ATM Network Impairment to Video Quality

Prepared by:
Alfred Ching Hong Wong

Supervised by:
Mr. M.J. Ventura

Department of Electrical Engineering
University of Cape Town
2000

This dissertation is submitted to the University of Cape Town in fulfilment of the academic requirements for the Degree of Master of Science in Electrical Engineering.

Declaration

I declare that this thesis is my own work. Where collaboration with other people has taken place, or material generated by other researchers is included, the parties and/or material are indicated in the acknowledgements or references as appropriate.

This work is being submitted for the Master of Science Degree in Electrical Engineering at the University of Cape Town. It has not been submitted to any other university for any other degree or examination.

Alfred Ching Hong Wong　　　　　　　　Date

Synopsis

Multimedia applications have begun to receive more and more attention since the 1990's. With the help of computer networking, multimedia technology allows individuals to retrieve information and to communicate visually and orally with one another. At this point in time, applications such as distance learning, video-on-demand and video-conferencing are already on the market. When these services are deployed by public service providers, it is important to consider issues such as infrastructure and bandwidth requirements, how these services will be charged, the quality of service expected by the general public, the network performance requirements etc. In general, users are more concerned with the quality of multimedia communication applications than the technologies that have enabled them. If the quality of these services is not satisfactory, it is very unlikely that they will attract many users or generate much revenue.

This study considers video traffic over ATM networks and investigates the effects of network impairments on the quality of video presented to end-users. A network impairment emulation architecture is developed and implemented in the form of an emulated ATM network. The video and audio distortions caused by ATM network impairments are investigated and an attempt is made to quantify the amount of degradation in video quality resulting from such distortions. The results presented in this research not only reveal how video traffic is impaired by ATM networks, but also provide the fundamentals to determine network performance requirements for supporting video communications.

Acknowledgements

I wish to express my sincere gratitude to the following people and organization, for direct and indirect assistance in the preparation of this dissertation.

Telkom, Siemens and THRIP, the sponsors of this research, for the wonderful opportunity to work on this project.

Mr. M.J. Ventura, my supervisor in the Department of Electrical Engineering at the University of Cape Town, for his time and effort.

My colleagues in the Communications Research Group (CRG), University of Cape Town, for numerous valuable discussions and constructive advice.

Matthew Roux and Patrick Vine, for developing the video-on-demand application and Matthew for his advice and support.

Concilium Technologies, for making a BSTS unit available.

Ricardo Sanchez and Sean House from the Information and Telecommunications Technology Center at the University of Kansas, for their expertise and advice on the virtual ATM switch.

Numerous experts from the ATM on Linux mailing list, for their wonderful work and their assistance on setting up ATM support in the Linux environment.

My Parents, to whom I dedicate every success of my life, for their unfailing support throughout my academic career.

Disclaimer

The video sequences described in this dissertation and included in the accompanying CD are for demonstration purposes only. The author has no intention of conveying any implications through the contents of these video clips whatsoever.

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1.

Introduction

1.1 Background on Multimedia Communications

Multimedia integrates audio, graphics, text, animation and video to improve the effectiveness of computer applications. With the help of computer networking, multimedia technology allows for information retrieval as well as visual and oral communications among people at different places. A multimedia communication application sends one or more flows of integrated information between two or more locations. Examples of such applications are video-on-demand (VoD), video-conferencing, distance learning, on-line library / shopping and e-mail incorporating video and audio. In general, these applications make use of digital video technology and require higher network bandwidth than that offered by conventional data networks. This need for broadband communications has driven further advances in computer networking, which has led to the development of Asynchronous Transfer Mode (ATM) and has enabled the above multimedia applications to be deployed as commercial services. Today, multimedia communication applications such as videoconferencing are widely used. There are also initiatives from around the world to carry out trials on multimedia communication services in the United States, Canada, Europe and Japan [TAT98].

Multimedia communication applications are useful in many areas. In the residential domain, individuals can gain access to entertainment and convenient information without leaving the home through a personal computer or a television with a smart set-top box. They are able to watch movies and daily news, interactively select and purchase goods through on-line shopping, look up electronic yellow pages and perform visual and oral communications with other individuals. In the AMUSE (Advanced Multimedia Services for Residential Users) project, various multimedia services are provided experimentally to home users in six European countries [ZAH97].

In education, students have the flexibility to attend lectures at any time and location by watching video of lectures. They also have advantages including the ability to control the learning pace and to review difficult sections in the lectures. A good example is the Centre for Distance Education at the University of Texas at Arlington[1] [UTA00]. The centre is proving to be beneficial to both on-campus and off-campus (distance learning) students. Another example can be found at the Knowledge Systems Institute [CHA98], where a Multimedia Micro-University has been set up to be an intelligent distance learning system. During a seminar or a conference presentation, an interactive session can be set up between audiences and the speaker at a different physical location to allow feedback and discussions.

In the business environment, video conferencing allows meetings to be held among companies at different locations, or between regional offices and the headquarters of a company. Staff members can attend meetings in their respective offices efficiently and conveniently, saving time and expenses on travelling. The effectiveness of presentations in these meetings can be enhanced by animated graphics, sounds, video clips and "whiteboards"[2] [MIN94]. In addition, important announcements within a company can be sent as e-mail incorporating video and audio using [KER97] or [HES98] in order to attract attention and ensure clarity.

Remote consultation and diagnosis are especially useful in rural areas where medical facilities are far away. If professional opinion from a specific field is required for a more accurate diagnosis, a general practitioner can consult with a specialist through the sharing of diagnostic information such as X-ray or tomography pictures, Magnetic Resonance Imaging (MRI), ultrasound images and electrocardiograms. A networked multimedia framework for medical imaging is presented in [WON97], which also shows that the average amount of data per medical examination varies from a few megabytes to 500 megabytes. In the future, Telepresence Surgery will enable the surgeon to perform remote operation on patients, who may be in areas of conflict, disease or natural disasters [KIT97].

In the hotel industry, added facilities such as video-on-demand, video-based city guide and electronic yellow pages can be provided in hotel rooms. Moreover, travelling company executives will find video conferencing facilities in hotels very helpful to keep in contact with business partners or staff members. In the manufacturing industry, high resolution computer aided design (CAD) images can be exchanged among members of a team in order to allow for collaboration in the design.

[1] Refer to news report on the UTA Center for Distance Education (UTA.mpg) in the */chapter1.1/* directory of theaccompanying CD for more details

[2] Whiteboards use a special input device that takes in the presenter's hard-writing and transfer the hand-written material to all the parties involved in the presentation.

1.1.1 Attributes of Multimedia Communication Applications

Successful deployment of multimedia services in the above areas depends on the requirements of the applications. These requirements vary among different applications depending on their characteristics. Before the factors contributing to the quality of multimedia communication services can be investigated, the attributes of these applications need to be analysed [KWO98].

Delivery of information by multimedia communication applications can be performed in a real-time or non-real-time fashion. Real-time information is consumed by the receiving party immediately and therefore requires instant delivery. Examples of real-time multimedia communication applications are electronic yellow pages and Video-on-Demand. Non-real-time information is used by the receiving party some time after it has been received. In this case, the receiving party does not even need to be present when the information is transferred, so instant delivery is not required. Electronic mail incorporating video and voice is an example of non-real-time multimedia communication application.

Information communicated by a multimedia application can be time-based or non-time-based in nature. Time-based information includes a sequential set of information units. Each of these units has an associated timing requirement relative to the other units. Video is an example of time-based information and a video frame constitutes an information unit. Video frames must be played back at a rate according to the timing information in order to make sense to the human eye. In contrast, still images, graphics and text are non-time-based and do not include a timing component as part of the information.

Multimedia communications can be conducted in one-way or two-way. In one-way communication, information flows only in one direction as in the case of cable TV distribution. The receiving party accepts information passively and does not have the ability to respond to the sending party. Bi-directional communications can be either symmetric or asymmetric in nature. For example, voice only communication between two human users is symmetrical because a similar type of information is transferred in both directions. An example of asymmetric communication is video-on-demand in which control messages are transferred in one direction from a client to a server and video transfer takes place in the opposite direction [ROU99].

In addition, communication applications can be classified as point-to-point or point-to-multipoint depending on the number of parties involved. Point-to-point communication refers to that between only two parties. Communication involving more than two parties are referred to as point-to-multipoint. For example, true video-on-demand is a point-to-point communication between an end-user and the video server. Multi-party teleconference [CLA92] and distributed distance learning are regarded as point-to-multipoint.

1.1.2 Criteria for Quality in Multimedia Communications

Having considered the attributes of multimedia communication applications, it is now possible to analyze the criteria upon which the quality of multimedia communication services is based. These criteria represent the requirements of multimedia communication applications from a user's perspective. The importance of user-oriented requirements cannot be overestimated because it is commonly agreed that one of the most important goals of constructing a network is to support networked applications. However, it is not possible for networks to provide satisfactory support for multimedia communication applications unless user-oriented requirements of these applications are clearly stated.

Since the multimedia services described above entail the transmission of multiple types of information over a communication network, one of the most important requirements is the proper synchronization among the temporal relations of various media types. This is referred to as inter-media synchronization [XIE99], which is often made more complicated by allowing user interaction as suggested in [LIA98] and [HUA98]. A good example is lip-synchronization. Another aspect of synchronization involves the management of information arriving from more than one source, such as during a videoconferencing session among three or more participants. This is referred to as inter-participant synchronization [ZAR96].

Although response time is irrelevant for one-way communication such as television broadcast, it is an important concern for applications with any level of interaction involved. For example, asymmetrical 2-way applications such as World Wide Web browsing are often criticised because of their slow response time. The response time of multimedia communication can be described by the time taken for the requested information to appear after such a request was issued. For symmetrical 2-way communications such as video-conferencing, response time has an added meaning of the round-trip delay experienced by the participants.

The quality of video and audio presented to the end-users has a significant effect on their impressions towards the service. Video and audio quality at the destination is dependent on many factors including the original quality at the source[3], the encoding techniques employed, the

[3] For example, the video clips in the */chapter1.1.2/* directory are all encoded in the MPEG format but in a wide range of quality (e.g. size and resolution, frame rate, sound quality, general picture quality etc).

characteristics of the transportation network and the decoding process within equipment at the destination.

The user-friendliness of a multimedia communication application is very important as explained in [ROS92]. The user-interfaces of these applications should aim to hide the technological complexities from the end-users as far as possible.

The reliability of information received by a multimedia communication application is especially important for mission-critical tasks. In remote medical diagnosis and telepresence surgery, high-resolution medical images are required to arrive at the destination free of errors. Errors in ultrasound images or electrocardiograms could result in incorrect interpretation of symptoms and possibly lead to aggravation of medical conditions or loss of life.

1.1.3 Influential Factors for Multimedia Communications

With an understanding of the attributes and the user requirements of multimedia communication services, the factors that influence the quality of these applications can now be examined. These factors are specific to particular implementations of these services and need to be considered when the quality of a multimedia communication application is evaluated. For example, video-conferencing can be implemented on an IP network or an ATM network. These two types of networks are based on different architectures and hence have very different characteristics. Since the architecture of a transportation network is fundamental to its suitability to support multimedia traffic, the choice of the transportation network has a direct effect on the quality of the multimedia service. Besides the choice of the transportation network, some of the other implementation specific factors that will be discussed are:

- the impairments being introduced by chosen network,
- the video and audio encoding technique being used,
- the scheduling of the different media types and the synchronization among them.

Before a certain multimedia communication service can be implemented, a transportation network needs to be chosen to carry the multimedia traffic. As outlined above, the attributes of a transportation network are critical and need to be analysed. A network can be circuit switching or packet switching. It can provide broadband or narrowband services. It can be connection-oriented or connectionless. It can provide a Quality-of-Service (QoS) architecture [AUR98] or a best-effort service. An ATM network is a packet switching broadband digital network. It offers connection-oriented services and provides Quality-of-Service (QoS) guarantees. Its suitability to support distributed multimedia applications is reported in [GUH95].

Secondly, digital communication networks in general are influenced by network impairments such as delivery delays and information errors. ATM networks are no exception. Although ATM networks are capable of providing a guaranteed QoS to traffic streams, QoS guarantees do not imply error-free transmission of information. The occurrence of network impairments is inevitable in most instances. Even in situations where network bandwidth is available and the physical transmission media is relatively reliable, multimedia traffic in ATM networks will still experience a certain degree of delays and errors.

