bursty traffic tends to have a large coefficient of variation (CoV) value (larger than the Poisson process). In the literature, there are several methods have been proposed in order to measure the traffic burstiness. Let $A_i, i > 0$ be the inter-arrival rate (in cell/GOP), $\mu = E[A_i]$ and $\sigma^2 = Var[A_i]$, where σ is the standard deviation. Burstiness degree can be estimated in different way as follows: (Table 3.7 shows these measures for several MPEG streams):

- The ratio between the peak and the mean (μ) rate (PMR) is one of the most widely used measures of burstiness . However, the definition and the applicability of a peak is not at all clear [Molnar97]. Is depends on the used time scale. Generally, PMR can be defined as follows:

$$PMR = \frac{Peak\,rate}{Mean\,rate}$$

If the burstiness (or PMR) is equal to one, then the traffic does not have burstiness.

- The ratio between the standard deviation (σ) and the the mean (μ) rate (Coefficient of variation):

$$CoV = \frac{\sigma}{\mu}$$

This measure gives more information than the peak to average ration since the traffic variation is a function of the CoV. Thus, the traffic is not bursty if the coefficient of variation is equal to zero.

- The density of bursts within the traffic cell stream. In this way, the burstiness is the statistical average of the burst length. Thus, this measure takes account only the first-order property of the traffic (it is a function of the marginal distribution of the interarrival times only).

- The index of dispersion for counts (IDC) parameter. It is related to the sequence of counts of arrivals in consecutive time units. The IDC shows the variability of a process over different time scales. It is defined so that it is constant (1) for the Poisson process. For a given interval, the index of dispersion is the variance-to-mean ration of the number of arrivals (in cell) in that interval (for instance a GOP time) [Frost94]; whereby:

$$IDC = \frac{Var(N)}{E(N)}$$

where $Var(N)$ and $E(N)$are the variance and the mean value of the arrival process.

- The Squared Coefficient of Variation($SCoV$). This measure is also widely used which includes information from the first two moment of the traffic process and is defined as:

$$SCoV = \frac{Var[A_i]}{E^2[A_i]} = \frac{\sigma^2}{\mu^2}$$

If $SCoV > 1$, then the traffic is more bursty.

3.6 Summary

Given a good understanding of the statistical behaviour of MPEG traffic will help to handle bursty traffic in an efficient way. In this chapter, we have

Sequence	PMR	CoV	ID	SCoV
Dino	4.00	0.401	65.86	0.161
Race	3.61	0.37	136.55	0.142
Settop	1.98	0.17	5.62	0.029
Movie	3.97	0.50	113.23	0.253

Table 3.7: Burstiness Measures for Several MPEG Streams.

presented a description of the statistical analysis of various MPEG streams in order to cover a wide range of video classes (in terms of stream activities). Empirical data sets have been obtained from different real video streams which are originally encoded using MPEG encoder. We have presented a general description of the main statistical parameters for various MPEG streams in order to show the variations of the bit rate of an MPEG stream. By comparing the statistical parameters, one can notice that there is a vast variety of the statistical properties even for streams belonging to the same video class (e.g. sport or movie). We have also shown that a high level of activity in the sequence results in large sizes of both 'P' and 'B' frames. The statistical analysis also addressed two main measures: distribution and correlation (dependency) properties of MPEG video sequences (based on an empirical data sets of encoded video streams).

We have studied the distribution of the empirical data sets in order to find the fittest distribution. Three distributions: *Gamma*, *Weibull* and *Lognormal* have been selected due to their similarities in shaping the typical histogram of the actual data set. The data sets have been tested using different statistical methods to assess how well these distributions represent the actual distribution of the MPEG empirical data sets. It has been observed that there is not yet a good agreement for the most distribution of all video traffic due to the amount of movements within the video stream.

We then studied the impact of the GOP structure, a strong correlation (roughly 0.9) has been shown between 'P', 'B', and a weaker correlation between (P, B and I). Therefore, 'P', 'B' and 'I' cannot be represented by independent processes. The ACF of the GOP sequence is also addressed in order to show the degree of dependencies between the GOPs within the same sequence. A sequence with a moderate activity exhibits a slow decay behaviour for the ACF curve (or long range dependency) while a sequence with large or low activities exhibits a rapid decay behaviour for the ACF curve (or short range dependency). As a result, we have found that scene changes have a substantial impact on the characteristic behaviour of an encoded MPEG video traffic.

Chapter 4

Scene Changes in MPEG VBR Video Traffic

4.1 Introduction

An important reason for fluctuations in the overall bit rate is the fact that scene changes take place within the video stream. As a result, an MPEG stream may have several spikes (peaks) due to scene changes. These may cause cell losses when multiple streams are multiplexed at an ATM switch. Furthermore, managing VBR video traffic is a very difficult problem due to the statistical properties of the video stream which, in turn, are dependent on the coding scheme and the content of the video sequence [Kara97]. Therefore, we need to analyse the magnitude of scene changes in order to achieve an efficient management for this type of traffic. This could be achieved by managing and classifying the video streams according to the amount of movement within the same stream. This chapter presents a technique designed to classify MPEG streams using the amount of activity within each stream. Firstly, we present two methods and associated algorithms to identify the scene changes within the MPEG stream. Based on the classification process of MPEG streams, we then introduce a '*Scene Change Scale*' (SCS) exhibiting (or grading) the amount of activity within the MPEG stream. The scale is used to demonstrate the impact of scene changes on QoS requirements. Our prime measurer of interest is the CLR at an ATM multiplexer. This means that we are considering the amount and magnitude of scene changes within the MPEG sequence in our analysis.

This chapter is organised as follows: In the next section, we present two methods to identify the scene changes within an MPEG stream. Then, SCS will be presented in section 4.3 . In section 4.4, we undertake several simulation experiments in order to demonstrate the extent of the performance of a multiplexed MPEG stream at a statistical multiplexer.

4.2 Scene Change Identifier

This section presents two methods that allow us to detect and measure the amount of overall bit rate fluctuations. We then use these methods to identify the amount (magnitude) of scene changes within an MPEG stream.

In a visual sense, a scene can be defined as that part of a movie which does not have sudden changes of view. As mentioned above, one of the more important reasons for fluctuations in the overall bit rate are the scene changes within the video stream. Thus, the scene change should be incorporated at the traffic characterisation process. In an empirical data set for a traced MPEG traffic, a significant change in the size of two consecutive GOPs is an indication of a scene change. We have used GOP sizes for our analysis because a GOP contains most of the picture's details (for our data trace, every GOP is composed of one 'I' frame which contains most of the picture information and three 'P' frames. In addition, there are 8 frames of type 'B' which contain information of any changes in relation to previous GOPs).

4.2.1 Scene Change Identification Using the Outlier Method

The basic idea of MPEG is that when a new scene begins, the size of the frames will be larger than the previous frame sizes. In other words, within the same data set, the changes in the size of two consecutive GOPs can be an indication of a scene change. The amount of change can be measured by using one of the most commonly used measurements of data variation (or variability), namely: the *variance* and the *standard deviation*. However, these parameters provide only the overall measurement of the data variation, and cannot be used as a measurement of location relating to the rest of the data set. In order to overcome this, we will use another statistical parameter, called *outlier*, to describe each element *relative* to the other elements in the same data set [Mendenhall94]. The *outlier* is an element value which seems to be unusual (or abnormal) compared to the other elements within the same data set. Within an empirical data set, outliers can be detected by using either a numerical method, based on a *z-score* measure, or a graphical technique called *Box Plot* [Groeneveld88].

The *z-score* measure can be used to describe the location of a Y_i relative to the mean in units of the standard deviation $\{Y_i\,,\, 0 \leq i \leq N-2\}$, where $Y_i = X_{i+1} - X_i$ and X_i is number of cells in the i-th GOP and N is the number of GOPs. As such, *z-score* can be calculated as follows:

$$z\_score_i = \frac{Y_i - E[Y]}{\sqrt{Var[Y]}}$$

$$E[Y] = \mu$$

$$Var[Y] = \sigma^2$$

According to the *z-score* definition, negative *z-score* values indicate that Y_i lies to the left of the mean, while positive values indicate that Y_i lies to the right of the mean. Therefore, Y_i is called *outlier* if Y_i is unusually large relative to the other values of $\{Y\}$ in the empirical data set. We use the Rule of Thumb for detecting outliers within an empirical data set [Mendenhall94]. The rule states that if the *z-score* value is greater than a *Threshold* (τ), usually $\tau \leq 3$, then an outlier is identified; thus:

$$\begin{array}{ll} If & z\_score_i \geq \tau \\ & Then\ Y_i\ is\ considered\ to\ be\ an\ outier\ at\ location\ i \end{array}$$

For the sake of illustration, we can demonstrate the technique using only one empirical set (the 'Dino' sequence). However, this technique can be adopted to any empirical data set. We have plotted the *z-scores* for the 'Dino' sequence (see Figure 4.1). The Figure shows many spikes that are caused by large changes in GOP sizes (i.e. large scene changes). Every spike over the τ value is considered to be an *outlier*. Consequently, each outlier is an indication of a significant scene change. It is important to notice that the smaller value of τ allows it to capture more scene changes, and vice versa. The following algorithm can be used to identify the starting point of a scene change within an MPEG sequence:

START
Let τ be a *Threshold* value
For i=1 To N-1
  Let X_i be number of cells in *i*-th GOP, $0 \leq i \leq N-1$
  $Y_i = X_{i+1} - X_i$
  Calculate z-score $(Y_i) \Longrightarrow \frac{Y_i - E[Y]}{\sqrt{Var[Y]}}$
  If $z\_score_i \geq \tau$ Then a starting point of a new scene is identified at location i.
End Loop i
END

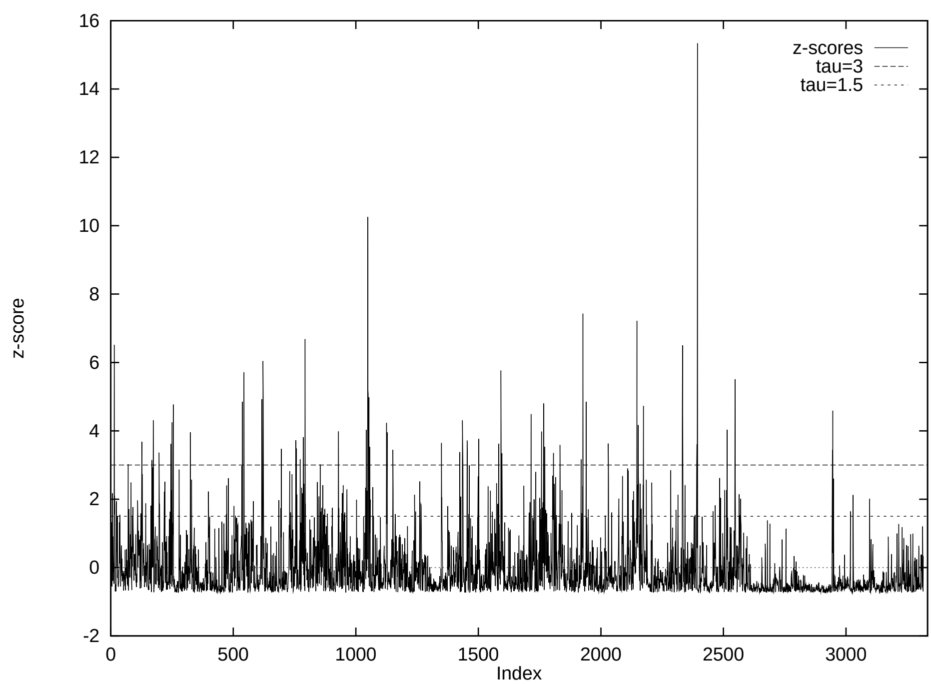


Figure 4.1: z-score Plots for the 'Dino' Sequence

A similar method can be used to detect an outlier by constructing the box plot of the empirical data set (see Figure 4.2). First, the method constructs

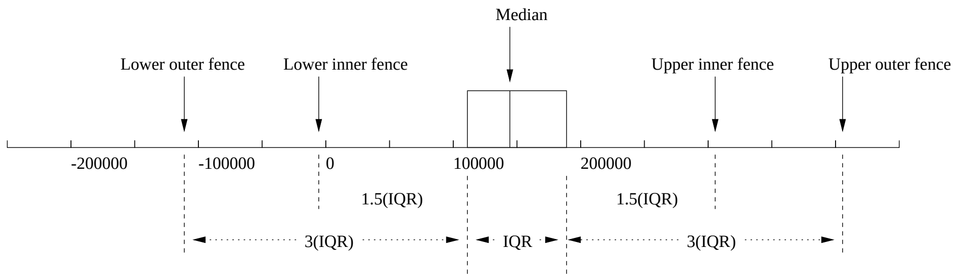


Figure 4.2: Box Plot for 'Dino' Sequence

two intervals based on a quantity value called the *Interquartile Range* (IQR):

$$IQR = Q_u - Q_l$$

where Q_u and Q_l are the *upper* and *lower* quartiles respectively.

Next, we construct two sets of limits out of IQR called *Inner* fences (If) and *Outer* fences (Of). *Inner* fence values are located a distance of $1.5IQR$ below Q_l and above Q_u, whereas *Outer* fence values are located a distance of $3IQR$ below the Q_l and above Q_u. Similarly, according to the statistical theory of the Rule of Thumb for detecting outliers, every $\{Y_i, i \geq 0\}$ which falls between the *inner* and *outer* fences is called *suspect outlier*. But, if Y_i is located outside the *outer* fences it is called *a highly suspect outlier*. In other words, every *suspect outlier* could be a moderate scene change while every *highly suspect outlier* could be a significant scene change.

The *outlier* method has been applied on various MPEG sequences. Table 4.1 presents several statistical parameters of *outliers* for three MPEG sequences representing various levels of activity.

Sequence	Mean(GOP) cell	Number of Outliers	Max(outlier) cell	Mean(outlier) cell	Stdev(outlier) cell
Dino	408	242	1634	809	160
Movie	446	262	1777	963	189
Video Conference	188	39	373	266	30

Table 4.1: Outlier for Various MPEG Sequences

These two methods produce similar results. However, the presence of one or more large *outlier* in a data set can inflate the value of the standard deviation (σ) used to calculate the *z-score*. Consequently, it will be less likely to be able to detect an element value with a high *z-score*. In contrast to this, the value of the quartiles used to calculate the fences for a box plot are not affected by the presence of outliers.

As a result, both methods can be used to identify the significant scene changes within an MPEG video stream. The scene change identifying process also helps to detect the most abnormal part of an MPEG video stream which may cause cell loss in most cases. This type of information is useful in classifying MPEG streams in terms of the amount of activity within each stream (as we will demonstrate later).

4.2.2 Scene Change Identifier Using Second Different Method

The last two methods identified scene changes based on the differences between only two consecutive GOPs. In order to impose more accuracy on the scene change identification results, it is desirable to compare a GOP with the previous and the next GOPs. Therefore, we will employ a method which is based on the *Second Difference*. The method can be used for any MPEG sequence to identify the scene changes within the sequence. Similarly, we shall use the 'Dino' stream to demonstrate our method. The time series plot (see Figure 4.3) indicates several spikes (peaks) due to possible scene changes. In order to determine which spike represents a true scene change, we need to analyse its magnitude. This can be achieved by relating each GOP spike with its neighbours (GOPs on both sides), according to the amount of movement within the same stream. As described before, a scene change occurs when a GOP size is abnormally larger than its neighbours. Based on this fact, we can quantify the scene change in the following way:

Let us assume that $\{X_i\}$ is the size of a GOP: $\{X_i : i = 1, 2, ..., N\}$. At a scene change, the second difference (Dif_2) will be large in magnitude and negative in sign [Hyman96]. The *Second Difference* is given by:

$$Dif_2 = ((X_{i+1} - X_i) - (X_i - X_{i-1}))$$

Figure 4.4 shows the plot of the second difference for the 'Dino' stream. Every large negative spike could be an indicator of a scene change. In order to quantify only the significant scene changes, we divide the second difference by the average of the past few seconds (t). The period of the last few seconds, t,

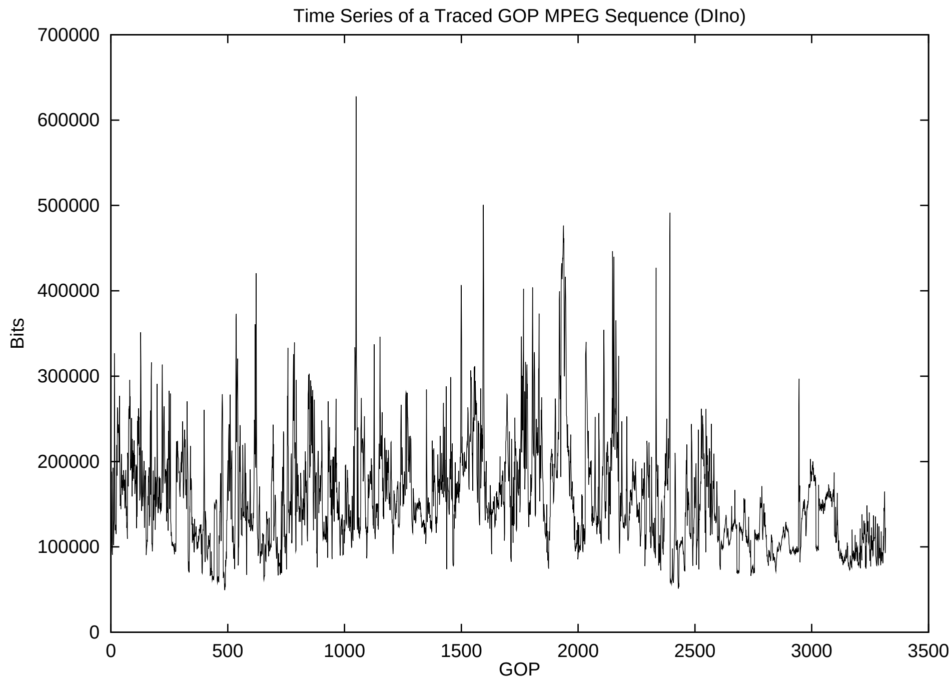


Figure 4.3: GOP Time Series (Dino)

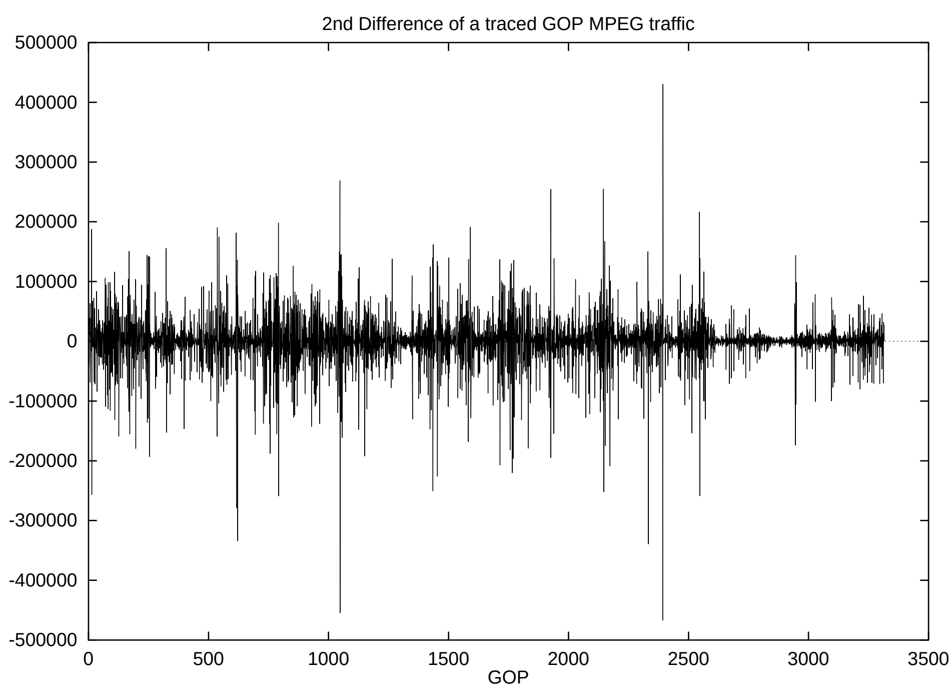


Figure 4.4: Second Difference

might vary. In some studies, the average length of a scene might range from 3 to 7 seconds [Krunz96]. We have tested various values for t, all of them giving similar results for $\{Y_i\}$:

$$Y_i = \frac{(X_{i+1} - X_i) - (X_i - X_{i-1})}{\frac{1}{t}\sum_{j=i-t}^{i} X_j} \ , \ i = 2, 3, , .., N - 1$$

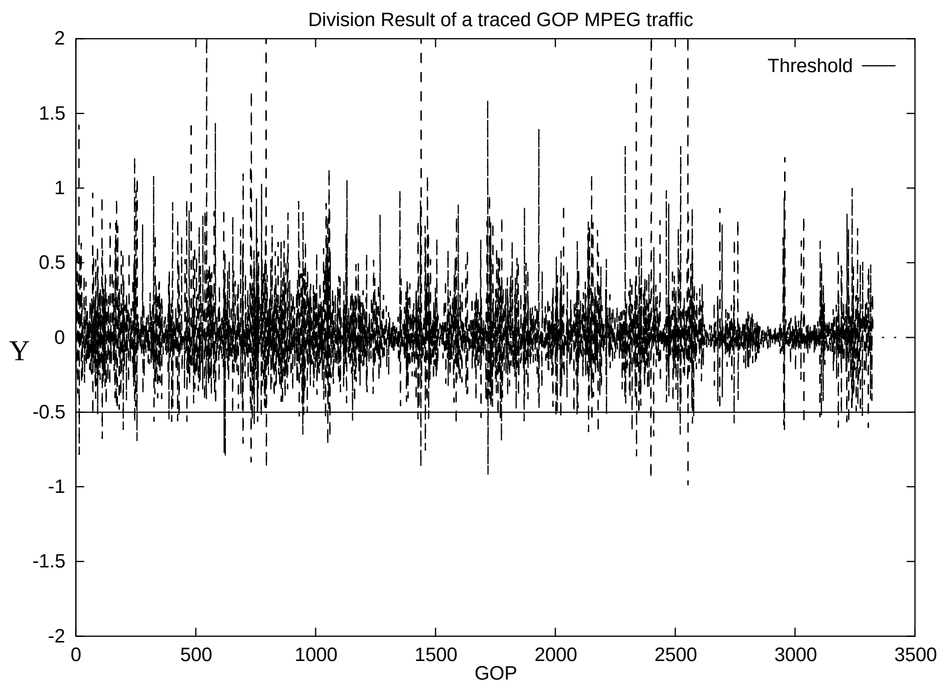


Figure 4.5: Scene Change Identification

A significant scene change can be identified with every negative large spike when we plot the division result Y_i from the above equation. We chose a *threshold* (T) as a critical value, where $0 < |T| < Max_Y$. The number of spikes below the *threshold* indicates the amount of large movements within the same MPEG stream (see Figure 4.5). Lower values for the *threshold* , $|T|$, capture more scene changes. In order to capture only the large scene changes, it is more obvious when T is below the mean value of $\{Y\}$. The following algorithm depicts the method which identifies the scene changes within an MPEG traced stream:

```
START
READ X_i from a traced GOP file of size N
Threshold = T
FOR i = 1 TO N
        Dif_2 = (X_{i+1} - X_i) - (X_i - X_{i-1})
        CALCULATE μ = (1/t) Σ_{i-t}^{i} X_i
        Y_i = Dif_2 / μ
        IF Y_i < T   THEN Scene Change is identified
End Loop i
END
```

In order to justify our criteria, see Figure 4.6 in which we plot $\{X_i\}$ time series and the second difference Dif_2 . It is clear that there is a good match between the two series. Every large and negative spike in the Dif_2 is associated with a large spike in the GOP time series plot.