Thirdly, a digital video compression technique is chosen to encode the video streams used within a multimedia communication application. When the quality of such an application is evaluated, the video encoding technique employed should be taken into account. This is because encoding schemes cope with network impairments differently. In other words, certain encoding techniques can be more vulnerable to the influence of network impairments than others. A video encoding technique can be reversible (lossless compression) or irreversible (lossy compression). Compression can occur within a single video frame (intraframe compression) or among multiple video frames (interframe compression). The nature of the compression technique and the presence of dependencies among video frames have an important bearing on how network impairments affect video quality at the destination [VER98b]. Today, many video-encoding standards exist and are employed by different multimedia communication applications. Among such standards, MPEG, motion JPEG and H.261 are commonly used to encode video in these applications.

Lastly, the scheduling and synchronization mechanisms employed by the multimedia communication applications have a direct impact on the quality perceived by end-users. As described above, synchronization is one of the more important user-level requirements. It entails the inter-media and inter-participant aspects which are determined by the scheduling and synchronization mechanisms respectively. With regard to video quality in multimedia communication applications, different packetization schemes for video information need to be considered since ATM networks are packet switching in nature. It is possible for the choice of packetization schemes to influence the ability of video encoding techniques to recover from the occurrence of network impairments. Video quality at the destination can be affected as a result.

1.1.4 The Quality-of-Service (QoS) Concept

The general concept of Quality-of-Service is important at two different levels in a multimedia communication application. At the application level, QoS is user-oriented. In the ITU-T Recommendation E.800, QoS is defined as the "Collective effect of service performances which determine the degree of satisfaction of a user of the service". Application QoS has an end-to-end significance. It can be described by subjective opinions or objective measurements at the point which the service is accessed by an end-user.

At the network level, QoS is often used to refer to Network Performance, which is defined at the boundary of the communication network and is therefore independent of terminal performance and user actions [I.350]. The ITU-T and the ATM Forum have defined a set of ATM layer QoS parameters to describe Network Performance. These QoS parameters are also known as network performance parameters and they constitute the Quality-of-Service at the network level. In other words, they define the Quality-of-Service guarantees provided by the ATM layer towards traffic streams.

Throughout this dissertation, "Quality-of-Service" is used as a generic term. However, the distinction between QoS perceived by end-users and QoS provided by the ATM network is extremely important. Therefore, specific use of this term to describe application level QoS or ATM layer QoS will be indicated accordingly.

1.1.5 The QoS Translation

In order to provide satisfactory video support for multimedia communication applications, it is important to translate the application level QoS requirements into ATM layer QoS parameters [JUN96]. Although the network level QoS determines the application level QoS perceived by end-users to a large extent, it is difficult to expect ordinary users to understand the meaning of the ATM layer QoS parameters. The mapping between QoS representations at different levels relieves users from considering network level QoS. It enables network level QoS to be translated to a more meaningful representation to the end-user during resource allocation and admission control.

This translation between application and network level QoS is by no means straightforward. Firstly, the user-oriented QoS described above cannot be used to specify the network performance requirements of a multimedia application. As stated in [I.350], although Network Performance ultimately determines the user observed QoS, it does not necessarily describe the quality in a way that is meaningful to end-users. Secondly, multimedia applications have diverse QoS requirements. For example, both video-on-demand and videoconferencing carry video traffic, but their QoS requirements are quite different. Thirdly, the effects of ATM network impairments on video quality vary from one multimedia application to another depending on the application attributes and the video encoding technique. Moreover, this translation depends significantly on the terminal equipment implementation as well as the protocol stacks and scheduling scheme employed. Therefore, it is not clear if a simple one-to-one relationship exists between the application QoS and the network QoS [GRI98]. Any attempt to establish the mapping between application QoS to network QoS needs to focus on one application at a time. The results of this mapping should apply to the application under consideration only.

1.2 Motivation and Dissertation Objectives

In an attempt to establish the mapping between the application QoS and the network QoS, the effects of ATM network impairments on video quality are studied. The study herein analyzes video quality in multimedia communication services by considering a Video-on-Demand application, which is a real-time asymmetrical communication application carrying time-based information. A VoD service can either be point-to-point or point-to-multipoint depending on its objective. Figure 1-1 shows the basic components of a VoD system.

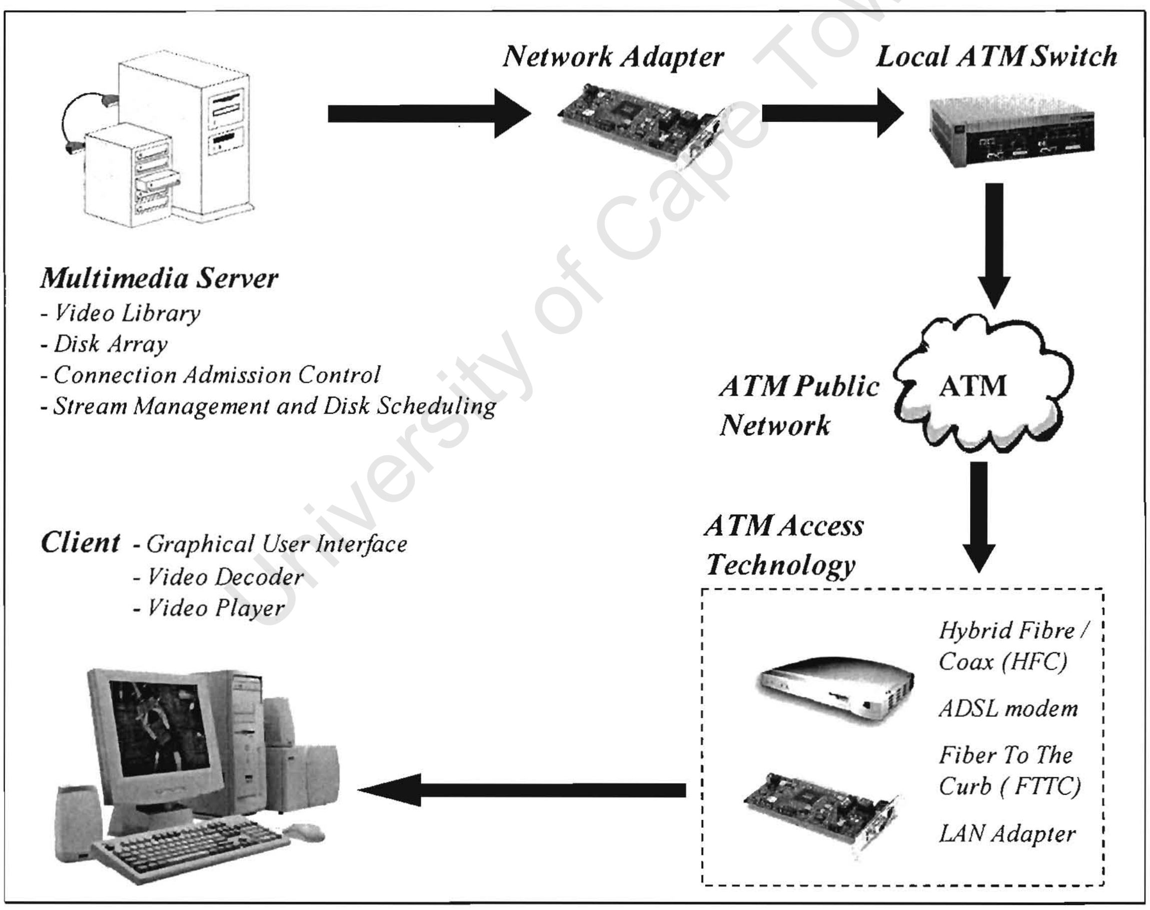


Figure 1-1 Basic components of a VoD system [ROU99]

The effects of ATM network impairments on video quality in a VoD system can be analyzed by first studying the different types of impairments in an ATM network. These effects can then be demonstrated by inserting controllable network impairments upon video traffic streams. The degradation in video quality can also be quantified by collecting individual opinions and subjective feedback from VoD users.

Although network impairments experienced by video traffic streams can be monitored and measured with the use of test equipment, controlling the level of network impairments is not a trivial task [HP95]. Depending on the amount of traffic carried by an ATM network, commercial ATM network products such as switches and multiplexers introduce impairments to passing network traffic. However, commercial ATM products do not provide the ability to directly control the level of the impairments. Even if deliberate overloading of commercial ATM equipment forces impairments to occur, this method of evaluating network impairments is inefficient and inflexible. Therefore, a specialized tool is required to allow network impairments to be introduced in a controlled manner. The implementation of this tool not only allows for video distortions caused by network impairment to be demonstrated, but also enables investigation on the relationship between application QoS and network QoS to be carried out.

This dissertation studies the effects of ATM network impairments to video quality[4]. It presents the design of the specialized architecture that is capable of providing cell switching functions and introducing controllable network impairments to digital video streams in the form of an emulated ATM network. Network impairments of a specific degree are imposed on traffic passing through the emulated network according to numerical impairment parameters. End systems connected to this emulated network send and receive video traffic normally as if connected to an ATM network. The effects of the network impairments introduced can be observed on the end-system receiving video information. This architecture is called NetIE, which stands for Network Impairment Emulation, meaning network emulation with impairment insertion.

NetIE aims to provide a general architecture which enables the evaluation of how video traffic traversing ATM networks is influenced by network impairments. It serves as a platform on which network impairments can be introduced in a controlled manner so that the connection between application QoS and network QoS can be established. Although these network impairments occur at the physical, the ATM and the ATM Adaptation layers as well as more than one physical location, they can be aggregated and categorized into five types of emulated network impairments within this architecture because their effects are additive. Through the use of the NetIE models, video quality degradation under the influence of network impairments can be demonstrated. In addition, the emulated network can also be used as a general test platform for studying the effects of network impairments on other types of traffic from a wide range of applications.

---

[4] The quality of audio associated with the moving pictures is examined occasionally, but the focus of this study remains with video quality.

1.3 Related Research

There are a number of research projects that focus on various issues addressed by this study including the translation of application QoS to network QoS, the quality of digital video over ATM networks, the effects of ATM network impairments to video quality etc. A list of research that are of close relevance to the main objectives of this thesis is presented in this section. Other researches that are related to this study are inserted in the appropriate sections.

Gringeri et al [GRI98] present a broad overview of issues related to the transmission of MPEG-2 video streams over ATM. They develop a platform with the ability to introduce network impairments to video traffic streams. Using this platform, they reveal the effects of ATM network impairments on video quality and describe various types of digital video and video artefacts. They also present quantitative results on the effects of CRC checking on video quality under the influence of physical bit errors and ATM cell losses. Although this work is similar to the NetIE architecture in principle, little detail with regard to the implementation of the platform is provided. Moreover, instead of the frame by frame viewing adopted in their studies, subjective opinion of end-users is used to measure video quality in this study.

Jung [JUN96] studies the Quality of Service (QoS) concept. He addresses the importance to translate user requirements into ATM layer QoS parameters as well as the translation of QoS parameters between layers. He confirms that user's QoS requirements depend on each application and user, and that application QoS can be greatly different from the ATM network performance. His work defines the generic QoS parameters in the ATM Adaptation Layer (AAL) and ATM layer as well as the ATM network performance parameters. QoS translations from Segmentation and Reassembly (SAR) sublayer to ATM layer, from Convergence Sublayer (CS) sublayer to ATM layer and from CS sublayer to SAR sublayer are illustrated in details. He further shows that QoS translation is a possible and good approach to end-to-end QoS guarantees.

Guha et al [GUH95] demonstrated the suitability of ATM technology for distributed multimedia applications through the construction of an ATM LAN testbed known as Mercuri. Experiments are performed to compare the use of native ATM, TCP/IP over ATM and TCP/IP over Ethernet in terms of throughput and delay, and to show how video frame delivery using native ATM was affected by the activities on the client and server processors and the network. End-to-end delay and jitter in video delivery are measured under no loading, source loading and destination loading conditions. From the results obtained, they conclude that performance bottlenecks are present in end-systems, such as the interaction of the adapter card with the host CPU, the video hardware and the host operating system.

Tsang et al [TSA96] constructed a video testbed in an ATM LAN and evaluated the end-to-end performance of digital coded video. They illustrated the use of a video testbed and the advantages of it over simulation models. Various experiments were carried out on throughput, frame loss and jitter with respect to the number of multiplexed video streams. Although the experiments are performed using TCP/IP and UDP/IP over ATM instead of native ATM, their results show that high burstiness in variable bit rate video streams causes high packet losses.

Lin et al [LIN96] consider the MPEG encoding technique and show that the packing of video information is an important factor in delivering digital video across the network efficiently. They identified the two main factors that determine maximum throughput: 1)software processing (depending on host CPU) and 2)information transfer speed (depending on I/O bus and the physical adapter). Different packing schemes for MPEG information are presented. They conclude that when small protocol data units (PDU) are sent, the throughput is limited by the ability of sender to send out PDUs. When large PDUs are sent, throughput is limited by performance of adapter and host bus.

Verscheure et al [VER98a] emphasize the importance of Quality-of-Service perceived by end-users in the delivery of packet video. They considered the entire video delivery process and identified three areas (the source coder, the network, and the decoder) that can impair video information. An experimental testbed is constructed to study the these three areas. They indicate that during transmission of video information, data loss that reduces the quality is dependent on the importance of the lost information type. They also point out that the effect of network impairments on user-perceived quality is not well understood and that the assessment of end-user quality is not trivial. Various quality metrics used to measure video quality are presented with the aim of measuring user perceived video quality. In contrast to their research, this study focuses on the impairment of video quality by an ATM network only.

The Broadband Series Test System from Hewlett Packard [HP-1] is a modular test platform for ATM, broadband and digital video transmission and protocol testing. One of the modules is the ATM network impairment emulator [HP-2], which emulates an imperfect ATM network. Together with a Broadband Series Test System base platform [HP-3] and a line interface module [HP-4], it allows five independently-controlled impairment types to be introduced to external ATM traffic of any type. The existence of such commercially available *test & measurement* equipment supports the Network Impairment Emulation architecture proposed and adopted in this dissertation.

1.4 Development Methodology

The study of ATM network impairment on video quality requires in-depth understanding of: 1) the characteristics of an ATM network, 2) why video compression is required, 3) what types of video compression technology can be used, 4) how video is sent over an ATM Network, 5) what are ATM network impairments, 6) why do they occur, and 7) how to emulate an ATM network. Knowledge in these areas helps understand and formalize the problem of video quality degradation resulting from ATM network impairments.

Having acquired the fundamental knowledge, the next step is to review previous works in the field of video transportation over ATM and previous studies on the effects of ATM network impairments on video quality. The review of literature helps to define the NetIE architecture as well as the functional requirements of the emulated ATM network. Conceptually, this emulated ATM network has five input parameters, each corresponds to one type of network impairments under study. The emulated network processes video information and introduces network impairments according to the values of the input parameters. The resulting video information is then decoded so that video artefacts and quality degradation can be studied.

Two designs of the emulated ATM network originated from the functional requirements defined in the previous phase. The first design approach involves the functional simulation of an ATM network as a whole including the network interfaces at end-systems, the intermediate ATM nodes and the network impairments. The second design approach makes use of existing end-systems that send and receive digital video and connects them via a virtual ATM switch to form an emulated network. The virtual ATM switch is an abstraction of a real-world ATM network and is based on the Linux operating system. It carries out the task of switching ATM cells and introducing network impairments. The two design approaches are compared and the second approach is preferred because of its flexibility and efficiency.

Network impairments can be introduced within the virtual ATM switch at three different levels, namely, the adapter driver level, the kernel level and the user level. The advantages and disadvantages of introducing network impairments at these levels are investigated. It became clear that a user level impairment insertion architecture should be adopted in the virtual ATM switch not only because it suits the requirements of NetIE, but also for its portability and simplicity.

Following the implementation of the emulated network, a multimedia communication application is used to send traffic over it in order to evaluate the effects of ATM network impairments on video quality. The Video-on-Demand application[5] chosen for this study generates ATM traffic containing digital video in the MPEG compression format over virtual channel connections (VCCs) that are established within the emulated network. One of the most important advantages in adopting the virtual ATM switch is that if the effects of ATM network impairment towards another compression format or another application needs to be studied, only the application running on the end-systems needs to be changed. No modification is required in the emulated network otherwise.

The integration of the VoD application into the emulated network enables the demonstration of video and audio degradations resulting from the occurrence of network impairments. During the testing phase, it became clear that the CRC checking on video information has a significant effect on video quality under the influence of network impairments as reported in [GRI98]. The spatial and temporal propagation of errors and losses in the video stream are studied in the emulated network and possible solutions against network impairments are investigated.

1.5 Plan of Development

Chapter 2 presents an introduction to network impairments and a system overview of the NetIE architecture. A comparison is made between network impairment emulation and in-service testing. The former is shown to be a more flexible and cost-effective means to investigate the effects of ATM network impairment to video quality. This chapter also explains the use of network impairment parameters in the NetIE model and their relevance to the ATM QoS performance parameters in a traffic contract. The functional requirements and desirable features of the NetIE architecture are summarized. Two NetIE models, the standard Statistical Model and the more comprehensive Application-Specific Model, are described and investigated. In the Statistical NetIE Model, network impairments are introduced solely according to statistical distributions and parameters. The Application Specific Model is an extension to the Statistical Model in that the type of information carried by the ATM cells is taken into account as well. This chapter concludes by presenting an outline of the background research required for studying the effects of ATM network impairment to video quality.

Chapter 3 introduces essential background research for the understanding and the development of the two NetIE models. As discussed earlier, there are a number of factors that influence quality of multimedia communication applications in general. Therefore, the investigation on how video quality is impaired in an ATM network cannot be complete without a thorough

[5] The Video on Demand application service infrastructure was jointly developed by Mr. M.J. Roux and Mr. P. Vine in the Communications Research Group (CRG).

understanding of these factors. This chapter highlights ATM networking issues that are critical to the design and implementation of NetIE. Since the multimedia service considered in this study utilizes MPEG encoded video, knowledge on the MPEG compression standard is necessary in order to understand why and how MPEG video quality degradation occurs under the influence of ATM network impairments. The syntax of the MPEG format is especially important to the Application-Specific model because of the need to categorize various information fields within a MPEG encoded stream. In addition, issues concerning the transportation of digital video over an ATM network will be examined. These include the transmission of video stream using constant-bit-rate (CBR) or variable-bit-rate (VBR) service and the support of MPEG video traffic using AAL 1 and AAL 5. Most importantly, this chapter defines and explains the fundamentals of ATM network impairments, which are essential to the investigation on how these impairments affect video traffic streams and the degradation in video quality that occurs as a result.

Chapter 4 is devoted to the design methodology for the emulated network that is implemented to fulfil the NetIE architecture. Two design approaches are presented. The first approach involves the functional simulation of an ATM network as a whole including the network interfaces at end-systems, the intermediate ATM nodes and the network impairments. The NetIE architecture is broken down into three main modules and the functions carried out by each module are described in details. The second approach tackles the problem differently and aims to construct an emulated ATM network to which physical ATM end-systems can connect. This emulated network is capable of providing cell switching functions as well as introducing controllable network impairments to video traffic passing through. The design of the emulated network in the form of a virtual ATM switch is presented in details. Three proposals on where the Impairment Insertion Modules should reside within the software architecture of the virtual ATM switch are illustrated and the implications of each scheme are explained. A comparison is made between the two approaches and the design considerations are examined.

Chapter 5 describes the implementation of the emulated network architecture designed in the previous chapter. The importance of implementation should not be underestimated because it validates the NetIE architecture and tests the design described in the previous chapter. The emulated ATM network is implemented in the form of a virtual ATM switch based on the Linux operating system. The Impairment Insertion Modules are capable of processing ATM traffic on a cell-by-cell basis. For the Statistical NetIE Model, each Impairment Insertion Module introduces one type of network impairment according to statistical parameters and a certain statistical distribution such as Gaussian and Uniform. The generation of the statistical events is explained in details and the mechanisms to insert the five types of impairments into the ATM cell stream are described. This chapter introduces the ATM Network Impairment Emulator within the Broadband Series Test System from Hewlett Packard. This Network Impairment

Emulator is functionally similar to the Statistical NetIE model. A comparison made between this emulator and the NetIE implementation reveals that there are some differences in the mechanisms used to generate impairment events. For the Application-Specific NetIE model, the information carried by the ATM cells needs to be examined and categorized so that impairments are inserted to specific type(s) of information only. This chapter also describes the procedures with which the virtual ATM switch and the end-systems are connected to form the emulated network.

Chapter 6 presents the effects of ATM network impairments to video quality. It characterises and describes various video and audio artefacts observed in MPEG digital video streams resulting from the influence of ATM network impairments. It examines the influence of the AAL error checking mechanisms on video quality and analyzes the effects of spatial and temporal propagation of errors and losses in video information on video quality. An attempt is made to investigate the relationships between the occurrences of the video and audio artefacts identified and the type of network impairment that causes the artefacts. This chapter also illustrates how user-oriented tests can be conducted to assess the overall quality of a particular video sequence under the influence of defined levels of network impairments. It outlines the procedures of these user-oriented tests and presents some preliminary test results to show the feasibility of this method. This chapter further describes how to establish the required network-level QoS that results in an acceptable user-level QoS and explains why this is not a trivial task.

Chapter 7 and Chapter 8 present conclusions and recommendations for future work respectively. It is concluded that the network impairment emulation (NetIE) models describe an efficient and flexible architecture to investigate the effects of ATM network impairments to video quality. The generic nature of the NetIE implementation using a virtual ATM switch ensures that it is not only applicable to MPEG video traffic, but also suitable to apply to video traffic in other encoding formats, as well as any type of ATM traffic in general. It has been shown that NetIE can be used to establish the mapping between network-level QoS and user-level QoS. It is therefore recommended that NetIE should be further developed and applied to other types of ATM traffic so that its potential can be fully exploited.

Chapter 2.

Introduction to Network Impairment Emulation (NetIE)

This chapter introduces ATM network impairments and network impairment emulation. It explains the relationship between QoS performance parameters used in an ATM traffic contract and network impairment parameters used in the emulated network. It then compares two methods to study the effects of network impairments on ATM traffic such as digital video. More importantly, this chapter also presents the Network Impairment Emulation (NetIE) architecture as well as the Statistical and the Application-Specific (AS) models. The AS model is an extension to the basic Statistical NetIE model and is developed specifically to experiment with MPEG video traffic. The functional requirements of the two NetIE models are described and the operations of both models are explained and compared. This chapter forms the basis for the development of the network emulation architecture which allows for the analysis in the quality of multimedia communication applications. The development process of this network emulation architecture is presented in Chapters 4 and 5.

2.1 ATM Network Impairments

This section introduces the concept of network impairment and the five types of network impairments that occur within an ATM network. Video traffic, as well as other types of traffic, is affected by varying degrees of impairments in the ATM network when it is travelling through the network. An impairment is defined to be an error or degradation in network performance. It includes the inaccurate and delayed delivery of network traffic from the source to the destination. It is important to distinguish between network impairments and equipment failures [HP95]. While the former occurs in an ATM network inevitably even under normal operations, the latter mainly results from mis-configured or defective network equipment. For example, an improperly configured ATM switch may send cells to an invalid connection or an incorrect destination. Moreover, ATM cells may be lost if a defective switch does not pass on traffic to the next network node. These two scenarios of equipment failures may be resolved by the network operator with appropriate corrective measures. Equipment failure is beyond the scope of this thesis, which focuses on ATM network impairment only.

Within an ATM network, impairments occur at the physical, the ATM and the ATM adaptation layers as well as more than one physical location. Network impairments experienced by ATM traffic can be classified into five types. They are defined by [I.356] and [TM4.0] as follows:

- Cell Error – refers to the occurrence of bit errors within an ATM cell and is characterized by the Cell Error Ratio (CER) parameter;
- Cell Loss – refers to the case when an ATM cell does not arrive at its destination and is characterized by the Cell Loss Ratio (CLR) parameter;
- Cell Transfer Delay – refers to the time taken for a cell to be transferred from its source to its destination and is characterized by the Cell Transfer Delay (CTD) parameter;
- Cell Delay Variation – refers to the variable Cell Transfer Delay experienced by different ATM cells and is characterized by the Cell Delay Variation (CDV) parameter;
- Cell Mis-insertion – refers to the arrival of ATM cells at the destination that do not originate from the source and is characterized by Cell Mis-insertion Ratio (CMR).

The descriptions of where these impairments occur in an ATM network and their formal definitions will be presented in Chapter 3.4.