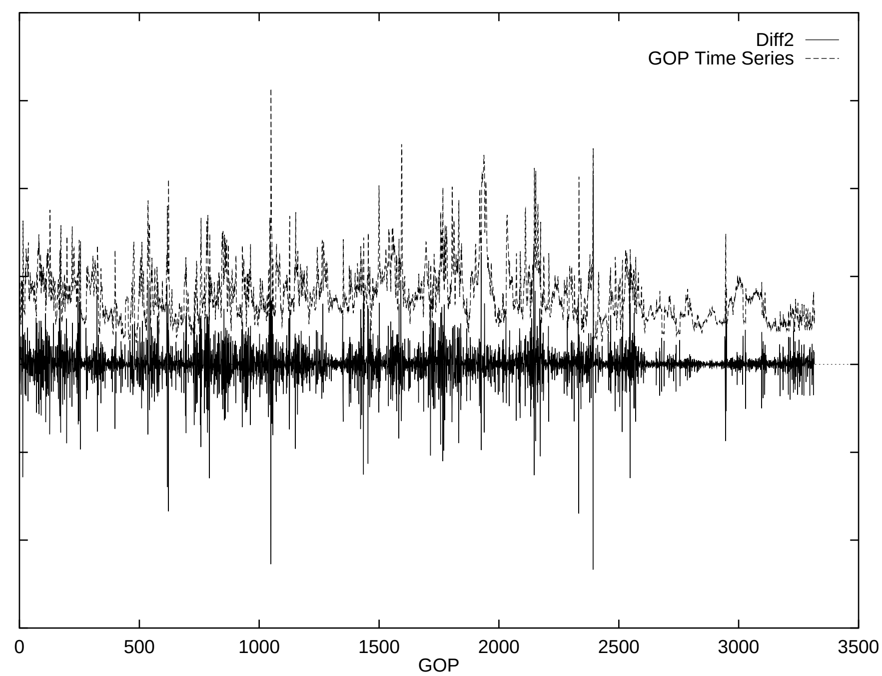


Figure 4.6: The Time Series Plot of GOP Associated With Dif_2

4.3 The Scene Change Scale (SCS)

In the last two sections, two methods were offered in order to provide indications on the fluctuations in the overall bit rate of an MPEG video stream. These indications could vary from one stream to another due to the amount of activities within the stream. In addition, the scene change identification process analyses the magnitude of each scene change. Thus, this type of information plays a crucial role in achieving an efficient modelling as well as management of MPEG traffic.

In [Kara97] and [Mashat98a], we showed that the amount of activities within an MPEG stream affects not only the traffic model but also the queueing performance at an ATM multiplexer. In order to classify MPEG streams according to the amount of activity and movement within the stream, we map the output of the scene change identification method to a scale which gives a more precise indication on the amount (or level) and the strength of the bit rate variation. This can be done by scanning the entire MPEG stream and thereby, detecting all significant scene changes $\{Sc_k, k > 0\}$ within the MPEG stream. We can then analyse and quantify their magnitude. By scaling the scene changes $\{Sc_k\}$ with the entire sequence, it is possible to calculate a *ration* (or scale) of the activities. According to the second difference method, the scale for an MPEG sequence can be found via the following equation:

$$SCS = \frac{\sum_{k=1}^{n} Sc_k}{\sum_{i=1}^{N-1} (X_{i+1} - X_i) - (X_i - X_{i-1})}$$

In addition, by using the outlier method, it is possible to derive the scale through the following:

$$SCS = \frac{\sum_{k=1}^{n} Sc_k}{\sum_{i=0}^{N-1} Y_i}$$

where n is the number of scene changes.

By using the last two equations, we are able to obtain and assess the level of activities within the sequence. Either of the two scene change identification methods can be used, as both give close results. However, the second difference method gives more accurate results in terms of identifying only the significant scene changes (as mentioned previously). Thus, the second difference method will be used for the purposes of our classification.

Based on the second difference method, 21 various traced MPEG streams have been tested and scanned using our method in order to define and classify the amount of activity for each stream. For a given *threshold* value, these streams are scaled to be presented in a '*Scene Change Scale*' (*SCS*), with a range from 0 to 1 (*i.e.* $0 < SCS < 1$). If the amount of activity within the stream is limited, then the stream will be allocated nearer to 0. Conversely, if the stream is highly active, then it will be allocated nearer to 1. Figure 4.7 shows these streams on the *SCS* with two different values for the *threshold*. It is important to note that with a lower *threshold* value, the strength value of the stream on the *SCS* will be increased.

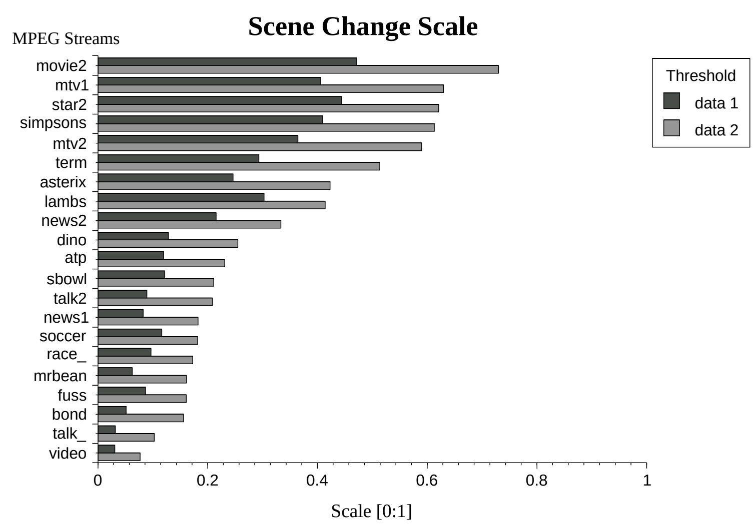


Figure 4.7: Scene Change Scale

4.4 Experimental Evaluation

In the ATM traffic management context, it is common to test the QoS performance at a statistical multiplexer, and to efficiently allocate the buffer size and bandwidth resources [Krunz96]. This section describes several simulation experiments, and presents the simulation results when multiple MPEG streams, with various scene activities, are multiplexed. The main objective of these experiments is to demonstrate the impact of scene changes activities on QoS requirements, and then relate these results to the *SCS*. This can be achieved by simulating the multiplexing of various streams, with different *SCS* values.

We have simulated the transmission of various video connections on an ATM multiplexer with a single link, and a buffer whose size B was determined by the delay constraints (D) on data transmissions out of the multiplexer: $B = D \cdot C$, where C is the link speed. In other words, the maximum queue length is bounded by the link speed and delay constraints [Zhang94]. In our simulation, the cells arrive at the multiplexer from a number of real video MPEG connections (based on the empirical data sets). Each connection generates a frame consisting of a variable number of cells (see Figure 4.8). For our sequence, the connection rate is 24 frames/sec. The FIFO service discipline policy is employed at the multiplexer. For each experiment, the link speed was adjusted to satisfy a system load of 80% (0.8 system utilisation).

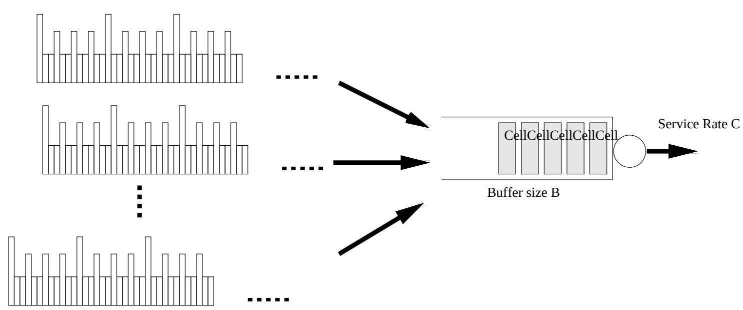


Figure 4.8: Multiplexing of Multiple MPEG Streams

If the system capacity (i.e. the buffer is full) is exceeded, then any incoming arrivals will be lost. Furthermore, larger buffer sizes will increase the waiting time for arrivals to be served. Therefore, a trade-off between the delay and cell loss requirements should be achieved. However, the cell losses in most cases are very important, because standard coding schemes (such as MPEG-I, H.261) are not designed for the compression of video, which are transmitted on a medium where a loss of data is possible [Rose94b]. Thus, our primary measure of interest is the CLR. However, the multiplexer may implement a particular frame (packet) discard policy, called *Pushout*, where in the event of one or more cell losses the whole frame (or packet) of which the lost cells are part of is dropped. Studies have shown that such a policy improves both the throughput performance and network efficiency [Romanow94] [Manthorpe96].

We have studied the case when multiple sources, at GOP process, are multiplexed into an ATM multiplexer with one server (link). We also assume that GOP sizes presents the same activity of the real trace at frame sizes [Chiotis97].

For the sake of explanation, results of three chosen traces from the real MPEG sequences are presented, namely: the 'Movie', 'Dino' and 'Talk' sequence. In order to obtain consistent results, we have selected only one *Threshold* value (-0.5) throughout the experiments for the calculation of the *SCS*. According to the *SCS*, the three chosen traces represent three various classes of VBR MPEG streams: high, moderate and low amounts of activity. It can be seen from Figure 4.9, for all buffer sizes studied, the losses resulting from the use of these sequences are compared. It is clear that the 'Movie' sequence produced the highest CLR while the 'Talk' sequence produced the lowest. For small buffers, the CLR is quite high. As the buffer increases, the CLR values start to decrease slightly. The delay bound (D) gives a similar performance, because the buffer size is a function of the delay constraints. Furthermore, it is important to note that the 'Movie' sequence has the highest value on the *SCS*. We have tested the same simulation using other various MPEG sequences. We have found that there is a strong positive correlation (about +0.73) between the CLR obtained from the sequence performance and its associated position on the *SCS*. The correlation has been calculated using the correlation coefficient factor as follows:

$$Corr(SCS, CLR) = \frac{Cov(SCS,CLR)}{\sigma_{SCS}\,\sigma_{CLR}}\ ,\ -1 \leq Corr(SCS, CLR) \leq +1$$

$$Cov(SCS, CLR) = E[(SCS - \mu_{SCS})(CLR - \mu_{CLR})]$$

In order to demonstrate the multiplexing behaviour for different system loads, we have compared the losses resulting from various loads. In other words, the service rate for the output link is adjusted to obtain different levels of utilisation (*U*) or system load. The system load can be defined as the ratio of the arrival rate to the service rate [Pitts96]. Figure 4.10 shows the performance of the 'Movie' sequence at different loads, namely U=50%, U=60% and U=70%. It is clear that the higher the load, the greater the value of CLR.

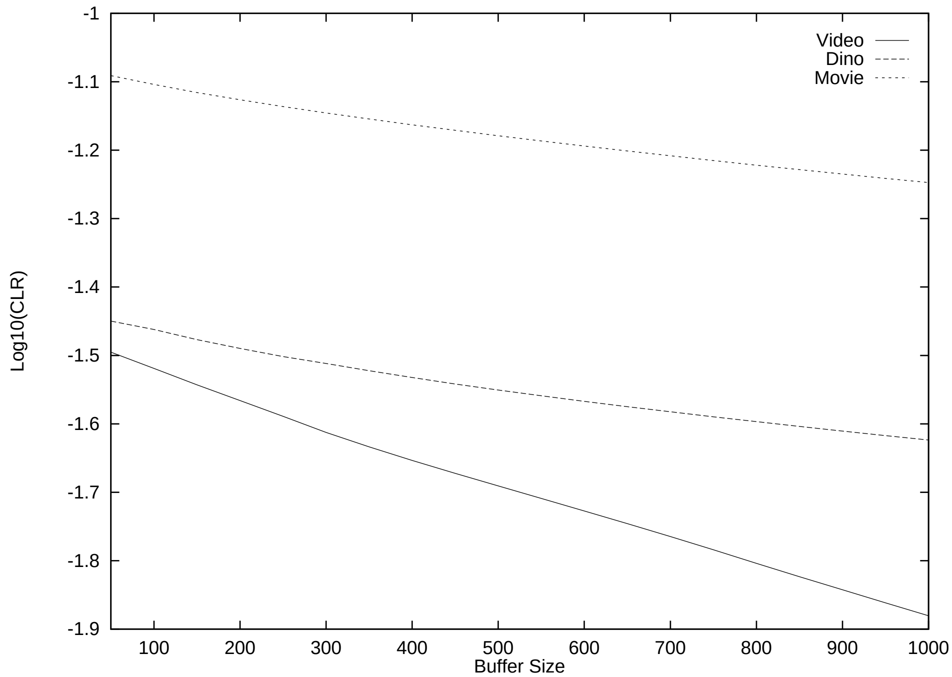


Figure 4.9: CLR for Multiplexed MPEG Streams

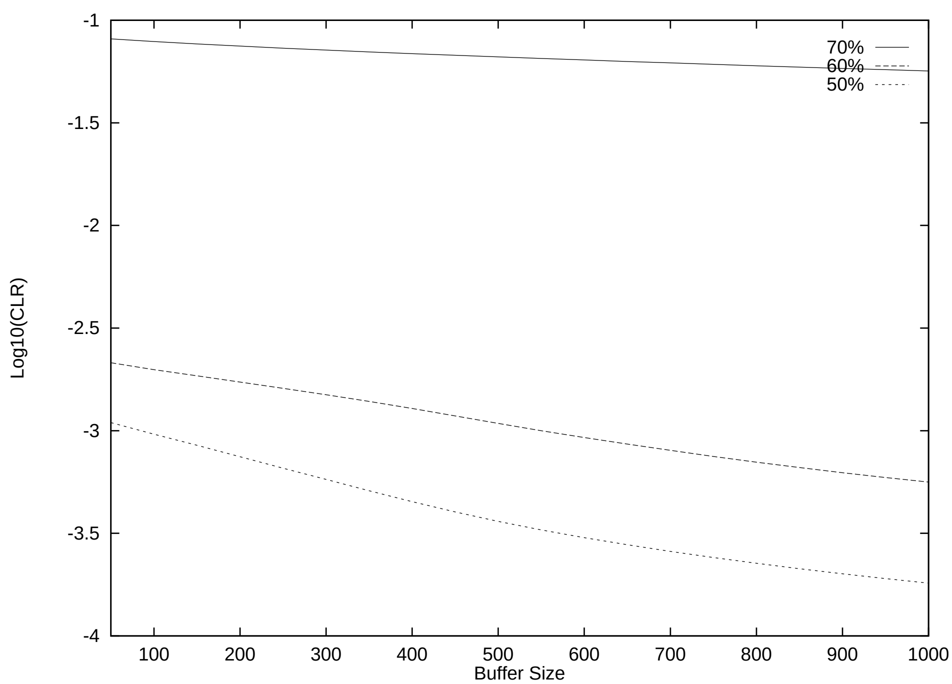


Figure 4.10: CLR For Multiplexed MPEG Streams at Different Levels of Utilisation

4.5 Summary

Because the scene change is the most important factor which affects not only the complexity of traffic characteristics, but also the long range dependency feature of MPEG streams, we need to analyse the magnitude of scene changes which cause the bit rate variations of an MPEG stream. In order to explore the fluctuations in the overall bit rate for an MPEG stream, we have introduced two simple methods and algorithms to identify the scene changes within an MPEG stream. As a result, we have mapped the amount of scene changes onto a Scene Change Scale (*SCS*) which can be used to exhibit the amount of activity within the overall MPEG stream. We have also explored the impact of scene changes on QoS requirements. The primary measure of interest was the CLR. We have related the CLR results obtained from several simulations with the SCS, and have found that there is a strong positive correlation between the CLR results and the Scale. In other words, SCS can be used as an assessment of QoS guarantees over ATM networks. For instance, a high CLR is associated with an MPEG stream with a high value on the scale. Therefore, more care should be applied in handling such a stream having a high value on the scale.

Chapter 5

VBR MPEG Statistical Modelling

5.1 Introduction

In order to achieve an accurate and effective evaluation of the performance of an ATM network, and to provide worthwhile guidance for the design of traffic management and control schemes, we need to have a good knowledge of various traffic sources. Generally speaking, there are two possible ways to achieve this aim [Ni96], by direct observation of a real video trace, or by constructing a mathematical model for the video source. The first option is quite simple. However, despite its simplicity, it is difficult to be formulated and applied to a relevant analysis. The second option helps to characterise the real video source more precisely. In addition, it produces effective and efficient mathematically analytical techniques.

An encoded video, as shown in the previous chapters, is not independent traffic. Thus, simple traffic model (such as Poisson) is not adequate to model video traffic. Video source models (e.g. MPEG traffic source models) play a variety of roles, including:

- The model can be used to identify effective sets of traffic descriptors for QoS parameters at call set-up. These descriptors are used to describe the traffic behaviour through the entire call (or connection);

- In order to test and compare different control schemes (for instance, end-to-end rate control), a source model can be used to exercise the degree to which QoS guarantees are provided [Rampal95]. In addition, the source model can be used to evaluate alternative policing mechanisms, such as the D-BIND traffic descriptors proposed in [Knightly96]; and

- It can also be used to predict the level of QoS that a particular application might experience at different levels of network congestion.

This chapter is organised as follows: we first address the main modelling approaches for VBR video traffic, including a Markovian based. Two simple source models for VBR MPEG, the Histogram-based and Detailed Markov chain model, are presented. Due to the significant role of scene changes, we will perform a further analysis on the scene changes. We present an extensive scene-based model and its performance. The models will be used to generate a synthetic workload representing VBR MPEG traffic. In order to validate these models, the statistical behaviour of the model as well as the queueing performance are compared with the original traffic. In turn, some experiments were performed to study the model's performance at an ATM multiplexer.

5.2 Modelling Approaches

A series of MPEG source models have been proposed in the literature, reflecting increasing insight into the nature and variety of MPEG source dynamics. Typically, there are three approaches that can be used to model MPEG video traffic [Izquierdo96]:

- **Markov chains** [Daigle86] [Chu95]: Video traffic can be approximated by a two-state Markov chain, one state represents the peak rate, while the other represents the minimum rate. A simulation experiment showed that the two-state Markov chain model is not accurate enough for statistical studies [Heyman92]. However, a detailed Markov chain model, with more finite states, provides a sufficient level of accuracy to be useful in traffic studies [Chu95].

- **A video model based on Autoregressive processes** [Maglaris88] [Doulamis96]: This type of model matches the features of video traffic: the distribution and the autocorrelation function. An autoregressive model of order 2 fits the data well in a statistical sense, but it does not produce enough large values to be useful for traffic studies.

- **Self-similar or fractal model**s [Izquierdo96] [Yao97]: These models are based on a self-similar process. Such a process is called self-similar if the samples for that process appear to be 'similar', regardless of the duration of the sampling interval. One of the most important characteristics of this process is long range dependency.

There are many proposed models for video traffic. Before we select a model for VBR video traffic, it is necessary to examine the attributes of the video source. Another matter requiring consideration is the purpose of the traffic model, which should be defined to determine the proper model. In this regards, many factors should be considered, including the following [Stamoulis94] [Izquierdo96]:

- The traffic (source) activities in terms of the amount of scene changes. If the traffic sequence contains significant scene changes, then a hierarchical model is required;

- The encoding scheme. For example, if the encoded sequence contains different frame types, then the traffic model should consider the attributes of each frame; and

- The selected level of modelling within the video sequence. It is possible to create a model for MPEG at different levels (GOP, frame, slide or even cell level). Having decided on the level, the statistical properties of that level should be defined. Furthermore, we need to lay down the way the levels depend on each other. For instance, if we want to use the cell level as the modelling level, we need to decide how frames are broken into cells. However, this depends on the considered ATM AALs, and on the existence of the shaping facilities between the video source and ATM network [Roberts96] .

5.3 Statistical Modelling of MPEG Using MC

This section describes the statistical models which are used to characterise an MPEG sequence. The main objective is to find a suitable and simple model to capture the statistical behaviour of VBR MPEG sequences. The model will be used to generate a synthetic workload representing VBR MPEG traffic. We describe two Markovian based models, namely Histogram-based and Detailed Markov Chain (DMC). These models are based on the results of the statistical analysis which were obtained in the previous chapters. The models will be used to approximate the statistical behaviour of the MPEG sequence. Both models are well known. However, with a degree of modification and the consideration of the scene change, it is possible to be tuned for a particular type of VBR MPEG sequencing in order to improve the actual traffic approximation. Another improvement can be added to these models is when they are used with the generation process or generation process (see section 5.5.1) to perform multiple levels of correlations (GOP-by-GOP and frame-by-frame).

The Markov chain method can be used to model different layers of an MPEG sequence (scene, GOP, frame or slide). It has been also used to model one layer codec stream [Murphy94]. It is very difficult to find a model that covers all layers [Rose95b]. Therefore, we have to decide which layer will be used. A higher layer will add more complexity to the model, but it will also improve the long-range dependency behaviour. The GOP layer can be used for our modelling without the need for modelling the frame-by-frame correlation and the only correlation used is the GOP-by-GOP (frame-by-frame correlation will be employed at the traffic generation process). In addition, an experimental result showed that frame-by-frame correlation has no influence on cell loss results [Rose95b]. Therefore, in some cases, it is adequate to use only one level of correlation. In addition, the GOP plays the most crucial role concerning the autocorrelation effects of an MPEG video sequence, due to the periodic nature of the MPEG sequence (which is caused by the GOP structure).

For the Histogram model, an 0-order Markov chain method has been used and an 1st-order Markov chain for the DMC. Both models have a finite number of states and will be used to generate a GOP size process. The range of GOP sizes of the empirical MPEG sequence will be divided into several quantization intervals. Each interval is related to a state of the Markov chain. Therefore, the number of states is equal to the number of GOP intervals. For each state, there is a mean value of the GOP interval associated with it. In the Markov chain model, the transition from one state to another is controlled by a transition matrix. With each state transition (entrance from its current state into the next state) a GOP size will be generated according to the mean value of the next state.

5.3.1 Histogram Model

The Histogram model can be described by a simple Markov chain with a finite number of states (M) which is equal to the number of the quantization intervals (see Figure 5.1). The transition from one state to another is completely independent, and the transition matrix of size $1 \times M$ for the Histogram model is defined as:

$$P_{ij} = \frac{Number\, of\, GOPs\, within\, interval\, i}{Total\, number\, of\, GOPs}$$

The Histogram model can be used to estimate the distribution of the empirical GOP size. However, this model does not approximate the GOP correlation. This is because each GOP sample is generated according to the histogram bins which are independent from each other. We found from our experimental work that there is almost no correlation between the generated GOPs (see Figure 5.3). Therefore a model which is based on the distribution function only cannot approximate the dependencies behaviour of the MPEG sequence. However, a probability density function (pdf) can be used in other models such as the Discrete Autoregressive (DAR) model [Doulamis96]. The DAR model requires to know the pdf of the source. It is very important to notice that the Histogram model can be useful when there is no agreement on the fitted distribution. This is the case for VBR video traffic, due to QoS requirements for the encoded scheme.

5.3.2 Detailed Markov Chain Model (DMC)

The Markov chain process has been used because its parameters can be identified easily and it can also be easily analysed. This is helpful in finding the most appropriate model. For long range dependency sequences the Markov chain is not adequate because its autocorrelation function decays exponentially. However, with some effort, it is possible to obtain a Markov chain model with a high coefficient of autocorrelation even for large lags. This can be done by increasing the number of states to a reasonable number (see Figure 5.2).