2.1.1 The ATM Traffic Contract

Since the occurrences of impairments is unavoidable in any type of communication networks including ATM, the Quality-of-Service principle and the Traffic Contract are introduced to ATM in an attempt to provide satisfactory services to users of networked applications and to protect network resources. A traffic contract is negotiated between end-systems and the network. It is an agreement between a user and a network regarding the Quality of Service (QoS) that a cell flow is guaranteed if the cell flow conforms to a set of traffic parameters [MCD95]. In principle, the traffic contract specifies the negotiated characteristics of a connection. It not only contains the characteristics of ATM traffic the network agrees to accept and the user promises not to exceed, but also defines the Quality-of-Service which the user is satisfied with and the network commits to provide. It is important to note that the negotiated QoS objective specified in a traffic contract is the worst case of network performance that the network will provide. During periods when network loading is much less than then engineered capacity, the network performance may be significantly better than the negotiated objective. In addition, the level of ATM traffic the network is prepared to accept specifies the upper limit of traffic the user can submit to the network but does not imply that the user must send traffic at that level all the time.

The traffic contract consists of three main sections [TM4.0]: 1)the Connection Traffic Descriptor; 2)the requested QoS, and 3)the definition of a compliant connection. The Connection Traffic Descriptor includes the source traffic descriptor, the cell delay variation tolerance (CDVT)[6] and the cell-by-cell conformance definition based on the Generic Cell Rate Algorithm (GCRA). The source traffic descriptor is a set of traffic parameters that describe the traffic characteristics of a source such as the traffic rate. It defines at least the Peak Cell Rate (PCR) (as specified in [I.371] and [UNI3.1]) and may optionally define a Sustainable Cell Rate (SCR), Maximum Burst Size (MBS) and Minimum Cell Rate (MCR) [TM4.0].

The QoS of a connection is described by the following Quality-of-Service performance parameters [TM4.0]:

- Cell Loss Ratio (CLR)
- Cell Error Ratio (CER)
- Severely Errored Cell Block Ratio
- Cell Mis-insertion Rate (CMR)
- Cell Transfer Delay (CTD)
- Cell Delay Variation (CDV)

The user of a virtual path connection (VPC) or a virtual channel connection (VCC) is provided with one of a number of QoS classes supported by the ATM network [I.150]. The ITU-T Recommendation I.356 and the ATM Forum User-Network Interface Specification 3.1 have defined a number of QoS classes with specified QoS performance parameters. Users can request and receive different QoS classes on a connection-by-connection basis in order to meet the distinct performance requirements of a vast variety of services and application. In addition to QoS classes, [TM4.0] states that QoS may also be negotiated in terms of individual numeric parameters using the procedures defined in [UNI4.0] and [PNNI-1].

The precise definition of a compliant connection is network specific and does not necessary require all cells to be conforming. In other words, while a connection for which all cells are conforming shall be identified as compliant, it is possible for non-conforming cells to exist in a

[6] CDVT is a mandatory parameter in any connection traffic descriptor and can either be explicitly specified at subscription time or implicitly specified. It specifies the tolerance of an ATM network for incoming traffic that exceeds the PCR value. CDVT is not the same as CDV, which is a QoS parameter.

compliant connection if the network operator finds them acceptable. According to the Conformance Definition, the agreed QoS objectives should be met for at least the number of cells equal to the conforming cells in a compliant connection. On the other hand, if a connection is determined to be non-compliant by the network, the agreed QoS objectives need not be respected by the network [UNI3.1].

2.1.2 ATM QoS Parameters

In order to validate whether the network is providing the QoS that it has committed to a connection, the transfer characteristics of the network are measured either on a single ATM cell or a sequence of ATM cells during the lifetime of an ATM virtual connection. The results of this measurement are compared against the ATM QoS parameters included in the traffic contract. At this point, it is important to realize that these QoS parameters are statistical in nature, i.e. they are measured over a sample of cells that is appropriately large for the particular QoS parameter. It is equally important to note that while the CLR, CTD and CDV QoS parameters can be negotiated, the CER and CMR QoS parameters are not negotiated. The difference is that it is possible for the negotiated QoS parameters to be different among connections depending on how the connections are handled. However, the non-negotiated QoS parameters are dependent on the inherent characteristics of an ATM network and are therefore the same across all virtual connections, i.e. cannot be changed during the operation of the network.

Depending on the capability of individual implementation of end-systems, QoS can be negotiated in terms of individual numeric parameters or at least QoS classes. Whichever way is used, it is very important for applications, especially video applications, to receive an appropriate QoS from the network.

From the point of view of an end-user or a video service provider, if the QoS requested for a video connection is too strenuous (i.e. the QoS requested is well above what is appropriate), the connection request may be refused due to a lack of enough resources in the network. Even if there is enough resources in the network, the connection may have to pay more due to the strict QoS. On the other hand, if the QoS request is too loose (i.e. QoS is below what is required), the quality of video delivered at the destination may be unsatisfactory.

From the point of view of a network provider, if the QoS of many connections are too strict, network resources will not be utilized efficiently. As a result, the network will not be able to provide services to as many users as it should and depending on the pricing strategy, the revenue generated might decrease as a result. Conversely, an unsatisfactory network service will be provided if the QoS is too low. Therefore, in order to utilize the network resources better and to ensure that video quality is acceptable at the receiving end, it is very important to determine the maximum level of network impairment video traffic can tolerate before quality drops below a satisfactory level. The ATM QoS parameters corresponding to this condition can be used as a guideline for the negotiation of QoS during ATM connection establishment. This process of mapping application level QoS to network level QoS by means of user-oriented video quality assessment will be briefly outlined in section 2.2. The procedures of these user-oriented assessments will be described in details in Chapter 6.

From the above discussion, the importance of mapping the application level QoS to the QoS in the network level is apparent if the interests of both the network providers and end-users are to be met. This is one of the goals the network emulation architecture presented in this dissertation aims to achieve.

2.1.3 ATM QoS performance parameters & network impairment parameters

At this point, an observant reader might have noticed that the five ATM QoS parameters are defined corresponding to the five types of network impairments experienced by ATM traffic. The ATM QoS performance parameters are used in a traffic contract to characterize the QoS commitment from the network towards the traffic stream in a virtual connection. In the opposite sense, these QoS parameters also characterize the amount of network impairments that the network is permitted to insert and that a traffic stream needs to expect and accept. This is because even when all the cells within an ATM connection are considered to be conforming, it is still possible for network impairments not exceeding the level define by ATM QoS parameter to occur for the ATM network to fulfil the traffic contract. QoS of a connection does not imply zero network impairment. It only specifies the 'worst-case' situation for the level of ATM network impairments. Therefore, it is appropriate for ATM network impairments to be characterized by the same parameters that define ATM QoS performance.

The values of the impairment parameters within the emulated network architecture are very important. This is because the use of the parameter values obtained in the emulated network in the traffic contract is very likely to produce video quality equal to or better than that shown in the emulated network. If quality of a certain type of video sequence is found to be good in the user-oriented video quality assessment survey introduced in chapter 1, the use of the same values of performance parameters in the traffic contract for an ATM connection is going to result in video of good quality with a high level of confidence. Moreover, the quality of video might even be better over an ATM network than through the emulated network if the ATM network carries a load well within its limit and its performance exceeds that of the committed level.

The mapping of application QoS to network QoS is the responsibility of the application to a large extent because it is the application which requests QoS for a VCC from the network during connection establishment. If the network provides the bandwidth and QoS that were requested for and the resulting quality of the application is unsatisfactory, the network should not be held responsibility as it is the application which has not provided adequate QoS requirements to the network. Therefore, the implementation of any ATM based applications, including multimedia communication applications, needs to consider the QoS performance requirements of their VCCs. The network emulation architecture proposed in this dissertation will help these applications to discover the QoS requirement of their ATM connection(s).

Despite the use of the same parameters, there is one main difference between QoS performance in an ATM network and network impairment in the emulated network. In an ATM network, performance parameters are measured and calculated from the events that occurred within the network[7]. On the other hand, network impairments can only be modelled and generated within the emulated network. Although the same methods can be used to measure the generated impairments, there is no standardized method of how to generate network impairment defined in either the ITU-T or the ATM Forum. Various methods are proposed in the literature in order to characterize the occurrences of network impairments in an ATM network so that these events can be generated accordingly in an emulated network. These methods are mainly based on the statistical nature of these parameters and define the specified impairment parameters as well as the way impairment events are distributed. Details about these methods employed in the implementation of the emulated network will be discussed in Chapter 5.

2.2 Network Impairment Emulation (NetIE)

In order to provide adequate support for multimedia applications over ATM, the underlying ATM network needs to satisfy the application level QoS requirements. However, the translation of application level QoS requirements of digital video streams into ATM network QoS performance parameters depends significantly on the end user, terminal equipment and protocols [GRI98]. Moreover, it is not clear if a simple one-to-one relationship exists between the application QoS and the ATM QoS performance parameters. Therefore, it is important to investigate how video quality is affected when digital video is transported over an ATM network. One of the ways to carry out this investigation is to carry out Network Impairment Emulation (NetIE).

[7] ITU-T I.356 and ATM Forum TM4.0 defines the methods to measure these ATM performance parameters

In general, Network Emulation involves the design of an emulation architecture which imitates a real-world network. For example, the emulated network should consist of network nodes connected by physical or logical links. It should also provide the normal functions that a network provides, such as transporting information traffic among network nodes over the transmission medium, providing services to higher-level applications etc. Although it is desirable for an emulated ATM network to be as similar to a real-world ATM network as possible, the focus of this particular emulated network is to emulate the impairments an ATM network imposes on traffic streams. In this case, the term Network Impairment Emulation (NetIE) is used to describe network emulation with impairment insertion.

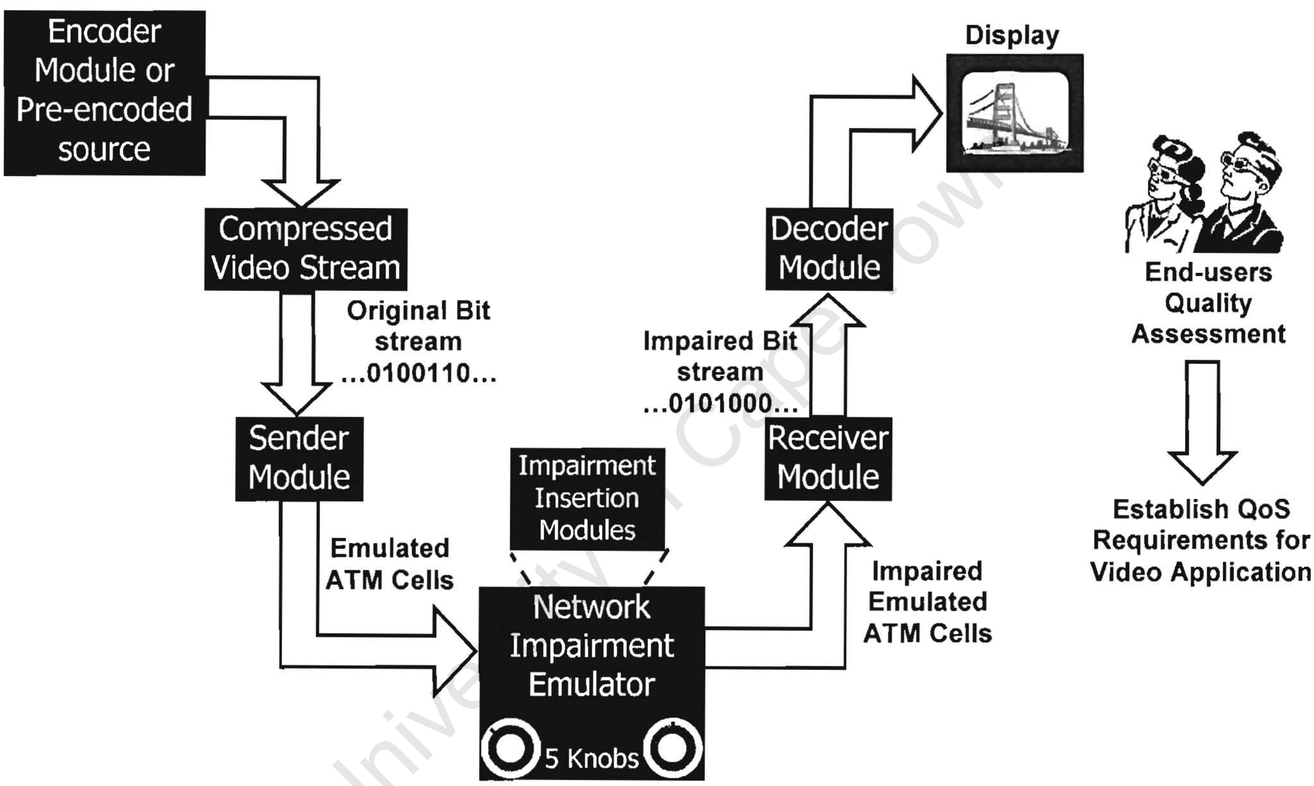


Figure 2-1 Function block diagram for the emulated network

In the context of this thesis, the primary purpose of the emulated network is to demonstrate how the quality of video presented to the end user is affected by the different levels of ATM network impairments. Figure 2-1 shows the functional block diagram for the emulated network. The core of the emulated network is the ATM network impairment emulation module. In a typical scenario, the sender module receives user data (in the form of compressed video bit stream made up of 1's and 0's) and performs AAL and ATM layer functions as if the use data were sent from an end system in an ATM network. The resulting emulated ATM cell stream is passed to the network impairment emulator, which possesses, conceptually, a few 'knobs' with each one acting as an user input to control one of the network impairment parameters: CER, CLR, CTD, CDV and CMR. Five impairment insertion modules (IIM), each one corresponds to one of the parameters, insert the respective type of network impairment at a level according to the values specified by the 'knobs'. Having traversed through the five IIMs, the emulated ATM cell stream is passed to the receiver module for end system AAL and ATM layer processing (including possible error correction). The reassembled bit stream is decoded by the video decoder module and displayed by the video-rendering module on either a computer monitor or a TV screen.

Once the implementation of the emulated network is completed, video and audio artefacts in the MPEG resulting from the influence of the emulated ATM network impairments can be observed. The relationships between video quality degradation and network impairments can be investigated by considering either individual or a mixture of impairments. The assessment of overall video quality can be achieved by conducting user-oriented tests in the form of a 'survey' to collect users' impression. The data collected during the survey can then be statistically analyzed in order to provide an insight to the mapping of network QoS and user QoS.

2.3 In-Service Testing

Since all networks impose some degree of impairment to traffic streams, another way to carry out an investigation on ATM network impairment is to connect data terminal equipment to each side of a network (or a single ATM switch) and perform in-service testing (IST). One of the ways to perform in-service testing involves 'load generation', where an ATM switch is loaded with background traffic near its carrying capacity [HP95]. A source of test traffic is then connected to the switch, making it handle an amount of traffic that approaches or even exceeds its capacity. As a result, impairments are forced to occur.

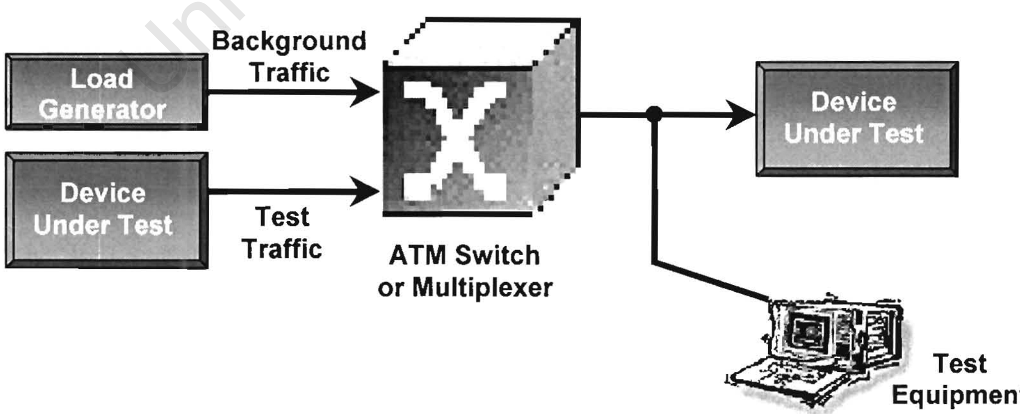


Figure 2-2 Load generation during In-Service Testing

At first sight, IST may seem more realistic because it involves the use of real ATM equipment. However, it has many drawbacks when compared to network impairment emulation. Firstly, in-service testing requires an existing and functional network, which may not be available.

Secondly, even if it is possible for a functional network to be built, the cost involved will be much higher than building an emulated network, if not prohibitive. Most importantly, during in-service testing, network impairments are neither directly produced nor controlled because the network equipment is not operating under controlled conditions. Although the use of test equipment within the in-service testing network may be able to quantify the levels of impairments the network exhibits, it is difficult to repeat any given impairment level because of the nature of the resulting impairments, which are by-products of the overloaded equipment. In addition, the network equipment may exhibit inter-dependencies among the different types of impairments, i.e. a specific level of one type of impairment cannot be achieved without the occurrence of another type(s). As a result, a desired level of a specific impairment type cannot be precisely controlled and generated. For example, overloading an ATM switch with a certain level of background traffic may result in a CLR of 10^{-6}. Now if a CLR of 10^{-5} is required, the amount of background traffic load needs to be reduced somewhat to a new level. This required level of background traffic is determined on a trial-and-error basis and may vary between different tests over time. This is because the ATM switch is not operating under a stable condition and the CLR resulted from the same level of background traffic is not likely to be constant. As a result, the level of background traffic needs to be adjusted for each test even if the same CLR is required. From the above discussion, it is obvious that this method is not only very time-consuming to set up, but also difficult to repeat.

On the other hand, Network Impairment Emulation provides a flexible and cost-effective way of evaluating ATM network performance. Since the testing environment in NetIE is designed and constructed specifically to perform impairment emulation, it provides the capability for different levels of various impairment types to be generated and controlled. Moreover, the occurrence each type of impairment can be made to follow the requirements accurately and the implementation of NetIE can ensure that there is no inter-dependency between two or more types of generated network impairments.

2.4 Functional Requirements and Features of NetIE

In a real-world ATM network, varying degrees of impairments are imposed on traffic streams depending on factors such as the current network load, the profiles of individual traffic streams, the QoS class etc. As a result, it is difficult to investigate the effect of network impairments on one particular type of traffic. In the emulated network, impairments are quantified so that their effects on each type of traffic can be studied in turn. The functional requirements and desirable features of the emulated network are summarized below:

- ATM network functions – the emulated ATM network has to provide the general functions performed by the ATM layers[8] (the AAL, the ATM layer and possibly the physical layer) as well as emulate the services that are offered by these layers to the application layer;

- Abstraction of real-world network impairments – network impairments occur at the physical, ATM and ATM Adaptation layers as well as different physical locations in an ATM network. In the emulated network, these impairments are aggregated and abstracted into the five types described in section 2.1 and defined in section 3.4.2.1 to 3.4.2.5 (represented by the parameters CER, CLR, CTD, CDV and CMR). The emulated network (Figure 2-1) provides the capability for introducing controllable impairments to the emulated ATM cells according to these parameters;

- Flexibility – although the objective of the emulated network is to investigate the ATM network impairment on video traffic and the resulting degradation in video quality, the network emulation architecture should be capable of being generalized with minor modifications to test other types of traffic as well;

- Modularity – a modular approach should be adopted during the design and implementation of the emulated network because it will not only break down the complexity of constructing the emulator, but also ensure no inter-dependencies exist among the generated network impairments. The modular approach breaks the emulator down into functional blocks, each contributing to the overall function of the emulator. The task of each functional block is carried out by a hardware or software module. These modules are integrated to form the emulated network;

- Scalability – the design of the emulator should allow for easy expansion in capabilities in the future. This is partly helped by the modularity approach described above: boundaries are defined among the modules and specified by a set of interface rules. If an extra capability needs to be added to the emulated network, the existing modules can be abstracted as 'black boxes' and the new module(s) with new features can be integrated into the existing ones with minimal effort.

2.5 The NetIE Models

Having reviewed the principles of network impairment emulation and the functional requirements of the emulated network, this section introduces two NetIE models, the Statistical

[8] One example of the requirements for the emulated network is to emulate switching functions at the ATM layer

model and the Application-Specific (AS) model. Although both models deal with one-way video traffic, applying these models in both directions can cater for bi-directional video communications.

2.5.1 The Statistical NetIE Model

The statistical NetIE model independently generates the five types of network impairments base on statistical parameters and distributions (Figure 2-3). The Sender and the Receiver can either be logical units or physical computer hosts depending on the design strategies. Their tasks are to send and receive MPEG video information respectively. As described in section 2.1.3, network impairments occur in the emulated network as generated events. In the NetIE model, the Impairment Controller is responsible for determining when each type of network impairment should occur according to a statistical parameter and a defined distribution. The Impairment Insertion Modules (IIM) follow instructions from the Impairment Controller and introduces network impairments to the MPEG video stream.

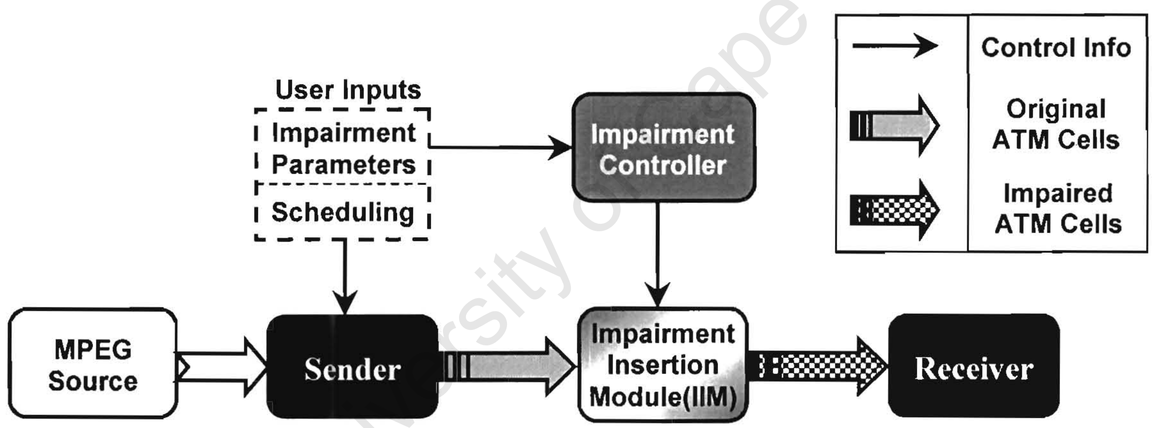


Figure 2-3 Functional Block Diagram of the Statistical NetIE Model

In a typical experiment, the sender takes in MPEG video streams from the source and schedules the information for transmission as physical ATM cells or logically simulated ATM cells. The Impairment Controller takes in values of impairment parameters, works out which ATM cells should be affected by the impairments and sends instructions to the Impairment Insertion Module. The IIM reads the instructions, inspects the passing cell stream and insert ATM network impairments accordingly. The impaired ATM cells are then passed to the receiver, which decodes and displays the MPEG video. The effects of network impairments on video quality can then be observed.

2.5.2 The Application-Specific NetIE Model

The Application Specific NetIE model is an extension to the Statistical NetIE model. The main difference between the AS and the Statistical model is that instead of introducing network impairments solely according to statistical distributions, the type of information carried by the ATM cells is taken into account as well. As outlined in chapter 1.1.3, it is possible for the implementation of a certain multimedia communication application such as VoD to choose among various video encoding techniques as well as different scheduling and packetization schemes. These choices determine how video information is packed into ATM cells and have an effect on how video quality is affected by ATM network impairments.

In terms of the MPEG encoded video, network impairments may affect specific information fields such as various header and synchronization fields, video (I, P, or B) frame content or audio packets. With the AS model, the effects of impairments occurring in each information field in MPEG video can be studied individually or combined to see how impairments in different fields interact with one another. It is reported in [GER98] and [VER98a] that video quality degradation resulting from data loss is dependent on the importance of the type of information being lost. For example, it is possible for information losses in different frame types to cause different degrees of degradation in video quality. Moreover, losses in various header fields affect video quality in a different way compared to losses of Discrete Cosine Transform (DCT) coefficients and motion vectors within video frames. Therefore, it is important to consider how the different types of information are allotted in ATM cells so that the influence of network impairments on different types of information can be studied. This is achieved by introducing network impairments to each type of information within the MPEG video stream individually.

Besides considering each type of information separately, a more general method is to classify the MPEG stream into three categories (video information, audio information and other data such as system information) and distribute the network impairments across the three types of information with defined percentages as described in [GRI98]. In this way, the effects of network impairments on particular type(s) of information can be evaluated. Since the choices of video encoding techniques, scheduling schemes and packetization mechanisms are specific to the type of application and the implementation of the application, this extended model is called the Application-Specific NetIE model.