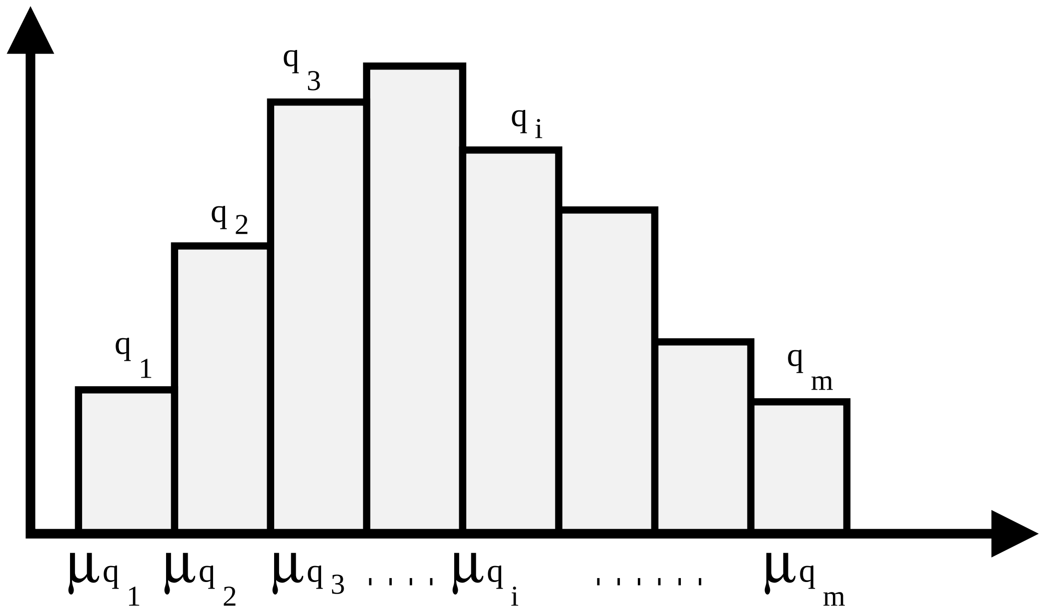


Figure 5.1: Histogram Model

The number of states (M) is derived by dividing the maximum value of the GOP size (GOP_{max}) by the standard deviation of GOP sizes ($STDEV_{GOP}$) [Rose95b] where $\lambda_i, i \geq 1$ is the associated bit rate. Therefore the quantization intervals of this model are dependent on the standard deviation value. It is possible to increase the number of states by dividing GOP_{max} by a smaller value than $STDEV_{GOP}$. The transition matrix of size $M \times M$ can be found as follows:

$$P_{ij} = \frac{Number\, of\, transition\, from\, state\, i\, to\, state\, j}{Total\, transition\, out\, from\, interval\, i}$$

It is important to note that the current state, i , depends on the previous state and not on neighbouring ones. Therefore, it is not necessary that the next state be a neighbouring state.

5.3.3 Model Validation (Markovian-based Models)

To examine both the appropriateness and the limitations of the presented Markovian-based models, we need to know whether or not the models are able to approximate the behaviour of the real MPEG sequence. This can be achieved by comparing the behaviour of the model and the original empirical data in terms of the statistical distribution and the sequence correlation

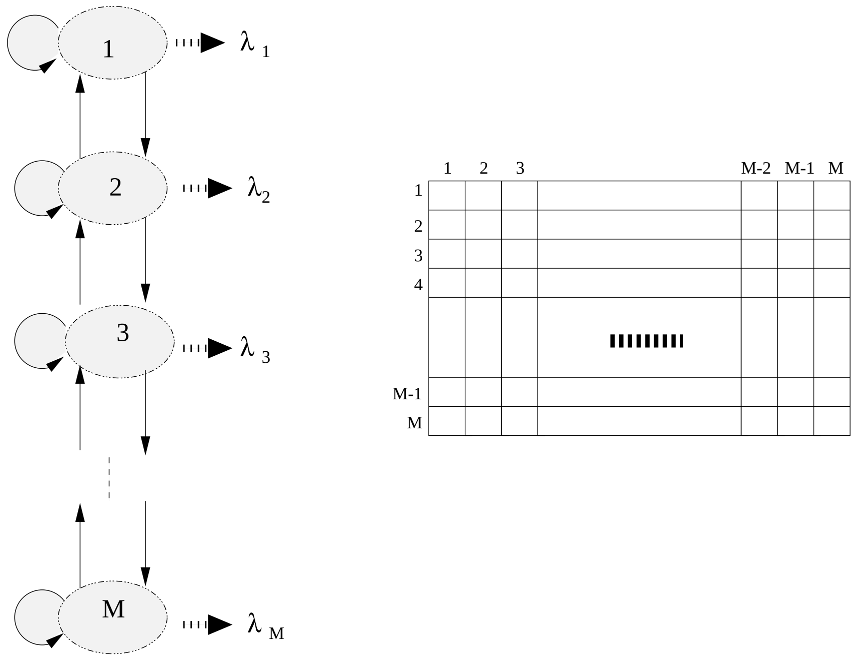


Figure 5.2: The Detailed Markov Chain Model (DMC)

[Izquierdo96]. For the GOP level, we first compare the distribution parameters for both models. Tables 5.1, 5.2 and 5.3 show that the parameters for both models have values that are close to those of the empirical data. However, the simulation results indicate that the DMC model is a good source model for approximating a sequence with a short range dependency feature even for large lags (about 15), as in the case for the 'Race' and 'video conference' sequences. This can be shown when we compare the curve of the autocorrelation function for the model with real traffic (see Figure 5.4 and 5.5). However, the DMC model shows better approximating for the 'Race' sequence than the 'video conference' sequence. In contrast, we found that there is almost no correlation between the generated GOPs in the case of the Histogram model (see Figure 5.3). Thus, the Histogram model does not approximate the autocorrelation function, because each GOP sample is based on the histogram bins which are independent from each other. However, the DMC model captures only few lags (about 5) for a long range dependency sequence (see Figure 5.3).

Data Sets	Mean(μ) Cells	STDEV(σ) Cells	Min Cells	Max (Peak) Cells
Real Sequence	408	164	129	1634
Histogram Model	410	160	173	1270
DMC Model	402	162	152	1634

Table 5.1: The simple statistical parameters for 'Dino' sequence and the models

Data Sets	Mean(μ) Cells	STDEV(σ) Cells	Min Cells	Max(Peak) Cells
Real Sequence	188	32	144	373
DMC Model	186	29	162	328

Table 5.2: The simple statistical parameters for the ''video conference'' sequence and the model

Data Sets	Mean Cells	STDEV Cells	Min Cells	Max Cells
Real Sequence	961	362	292	3470
DMC Model	969	350	538	3019

Table 5.3: The simple statistical parameters for the 'Race' sequence and the model

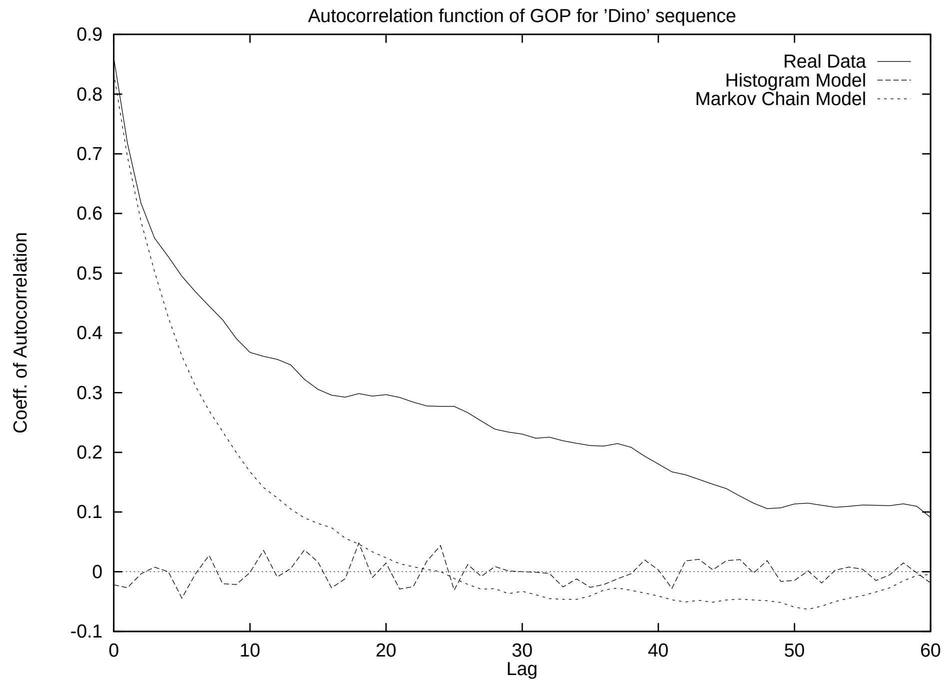


Figure 5.3: Autocorrelation Function Comparison (Dino)

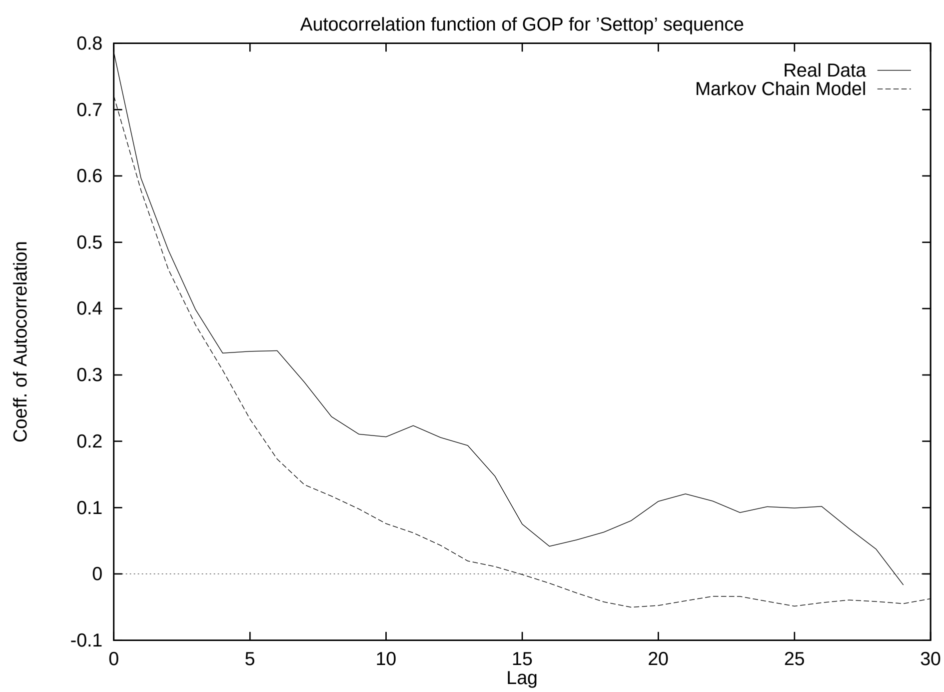


Figure 5.4: Autocorrelation Function Comparison (Settop)

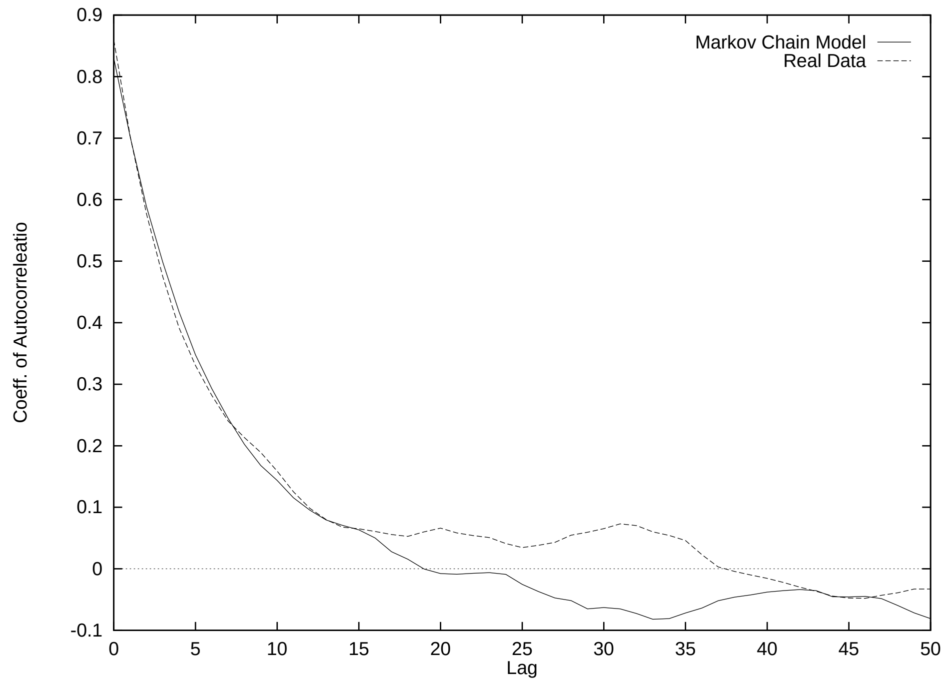


Figure 5.5: Autocorrelation Function Comparison (Race)

5.4 Scene Change Based Model for VBR MPEG Traffic

Various models have been proposed for VBR MPEG traffic [Kara97] [Rose95b] [Stamoulis94] [Heyman92] and [Frost94], but there are few models that incorporate scene changes [Krunz96] [Lazar93] and [Heyman96]. As shown in the previous section, the Markovian models were not adequate to capture the long range dependency feature. In this section, a new form of scene-based model is introduced, based on characterising MPEG traffic as a collection of scenes. The model is used to capture the long range dependency feature of MPEG traffic.

As shown in the last chapter, the main reason for the overall bit rate variation is scene change, which leads to dramatic increases in the queue length statistics of the multiplexer . Consequently, scene changes should be incorporated in such traffic modelling approach. In order to model VBR MPEG traffic based on the scene change, we need a functional definition of scene duration based on the bit rate variations. From chapter 4, we are able to identify, and quantify, scene changes within an MPEG sequence. As illustrated in Figure

5.6, an MPEG stream can be characterised using two processes, namely: scene length $\{Sl_i, i > 0\}$ and the scene fluctuations process $\{Sc_i, i > 0\}$.

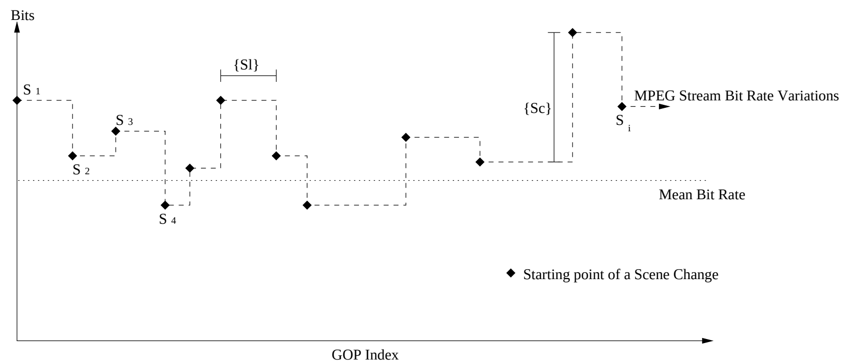


Figure 5.6: Scene Change Model

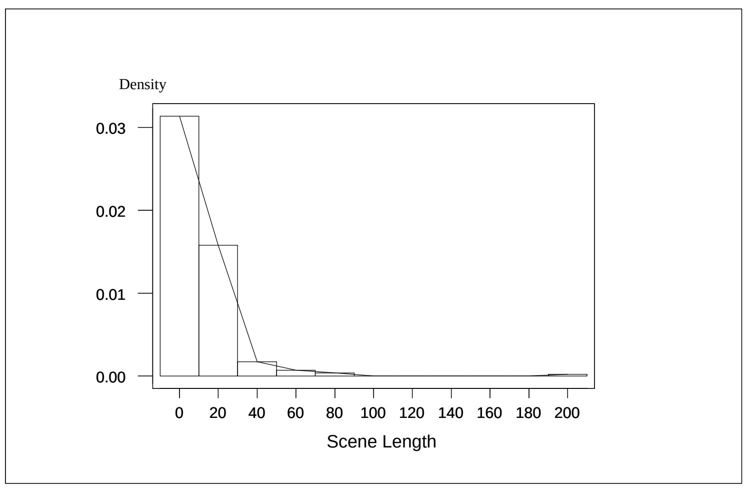


Figure 5.7: The Fitted Curve for Scene Length Distribution

The first process $\{Sl_i,\ i = 1, 2, 3, ..., n\}$ can be defined as the scene length duration within the MPEG sequence, where n is the number of scene changes. In other words, Sl_i is the number of GOPs in the i-th scene. The scene length can be easily calculated (as showed in chapter 4) by adding the number of GOPs between the starting points of two consecutive scenes. The statistical characteristics of the scene length were examined. Then, the fitted curve for the histogram of the scene change length process was drawn (see Figure 5.7). The figure shows that the scene length can be modelled by a geometric distribution. The shape of the probability density function (*pdf*) in the figure was also observed on all other analysed MPEG sequences. In [Lazar93], it has also been confirmed that scene length for VBR video traffic can be modelled as a geometric distribution. Therefore, we use the *pdf* with a parameter q for this distribution as the basis for our model. The scene length process can be described as follows:

$$Pr\{Sl_i = n\} = q^{n-1}p\,, \;\; p = 1 - q \quad for\ n = 1, 2, 3, ...$$

$$E[Sl_i] = \frac{1}{p}$$

Sequence	Number of Scenes	Mean (GOP)
Dino	569	11.27
Race	579	13.41
Movie	670	5.3
Video Conference	65	20.2

Table 5.4: Statistical Parameters of Scene Change Durations

The average size of the scene changes can be used as another indicator of the video stream activity. Table 5.4 presents simple statistical parameters of the scene change duration for various MPEG sequences. For our analysed trace (Dino), the average of the scene length is about 11 GOPs (5 seconds). In addition, we calculated the autocorrelation function (ACF) for scene lengths. Figure 5.8 shows that the shape of the ACF for $\{Sl\}$ alternates very closely on either sides around the 0-line. In other words, there is a very weak presence of correlations among scene lengths (we could say it is uncorrelated). Therefore, the main issue is characterising only the distribution of the scene change. Consequently, the scene lengths constitute a sequence of *idd* random variables with a geometric distribution.

The second process $\{Sc_i\}$ is based on the fact that a significant difference between two consecutive GOPs is an indication of a scene change. Hence, we modelled the GOP variation using $\{Sc_i\}$ and the *mean* value of GOPs for the previous scene ($\widetilde{gop}_k\,,\ k > 0$). From our observation of the $\{Sc\}$ histogram shape (see Figure 5.9), we have found that the GOP variation could possibly be described using one of the three distributions (or even a Normal distribution) which are presented in section 3.4.1. Furthermore, the autocorrelation function of scene variations process exhibits weak correlations a part from the first few lags (see Figure 5.10). Therefore, the process could be modelled based on the

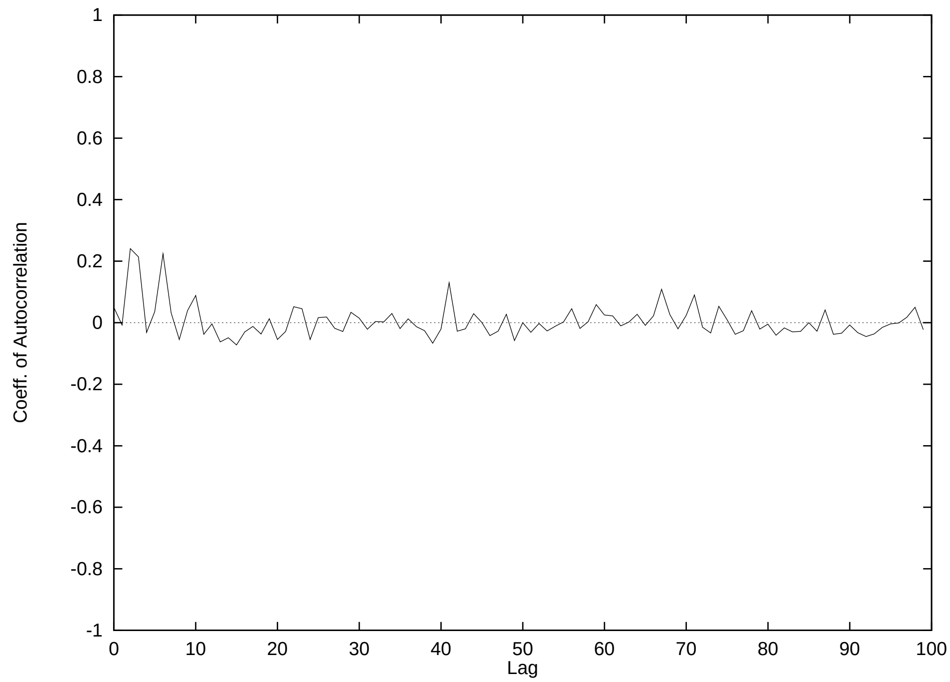


Figure 5.8: Autocorrelation Function of The Scene Length

distribution properties only. However, due to the difficulties in finding the most fitted distribution, $\{Sc\}$ process is modelled using the *Histogram-based* model with a transition matrix $[P]$ and M states (see section 5.3.1 for more details on the *Histogram-based* model). Thus, the i-th GOP size (within the k-th scene) can be found as follows:

$$X_i = \widetilde{gop}_{k-1} + Sc_i \ , \ \ 0 < i < N \ \ and \ k = 1, 2, ..., n$$

where N is the number of GOPs in the sequence and n is the number of scenes within the sequence.

5.4.1 Model Validation (Scene-based Model)

As shown in section 5.3.3, we performed the same method for the model validation in order to examine the appropriateness and the limitations of the scene-based model. A synthetic sequence should be generated to allow us to examine whether the model is able to approximate the behaviour of the real MPEG sequence or not.

The synthetic GOP sequence has been generated using the proposed model.

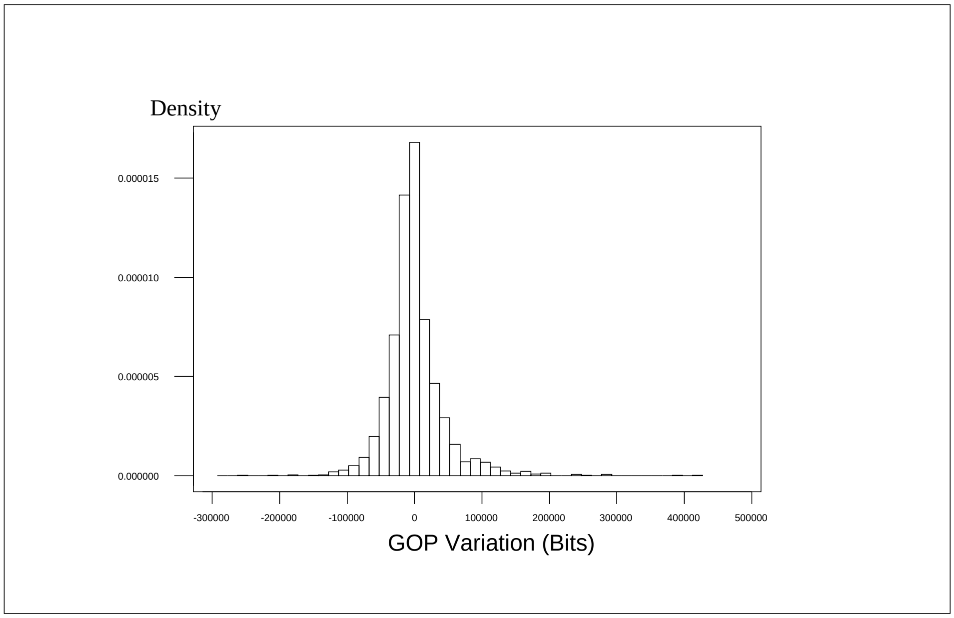


Figure 5.9: The Histogram Graph of the Scene Variations

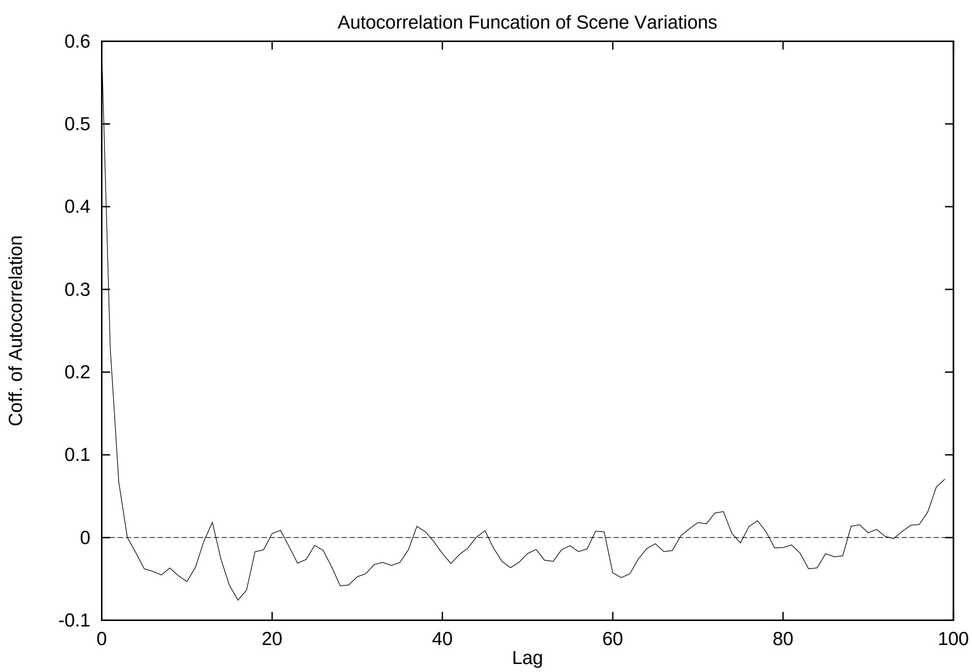


Figure 5.10: ACF of the Scene Variations

The generated sequence was based on the fitting of the original 'Dino' sequence[1]. Firstly, a scene length is determined using a given geometric distribution. The GOPs are then generated with association of the scene variation process. The following algorithm describes the method which generates a GOP sequence of size N :

$START$
$N = number\,of\,scenes$
$Scene\,variation\,\{Sc\} \equiv Histogram(\mu_{q_t}; 1 \leq t \leq M)$
$For\,i = 1\;\; To\;\; N$
  $Get\,Scene_i\;\; length\; = sl$
  $For\,k = 1\;\; To\;\; sl$
    $\{Sc\} = Histogram(\mu_{q_t})$
    $GOP_{i+k} = \widetilde{gop}_{i-1}\; + \; \{Sc\}$
  $End\,Loop\,k$
$End\,Loop\,i$
END

In order to validate the proposed model, we need to compare the most important statistical behaviour of the original sequence, with that of the generated sequence. Table 5.5 shows the important statistical attributes (or parameters) for both sequences. By comparing the statistical behaviour of the model-based sequence with the original sequence, we can see that the model captures most of the main statistical parameters of the original sequence accurately.

Because the long range dependency (LRD) is another important feature of an MPEG sequence which can be described by the ACF [Izquierdo96]. Thus, we can examine the ACF of the generated sequence and the original one (see Figure 5.11). By comparing both curves, it is clear that the model captures the ACF of the actual sequence (the curves follow the same pattern for lags over 60). However, ACF values (but only for the first lags) in the model are slightly less than the ACF values in the original sequence, mainly because of simplification in the model. Apart from that, the model provides a good approximation of the LRD feature.

[1]The Dino sequence has been selected due to its long range dependency feature.

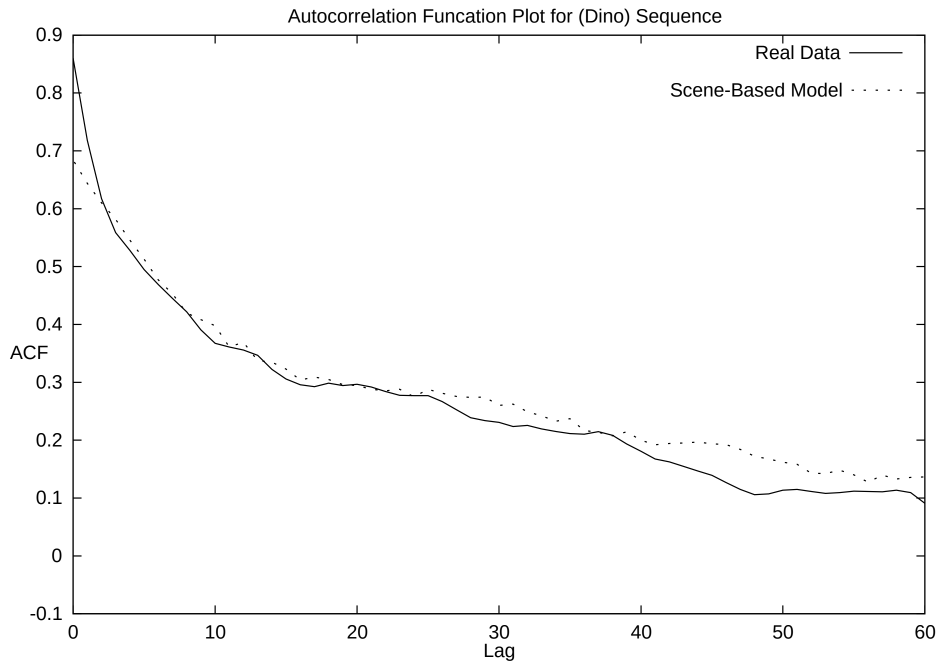


Figure 5.11: ACF for Real and Model-based Sequence (Dino Stream)

Video Sequence	Mean (μ) Cell/GOP	CoV($\frac{\sigma}{\mu}$)	Stdev(σ) Cell/GOP	Peak Cell/GOP	Peak/Mean
Real Sequence	408	0.40	164	1634	4.001
Model-Based	408	0.42	173	1845	4.5

Table 5.5: Simple Statistical Parameters For The Real and Model-Based Sequences (GOP-Based)
Where : CoV is the Coefficient of Variation$\frac{\sigma}{\mu}$ And Stdev is Standard deviation σ.

In order to add more accuracy to the model validation, simulation experiments were conducted to study the performance of an ATM multiplexer using the generated (or synthetic) sequence based on the scene-based model and the original traffic This will help us to examine the performance of the model in terms of the queueing performance. The simulation results obtained from the performance of the model should then be compared with that obtained from the original sequence. An ATM simulator called YATS has been used to perform our experiments [Baumann97]. YATS is a small discrete-time simulation tool tailored for investigations of ATM networks. It expects the name of an input file, which contains a description of the simulation configuration and commands to the simulation kernel and to the simulation objects. The experiment (simulation) model can be described as an ATM multiplexer, with a finite buffer size that accepts ATM cells from MPEG sources, and then transmits them through a link speed C (i.e. single server). First, the MPEG frame sizes are packetised into ATM cells with a payload of 48 bytes. We have used AAL5 for the transmission of MPEG stream because the ATM Forum recommends using AAL5 as the preferred transport protocol [Gringeri98]. The cell stream is assumed to be suitably spaced during a frame duration. The spacing between the cells within a frame duration in turn should vary based on the frame bit rate. However, the spacing within each frame duration is the same. At the multiplexer, the incoming cells are served based on the FIFO manner. A cell loss occurs when the multiplexer buffer becomes full. Incoming cells which arrive during such a buffer-full condition are discarded.

The performance of the simulation system was studied for different buffer sizes at 0.8 system load (system utilisation). Since networks under congestion are the subject of interest, a sufficient load is needed to result in congestion. Figure 5.12 depicts the cell loss ratio against different buffer sizes; it was measured in accordance with the simulation of the generated and original sequences based on the 'Dino' stream. For most buffer sizes, it was observed that the generated sequence shows a good agreement of the losses curve especially with a moderate buffer sizes (300-400 cells). In other words, the model approximates the queueing performance of the original MPEG traffic.

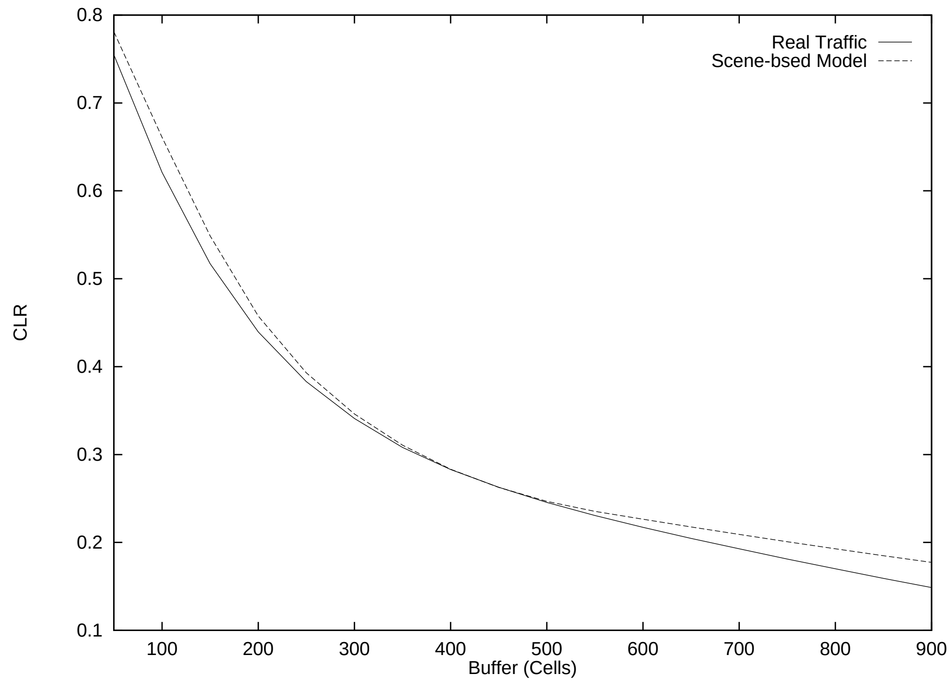


Figure 5.12: Comparison of CLR for 10 Multiplexed Stream (Dino)

5.5 The Overall Model's Comparison

In this section, we summarise the overall comparison of the three presented MPEG models to examine whether the models are able to approximate the long range dependency behaviour of the real MPEG sequences or not. In other words, we consider the relative strengths and the limitations of the model with respect to several statistical properties and the ability to capture ACF of VBR MPEG traffic. In order to validate a model, the statistical analysis results of the model should be compared with the empirical data in terms of the statistical distribution and the sequence correlation (ACF). The 'Dino' sequence has been selected for our comparison because it has the long range dependency feature.

Table 5.6 shows the comparison of the most important moments of GOP sizes for the models and the original sequence. It is interesting to observe that scene-based model matches exactly the original sequence in terms of the mean value. However, for network dimensioning purposes, it is more convenient to use a model which behaves worse than the real traffic [Rose95b]. This was the case for the scene-based model which overestimates the peak value. Thus, it gives a higher value of burstiness parameter.

Data Sets	Mean(μ) Cells	STDEV(σ) Cells	CoV ($\frac{\sigma}{\mu}$)	Max (Peak) Cells	Peak/Mean Burstiness
Real Sequence (Dino)	408	164	0.401	1634	4.00
Histogram Model	410	166	0.404	1270	3.09
DMC Model	402	162	0.402	1634	4.06
Scene Based Model	408	173	0.420	1845	4.52

Table 5.6: The Simple Statistical Parameters for the 'Dino' Sequence compared to the Models

The main difference between the models is their capability to approximate the autocorrelation function (or dependence feature) of the original MPEG sequence (empirical MPEG data set). Figure 5.13 plots the ACF curves of the generated sequences using the three proposed models at the GOP level. Comparing these curves leads us to a conclusion that the scene-based model with a multiple level of modelling shows a better agreement with the empirical (or original) data sets. The simpler models, on the other hand, are unable to model the correlation feature of the data sets over a longer period of time (lags).

In order to compare the queueing performance of the generated traffic using the presented models, we have simulated the multiplexing of multiple sequences at an ATM multiplexer. Simulation experiments have been conducted to examine the CLR performance based on the three models. We have used a similar simulation model used in section 5.4.1. However, the link speed (output link) was fixed (50 Mbits) to achieve consistent results from the simulation of the different models. The number of stream were selected to adjust the system load at 0.8.

The simulation results obtained from the performance of each model are shown in Figure 5.14. This Figure depicts the cell loss ratio (CLR) against different buffer sizes. The figure shows only the *mean* values for the several simulations which have been conducted in order to achieve a 95% confidence interval. The cell loss ratio results were obtained from the simulation of the generated IPB sequences which were originally generated based on the three presented models (The IPB sequence has been generated from the GOP sequence using the scaling factor process which will be described in 5.5.1). The

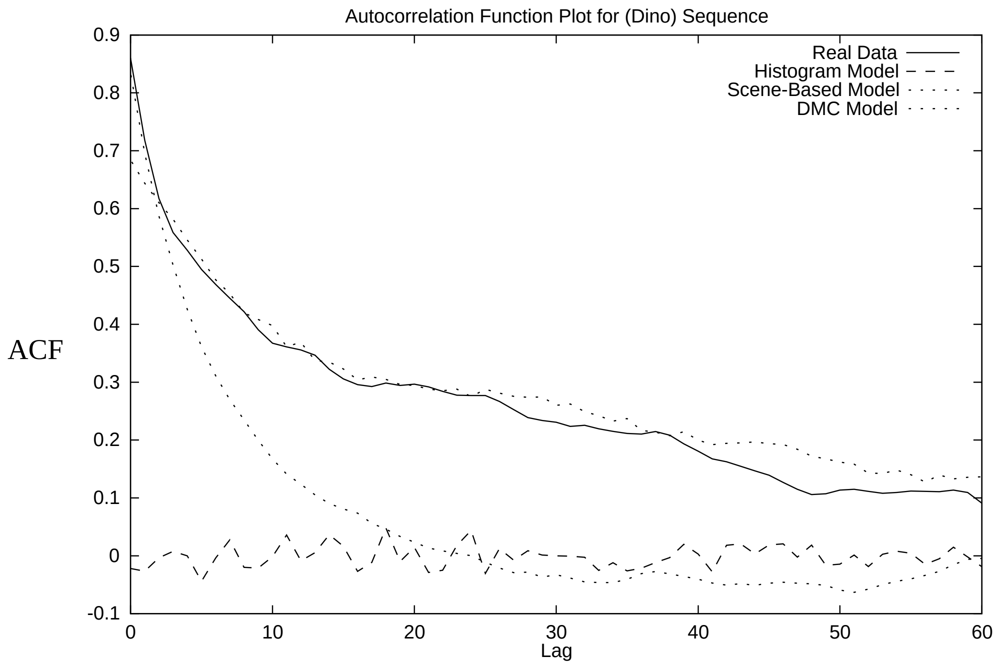


Figure 5.13: ACF for Real and the Three Models for 'Dino' Stream

CLR curve for the original sequence is also plotted based on 'Dino' stream for the sake of overall comparison. For small buffer sizes, it was observed that all generated sequences show close/good agreement of the losses curve of the original sequence. However, for larger buffer sizes, the CLR curve for the scene-based model is performed more closer to the original sequence than others.

As results, we point out that all traffic models have their advantages and disadvantages, and that some care has to be taken as to what type of model is chosen for the performance analysis.

5.5.1 The Scaling Factor Process

This process is used to derive a periodical sequence of frame sizes for different types (I, P and B) from the generated GOP sequence. Figure 5.15 depicts the overall picture of the scaling factor process (generation process). It has been observed that this method gives a reasonable approximation for cell loss probability [Rose95a].

In order to produce a synthetic sequence of MPEG frame sizes (I, P and

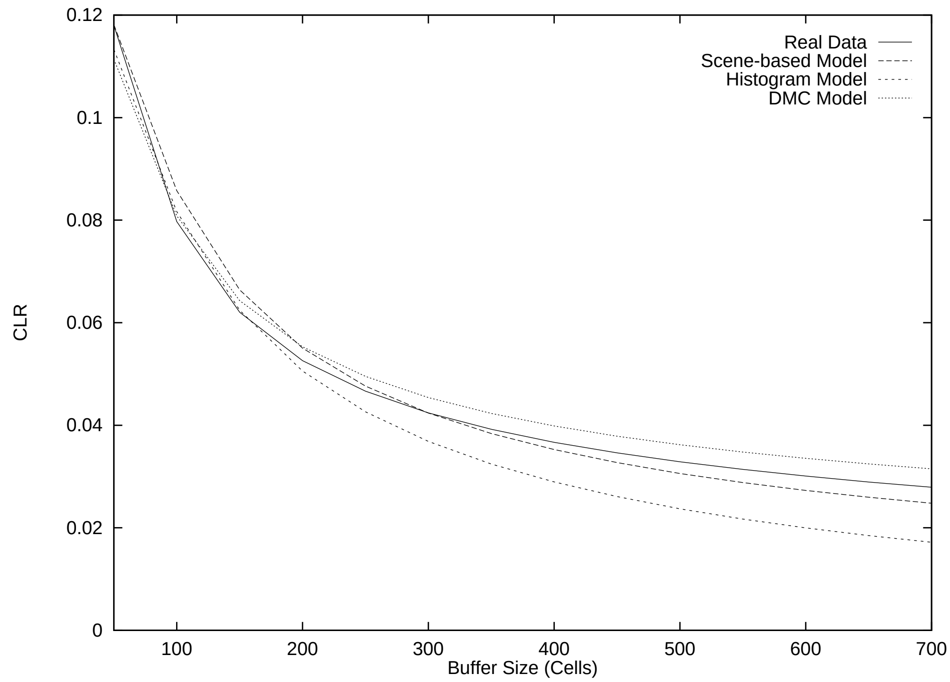


Figure 5.14: CLR Performance Evaluation (Fixed Link Capacity)

B) similar to a real MPEG frame sequence, we use the following method: the frame sizes could be derived from GOP sizes by scaling the frame sizes of different types with GOP sizes. This process uses a scaling parameter called the *Scaling Factor* (f). The parameter, $f_x\, x\ :\ I,\ P\, or\, B$, is calculated by dividing the *mean* value for the frame type by the *mean* value of the GOP:

$$f_I = \frac{E[\{I\}]}{E[\{GOP\}]} \quad f_P = \frac{E[\{P\}]}{E[\{GOP\}]} \quad f_B = \frac{E[\{B\}]}{E[\{GOP\}]}$$

For each generated GOP, a frame of type I, P or B is multiplied by the corresponding f parameter. We have generated frame sequences from the generated GOP sequences (based on the three presented models). Table 5.7 shows two examples of the scaling factors (f) for both the 'Dino' and 'video conference' sequences. It is clear that the scaling factor for B frames (f_B) is the smallest, while the scaling factor for I frames (f_I) is the largest. This is due to the fact that I frames represent most of the GOP size. Another observation that can be seen from the table, is that f_B for the 'video conference' sequence is very small compared to the f_B for the 'Dino' sequence. In contrast, f_I for the 'video conference' is larger (twice as large) than the f_I for the 'Dino' sequence. This is due to the limited amount of activities within the 'video conference' sequence, which makes the I frame sizes very large and the B frames somewhat smaller.

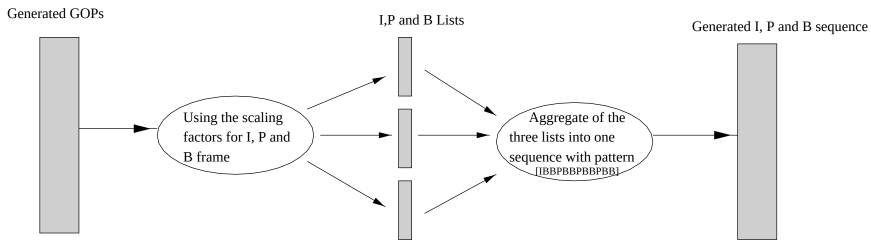


Figure 5.15: The Scaling Factor Process

Previously, the model has been validated at the GOP level. Now, we validate the model at the frame level. Because the dependency feature is exist among MPEG frames (as we discussed in the previous chapters), we have focused on the evaluation of this feature in this section. The correlations have been calculated between the generated sequence and the empirical sequence (original sequence). Table 5.7 depicts a strong correlation in the case of the 'Dino' sequence, with an even stronger one for the 'video conference' sequence (based on the DMC model). The correlations have also been calculated between the three generated sequences based on both Markovian models and the scene-based model, and the original 'Dino' sequence. Table 5.8 shows a strong correlation for both Markovian models, but an even stronger one for the scene-based model. On the other hand, we plotted the autocorrelation function of I, P and B frames for both the generated and the actual data. Figures 5.17 and 5.16 show the ACF of the 'video conference' and the 'Dino' sequences for both: the generated and the original data. The top points represent I frames, the middle points represent P frames, while the bottom points represent B frames. In the case of 'Dino', there is a good agreement between the ACF of the three frame types (I, P and B) for the generated and the original sequences. In contrast, a stronger agreement could be shown in the case of the 'video conference'.