Figure 2-4 shows the functional block diagram of the Application-Specific NetIE Model. In order to find the location of the various information fields in the MPEG stream and what type of information is situated in each ATM cell, the MPEG video stream from the MPEG source is passed to the Analyzer module and then the Segmentation module. The Analyzer module reads the MPEG streams and determines where each type of information resides by looking for various 'start codes' of MPEG [M13818]. Information is categorized as follows: video information (including I, P, B frames), audio information and system level information (including certain header fields). This information is then passed to the segmentation module, which works out how the MPEG stream will be segmented into ATM cells. The segmentation of the MPEG stream depends on the AAL type being used, the software scheduling employed by the sender and the size of AAL PDUs being sent.

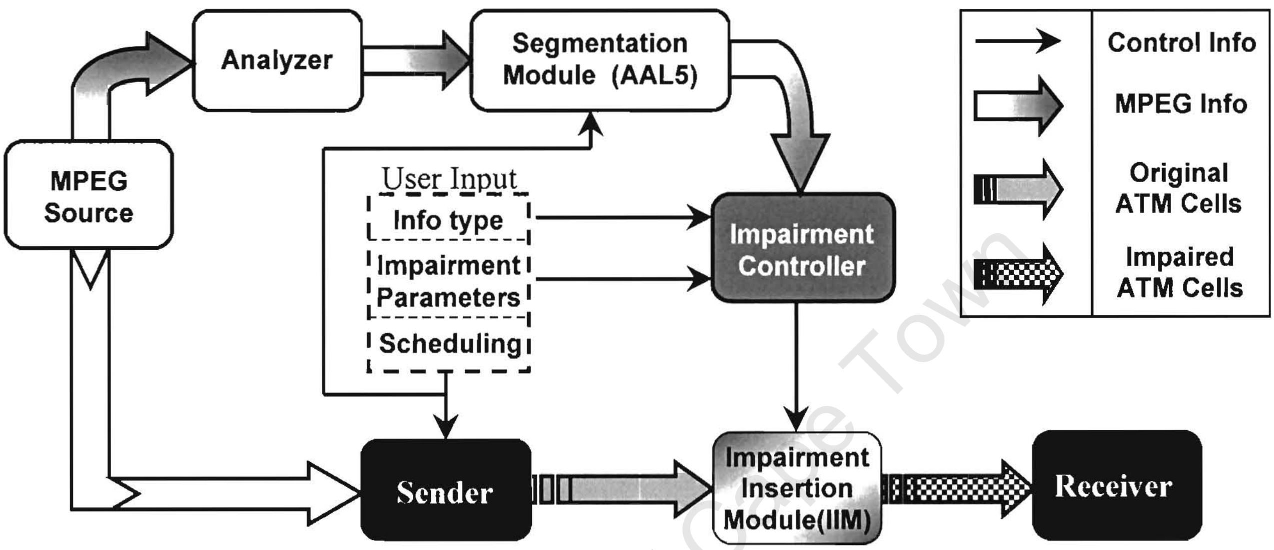


Figure 2-4 Functional Block Diagram of the Application-Specific NetIE Model

The Impairment Controller receives information about the type of content in each ATM cell. It also takes in values of impairment parameters in the same way as in the Statistical NetIE model. However, since the key to the application-specific NetIE model is to insert impairments to specific type of MPEG information, the Impairment Controller requires its user-input to indicate which type of information is under examination so that only those type of information in the MPEG stream is affected by the impairments. If more than one type of information need to be affected, the user-input also requires to indicate the percentages with which network impairments should be distributed. The rest of the application-specific model is the same as the statistical model.

2.5.3 Comparing the two NetIE models

The two NetIE models presented above have their share of advantages and disadvantages. They are used for different purposes and are equally important because they complement the weaknesses of each other. The statistical NetIE Model is applicable to all types of ATM traffic because it is essentially a general abstraction of an ATM network. The limitation of this is that network impairments are introduced solely according to statistical parameters and distributions without being aware of the type of information being affected by the emulated network impairments. In the Application-Specific Model, the type of information carried by the ATM cells is also taken into account during the insertion of network impairments. However, the AS model is not as flexible as the statistical NetIE model. If the AS model is implemented to target MPEG video traffic, a considerable amount of changes is required in the implementation before it can be applied to video in another format or traffic of another type.

The choice between these two NetIE models depends on the purpose of the study. If it is required to carry out a comprehensive study on ATM network impairment to the quality of video in a particular format, the AS model should be preferred because of the insight provided. Conversely, since network impairments in a real-world ATM network are imposed on cells irrespective of the type of payload they carry, the statistical can be used to study the effects of network impairments on any type of traffic in a general manner.

Having introduced the network emulation architecture in this chapter, the next chapter covers technical materials that are essential to the design and implementation of the emulated network such as technical knowledge on ATM networking, ATM network impairments, video encoding techniques and video transportation over ATM.

Chapter 3.

Background Research

This chapter aims to pave the path for the design and implementation of the emulated network by presenting essential concepts on ATM networking, MPEG video compression standard, transportation of video over ATM and ATM network impairments. A comprehensive study of ATM networking is necessary because this research aims to study the quality of video transmitted over ATM by developing the NetIE architecture, whose objective is to emulate an ATM network and the impairments it introduces. This cannot be achieved without a good understanding on the functions and behaviour of ATM networks. Moreover, a thorough analysis of ATM network impairments is particularly critical to this study as it lays the foundation for the emulation of these impairments within the emulated network.

Since this study is centred around the degradation of video quality under the influence of ATM network impairments, and video is generally sent over an ATM network in compressed format, digital video compression needs to be examined in order to understand why and how quality degradation occurs. Because the multimedia service considered in this study, Video-on-Demand, supports video in MPEG format, a broad understanding on the MPEG compression standard is required. The syntax of MPEG is especially important to the Application-Specific NetIE model because of the need to categorize and identify various information fields within a MPEG encoded bit stream. Lastly, as pointed out by section 1.1.3, it is possible for video quality to be affected by different transportation schemes. These schemes have a varying ability to compensate for network impairments and they influence the ability of the video decoder to recover from the occurrence of network impairments. Therefore, issues related to the transportation of digital video over an ATM network also need to be reviewed.

3.1 ATM Networking

This section highlights ATM networking issues that are fundamental to the design and implementation of the NetIE models. It does not intend to serve as an introduction to B-ISDN and ATM because it is assumed that the reader has a prior knowledge of ATM. An overview of the B-ISDN and ATM standard is presented in Appendix A.

Since the NetIE architecture aims to emulate an ATM network, thorough understanding of the following aspects in ATM networking is required.

- General operations and behaviour of ATM networks (including the concepts of an ATM cell, virtual channel connections, switching and multiplexing etc);
- Functions of the Physical and ATM Layers at intermediate network nodes and end hosts;
- Functions of the ATM Adaptation Layer in end-systems and the different types of AALs;
- The importance of the traffic contract and the Quality-of-Service commitment from an ATM network;
- Service categories within an ATM network including Constant Bit Rate (CBR) and Variable Bit Rate (VBR) services.

Although the design and implementation of the NetIE architecture in this study are only concerned with the above, other ATM network functions, such as, User-to-Network (UNI) signalling, Call Admission Control (CAC), Usage Parameter Control (UPC) etc, need to be considered for NetIE to become a closer representation of ATM networks.

3.2 MPEG Video Compression

This section provides a general introduction to the MPEG video compression standard because it is employed by the Video-on-Demand (VoD) application in the emulated network. The effects of ATM network impairments on compressed digital video streams presented in this dissertation are specific to video encoded in the MPEG format as a result. A general introduction to digital video and digital video compression is presented in Appendix B for the benefit of the reader.

The MPEG standard is a widely used format for coding digital video and associated audio information. It makes use of the temporal and spatial redundancies found in video frames to achieve a high degree of compression (ranging from 30 to 1 to as high as 100 to 1). The first and the second of MPEG standards are known as MPEG-1 (ISO11172) and MPEG-2 (ISO13818) respectively. MPEG-2 is an extension to MPEG-1 in terms of providing a wide range of resolutions, bit rates and encoding options. The MPEG standards have three key parts:

- **'Systems'** – addresses the synchronization of video and audio information, initial and continuous management of coded data buffers to prevent over or underflow, absolute time identification etc;
- **'Video'** – addresses video coding;
- **'Audio'** – addresses the compression of digital audio information.

The hierarchical structure of a MPEG bit stream is illustrated in Figure 3-1. A MPEG video sequence starts with a sequence header that describes the basic parameters of the coded sequence such as the dimensions, resolution and frame rate of the sequence. At the Group of Picture (GOP) level, coded frames are grouped together with the header containing time and reconstruction information. Within each GOP is a number of frames or pictures, each of which is composed of a series of slices. The slice header can be used by the decoder to re-synchronize with the coded bit stream when an error occurs. In this case, the decoding process skips to the beginning of the next slice and continues decoding. Each slice contains macroblocks that are made up of luminance components and the spatially corresponding chrominance components. In the macroblock layer, blocks of Discrete Cosine Transform (DCT) coefficients are used to represent these luminance and chrominance components.

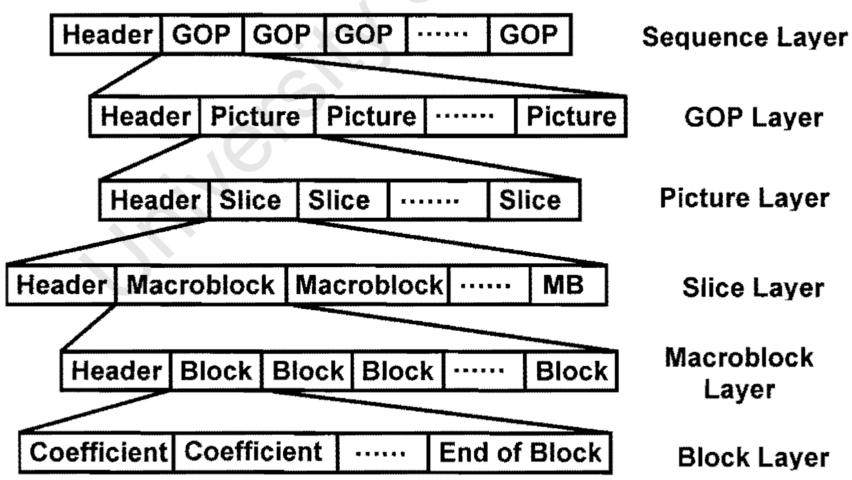


Figure 3-1 Structure of MPEG stream

Knowledge of the MPEG bit stream structure demonstrates that information fields in the MPEG bit stream carry a different level of importance. It justifies the development of the Application-Specific NetIE model and enables its implementation. It also enhances the analysis of the video distortions that appear under the influence of network impairments in Chapter 6.

Three types of picture frames are defined in the MPEG standard [M13818]: I frame, P frame, and B frame. A mixture of these three types of frames is present in a MPEG Group of Pictures (GOP). The structure and size of each GOP is not specified in the standard and can be chosen to suit the application. The prediction dependencies in a GOP illustrate the difference in importance among MPEG information fields and account for the temporal propagation of video distortions described in Chapter 6.

Figure 3-2 shows an example of a GOP structure as well as the prediction dependencies among the different types of frames. Each type of frame uses different coding methods described as follows:

- An Intra-coded picture (or I-Frame) is coded using information only from itself and only makes use of intraframe coding techniques;
- A Predictive-coded picture or (P-Frame) is a picture which is coded using motion compensated prediction from a past reference frame (i.e. a previous I or P frame in the sequence);
- A Bi-directionally predictive-coded picture (or B-Frames) is interframe coded using interpolated motion prediction between the previous I or P frame and the next I or P frame in the sequence.

Generally, B-Frames can achieve a higher degree of compression than P-Frames because of its bi-directional prediction nature, resulting in a smaller frame size. P-Frames are in turn smaller than I-Frames as I-Frames only makes use of intraframe compression. As a result, the generated bit rate is not constant and MPEG videos are generally considered to constitute variable bit rate traffic in ATM networks.