In order to compare the ACF of the three formulated models, Figure 5.18 shows the ACF curves for the actual 'Dino' sequence and the generated sequence using the presented models. In the case of the DMC and the scene-based models, there is a good agreement between the ACF curves of the three frame types (I, P and B) for the generated and the actual sequence. However, a weaker agreement can be shown in the case of the Histogram model (especially

Sequence	Scaling factor for I frame	Scaling factor for P frame	Scaling factor for B frame	Corr(Empirical,Generated)
Dino	0.351	0.092	0.047	0.768
Video Conference	0.611	0.053	0.029	0.977

Table 5.7: Correlation factors between the generated and original sequences

	$f_I = 0.351 \quad f_P = 0.092 \quad f_B = 0.047$
Model	Corr(Empirical,Generated)
Scene-based	0.864
DMC	0.768
Histogram	0.771

Table 5.8: Correlation factor between the generated and actual sequence (Dino)

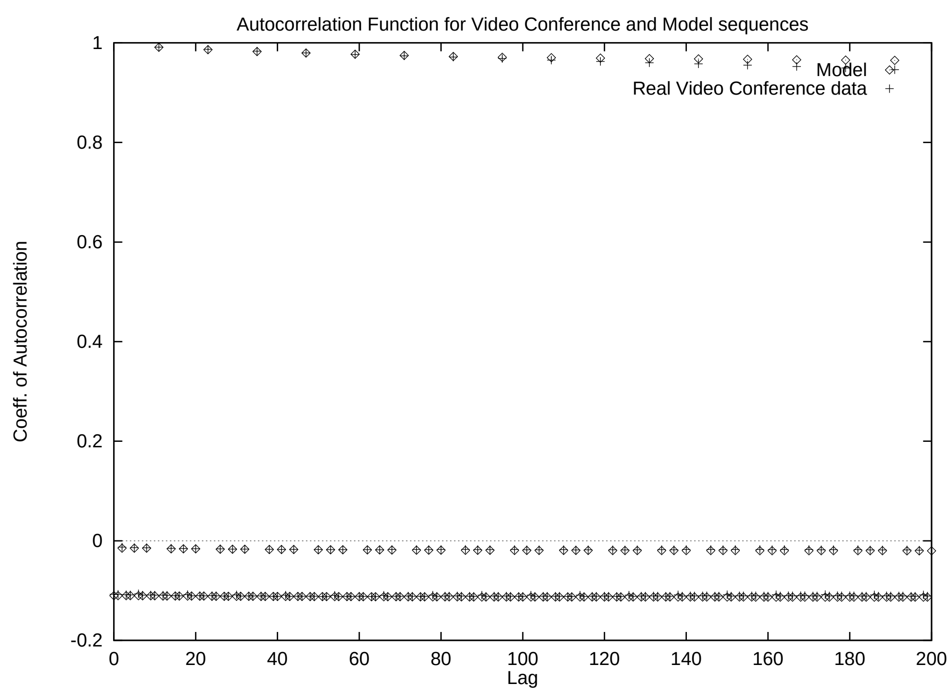


Figure 5.16: Autocorrelation Function Validation (Video Conference)

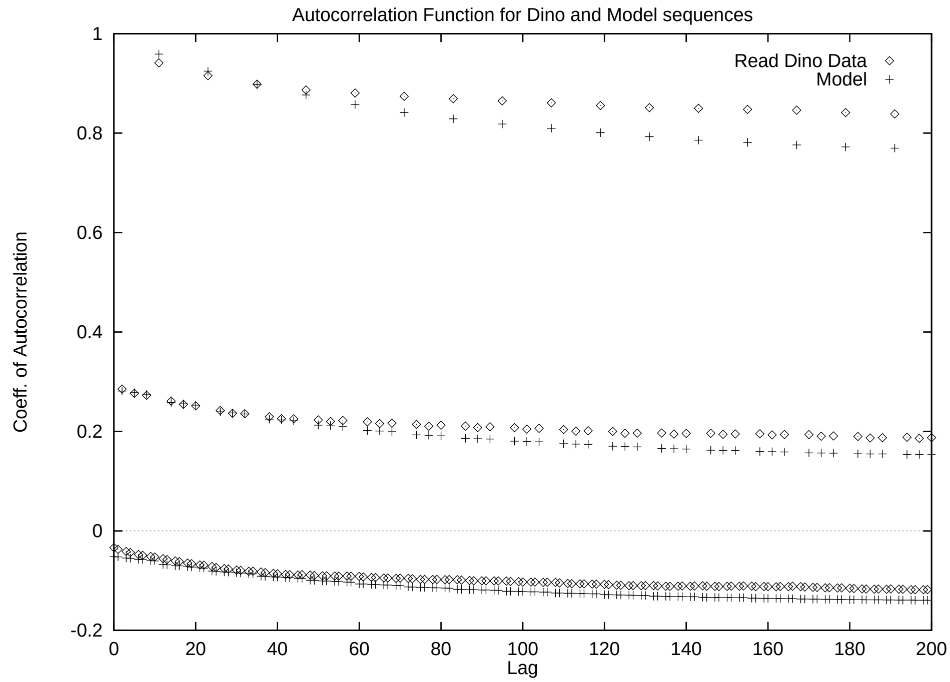


Figure 5.17: Autocorrelation Function Validation (Dino)

in the capturing the ACF of I frames).

5.6 Integrating MPEG Stream into the Workload Model

In order to test and evaluate the performance of a network system or an ATM switch, it is necessary to provide an artificial traffic (as a workload into the system) closely resembling the real traffic. Recording and producing the real traffic can be difficult and expensive because this process requires large storage space and high processing speed [Chu95]. The alternative is to use a traffic source model to imitate the behaviour of the real traffic. In this section, we describe the steps required to integrate the MPEG sources onto a realistic multimedia workload model. In order to do so , we have to use a traffic model that is able to describe the statistical behaviour of the empirical data sets.

The traffic modelling is an important aspect of any network simulation study. Outputs from a simulation model are highly dependent on the inputs provided to the model, and without realistic input workload models, the simulation results are of little value. The traffic characterisation and the analysis

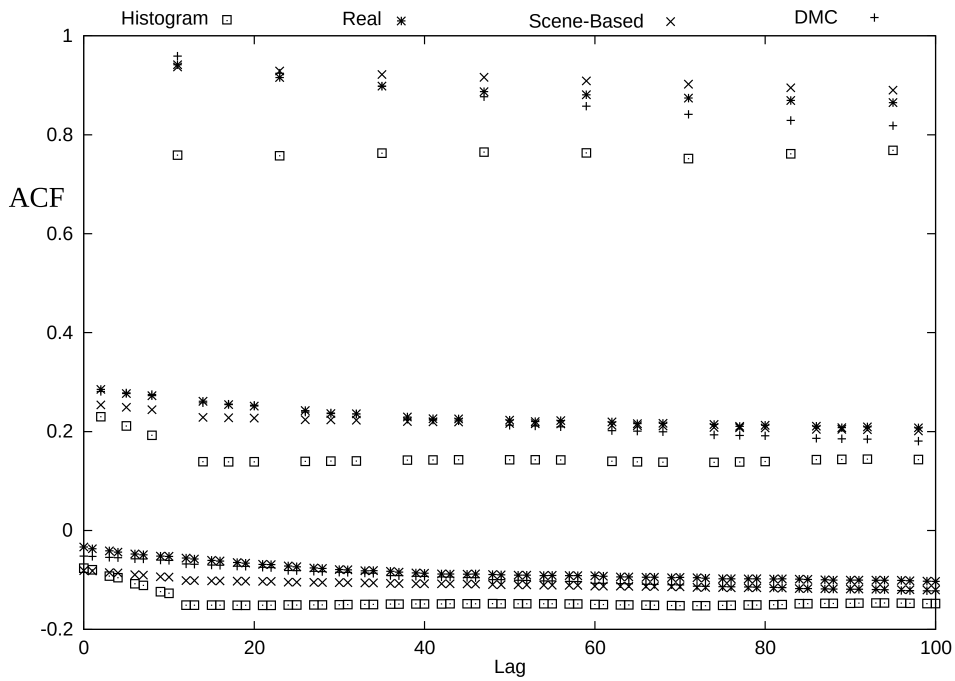


Figure 5.18: ACF for I, P and B Frames (Generated-based and Actual Sequences)

process (which are part of the traffic modelling) can be used to develop a traffic model which captures the main features of that traffic. There are many traffic models which have been constructed to be used within a workload model as representative examples of multimedia traffic. It is important to mention that, as a synthetic workload , we need a simple and efficient model with a small number of parameters. More complex models require more accurate parameters. Therefore, the choice of the models to be used will depend entirely on the performance measurements of the realistic traffic.

The goal is to provide a simple source model for VBR MPEG traffic in order to generate a synthetic traffic for the *application level traffic generator*. Generating a synthetic workload eliminates the need to store voluminous frame-level traces representing MPEG traffic. Such an approach also offers flexibility, tunability and reproducibility in the generated traffic. Figure 5.19 depicts the complete overview of the modelling process, as well as, the generation process which can be described through three main tasks as follows (the first two tasks were discussed in the previous chapters):

1. Analysing empirical data sets for various MPEG sequences to define

the statistical characteristics. This included defining the characteristics of VBR MPEG streams in terms of their statistical behaviour. We used three statistical measures: Distribution, Autocorrelation Function (ACF) and Scene change;

2. We then formulated the statistical results, using a suitable model which captures the statistical behaviour of the empirical data (at GOP level). In order to evaluate the model, we needed to examine whether the model was able to approximate the statistical behaviour of the real MPEG sequences or not. We then compared the simulation results of the model with the real empirical data in order to examine the performance impact; and

3. Lastly, we used the model to generate a synthetic traffic (at frame level) with similar characteristics of the real traffic. This process included an integration of VBR MPEG sources to the workload model. Then, a periodical frame sequence is generated representing an actual video stream. The output of this task is a synthetic traffic which can be used for any performance evaluation study as a system workload

5.6.1 An Example of Generating a Synthetic MPEG Traffic

This section illustrates a full process of generating a synthetic MPEG sequence (or a pattern, including I, P and B frames) based on a traffic model. The main objective is to give (demonstrate) an idea on how a source model can be used in order to provide synthetic traffic representations of a realistic workload. In this section, the Markovian based model are used to illustrate the example of generating a synthetic MPEG traffic. However, the generation using the scene based model was discussed previously in section 5.4.

Firstly, we use the simple traffic models (see sections 5.3 and 5.4) to generate the GOP layer. Then, the scaling factor process is used to generate an MPEG video pattern (stream) from the GOP sequence. As an application level traffic generator, it is important to notice that we have to use simple models in order to characterise an MPEG sequence. Consequently, the models

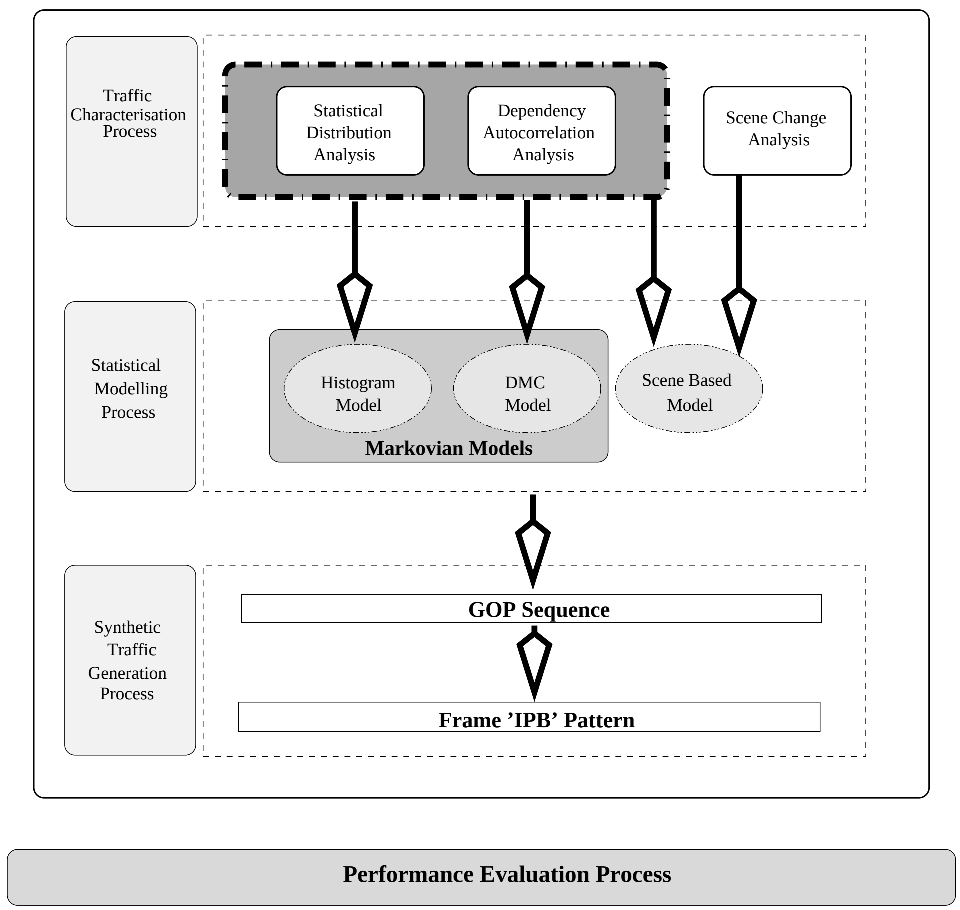


Figure 5.19: Generation Process

will be parameterised to generate synthetic workloads that closely match the actual MPEG stream (the original empirical data sets).

The integration of the MPEG source process starts with an empirical data set of an MPEG sequence. The empirical data set contains frame sizes for the MPEG sequence. A GOP layer will be used for our modelling as a first level of capturing the GOP-GOP correlation. The GOP sizes could be calculated by summing up every 12 consecutive frames. First of all, we need to find some statistical parameters (for example mean, variance, peak ...) to describe the data distribution. The range of GOP sizes is divided into several quantization intervals $\{q_i : i = 1, 2, 3, ..., M\}$ where M is the number of quantization intervals and the number of the intervals depends on the formulated model. The *mean* value of the interval i , μ_{q_i}, should be found to represent the size of the quantization interval.

In the Histogram model, the number of quantization intervals is based on the selected number of the histogram bins. As shown in section 5.3, the Histogram model can be described as 0-order Markov chain process. The number of Markov states is equal to the number of the quantization intervals. It is possible to improve the distribution feature of the model by increasing the number of quantization intervals, but this will led to an increase in the number of states. Each state i is associated with the Mean value μ_{q_i} of that interval. The transition from a state to another is controlled by the transition matrix. We estimate and define the transition matrix as follows:

$$P_{ij} = \frac{n_i}{N} \, where \; N = \sum_{i=1}^{M-1} n_i \;\; and \, n_i = number \, of \, GOP's \, in \, q_i$$

With every transition from i to $i+1$ state, the μ_{q_i} value of next state is generated. It is very important to mention that the transition from a state to another is completely independent from the previous transition.

The DMC differs from the Histogram model in two main ways: the number of quantization intervals and the transition matrix. The number of quantization intervals is based on the standard deviation value of GOP sizes. The larger the number of states, the better the results, but a large number makes the model more complicated, and needs more memory in the generating process. Therefore, a careful selection of the number of states should be made. We used the standard deviation of the GOPs as a scale to find a suitable number of states. The first quantization interval, q_1, starts with the minimum value of GOP. Another way of finding the size of the quantization interval could be achieved by employing the following calculation:

$$Size_q = \frac{Max_{GOP} - Min_{GOP}}{k}$$

where $Max_{GOP}\, and\, Min_{GOP}$ are the maximum and the minimum value of the GOP and k is a selected quantization value.

The calculation of the transition matrix shows that the transition from one state to another depends on the previous transition. With every transition from state $i\, to\, j$, μ_{q_i} is generated. The transition matrix is defined and estimated as follows :

$$P_{ij} = \frac{n_{ij}}{N_i} \quad where \quad N_i = \sum_{i=0}^{M-1} n_{ij} \quad and \quad \sum_{j=0}^{M-1} P_{ij} = 1 \quad for\ i = 0, 1, ..., M-1$$

where n_{ij} is the number of transitions from state i to state j , N_i is the total number of transitions out from state i. For instance, Table 5.9 shows the transition matrix for the 'video conference' sequence.

M	0	1	2	3	4	5	6	7	8	9
0	0.68	0.29	0.03	0	0	0	0	0	0	0
1	0.20	0.63	0.13	0.02	0.02	0	0	0	0	0
2	0.02	0.41	0.33	0.19	0.03	0.02	0	0	0	0
3	0.03	0.15	0.29	0.09	0.06	0.09	0	0	0	0
4	0	0	0.14	0.33	0.24	0.19	0.05	0.05	0	0
5	0	0	0.11	0.11	0.56	0.22	0	0	0	0
6	0	0	0	0.14	0.43	0	0.29	0	0.14	0
7	0	0	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	1	0	0	0
9	0	0	1	0	0	0	0	0	0	0

Table 5.9: Transition Probability Matrix for 'video conference' sequence

Consequently, a sequence of GOPs is generated using the above process either based on the Histogram-based or DMC model. In order to integrate the statistical model to our workload model, we need to produce a sequence of frame sizes. The frame sizes could be calculated from GOP sizes by the *Scaling Factor* process which has been presented in section 5.5.1.

Because the traffic generator is expecting to be worked at an application level, the frame rate, at particular rate (for instance 25 frames/sec), could be generated in order to represent an MPEG source. For our purpose, we could have three traffic sources which represent the three classes of the MPEG sequences: a sequence with a large number of movements (a high activity sequence); one with an average number of movements (a moderate activity sequence) and a third sequence with a small number of movements (a low activity sequence). These three types of traffic can be integrated into the workload model for the video traffic representation. For the traffic generation process, there are two ways to transmit traffic units: either at the maximum rate of the input link, to the network, or they could be transmitted with a constant interarrival time. [Heyman92] used the first one while [Rose95b] used the second.

In order to compare the impact of the generated sequence and the actual sequence on the system performance, simulation experiments are conducted to study the performance of an ATM multiplexer using the generated (or synthetic) sequence based on the traffic models (DMC has been selected to demonstrate this simulation experiment). The multiplexer is modelled as a finite capacity queueing system with, buffer size B, and one server with service rate C. A FIFO service discipline is assumed. The input of the multiplexer consists of MPEG sequences (or frame level). Two simulation experiments are conducted. The simulation uses the actual MPEG sequence, while the second experiment employs the generated sequence based on the proposed models. In both experiments, the traffic source (multiplexing input) consists of a large number of frames arranged according to the compression pattern [IBBPBBPBBPBB]. Bits in each frame are packetised into ATM cells (in addition to 5 bytes as a cell header) and the cells are transmitted using fluid flow approach [Knightly96].

To examine the performance of the simulation operating at heavy load, a sufficient load is needed. A low load level is not interesting because the probability of congestion is too low, and a very high load level is an unrealistic network operation [Liu92]. Thus, the service rate is adjusted in order to obtain an 80% of system load (or 0.8 load). Figure 5.20 shows the cell loss ratio (CLR) versus the buffer size for both sequences (actual, generated). For the sake of elucidation, we show the simulation results for the 'video conference' sequence

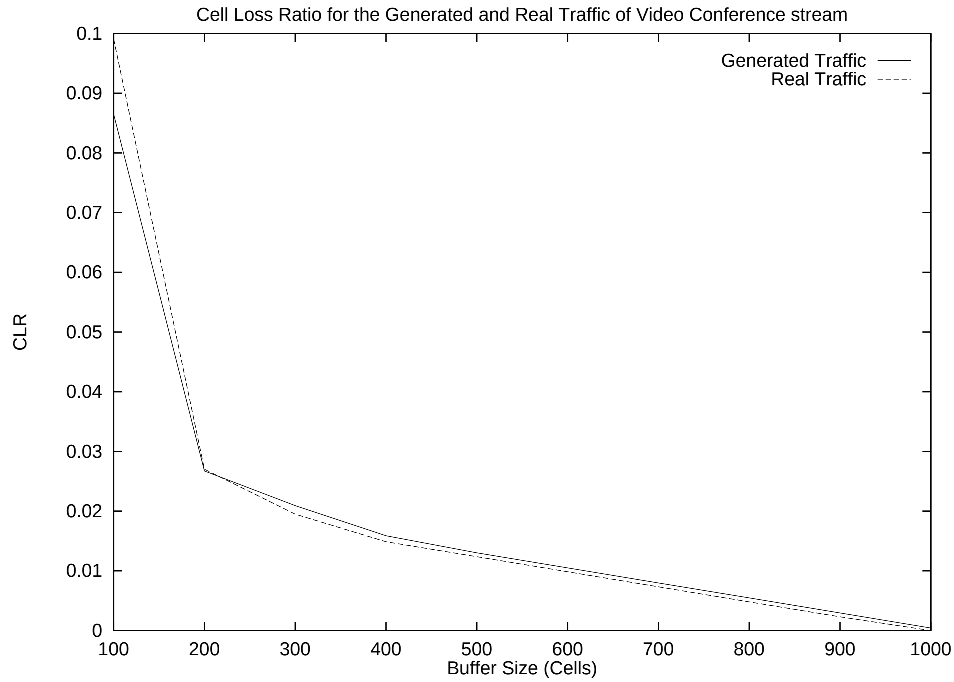


Figure 5.20: Cell Loss Ratio Results

only. It is clear that there is an improvement in the CLR performance with a larger buffer size. Another observation is a clear agreement between the results obtained using the model and the results from the actual sequence. Thus, the traffic model can be used to emulate a realistic MPEG traffic for simulation experiments for the sake of the performance evaluation testing.

5.7 Summary

In this chapter, we have described the statistical analysis and modelling of MPEG sources. The aim of this analysis and modelling is two-folds (1) to characterise an MPEG sequence using an appropriated model to determine a simple model to capture the statistical behaviour of a VBR MPEG sequence, and (2) to derive traffic models to be generated by programs (library of traffic model functions). In turn, these programs are used for a multimedia workload model. In the statistical analysis and modelling, we have used several MPEG data sets (e.g. ‘Dino’, ‘Race’, ‘video conference’ sequences), to cover as wide range of scenario types as possible, and to ensure that the workload model delivers realistic scenarios.

Based on the results of the statistic analysis, we have used two Markovian based models (the Histogram based model and the Detailed Markov chain model (DMC)) to approximate the statistical behaviour of the MPEG sequence. We have shown that the DMC model can be used to approximate two classes of MPEG sequences (high activity and low activity sequences). Based on the analysis of scene change within an MPEG stream, we presented an extensive scene change based model. We used the scene change identifying techniques and the characteristics of the scene changes to construct an MPEG source model. We have examined the statistical behaviour of the generated sequence from the model and the original sequence. We observed that the model captures the main statistical parameters of the actual sequence at different time scales (for both GOP and frame). We also showed that the model captures the dependency features of the original sequence. We have also described the modelling process component of this work; that is, once the models are identified, what is the process of translating these into programs. In addition, we have explored the steps of integrating the MPEG traffic model within the multimedia workload model.

Chapter 6

Multiplexing of VBR MPEG Traffic

6.1 Introduction

As mentioned before, an encoded video traffic is a correlated and bursty traffic with a high degree of peak to mean ratio (or burstiness). For ATM networks, the challenge of QoS guarantees is to allocate an effective bandwidth for each video connection; a tradeoff should be achieved between improving the network utilisation and providing QoS guarantees. A number of algorithms have been proposed to calculate the effective bandwidth for VBR traffic [Belhaj97] [Zhang94] or for a general traffic for high speed networks [Guerin91]. Generally speaking, the effective bandwidth allocation can be defined as the service rate corresponding to the cell loss probability. The effective bandwidth is computed to be close to the long range average (*mean*) rate and far from the peak rate. The bandwidth allocation algorithm works as a part of the call admission control and should be done on the fly. Therefore, the algorithm should be kept simple in order to meet real time requirements. The allocated bandwidth could be based on either deterministic multiplexing or statistical multiplexing. Deterministic multiplexing provides stringent bounds on QoS guarantees, while statistical multiplexing provides probability based bounds on QoS guarantees. However, statistical multiplexing improves the network utilisation, while deterministic multiplexing improves performance guarantees. Thus, a tradeoff should be achieved between the network utilisation, and delivering the desired level of QoS guarantees.

In this chapter, we use the characteristics of VBR MPEG streams to study and then improve the network utilisation. Several models have been proposed for the characterisation of MPEG traffic. The model can be based either on stochastic processes [Kara97] [Rose95a] [Venturin95] or deterministic processes (or models) [Knightly95] [Krunz97]. A deterministic model for MPEG traffic will be used for our study. An MPEG stream is a deterministic periodical pattern. Thus, the advantage of this deterministic pattern will be used to achieve a stream synchronisation (stream arrangement) for the multiplexed MPEG connections. In other words, we define an arrangement[1] to multiplex multiple MPEG streams before entering the network. The process of the arrangement will be analysed and described for multiplexing MPEG connections. The multiplexing gain (mg) factor is used to quantify the network gain. This factor will be used to measure the improvement in the network utilisation. To demonstrate the multiplexing gain, we use several sequences of real MPEG video streams which contain empirical data sets of MPEG frame sizes. In order to study the impact of the scene changes on the multiplexing gain (mg) for synchronised streams, the data sets represent a variety of video streams including video conference, sports events and TV programs.

This chapter is organised as follows: In the following sections, we describe a deterministic multiplexing framework based on an arrangement for multiplexed VBR MPEG streams. We then describe a method by which we can estimate the allocated bandwidth for each multiplexed stream. Lastly, we undertake several simulation experiments in order to show the impact of the stream arrangement and the scene changes on the allocated bandwidth, and the multiplexing gain.

6.2 Deterministic Multiplexing of MPEG Streams

In this section, we present a deterministic multiplexing framework with no data loss and very small delay bound at an ATM multiplexer. In chapter 4, we explored the characteristics of the burstiness behaviour for MPEG traffic, which enabled us to define the amount of activity within the MPEG stream. In this section, we relate the traffic behaviour and the deterministic multiplexing. This can be achieved by describing the deterministic model that we used to

[1] In this chapter, the terms 'arrangement' and 'synchronisation' are used interchangeably.

characterise an MPEG sequence. The main objective of this is to find a suitable model to capture the deterministic behaviour of a VBR MPEG sequence (or pattern). We then present a stream synchronisation way for multiplexing MPEG streams. Based on this synchronisation, we will be able to calculate and allocate the effective bandwidth for a number of multiplexed VBR MPEG sources with guaranteed QoS. Lastly, we demonstrate the impact of both the synchronisation, and the scene changes, on the allocated bandwidth and the multiplexing gain.

6.2.1 Deterministic Model

Several traffic models were proposed for the characterisation of compressed video traffic starting from a simple model (such as stochastic) up to a more sophisticated one (such as self-similar). Because most of these are probabilistic in nature, they cannot be used to provide deterministic guarantees.

We explore the case when N homogeneous VBR video sources, $\{Sr_i : 0 \leq i \leq N-1\}$, are multiplexing and transmitting VBR MPEG streams. We consider a network with non-blocking switches, where queueing happens at the output link of each switch. The service rate (or link speed) is constant and the arrival rate for each stream is equal to the frame rate of MPEG encoder. We use a deterministic model for the MPEG source [Krunz97] which was first mentioned in [Knightly95]. We extend the analysis of the deterministic multiplexing in terms of providing a guaranteed QoS. The model uses a traffic envelope which provides an upper bound on the bit rate. VBR MPEG traffic is very bursty. Thus, the traffic envelope is varied for every stream according to the statistical behaviour of the traffic and the amount of activity within the stream. Therefore, we need to know in advance the upper bound of each multiplexed stream.

The deterministic model can be defined, using five simple parameters (see Figure 6.1) for characterising VBR MPEG traffic:

- L: is the size of GOP or the number of frames between two 'I' frames within the same MPEG pattern.
- Q: is the number of frames between 'P' frames within the same MPEG pattern.

- I_{Ub}: is the maximum size of 'I' frames.
- P_{Ub}: is the maximum size of 'P' frames.
- B_{Ub}: is the maximum size of 'B' frames.

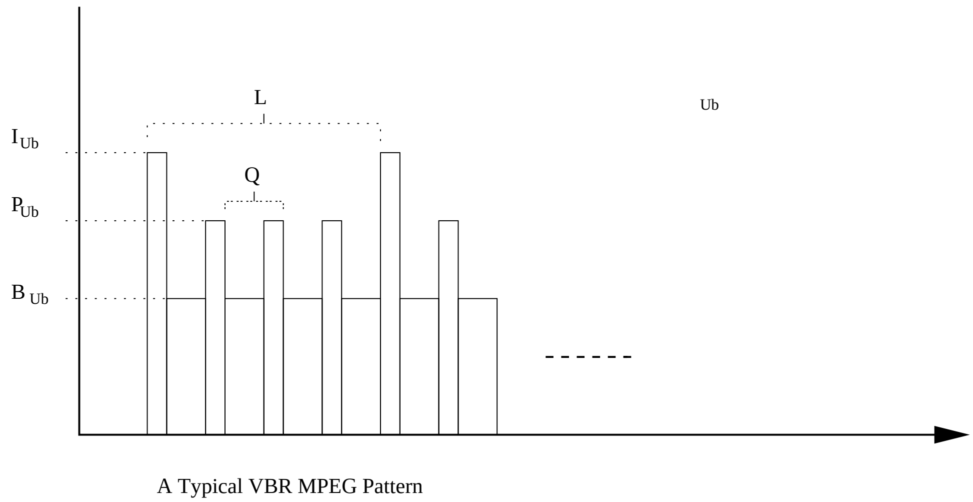


Figure 6.1: The Deterministic Model For VBR MPEG Streams

A traffic constraint function, $Ub(t)$, bounds the actual bit rate where t is measured in frame period. In the case of one MPEG stream, by using these parameters, $Ub(t)$ can be defined as I_{Ub}, $P_{Ub}\, or\, B_{Ub}$ for a given t. As a result, the traffic can be characterised by a traffic envelope $(I_{Ub}, P_{Ub}, B_{Ub}, L, Q)$. In the next section, we will show that the constraint function and the periodical deterministic structure of the MPEG stream can help to provide and support the deterministic QoS guarantees with conjunction with the statistical multiplexing.

6.2.2 A Synchronisation Structure for MPEG Streams

This section describes the possible arrangements for multiplexed MPEG streams. There are various ways to arrange the starting time of each multiplexed stream. For instance, all streams may start at the same time (i.e. all streams transmit their 'I','P' and 'B' frames at the same time). Alternatively, the starting time of each stream could be chosen randomly. However, when a CAC function is executed without considering the starting time, there always exists the possibility of not meeting the QoS requirements [Roh97]. Thus, the cell losses are provided only as a function of the number of connected sources, regardless starting time arrangements.

For the statistical multiplexing, [Rose95b] presented the impact of the various ways of multiplexing on cell losses. Generally speaking, the stream arrangement could be achieved by enforcing the starting time of multiplexed MPEG streams.

Now, let us assume that N homogeneous MPEG streams are multiplexing. Each stream is characterised, as specified in the last section, by the following traffic descriptor: $(I_{Ub}, P_{Ub}, B_{Ub}, L, Q)$

If the starting time for the stream i is u_i where $u_i \in U$ which is a set of possible starting times (frame based), the starting time for each stream could be one of the following integer values: $\{0, 1, 2, 3, ..., L-1\}$. The starting time, u_i, could be specified as a counter which increases with every incoming stream. If $N \geq L$, then the starting time for stream J, where $Mod_L J = 0$, restarts again to 0 and so on. The $\widetilde{U}$, which is the synchronised set of N streams, can be defined as follows:

$$\widetilde{U} = \{\underbrace{u_1, u_2, u_3, u_4, \ldots, u_L,}\ u_{L+1},\ \ldots\ , u_N\}$$

$$m = \{u_1, u_2, u_3, u_4, ..., u_L\}$$

Where $u_1 = 0, u_2 = 1, u_3 = 2, u_4 = 3, ..., u_{L-1} = L-2, u_L = L-1, u_{L+1} = 0, ..., u_N = N-2$. The set m can be defined as a repeated sequence, $m \subseteq U$ where $1 \leq number\, of\, m\, sets \leq \frac{N}{L}$.

6.2.3 Bandwidth Allocation For Multiplexed MPEG Streams

In this section, the allocated bandwidth for each multiplexed MPEG stream (deterministic multiplexing is considered) is described when multiple MPEG connections are multiplexed. We will then demonstrate how we can achieve bandwidth gains when a number of MPEG streams are multiplexed with the stream synchronisation $\widetilde{U}$.

Most of the compressed video traffic requires very restricted QoS guarantees with no losses and very short queueing delays. Such deterministic guarantees can be provided by allocating a bandwidth based on the peak bit rate for each traffic source. However, this methodology reduces the network utilisation, which can be improved by using the statistical multiplexing technique. Moreover, statistical multiplexing is typically used for statistical guarantees not for deterministic guarantees. However, if the stream synchronisation approach is employed, it is possible to maintain deterministic guarantees in conjunction with statistical multiplexing.

For the deterministic guarantees, an effective bandwidth (C) can be allocated using the constraint function $\{Ub(t),\ t \geq 0\}$ which is the maximum boundary of the stream [Krunz97]. In order to estimate the effective bandwidth for multiplexed MPEG streams, let us first demonstrate only two multiplexed streams. The following example demonstrates the benefits of the statistical multiplexing with two MPEG streams. The two MPEG streams are homogeneous and the second stream starts just one frame duration after the first one. The maximum value of the constraint function for the aggregated stream can be defined as:

$$Ub_{aggregated}(t) = Ub(t) + Ub(t+1) < 2Ub_{max}(t \geq 0)$$

where $Ub_{max}(t), t \geq 0$ is the maximum bound of the stream (which is frame based). For deterministic guarantees (no loss and very short queueing delays) an effective (or *equivalent*) bandwidth should be allocated for each stream. The effective bandwidth, C, can therefore be defined as:

$$C = \frac{Max\{Ub(t), Ub(t+1)\}}{2} = \frac{I_{Ub} + B_{Ub}}{2} < I_{Ub}$$

C could be measured in cells per frame period. It is clear that the effective bandwidth, which needs to be allocated for each stream is less than the peak. In some cases, a small buffer is needed when two cells arrive at the multiplexer simultaneously. It is important to note that the bandwidth gains from the statistical multiplexing obtained from the spatial averaging (or pattern synchronisation), not from temporal averaging (or buffering). Therefore, the arrangement of the starting time of each MPEG stream has an impact on the effective bandwidth allocation. The arrangement could be placed either at the MPEG source before the multiplexing (such as at the Video On Demand (VoD) server node) or at the intermediate network node (such as the intermediate switch). In the first case, a small amount of delay (in term of frame period) could be imposed on the starting time of the MPEG stream. In the second case, a small buffer could be needed to arrange the frame.

In order to estimate the effective bandwidth of N multiplexed homogeneous streams associated with a stream arrangement, let us suppose that u_i is the difference between the arrival time of the 'I' frame for stream i and the arrival time of an 'I' frame for the last recent stream (e.g. $i-1$). By doing so, the synchronisation of the streams can be specified as:

$$u_i \in \{\widetilde{U} : (u_1,\, u_2,\, u_3,\, ...,\, u_N)\}\; where\, u_1 = 0$$

Consequently, the effective bandwidth (cells/frame) for each stream with a given $\widetilde{U}$ set can be defined as [Krunz97]:

$$C(u,N) = \frac{Max_{t\geq 0}(\sum_{i=1}^{N} Ub(t+u_i))}{N}$$

Let $n_I,\, n_P,\, and\, n_B$ be the number of MPEG streams that send 'I', 'P' or 'B' frames simultaneously, where $n_I + n_P + n_B = N$. Then, the effective bandwidth can be written as:

$$C(u,N) = \frac{n_I\, I_{Ub} + n_P\, P_{Ub} + n_B\, B_{Ub}}{N}$$

Similarly, C is measured in cells per frame period. It is clear that $C(u,N)$ is less than the peak $\sum_{i=1}^{N} I_{Ub}$.

In the case of multiplexing heterogeneous streams, I_{Ub}, P_{Ub} and B_{Ub} are determined as the maximum values of these parameters from the various multiplexed streams. However, this will reduce the utilisation if there is a vast variation between the peak values of the multiplexed streams. For instance, when the 'video conference' streams are multiplexed with the 'Movie' streams.

6.2.4 The Impact of The Stream Arrangement and Scene Changes

In this section, we explore the impact of scene changes within the same MPEG stream on the allocated bandwidth when several MPEG streams are multiplexed with arbitrary arrangements. First, in order to show the impact of the stream arrangement on the bandwidth gains, we multiplex three MPEG streams with various starting times u_i (except the first stream, u_1 which is always 0). To demonstrate a realistic MPEG stream, and for the sake of illustration, we employ a real traced data for only two MPEG video streams 'Dino' and 'video conference'. These streams have been chosen because they represent different classes of VBR MPEG sequences (see chapter 3).

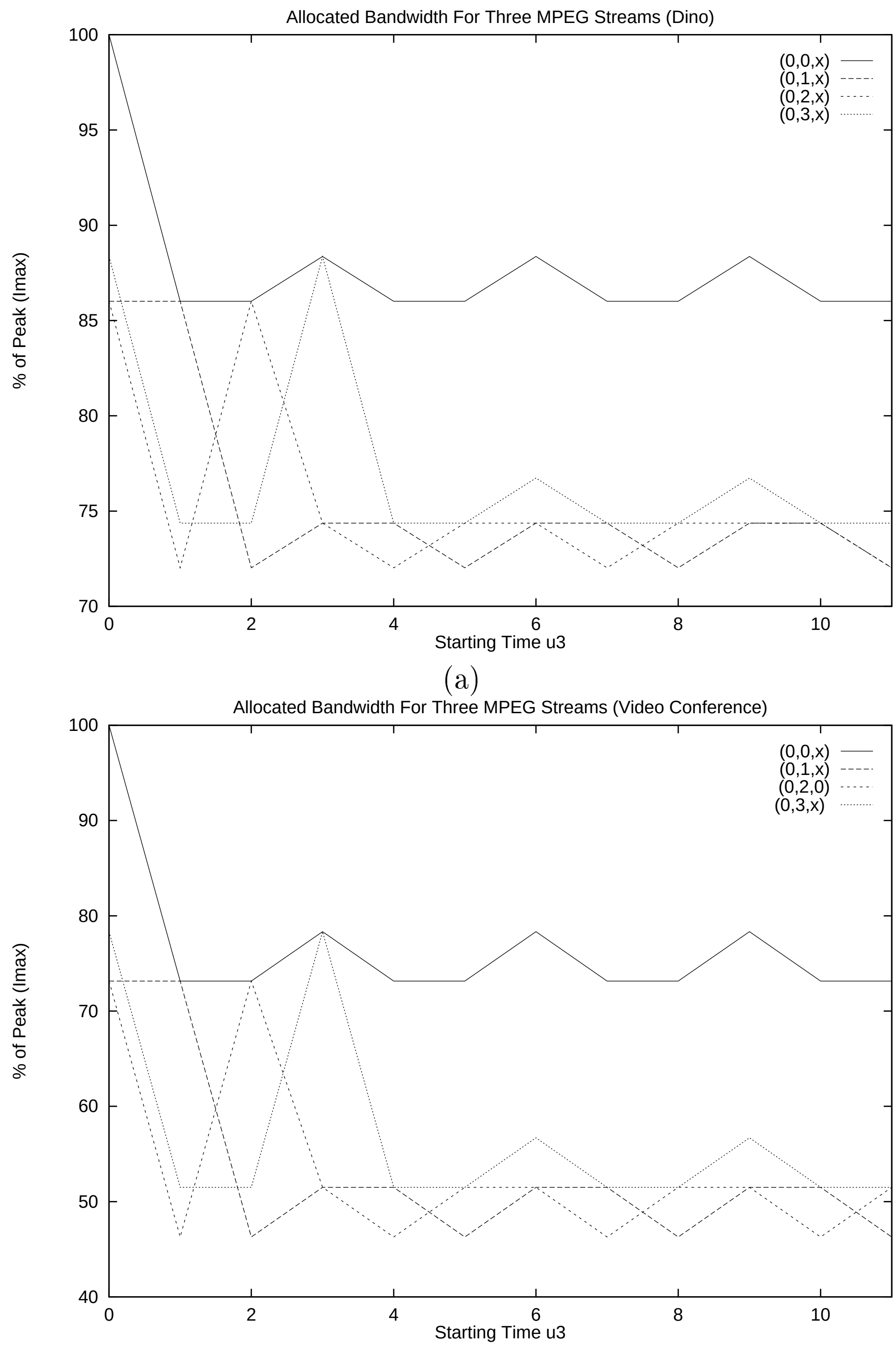
We will demonstrate how much bandwidth gain can be achieved with various arrangement sets. When three MPEG streams are multiplexed with various starting times, u_i, the arrangement set can be presented as the following:

$$\widetilde{U} = \{0\,,\, j\,,\, i\}\, where\,,\; 0 \leq i \leq L-1 \; and \; 0 \leq j \leq 3$$

In the case of the 'Dino' sequence, see Figure 6.2(a), it is clear that with some stream arrangements we can reduce the bandwidth requirements even when bounded deterministic guarantees are provided. It is possible to reduce the allocated bandwidth for each stream up to 28% off peak (i.e. 72% of peak). There is only one exception, (0, 0, 0), when all streams have the same starting time (0), the allocated bandwidth is equal to the peak. Therefore, the amount of reduction depends first on the arrangement set $\widetilde{U}$. The scene change is another factor which has an impact on the amount of reduction on the bandwidth requirement. This can be observed when we demonstrate the last example with the 'video conference' stream. Figure 6.2(b) shows the amount of reduction on the bandwidth requirement. In the case of this class of sequence, it is clear that we can achieve up to 54% off peak (i.e. 46% of peak). This amount of reduction has been achieved because of the ratio between I_{Ub} and B_{Ub}. Due to the amount (or level) of activity within the sequence (see section 4.3), the 'video conference' sequence has a small amount of scene change. Therefore, the size of its 'I' frames are much bigger than the size of its 'B' frames; while in the case of the 'Dino' sequence, there are a moderate number of movements. As a result, the amount of activity within an MPEG stream has a noticeable impact on the bandwidth allocation.

In another experiment, we observed that the amount of allocated bandwidth for each multiplexed stream decreases when the number of the streams increases. It is important to notice that there is a limit to the number of multiplexing streams needed to achieve the minimum allocated bandwidth; therefore, we have

$$C_\theta \propto \frac{1}{n}\,,\, 1 \leq \theta \leq n$$



(b)

Figure 6.2: Bandwidth Gain (the 'Dino' and 'video conference' Streams)

where n is the maximum number of streams to achieve the minimum bandwidth C_θ. The minimum allocated bandwidth is limited by the relation between the VBR MPEG stream descriptors and the number of multiplexed streams. Figure 6.3 depicts a decay in the amount of the allocated bandwidth in relation to the increasing number of multiplexed streams. This shows the efficiency sharing of the link capacity. However, if there is only one stream, the peak rate is allocated. Furthermore, the curve decays very rapidly until it reaches the minimum bandwidth. There are other smaller peaks in the curve, which appear when $Mod_L N = 0$. As a result, the allocated bandwidth can be related to both the stream descriptor and the number of the multiplexing streams, as follows:

$$C = \frac{Int(\frac{N}{L}+1)I_{Ub} + \frac{N}{Q}P_{Ub} + N(\frac{1}{L}+\frac{1}{Q})B_{Ub})}{N}$$

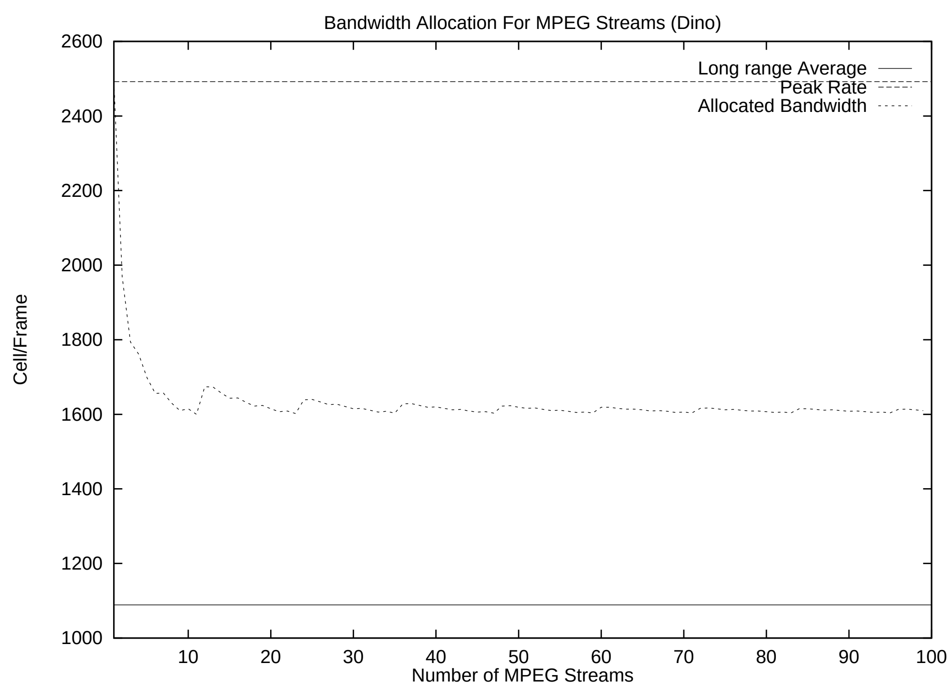


Figure 6.3: Allocated Bandwidth Gain For N MPEG Streams

6.2.5 Multiplexing Gain

This section evaluates the statistical multiplexing for VBR MPEG streams and describes another factor (mg) that can be used to measure the gain in the utilisation above the peak rate allocation. This multiplexing gain factor will be used to quantify the improvement of the network utilisation. The multiplexing gain (mg) could be defined as:

$$mg = \frac{NI_{Ub}}{C}$$

In order to show the amount of multiplexing gain, we have multiplexed several MPEG streams with a given $\widetilde{U}$ arrangement. Figures 6.4 (a) and (b) depict the multiplexing gain for the 'video Conference' and the 'Dino' sequences. The value of the multiplexing gain starts with 0 , then it starts to increase with the increasing number of multiplexed streams. Another observation from these figures is that the value of mg_{video} is higher (up to 3.2) than that for mg_{Dino} (up to 1.56). Consequently, the amount of activity has also an impact on the value of mg. In addition, the figures show several peaks when $Mod_L N = 0$. In other words, when 'I' frames are overlapped. Furthermore, there is a limit to the multiplexing gain value even though the number of the multiplexed streams is increasing.

6.3 Influence of Stream Arrangement on Queueing Performance

In order to explore the performance of the stream arrangement on an ATM multiplexer, simulation experiments are presented in this section for the multiplexing of multiple 'Dino' sequences at different level system loads, and through different buffer sizes. Then, cell losses result for the multiplexer buffer are also presented in this section.

The simulation results showed which property of MPEG video stream has a major impact on the multiplexer performance (YATS simulator was used). For example, these experiments helped to show the influence of GOP pattern on the cell losses. We have chosen two approaches of multiplexing: arranged and non-arranged streams. The simulation experiments have been conducted with both, arranged and non-arranged streams, approaches. In the arranged streams, the offset (or the starting time) of each stream was synchronised one frame period after the previous stream, thereby ensuring a minimum overlapping of 'I' frames (which did not have identical multiplexing). While in the non-arranged streams, all streams start the GOP at the same time, i.e. all streams transmit their I, P and B frames during the same time intervals.