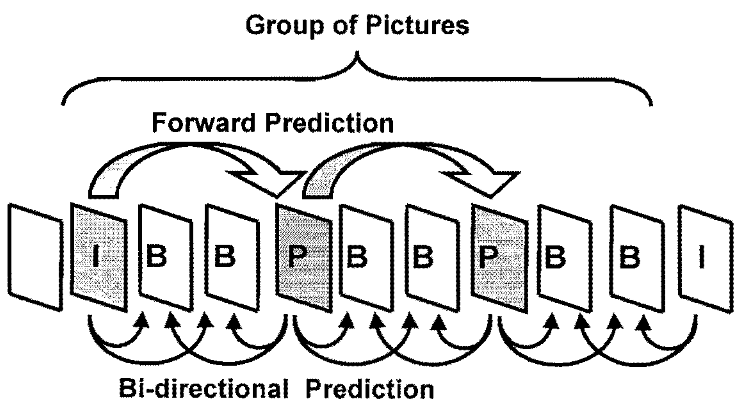


Figure 3-2 MPEG Group of Picture (GOP) structure and prediction dependencies

Typically, MPEG-1 encoders support the coding of video and associated audio into a single data stream of bit rates around 1.5Mbps with 30 frames per second at a resolution of 352x240 pixels. The video quality is comparable to that produced in videocassette recorders (VCR). Amongst the MPEG encoding parameters, changing the interframe to intraframe ratio and the quantization scale can lead to a significant change in the characteristics of the resulting video sequence [PAN94]. MPEG-2 builds on the MPEG-1 video standard as a compatible extension in terms of providing a wide range of resolutions, bit rates and encoding options. Differences between MPEG-1 and MPEG-2 are described in Appendix B.

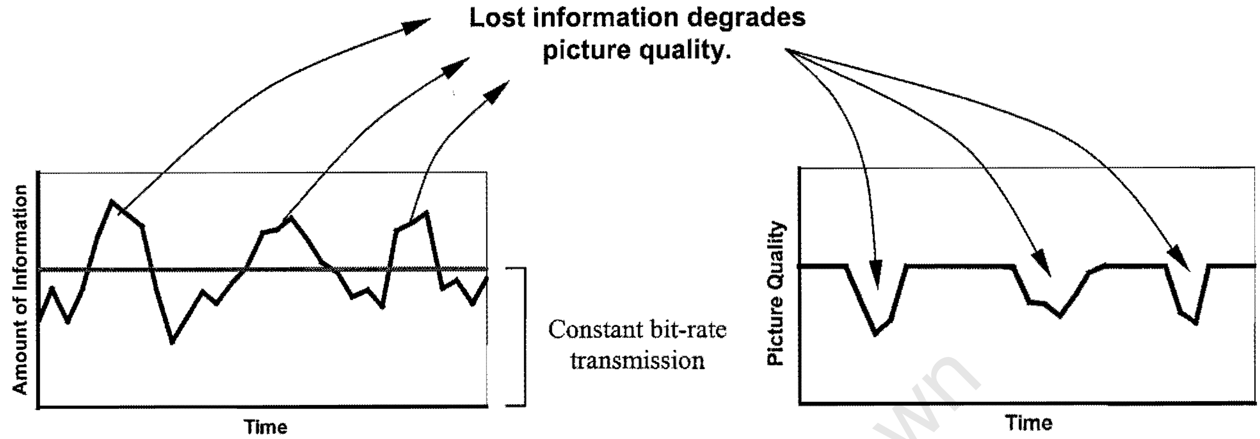
3.3 Video Transport over ATM

This section describes how MPEG video traffic is supported by an ATM network. It first explains the implications of Constant Bit Rate (CBR) and Variable Bit Rate (VBR) video transmission. The encapsulation of video bit streams into AAL 1 and AAL 5 service data units (SDUs) will then be illustrated and the advantages and disadvantages of using these two types of AAL for the transportation of video streams will be discussed. Lastly, the use of one or several ATM connections to carry video traffic is explained.

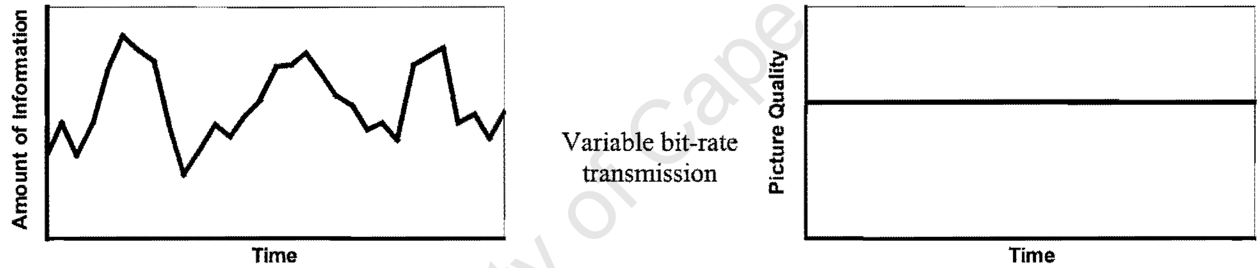
3.3.1 CBR and VBR

In a compressed video sequence, the instantaneous amount of information in the video signal may be highly varying depending on the type of picture (such as the different frame types in MPEG) and the amount of activity in the captured scene. This type of traffic is referred to as Variable Bit Rate (VBR). If video information needs to be transmitted at a constant bit rate (CBR) as shown in Figure 3-3, data exceeding the transmission bit rate needs to be discarded, and the picture quality is degraded at the receiving end. Another way of achieving CBR video stream is to enforce a constant output rate at the video encoder. When there is more video data waiting for transmission than available bandwidth present, the amount of compression is increased in order to keep the transmitted rate within the prescribed limit and vice versa. Since this is done regardless of the information content in the video source, the quality of the received signal may vary with the compression [LAP95]. A lot of work has been done in the area of CBR MPEG video with the main objective of improving techniques to enhance video quality. Variable bit rate video can also be transported over ATM as CBR traffic using circuit emulation [CIS98]. This is achieved by buffering the compressed video stream before transmitting it through the network. However, this introduces undesirable delay variation in the traffic stream. In addition, buffering impacts the video quality when the buffer becomes congested [LEB92].

On the other hand, variable bit rate transmission is able to match the varying amount of information of compressed video stream. As a result, picture quality should remain constant over time as shown in Figure 3-3b. It is shown in [LEB92] that in ATM, variable bit rate connection can guarantee a better video quality provided that cell loss priority is used[9].



a) Picture quality in constant bit-rate communication



b) Picture quality in variable bit-rate communication

Figure 3-3 Picture Quality in CBR and VBR Transmission

Variable bit rate communication allows statistical multiplexing among two or more VBR sources. This improves bandwidth utilization and is suitable for bursty traffic. Figure 3-4 shows how bandwidth can be utilized more efficiently by multiplexing three VBR sources [LAP95]. Figure 3-4a shows the transmission of information with each information source assigned a fixed amount of bandwidth, which is greater than or equal to the peak level of bandwidth for each source. In this case, there are gaps in which the bandwidth is not fully utilized. Figure 3-4b shows that with statistical multiplexing of VBR sources, the aggregate bandwidth required for a number of VBR source is less than the sum of the peak level of bandwidth for each source. This results in a more efficient utilization of the network bandwidth. As the number of VBR sources increases, the usage of bandwidth becomes more efficient. The saving in bandwidth is characterized by a parameter known as the multiplexing gain.

[9] Cell loss priority is implemented with the use of the CLP bit in the header of an ATM cell.

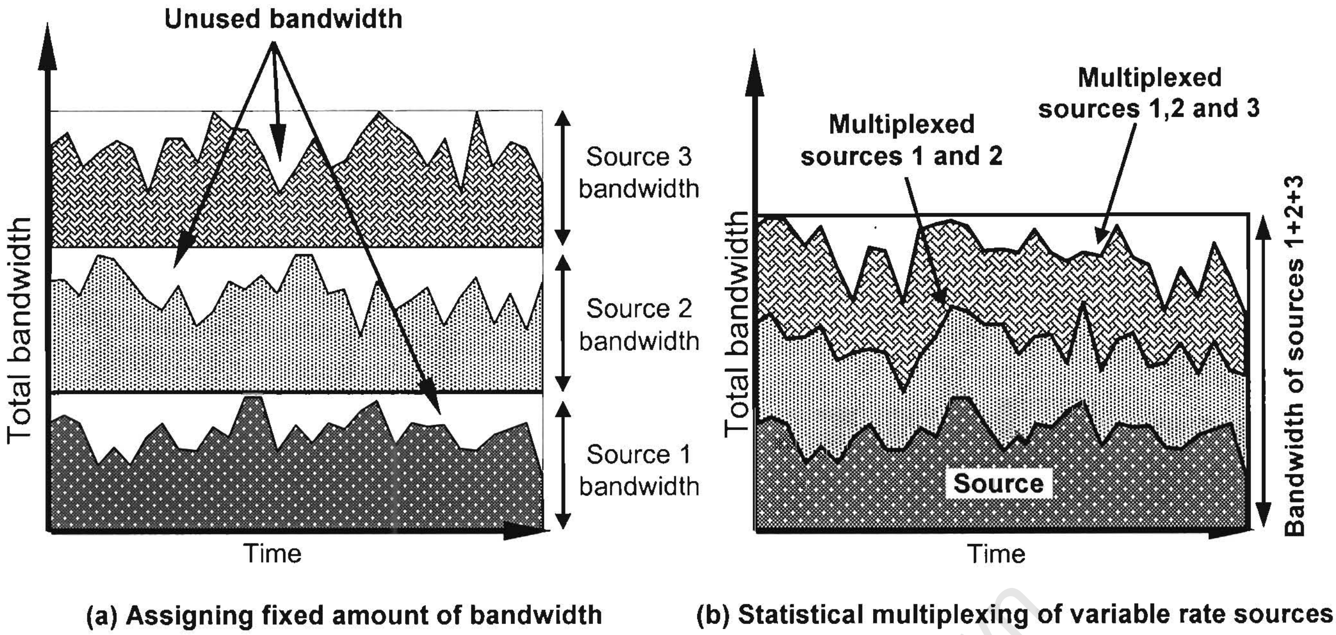


Figure 3-4 Statistical Multiplexing of VBR sources

Figure 3-4 also illustrates that the more bursty the traffic is, the higher the bandwidth peak-to-average ratio becomes and the more bandwidth is unused. Therefore, the effectiveness of the statistical multiplexing of VBR sources depends on the bandwidth peak-to-average ratio of each individual source as well as the number of VBR sources. Table 3-1 shows the level of burstiness for moving pictures for several video types [CIS98]. The level of burstiness is dependent on various factors such as the compression scheme used and the level of motion between frames.

Video Type	Bandwidth Peak-to-Average Ratio
Studio-quality video	1.9
Broadcast-quality TV	2.7
Video conference	3.1
Video telephone	4.4

Table 3-1 Bandwidth Peak-to-Average Ratio for various types of video.

The emulated network based on the NetIE architecture is capable of handling both CBR and VBR traffic class because it does not modify the service class of the traffic stream passing through. The transportation of video information as CBR or VBR traffic is dependent on the implementation of the multimedia communication application. In the context of this research, the VoD application uses variable bit rate service for the video stream.

3.3.2 MPEG Transport over AAL1

According to ITU-T Recommendation H.222.1, the Program Stream and the Transport Stream may use the service provided by AAL type 1 at the AAL Service Access Point (SAP). The following figure shows the mapping of a 188-byte TS packet into exactly four ATM cells using AAL 1.

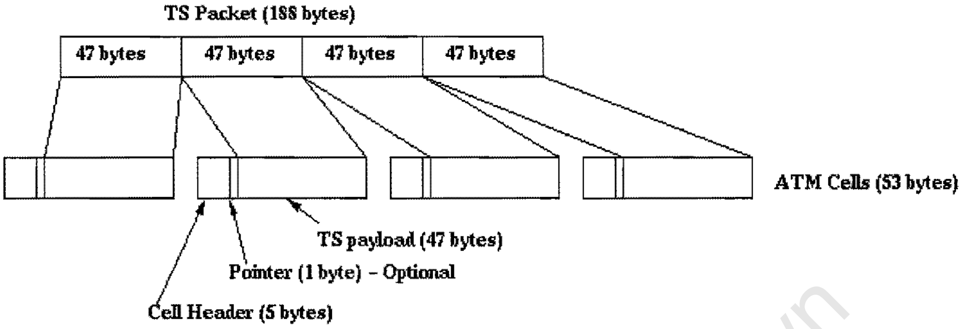


Figure 3-5 TS packet to AAL 1 cell mapping [H.222.1]

AAL 1 is designed for real-time applications and is therefore suitable for supporting video streaming over ATM. Figure 3-6 shows the SAR-PDU format for AAL 1. The other advantageous features provided by the Convergence Sublayer (CS) of AAL type 1 to transport video signals for interactive and distributive services include the following:

- Handling of possible rate mismatch between the sender and receiver end-systems – this is handled by the Synchronous Residual Time Stamp (SRTS) or the adaptive clock recovery method and prevents buffer to overflow or underflow during the delivery of video streams;

- Handling of lost and misinserted cells – the sequence count values are used to detect and locate lost and misinserted cells at the receiver. While detected misinserted cells are discarded, it may be necessary to insert appropriate dummy SAR-PDU payloads (maintaining the bit count integrity) or set the error indication bit in the TS packets (activating error concealment in the decoder) to compensate for lost cells. The Sequence Counter is also protected against bit errors[10];

- Handling of cell delay variation – a buffer is used to support this function. In the event of buffer underflow or overflow, it may be necessary for the CS to maintain bit count integrity by inserting or dropping an appropriate number of bits;

[10] The 4-bit SN field, including the CSI and the Sequence Counter (SC) fields, is protected by a 3-bit CRC code in the Sequence Counter Protection (SNP) field. The resulting 7-bit codeword is protected by an even parity bit.