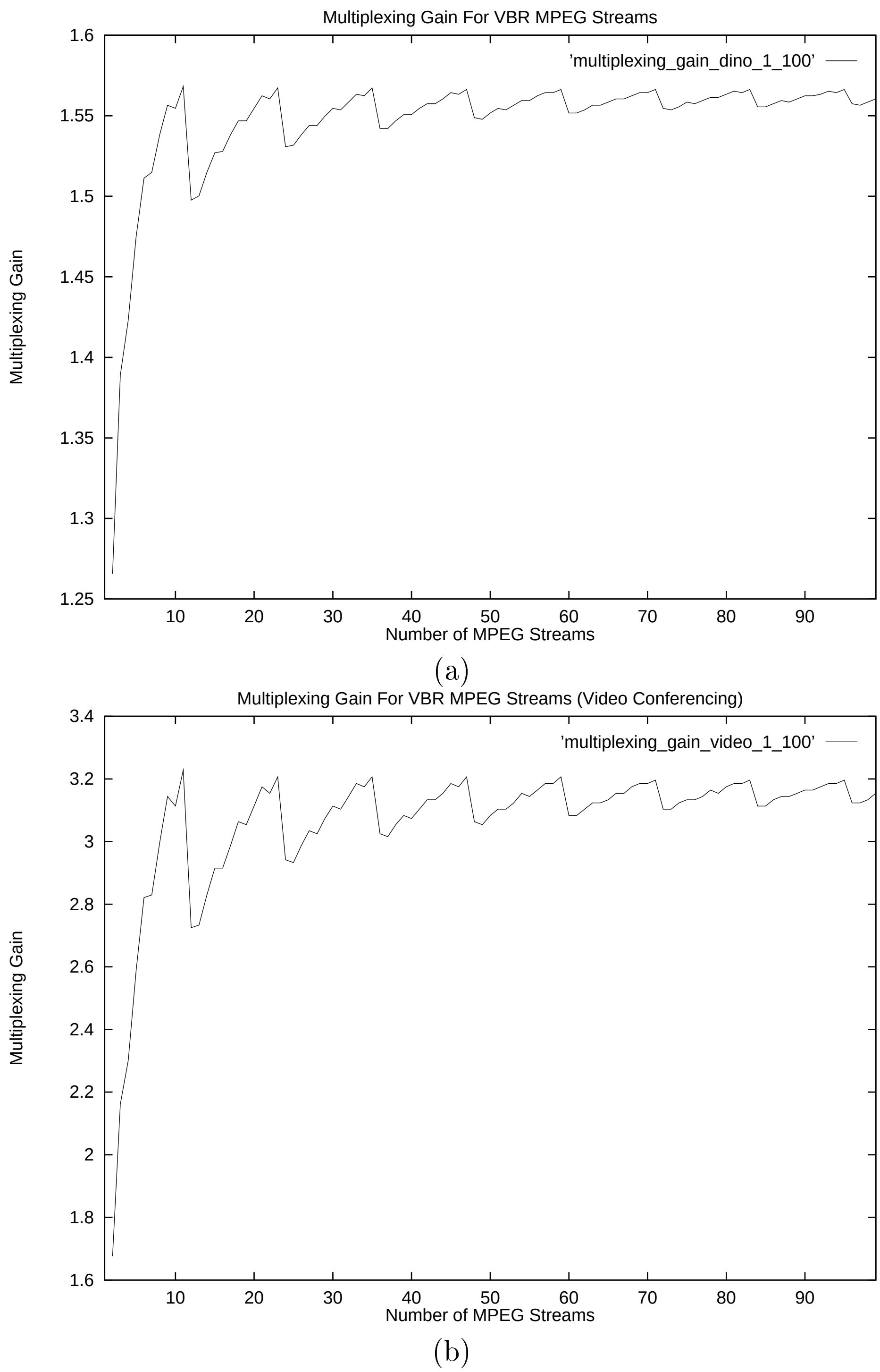


Figure 6.4: Scene Change and Multiplexing Gain

The simulation model can be described as N MPEG streams which are multiplexed through a single multiplexer with an output link of 50 Mbits capacity. The number of multiplexed streams determines the system load. Various buffer sizes were used for the multiplexer in order to smooth out the fluctuations of the multiplexed streams when the system's load is increased (i.e. relating the buffer size with the maximum delay). An FIFO discipline was considered. Each frame was packetised into the payload of ATM cells with an evenly distributed (within the same frame period). Thus, the interarrival times are equal within the frame duration.

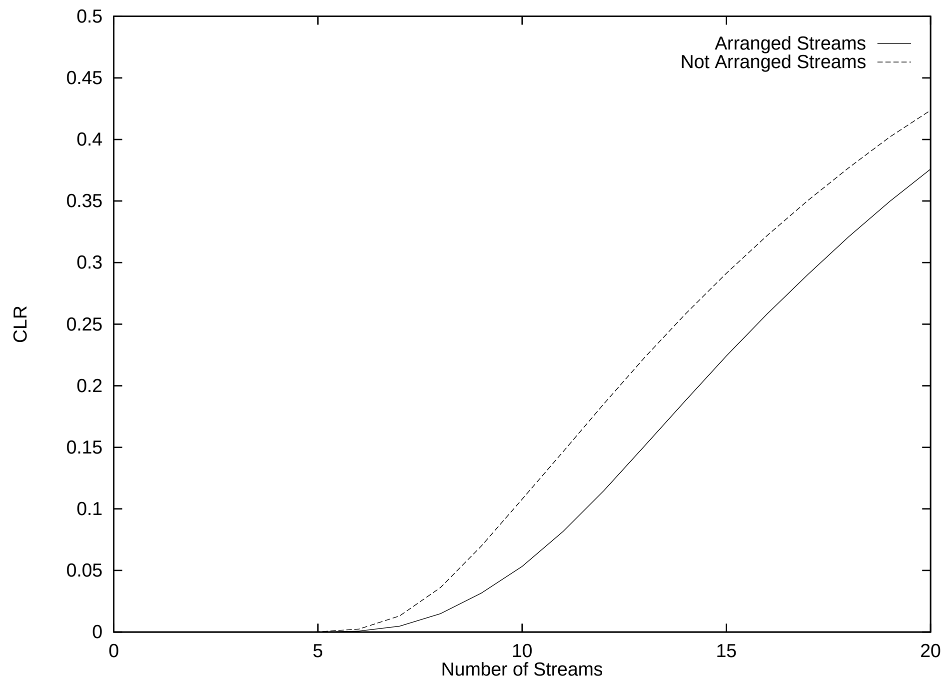


Figure 6.5: CLR for N Multiplexed Streams (Dino)

Figure 6.5 illustrates the cell loss ratio against the number of multiplexed streams for both arranged, and identical stream, multiplexing at buffer size 1000 cells. It is clear that the CLR curve in the case of arranged streams is below the CLR curve for the identical stream multiplexing. This occurs mainly because of 'I' frame overlapping, which leads to bursts of cell transmissions during 'I' frame durations. On the other hand, the CLR is reduced when the multiplexed streams are arranged. Therefore, the number of multiplexed streams can be increased with the arranged multiplexing approach.

Increasing the buffer size reduces the CLR, but at the cost of increasing queueing delay. We have also conducted simulation experiments on the multiplexing of N MPEG streams, with the identical approach at different buffer sizes. Figure 6.6 shows the effect of buffer size on cell loss performance. Clearly, buffers can reduce the cell loss to great degree. For instance, with buffer sizes of 500 and 200 cells, the CLR increases with a smaller buffer size. However, a larger buffer size increases the waiting time in the buffer (i.e. maximum delay). The cell loss increases sharply when the number of stream is increased due to such bursty traffic (particularly from 'I' frame transmission). However, when the number of streams is more than 10, the curve of the cell loss starts to take more horizontal shape. Thus, the buffer size increase has very little impact on cell loss performance when the increase passes a certain point. This means that just increasing the buffer size, for instance at an ATM switch, cannot solve the problem of cell loss satisfactory when traffic is highly bursty such as MPEG and burst durations are long ('I' frame).

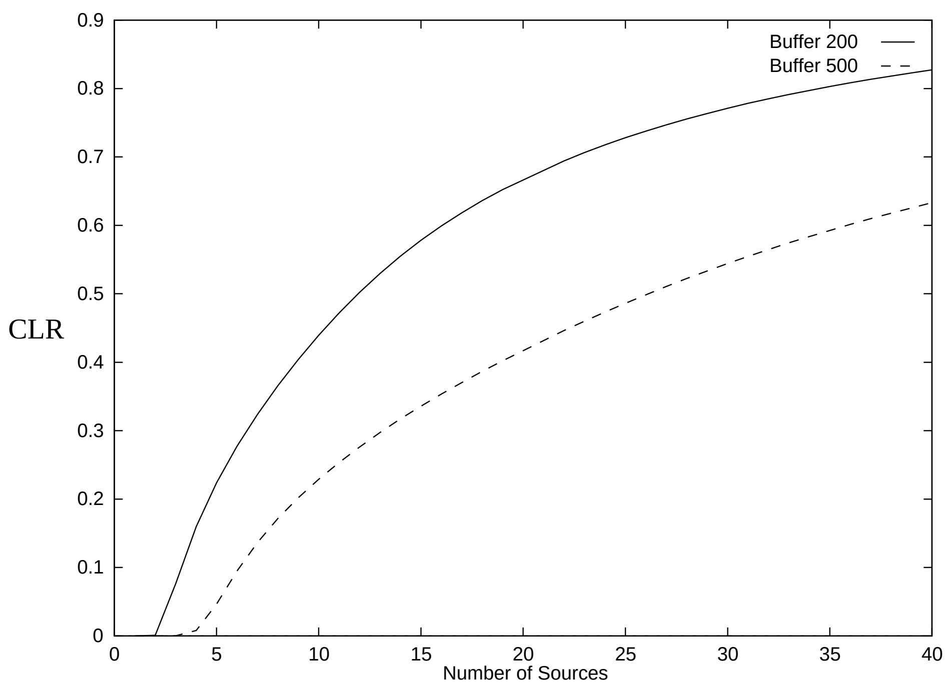


Figure 6.6: CLR for N Multiplexed Streams at Different Buffer Sizes (Dino)

Consequently, ATM uses statistical multiplexing as a means for resource sharing. For instance, the link capacity, at an ATM multiplexer, is shared by cells from various traffic streams. Thus, the bandwidth is dynamically allocated so that if a stream is temporarily idle, its bandwidth is given to other active streams. As a result, statistical multiplexing improves significantly the bandwidth utilisation. However, the QoS guarantees are offered only on a statistical basis.

In order to investigate the feasibility and efficiency of resource sharing in the case of the statistical multiplexing, we have focused on an ATM multiplexer with multiple VBR MPEG video streams at its input. Although a general ATM network will be more complex, the consideration of a simple ATM multiplexer model is useful to understand the behaviour and the impact of the statistical multiplexing of VBR MPEG streams with respect to the system configuration (i.e. system capacity and link speed). Other important issues which can be considered are the measurement of the cell loss ratio and the throughput estimation (allocated bandwidth) in such systems.

6.4 Summary

In this chapter, we showed that it is possible to achieve multiplexing gains even when we provide deterministic QoS guarantees. We described an arrangement for multiple multiplexed MPEG streams called stream synchronisation. The arrangement could be achieved by enforcing the starting time of each multiplexed stream. Taking advantage of MPEG coding (spatial averaging), we are able to reduce the amount of the allocated bandwidth for each multiplexed stream. In addition, we showed that the synchronisation of the streams affects the calculation of an effective bandwidth.

We also showed that it is important to notice that multiplexing gains do not only depend on the stream's synchronisation, but also depends on the amount of activities within the multiplexed streams. Consequently, the results obtained in this work are important in terms of demonstrating the effect of stream activities on the network utilisation. This has be achieved when we employed the multiplexing gain factor, mg, which quantifies the network gain. Several simulation experiments are performed to show the influence of MPEG properties on the performance of an ATM multiplexer. The simulation results showed that GOP structure has a major impact on the multiplexer performance in terms of increasing the cell losses. In other words, when multiple MPEG stream are multiplexed with an arrangement approach, the cell loss performance improves. The impact of increasing the buffer size is also studied. Due to the burstiness of MPEG traffic, the buffer size increase does not have much impact on the cell loss performance when the increase passes a certain point. Furthermore, a tradeoff should be achieved between having a large buffer and increasing the waiting time at the buffer (corresponding delay).

Chapter 7

Evaluation

7.1 Introduction

This thesis deals with the characterisation, modelling and multiplexing of VBR MPEG traffic over an ATM network. The encoded MPEG traffic introduces several issues that must be addressed in order to attack the problem of the traffic modelling. This includes analysing the statistical characteristics of MPEG, predicting the traffic behaviour, and then formulating the characteristics as a traffic model. Modelling the VBR MPEG traffic requires an understanding of its statistical characteristics. The complexity involved in modelling MPEG coded video traffic increases as the movement activity. Furthermore, the periodic characteristics associated with the MPEG encoder increase the complexity involved in modelling as well. However, from the way MPEG is designed, network utilisation improvement and multiplexing gains can be achieved from the MPEG sequence by forcing a sort of arrangement for the multiplexed streams. This chapter presents an overall evaluation of the results which were obtained from our analysis and modelling of MPEG throughout the thesis.

7.2 Evaluation

In this thesis, extensive traffic characterisation and modelling processes were introduced. An evaluation of this work can be presented through three major topics: statistical analysis of VBR MPEG traffic, VBR MPEG modelling and the statistical multiplexing of multiple MPEG streams at an ATM multiplexer.

7.2.1 Statistical Analysis of VBR MPEG

Supporting such a complex traffic (as MPEG) requires a good understanding of traffic behaviour. Thus, properties of the analysed MPEG streams have been taken into consideration in order to achieve an accurate traffic characterisation.

7.2.1.1 Video Streams

In many previous studies, the video streams used for the analysis were either short (such as [Gruenefelder91]) or employed only one type of traffic activity (such as [Heyman92]). However, the study in this thesis was based on long (40000 frames) and various video streams (21 different streams) obtained from actual *Movies, News* and *Sports* events. This means that a wide range of traffic activity was covered in order to achieve a better understanding of the stream behaviour, especially the long range dependency feature. Moreover, many characterisation studies of MPEG video traffic tended to characterise MPEG at low levels, such as cell or frame [Venturin95] [Reininger94], while ignoring the higher levels, such as GOP, which play the most important role concerning the autocorrelation effects of an MPEG video stream behaviour (as shown in chapter 3). However, the characterisation process in this thesis considers the higher levels of MPEG as well as the lower ones. Therefore, it is challenging to introduce a traffic model which captures the behaviour not only of one time scale, but several ones.

7.2.1.2 Statistical Measures

To ensure a proper traffic characterisation, the most significant properties of video traffic were considered for the characterisation process . Three statistical measures were explored, namely distribution, autocorrelation function and scene changes. As discussed previously, each of these measures has an impact on the traffic behaviour. Therefore, extensive studies were performed on various MPEG streams in order to gain a degree of insight as to how such compressed video traffic behaved.

- **Distribution:** The probability function of video traffic is believed to play an essential role in approximating the queueing behaviour [Casilari98]. However, in the case of MPEG traffic, it has been observed that there is no good agreement yet on which the ideal distribution could be used to approximate the GOPs. Thus, the probability density function can not be used as a basis to characterise GOP sequences. However, *Gamma* distribution was found to be the closest one among the other (Lognormal and Weibull). The observations of the main statistical parameters, such as the mean and standard deviation values, on various MPEG sequences showed that their characteristics vary vastly from one to another depending upon their contents.

- **Dependency:** The dependency feature was explored at different time scales of MPEG traffic. Correlations (or dependencies) between arrivals were found to cause considerable degradation in network performance (as measured by cell loss rate). We have demonstrated that an MPEG traffic sequence exhibits a periodical and complicated correlation structure due to the presence of three different types of frame in one sequence. The dependency was also explored at the GOP level. It was observed that the autocorrelation function of the GOPs sequence may has various decaying shapes. For example, some sequences, such as 'Race' and 'video conference' have a negative exponential shape (exponentially decaying), while some others, such as 'Dino', exhibit a slowly decaying shape (hyperbolically decaying) as lag increases. Given these observations, we have assumed that a sequence with a high or low amount of activity exhibits a short range dependency, while a sequence with a moderate amount of activity experiences a long range dependency.

- **Scene Change:** Unlike that of previous studies, the scene change measure was analysed here in great detail because it is seen as an important reason for fluctuations in the overall bit rate within the video stream. Thus, an extensive scene change analysis was undertaken in order to detect the impact of scene changes on the traffic behaviour. A heuristic approach was used for the scene change characterisation, based on the fact that a significant change in the size of consecutive GOPs is an indication of the start of a new scene. Two techniques were introduced to identify a scene change within an MPEG sequence, namely *Outlier* and *Second difference*. These techniques were used in order to view an MPEG sequence as a collection of scenes. As the result, the MPEG sequence can be broken down into smaller portions (scenes) which reflect the amount of activity within each MPEG sequence. A large number of scenes means a high level of activity. The identification technique by the way of conclusion, is used to classify various MPEG streams in terms of the sequence activity.

7.2.2 VBR MPEG Modelling

Although the modelling of VBR video sources has recently received significant attentions, there is still no commonly acceptable model which captures the behaviour of a wide variety of video streams, ranging from a very low active stream to higher ones [Izquierdo96]. The main reasons behind this are the existences of different encoding schemes and various activity sequences. We reviewed the contribution of our research with respect to VBR MPEG modelling, using the criteria of an appropriate model defined in section 2.7.3.

7.2.2.1 The Modelling Approach

In this thesis, we have focused on different modelling approaches for VBR MPEG traffic in order to find an adequate model with respect to the simplicity and generality properties of the model. The Markov chain process is widely used in several studies to capture the behaviour of video traffic [Heyman92] [Stamoulis94]. However, a simple Markovian model shows a lack in capturing the dependency feature of video traffic at various time scales [Rose94a]. Therefore, the ability of the Markov chain process was examined in order to better capture the behaviour of MPEG traffic at GOP time scale. Two Markovian based models which are of different complexity have been studied here, namely *Histogram* based model and *Detailed Markov Chain* model (*DMC*). The Markov chain approach was also considered in [Heyman96] to formulate a video model. This model overestimated cell losses, due to the simplicity of the model and to the level of modelling (frame). As extracted from the statistical characterisation analysis of a variety of MPEG sequences, the latter may exhibit two types of dependency: both short range and long range. The two Markovian based models were employed in order to approximate the statistical behaviour of the MPEG sequence. With some effort, we have shown that the *DMC* model can be used to approximate two classes of MPEG sequences (high activity and low activity), even with a small number of states (about 9 states).

We have also presented a better assessment on the number of states required by the Markov chain model. This was done in order to ensure a greater degree of accuracy.

Most modelling studies have focused mainly on the statistical behaviour of MPEG traffic as a sequence of frames (at a frame level) without considering the impact of scene changes within the video stream. These are believed to introduce a large variation in the overall bit rate [Lazar93]. The scene change characterisation process was used to formulate a new scene-based model. The novel element of our modelling approach is the introduction of a new model for VBR video traffic which captures MPEG traffic behaviours at three time scales: frame, GOP and scene. The importance of the scene changes was incorporated into the traffic modelling because of its impact on the traffic behaviour (as stated previously). For the sake of comparison, scene changes were also considered for modelling MPEG traffic in [Krunz96]. However, the scene changes in that study were based only on 'I' frames, while changes in all frame types were considered in this thesis. This has improved the behaviour of the model, especially in relation to queueing performance.

7.2.2.2 Ease of Implementation

The simplicity of the introduced models helped to keep the models easy to implement in case of generating a synthetic traffic. The number of parameters used to describe the model was deliberately kept low. For instance, nine Markov chains as well as one dimension transition matrix were used to describe the Markovian models, while the scene based model is described using two processes, namely the scene length and scene variation. The former process needs only one parameter (q) while the latter one can be described as a Markovian model.

7.2.2.3 Model Appropriateness and Limitations

The suitability as well as the limitations of the models have been validated by comparing the behaviour of the generated traffic and the original one in terms of the statistical and queueing performance [Lazar93]. Due to the different approaches of modelling, it is possible for the model to predict one behaviour accurately, while inaccurately predicting another. The two Markovian models as well as the proposed one were validated to assess their ability in capturing the desirable traffic behaviour. The ability of models can be assessed using three main criteria as follows:

- **Statistical Parameters:** Statistical analysis was performed on generated sequences (GOPs) using both Markovian traffic models. Despite the simplicity of the modelling approach, both models have shown a good agreement in terms of capturing the main statistical parameters of the original MPEG sequence. The scene-based model was validated by analysing their suitability to capture the statistical behaviour of the original MPEG sequence, as well as the multiplexing performance. The model was used to generate a synthetic MPEG sequence. By comparing the statistical behaviour of the sequence (based on the proposed model) and the original one, it has been observed that the model can capture the statistical behaviour of the actual sequence accurately. The main statistical (or distribution) parameters of the generated sequence were much closer to that of the original sequence.

- **Long and Short Dependency:** The correlation feature of MPEG was also examined at different time scales (GOP and frame). The *DMC* model showed an ability to capture only a sequence with a high or low level of activity (i.e. sequences with short range dependency), while the *Histogram* model showed no correlation at all, due to its concept, which is based only on the distribution properties. However, both models demonstrated a good agreement in capturing the periodical correlation shape of the MPEG sequence at frame level. The frame sequence was obtained from the generated GOP sequence by using a process called *Scaling Factor*. Therefore, an improvement is achievable when the modelling approach is based on a higher level (such as GOP), while a smaller level (frame) can, accordingly, be derived. Due to the limitations of the Markovian based models in capturing the dependency feature, a new scene-based model has been proposed to capture the statistical behaviour of MPEG traffic, especially the sequence with a long range dependency feature. The importance of the scene changes in incorporating traffic modelling is due to its impact on the traffic behaviour (as stated previously). In addition, the model exhibits ideally the long range dependency feature of the traffic. For the sake of comparison, it has also been found that the ACF curve of the generated sequence fits the ACF curve of the original sequence.

- **Queueing Performance:** In most cases the validity or 'goodness' of a model is determined by comparing the multiplexing performance of the model to the original traffic at an ATM multiplexer [Lazar93]. Simulation experiments were conducted to study the performance of an ATM multiplexer using the generated (or synthetic) sequence based on the proposed model and the original MPEG sequence. The outcome of these simulation results showed that the scene based model showed a better performance in capturing the queueing behaviour of the original MPEG traffic at an ATM multiplexer over the two Markovian models.

7.2.2.4 Model Usability

Another aim of the analysis and modelling of MPEG traffic was to derive a simple, as well as an efficient, traffic model to be employed by simulation programs (through the library of traffic model functions). In turn, these programs are used for a multimedia workload model to cover as wide a range of scenario types as possible, and to ensure that the workload model can deliver realistic scenarios. We have explored the steps required for integrating the MPEG traffic model within the multimedia workload model. An example of generating a synthetic MPEG traffic was presented in this work, based on one of the MPEG traffic models. The generated traffic showed a massive performance compared to the original traffic. Therefore, the synthetic traffic can be used for the performance evaluation process of an ATM multiplexer when multiple MPEG streams are multiplexed.

7.2.3 Multiplexing Gains

With such bursty and complex traffic, we have shown that it is possible to achieve a multiplexing gain (up to 3.4 over peak) even when deterministic QoS guarantees are provided. The beauty of MPEG compressed traffic is that it is more 'regular' and 'structured' than any other data types. Another significant benefit of the way MPEG is designed is that a large 'I' frame is followed by a small 'B' frame. This reduces the traffic burstiness caused mainly by the presence of 'I' frames.

An arrangement was described for multiple multiplexed MPEG streams called stream synchronisation. This arrangement was achieved by enforcing the starting time of each multiplexed stream. By taking advantage of MPEG coding (spatial averaging), the amount of the allocated bandwidth can be reduced for each multiplexed stream (up to 54% off peak). In addition, it has been observed that the synchronisation of the streams affects the calculation of the effective bandwidth.

Furthermore, it has been shown that the multiplexing gain does not depend only on the stream synchronisation, but also on the amount of activities within the multiplexed streams. Another factor which has an impact on multiplexing gains is the scene changes and their magnitude within the VBR MPEG stream. We have shown the amount of saving in terms of bandwidths over peak. Consequently, the results obtained in this work are important in terms of demonstrating the effects of the stream activities on the network utilisation. Beside, the multiplexing gain factor (mg) was used in order to quantify the overall network gains.

Chapter 8

Conclusions and Further Work

8.1 Conclusion

This thesis has covered three main issues regarding the management of VBR MPEG traffic over ATM networks. The traffic characterisation process is an important issue for studying the performance of traffic transmission throughout the network. It can also be used to develop appropriate traffic management and control schemes. In addition to analysing the statistical behaviour (distribution, dependencies) of various MPEG sequences, the scene changes have been explored within an MPEG sequence. Long and various video streams have been studied in order to cover a range as wide as possible of MPEG sequences, including real *Movies, News* and *Sports* events.

The results of the statistical analysis carried out in this study suggest that the MPEG traffic is correlated, bursty and exhibits complex patterns which vary from one stream to another. MPEG traffic cannot, therefore, be described as independent traffic. In addition, it has been observed that one of the most important reasons for the fluctuations in the overall bit rate of MPEG is because of the scene changes within the video stream. Consequently, scene changes have been analysed in details in order to explore the behaviour and the impact of scene changes within an MPEG stream. Two methods have been presented to identify the scene changes: the *Outlier* and *second difference*. Once the amount of scene changes are identified, it would be possible to assess the degree of activity within the video stream. A '*Scene Change Scale*' (SCS) has been introduced to exhibit the amount of activity within the MPEG stream. The scale has been used to demonstrate the impact of scene changes on QoS requirements. Several experiments have been performed to show the impact of multiple MPEG streams multiplexing at an ATM multiplexer, relating the simulation results to SCS. The scale has been correlated with the results obtained from simulation experiments. It has been shown that this scale can be used to assess the impact of scene changes on the QoS guarantees in terms of cell loss ratio (CLR) for multiplexed MPEG video streams.

Another thread that has been explored is a source model construction for VBR MPEG traffic based on our statistical analysis. The GOP level of modelling has been chosen in order to improve the model performance. From the statistical analysis of various MPEG streams, it has been shown that both long and short range dependencies can be seen as being based on the amount of activity within the stream. Due to the fact that the Markovian approach exhibits short range dependency, two Markovian based models have been introduced incorporating the *Histogram based* model and the *Detailed Markov Chain* model (*DMC*). To examine the appropriateness and limitations of the models, simulation experiments have been conducted to study the performance of an ATM multiplexer using the generated (or synthetic) sequence based on the models and the real traffic. It has been observed that the generated sequences using these models capture the main distribution parameters accurately. However, in terms of dependency feature, it has been shown that the *DMC* model can be used to approximate two classes of MPEG sequence (high activity and low activity sequences). Furthermore, this thesis has presented two steps necessary to integrate the MPEG streams to the multimedia workload model. The first step was a statistical modelling process which captured the individual behaviour of MPEG sources. The second step was the integration process (through a histogram based and/or Detailed Markov Chain model) to link these (MPEG) statistical distributions to the workload model.

Based on the extensive characterisation of the scene change within an MPEG stream, this thesis proposed a scene change based model. The scene change identifying techniques and the characteristics of the scene changes have been used to construct a composition MPEG source model. The statistical behaviour of the generated sequence from the model and the original sequence have been examined. It has been observed that the model captures the main statistical parameters of the actual sequence at different time scales (GOP and frame). It has also been shown that the model exhibits the long range dependency feature of the original sequence due to the incorporation of scene change characteristics.

In order to compare the overall achievements of MPEG traffic modelling, it has been shown that the *Histogram* model does not approximate the dependency feature while *DMC* is only adequate for capturing the short range dependency feature of the real MPEG stream. By contrast, the proposed model (*Scene-based*) is capable of capturing the long range dependency feature of such traffic.

This thesis has explored the tradeoff between providing QoS guarantees and improving the network utilisation when multiple VBR MPEG streams are multiplexing at an ATM multiplexer. A deterministic framework has been presented for MPEG traffic. An allocation bandwidth approach based on a deterministic model for multiplexed VBR MPEG streams has also been presented. Different arrangements have been described for the multiplexed VBR MPEG streams. The impact of such arrangements on the allocated bandwidth has been shown as well. As a result, it has been concluded that it is possible to improve the network utilisation with statistical multiplexing, even in the case of providing deterministic QoS guarantees. Finally, the impact of the stream activity (the amount of scene changes within MPEG stream) has been explored on the allocated bandwidth and the network multiplexing gains. Consequently, it has been shown that it is important to notice that multiplexing gain does not only depend on the stream's synchronisation, but also on the amount of activities within the multiplexed streams; this is due to the way MPEG is designed.

8.2 Further Work

The research presented in this thesis can be extended in many ways. These may include:

- The statistical multiplexing helps to improve the network utilisation, especially in the case of bursty traffic. As shown in this thesis, MPEG traffic exhibits various degrees of burstiness, based on the video class (activity). The management process of such a complex traffic is an important aspect for any efficient use of the available bandwidth. One way of gaining greater insight into the proposed model *(scene-based)* would be by deriving a better way of calculating the effective bandwidth required for each multiplexed MPEG connection, with respect to the desired QoS requirements.

- The study of the effect of traffic behaviour in this thesis shows that traffic characteristics have a significant impact on the performance of an ATM multiplexer. It might be desirable to study the multiplexing under a queueing system more complicated than the discipline FIFO, as presented in this thesis. For example, the *Weighted Fair Queueing* (WFQ) or *Strict Priority Control* (SPC) could be of interest. The latter provides guarantees on bandwidth and delay for real time VBR traffic, while the former only provides guarantees on the share of available bandwidth [Kara99]. It might be also desirable to conduct a simulation with a more complex network system, including multiple switches and cross sources, in order to explore the impact of the cross traffic on end-to-end performance.

- This thesis study has concentrated on video traffic characterisation and modelling, without taking into account any smoothing technique. This would reduce traffic variations and hence improve the traffic handling methods. Therefore, it would be desirable to explore the impact of the traffic smoothing on the performance of the Markovian based models.

- One way of gaining greater insight into the delivery of QoS guarantees could be by studying end-to-end QoS for the transmission of MPEG streams, and then comparing the actual scene errors (the relation between the QoS parameters and the user view). Exploring the effects of CLR and jitter on the picture quality from the end user visual point of view would help to improve QoS guarantees.

- Due to the existence of various encoding schemes, it is difficult to find a generic source model representing an encoded video traffic. The scene change based model exhibits an accurate approximation of MPEG video traffic. It would be possible to make an extensive use of the scene change characterisation processes so that they can be applied to other encoding schemes, such as H261.

- In this thesis, it has been observed that the scene changes affect not only the complexity of the traffic characteristics, but also the long range dependency feature of the video traffic. There is a general belief that the long range dependency has an effect on the queueing performance at an ATM multiplexer. However, some high activity MPEG streams exhibit a short range dependency feature. The relationship between the impact of the dependency features, long and short, of various MPEG streams on the queueing performance should be investigated in more detail in order to demonstrate the relationship between the stream activity and the dependency feature.

- To test and evaluate the performance of a network system, or an ATM switch, it is necessary to provide artificial traffic which closely resembles real traffic. This thesis has presented the steps needed to integrate the MPEG source model into a multimedia workload model. It would be possible to build a prototype for the workload model as a representative example of multimedia traffic.

Bibliography

[Baumann97] M. Baumann, ‘YATS - Yet Another Tiny Simulator User's Manual for Version 0.3’, Available in http://www.inf.et.tu-dresden.de/TK/yats/.

[Belhaj97] A. Belhaj and L. Pap, ’Efficient Bandwidth Assignment Algorithm for Guaranteed Quality of Service Traffic in ATM Networks’, Proceedings of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 49, July 1997.

[Bunzel94] Mark J. Bunzel and Sandra K. Morris, ‘Multimedia applications development’, McGraw Hill Inc, 1994.

[Casilari98] E. Casilari, A. Reyes, A. Estrella and F. Sandoval, ‘Characterisation and Modelling of VBR Video Traffic’, Electronics Letters, Vol. 34, No. 10, 14th May 1998.

[Celandroni97] N. Celandroni, E. Ferro and F. Potorti, ‘A Traffic Generator for Testing Communication Systems: Presentation, Implementation and Performance’, Real Time Systems, Vol. 13, No.1, pp 5-24, July 1997.

[Chiotis97] T. Chiotis, F. Stamatelopoulos and B. Maglaris, ‘Traffic Source Models for Realistic ATM Performance Modelling’, 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 17, July 1997.

[Chu95] P. Chu, ‘ATM Burst Traffic Generator’, Proceedings of the 5th Great Lakes Symposium on VLSI, pp 262-265, March 1995.

[Conti96] M Conti, E Gregori and A Larsson, 'Study of the Impact of MPEG-I Correlations on Video Sources Statistical Multiplexing, IEEE Journal on Selected Areas in Communications, Vol. 14, No. 7, pp 1455-1471, September 1996.

[Daigle86] J. Daigle and J. Langford, 'Model for Analysis of Packet Voice Communications Systems', IEEE Journal of Selected Arias in Communications, Vol. 4, No. 6, pp 847-855, 1986.

[Danzig92] P. Danzig, S. Jamin, R. Cacerest, D. Mitzel and D. Estrin, 'An Empirical Workload Model for Driving Wide-Area TCP/IP Network Simulations', Journal of Interworking, 1992. Available in http://irl.eecs.umich.edu/jamin/publications.html.

[Doulamis96] A. Doulamis, N. Doulamis and G. Konstantoulakis, 'Traffic Characterisation and Modelling of VBR Coded MPEG Sources', Proceedings of the 4th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 6, July 1996.

[Frost94] V. Frost and B. Melamed, 'Traffic Modelling For Telecommunications Networks', IEEE Communications Magazine, pp 70-81, March 1994.

[Gall91] D. Gall. 'MPEG: A Video Compression Standard for Multimedia Applications', Communications of the ACM, Vol. 34, No. 4, April 1991.

[Gallassi90] G. Gallassi, G. Rigolio and L. Verri, 'Resource Management and Dimensioning in ATM Networks', IEEE Network Magazine, pp 8-17, May 1990.

[Gringeri98] S. Gringeri, B. Khasnabish, A. Lewis, K. Shuaib, R. Egorov and B. Basch, 'Transmission of MPEG-2 Video Streams over ATM', IEEE Multimedia Journal, Vol. 5, No. 1, pp 58-71, January-March 1998.

[Groeneveld88] R. Groeneveld, 'Introductory Statistical Methods: An Integrated Approach Using Minitab', PWS-KENT, 1988.

[Gruenefelder91] R. Gruenefelder, J. Cosmas, S. Manthrope and A. Okafor, 'Characterisation of Video Codecs as Autoregressive Moving Average Processes and Related Queueing System Performance', IEEE Journal on Selected Areas in Communications, Vol. 9, pp 284-293, April 1991.

[Guerin91] R. Guerin, H. Ahmadi and M. Naghshineh, 'Equivalent Capacity and Its application to Bandwidth Allocation in High-Speed Networks', IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, pp 968-981, September 1991.

[Habib92] I. Habib and T. Saadawi, 'Multimedia Traffic Characteristics in Broadband Networks', IEEE Communications Magazine, Vol. 30, No. 7, pp 48-54, July 1992.

[Helvik95] B. Helvik, 'Synthetic Load Generation for ATM traffic measurements', Telektronikk, Vol. 91, No. 2/3, pp 174-194, October 1995.

[Herrtwich91] R. Herrtwich, 'The Roll of Performance, Scheduling and Resource Reservation in Multimedia Systems', Operating Systems of the 90s and Beyond, International Workshop Dagstuhl Castle, Germany, pp 279-284, July 1991.

[Heyman92] D. Heyman, A. Tabatabai and T. Lakshman, 'Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks', IEEE Transactions on Circuits and Systems for Video Technology,Vol. 2, No. 1, pp 49-59, March 1992.

[Heyman96] D. Heyman and T. Lakshman, 'Source Models for VBR Broadcast-Video Traffic', IEEE/ACM Transactions on Networking, Vol. 4, No. 1, pp 40-48, February 1996.

[ISO/IEC97] ISO/IEC working group, Available in http://drogo.cselt.stet.it/mpeg/

[ITU93] ITU-T Recommendation I.211. B-ISDN Service Aspects. Rev.1 1993.

[Izquierdo96] M. Izquierdo and D. Reeves, ‘A Survey of Source Models for Variable Bit Rate Encoded Video’, Multimedia Systems Journal, August 1996.

[Kara95] Mourad Kara and Peter Dew, ‘A Configurable Traffic Generator: the Performance Evaluation of Resource Scheduling Algorithms and Protocols in ATM Networks’, Proceedings of the 3rd IFIP workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, July 1995.

[Kara97] M. Kara and A. Mashat, ‘Statistical Analysis and Modelling of MPEG sources for Workload Characterisation of Distributed Multimedia Application’, 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, July 1997.

[Kara99] M. Kara and M. Rahin, ’Comparison of Service Disciplines in ATM Networks Using a Unified QoS Metric’, To appear in the Proceedings of the 16th International Teletraffic Congress (ITC-16), Edinburgh, UK, June 1999.

[Knightly95] E. Knightly and H. Zhang, ‘Traffic Characterisation and Switch Utilisation using a Deterministic Bounding Interval Dependent Traffic Model’, Proceedings of the IEEE INFOCOM’95, 1995.

[Knightly96] E. Knightly and H. Zhang, ‘D-BIND: An Accurate Traffic Model for Providing QoS Guarantees to VBR Traffic’, IEEE/ACM Transactions on Networking, Vol. 5, No. 2, pp 219-231, May 1996.

[Krunz95] M. Krunz and H. Hughes, ’A Traffic Model for MPEG-Coded VBR Streams’, Proceedings of the ACM SIGMETRICS’95, 1995.

[Krunz96] M. Krunz and s. Tripathi, ‘Scene-Based Characterization of VBR MPEG-Compressed Video Traffic”, Technical Report TR-3573, Institute for Advance Computer Studies, Department of Computer Science, University of Maryland, 1996.

[Krunz97] M. Krunz and S. Tripathi, 'Impact of Video Scheduling on Bandwidth Allocation for Multiplexed MPEG Streams', ACM Multimedia Systems Journal, Vol. 5, No. 6, pp 347-367, November 1997.

[Kuo98] F. Kuo, W. Effelsberg and J. Garcia-Luna-Aceves, 'Multimedia Communications: Protocols and Applications', Prentice Hall Inc., 1998.

[Law91] A. Law and W. Kelton, 'Simulation Modeling and Analysis', 2nd Edition, MacRAW Hill Inc., 1991.

[Lazar93] A. Lazar, G. Pacifici and D. Pendarakis, 'Modeling Video Sources for Real-time Scheduling', Technical Report 324-93-03, Columbia University, Department of Electrical Engineering and Center for Telecommunications Research, April 1993.

[Li95] S. Li, T. Yang and J. Ni, 'Source Modelling and Queueing Analysis of VBR Video Teleconference Traffic in ATM Networks', Proceedings of the 3rd IFIP workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 4, July 1995.

[Liu92] H. Liu, 'Traffic Shaping for Congestion Control in High Speed ATM Networks', Msc Thesis, Department of Computation of Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, August 1992.

[Maglaris88] B. Maglaris, D. Arastasion, P. Sen, G. Karlsson and J. Robbins, 'Performance Models of Statistical Multiplexing in Packet Video Communications', IEEE Transactions on Communications, Vol. Com-36, No. 7, pp 834-844, July 1988.

[Manthorpe96] S. Manthorpe, 'STCP User Manual', Tech Report #1.0, LRC-DI-EPFL, Lausanne, Switzeland, 1996.

[Mashat95] A. Mashat and K. Jambi, 'Distributed Multimedia: A Research Point of View', Proceedings of the Symposium on Multimedia: Tools, Applications and Trends, Dhaharan, Saudi Arabia, December 1995.

[Mashat98a] A. Mashat and M. Kara, 'The Impact of Synchronising MPEG Streams on Bandwidth Allocation', Proceedings of the 14th UK Performance Engineering Workshop, July 1998.

[Mashat98b] A. Mashat and M. Kara, 'Scene Change Scale: An Indicator of Multiplexing MPEG Streams for The Assessment of QoS Guarantees Over ATM Networks', Proceedings of the 6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 101, July 1998.

[Melamed92] B. Melamed and B. Sengupta, 'TES Modelling of Video Traffic', IEICE Transactions Communications, Vol. E75-B, No. 12, pp 1292-1300, December 1992.

[Melamed94] B. Melamed, D. Raychaudhuri, B. Sengupta and J. Zdepski, 'TES-Based Video Source Modeling for Performance Evaluation of Integrated Networks', IEEE Transactions on Communications, Vol. 42, No. 10, pp 2773-2777, October 1994.

[Mendenhall94] W. Mendenhall and T. Sincich, 'Statistics For Engineering and The Sciences', Fourth Edition, Prentice-Hall Inc, 1994.

[Minitab89] Minitab Reference Manual: Release 7, Minitab Inc., 1989.

[Molnar97] S. Molnar and G. Miklos, 'On Burst and Correlation Structure of Teletraffic Models', Proceedings of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 22, July 1997.

[Monk91] A. Monk, 'Exploring Statistics With MINITAB: A Workbook For The Behavioral Sciences', John Wiley & Sons Ltd, 1991.

[MPEG99] MPEG Video Webpage, Fundamentals of MPEG Video Compression Algorithms, Available in http://bs.hhi.de/mpeg-video/papers/.

[Murphy94] J. Murphy and J. Teahan, 'Video Source Modelling for ATM Networks', Proceedings of the 11th UK IEE Teletraffic Symposium, Cambridge, UK, March 1994. Available in http://www.eeng.dcu.ie/ murphyj/pub1/.

[Nahrstedt95] K. Nahrstedt and R. Steinmets, ‘Resource Management in Networked Multimedia Systems’, Computer, pp 52-63, May 1995.

[Ni96] J. Ni, T Yang and D. Tsang, ‘Source modelling, queueing analysis, and bandwidth allocation for VBR MPEG-2 Video traffic in ATM Networks’, IEE Proc-Communications, Vol. 13, No 4, pp 197-204, August 1996.

[Nomura89] M. Nomura, T. Fujii and N. Ohta, ’Basic Characteristics of Variable Rate Video Coding in ATM Environment’, IEEE Journal on Selected Areas in Communications, Vol. 7 , No. 5, pp 752-760, June 1989.

[Onvural95] R. Onvural, ‘Asynchronous Transfer Mode Networks: Performance Issues’, Artech House INC, 1995.

[Pancha93] P. Pancha and M. El Zarki, ‘Bandwidth Allocation Schemes for Variable Bit Rate MPEG Sources in ATM Networks’, IEEE Transactions on Circuits and Systems For Video Technology, Vol 3, No. 3, pp 190-198, June 1993.

[Paxson95] V. Paxson and S. Floyd, ‘Wide Area Traffic: The Failure of Poisson Modeling’, IEEE/ACM Transactions on Networking, Vol. 3, No. 3, pp 226-244, June 1995.

[Perros96] H. Perros and K. Elsayed, ‘Call Admission Control Schemes: A Review’, IEEE Communications Magazine, Vol. 34, No. 11, pp 82-91, November 1996.

[Pitts96] J. Pitts and J Schormans, ‘Introduction to ATM Design and Performance’, Wiley and Sons Ltd, 1996.

[Rampal95] S. Rampal, ‘Routing and End-to-end Quality of service in Multimedia Networks’, PhD Thesis, Department of Electronical and Computer Engineering, North Carolina State University, August 1995.

[Reininger94] D. Reininger, B. Melamed and D. Raychaudhuri, 'Variable Bit Rate MPEG Video: Characteristics, Modeling and Multiplexing', Teletraffic Science and Engineering, Vol.1, Ch. 143, No. A and B, 1994.

[Reyes97] A. Reyes, E. Casilari, D. Estrela and Sandoval, 'Simulation of MPEG Video Traffic Using Neural Networks', Proceedings of IWANN'97, 1997.

[Roberts91] J. Roberts: editor, 'COST 224: Performance Evaluation and Design of Multiservice Networks', Management Committee of the COST 224 Project, Commission of the European Communities, Final Report, 1991.

[Roberts96] J. Roberts, U. Mocci and J. Virtamo, 'Broadband Network Teletraffic: Performance Evaluation and Design of Broadband Multiservice Networks, Cost 242', Springer Verlag, Berlin, Heidelberg, 1996.

[Roh97] B. Roh, H. Ahn and J. Kim, ' A Connection Admission Control With Picture Type Dependent Effective Bandwidth for VBR MPEG Video Traffic in ATM Networks', IEE Electronics Letters, Vol 33, No. 23, November 1997. Available in http://vishws.kaisk.ac.kr/ bhroh/.

[Romanow94] A. Romanow and S. Floyd, 'Dynamics of TCP Traffic over ATM Networks, ACM SIGCOMM Computer Communications Review, October 1994.

[Rose94a] O. Rose and M. Frater, 'A Comparison of Models for VBR Video Traffic Sources in B-ISDN', Proceeding of IFIP Transactions C-24: Broadband Communications II, North Holland-Elsevier Science Publisher, pp 275-287, 1994, Available in http://www-info3.informatik.uni-wuerzburg.de/html/publications.html.

[Rose94b] O. Rose and M. Ritter, 'MPEG-Video Sources in ATM-Systems: A new Approach for The Dimensioning of Policing

Functions', IFIP 3rd International Conference on Local and Metropolitan System, Kyoto, December 1994.

[Rose95a] O. Rose, 'Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modelling in ATM Systems', Proceeding of the 20th Annual Conference on Local Computer Networks, Minneapolis, MN, October 1995.

[Rose95b] O. Rose, 'Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic', Report No. 120, Institute of Computer Science, University of Wurzburg, Germany, July 1995.

[Rose96] O. Rose, 'Estimation of the Hurst Parameter Long-Rang Dependent Time Series', Report No. 137, Institute of Computer Science, University of Wurzburg, Germany, February 1996.

[Ryu96] B. Ryu, 'Fractal Network Traffic from Understanding to Implementations', PhD thesis, School of Arts and Science, Columbia University, 1996.

[Schuler96] C. Schuler, 'Realtime ATM Traffic Generation', Proceedings of the 3rd International Workshop on Protocol for Multimedia Systems (PROMS'96), Madrid, October 1996.

[Sikora98] T.Sikora, 'MPEG Digital Video Coding Standards', In Digital Electronics Consumer Handbook, McGraw Hill Company, Available in http://bs.hhi.de/mpeg-video/papers/sikora/mpeg1_2/mpeg1_2.htm.

[Skelly93] P. Skelly, M. Schwartz and S. Dixit, 'A Histogram-based Model for Video Traffic Behaviour in an ATM Multiplexer', IEEE ACM Transactions on Networking , Vol. 1, No. 4, pp 446-459, August 1993.

[Sriram86] K Sriram and W. Whitt, 'Characterization Superposition Arrival Processes in Packet Multiplexers for Voice and Data', IEEE Journal on Selected Areas in Communications, Vol 4, No. 6, pp 833-846, September 1986.

[Stamoulis94] G Stamoulis, M Anagnostou, 'Traffic Source Models for ATM Networks: a Survey', Computer Communications, Vol 17, No. 6, pp 428-437, June 1994.

[Venturin95] Rama Venturin, 'Traffic Source Modelling of Multimedia Application for ATM', Informacija Telekomnni Kacije Automati, Vol 14, Iss 1-3, pp 235-243, June 1995.

[Wurzburg95] Wurzburg ftp Site for MPEG Traces. ftp://ftp-info3.informatik.uni-wuerzburg.de/pub/MPEG/

[Yao97] J. Yao, F. Chen, J. Mellor, M. Ritch and P. Mars, 'Simulation of Cell Scheduling Algorithms Using Conventional and Self-Similar Traffic Models', Proceedings of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, Paper No. 41, July 1997.

[Yousef97] S. Yousef and C. Strange and A. Schormans, 'Resource Allocation of the Effective Bandwidth by Simulation in the ATM-Based B-ISDN', Proceedings of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Ilkely, UK, July 1997.

[Zhang94] H. Zhang and E. Knightly, 'Providing End-to-end Statistical Performance Guarantees with Bounding Interval Dependent Stochastic Models', Proceedings of ACM SIGMETRICS'94, Nashville, TN, May 1994.