- Correction of bit errors and lost cells – this is an optional function provided by the AAL 1 Convergence Sublayer. For correcting bit errors, the Forward Error Correction (FEC) technique using Reed-Solomon (128, 124) codes, which are able to correct up to 2 errored octets, is used. For the correction of bit errors and cell losses with delay restrictions, a method that combines FEC (Reed-Solomon (94, 88) codes) with octet interleaving of data (using a 16-cell interleaver) is used. This method can correct one cell loss occurrence in the group of 16 cells, or three errored octets in a row of 94 octets but has an overhead of around 6% [I.363.1].

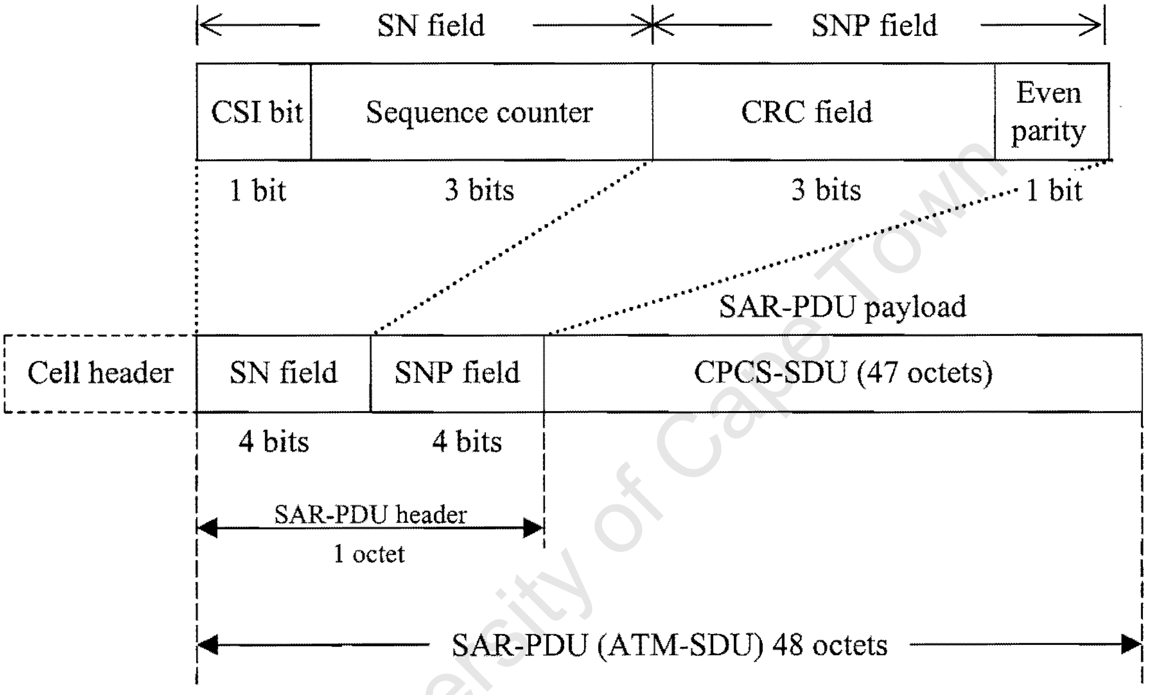


Figure 3-6 SAR-PDU structure for the AAL type 1 [I.363.1]

On the other hand, the transportation of MPEG streams over AAL 1 has the following disadvantages:

- AAL 1 is designed to support Constant Bit Rate (CBR) applications. However, MPEG bit streams are generally considered as Variable Bit Rate (VBR) in nature as explained in section 3.3.1. As a result, special encoding technique is required to generate CBR MPEG streams;

- In providing some of the function to support video signals listed above, the use of the SN and SNP fields (1 byte out of 48 in the ATM cell payload) in AAL 1 results in higher transmission and processing overheads compared to AAL 5;

- Support for AAL 1 in end-user equipment is not very widespread at this point in time yet.

3.3.3 MPEG Transport over AAL 5

The Video-on-Demand specification approved by the ATM Forum [VOD97], as well as the ITU-T Recommendation H.222.1, defines the mapping of MPEG-2 Transport Stream packets into AAL 5 with a NULL Service Specific Convergence Sublayer (SSCS). One to N TS packets are mapped into an AAL 5 SDU and the value of N depends on the type of virtual circuit used. For Switched Virtual Circuits (SVCs), the value of N is established during the ATM user-to-network connection setup phase as described in ATM Signalling 4.0 of the ATM Forum and ITU-T Recommendation Q.2931. For Permanent Virtual Circuits (PVCs), the default value of N is 2. Figure 3-7 shows the mapping of two TS packets into an AAL 5 CPCS-PDU which eventually breaks up into 8 ATM cells by the SAR sublayer.

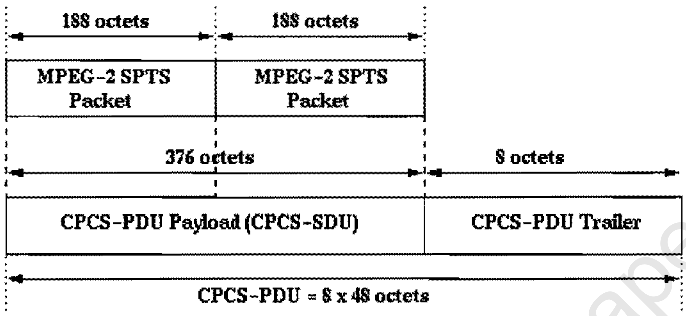


Figure 3-7 Format of AAL 5 PDU containing 2 TS Packets [COF97]

The functions offered by AAL 5 and its advantages to carry MPEG streams includes the following:

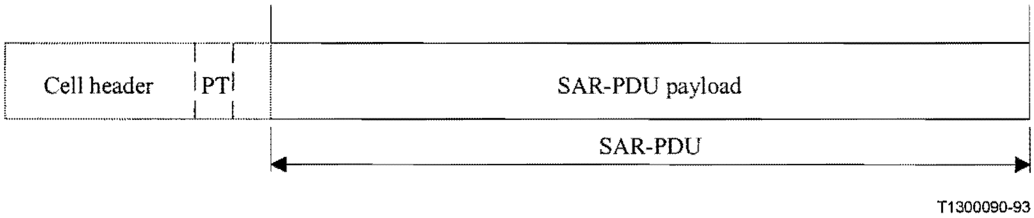
- AAL 5 is designed to be a simple and efficient. It carries relatively low transmission and processing overhead. There is only 8 octets transmission overhead per AAL 5 PDU and no per cell SAR protocol processing;

- Because a Service Specific Convergence Sublayer (SSCS) has been defined in AAL 5 to support ATM signalling protocols, it has become widely available to all ATM switches and end-stations that implement SVCs. For video application running on end-stations that posses signalling capability, AAL 5 is already implemented and ready to be utilized;

- The adoption of a NULL Convergence Sublayer (CS) requires no additional network functionality to be defined;

- The size of the CPCS-SDU can be up to 65535 octets (64 kilobytes – 1), which allows for a high degree of flexibility in the encapsulation of MPEG encoded bit stream;

- End of SAR-SDU indication – the SAR sublayer utilizes the ATM-User-to-ATM-User indication (AUU) parameter in the PT field of the ATM header (Figure 3-8) to indicate whether the payload of the current ATM cell contains the end of a SAR-SDU[11] (AUU = 1) or not (AUU = 0);



PT Payload Type

NOTE – The Payload Type field belongs to the ATM header. It conveys the value of the AUU parameter end-to-end.

Figure 3-8 SAR-PDU format for the AAL type 5

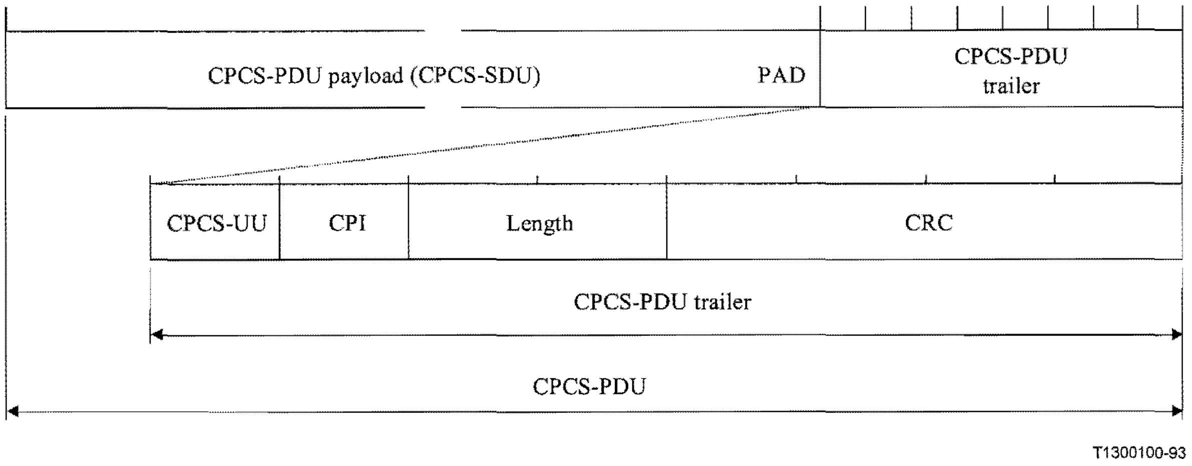
- Detection of lost or misinserted cells – the 8-octet CPCS-PDU trailer contains a Length field and a CRC field (Figure 3-9). The Length field is used to encode the length of the CPCS-PDU payload field and enables the receiver to detect cell loss or cell mis-insertion. As in the case for AAL 1, error concealment in the decoder at the receiver can be activated by the detection of a loss cell;

- Detection of bit errors – the CRC field is used to detect bit errors in the CPCS-PDU. It contains the value of a CRC-32 calculation that is performed over the entire contents of the CPCS-PDU, including the CPCS-payload, the PAD field[12], and the first four octets of the CPCS-PDU trailer.

- Error handling – corrupted CPCS-SDU can either be discarded or optionally delivered to the SSCS.

[11] The reassembly of the CPCS-PDU on the receiver starts when the end of SAR-SDU is detected

[12] The PAD field is used to fill up the entire CPCS-PDU (the CPCS-PDU payload, PAD field and CPCS-PDU trailer) to be an integral multiple of 48 octets. Figure 3-7 shows that the PAD field is not needed in this case.



PAD	Padding	(0 ... 47 octets)
CPCS-UU	CPCS User-to-User Indication	(1 octet)
CPI	Common Part Indicator	(1 octet)
Length	Length of CPCS-SDU	(2 octets)
CRC	Cyclic Redundancy Check	(4 octets)

Figure 3-9 CPCS-PDU Format for the AAL type 5 [I.363.5]

On the other hand, the transportation of video signals over AAL 5 has the following drawbacks:

- There is no mechanism built into AAL 5 for timing recovery;
- AAL 5 does not provide native support for Forward Error Correction (FEC);
- Although the procedures for the delivery of corrupted CPCS-SDU are defined in the ITU-T Recommendations I.363.5, they have not been widely implemented in end-user equipment. The effects of delivering corrupted CPCS-SDU to applications will be further investigated in section 6.2;
- While the Length Field of AAL 5 is only capable of detecting the presence of a cell count error, the Sequence Counter in AAL 1 can also locate the lost or mis-inserted cell.

The comparisons between the use of AAL 1 and AAL 5 to carry MPEG video streams over ATM networks can be highlighted by the fact that more error handling functions in AAL 1 results in higher transmission and processing overheads. On the other hand, a lower overhead in AAL 5 implies less error handling ability. The other critical difference between these two AAL types is that AAL 5 is already widely implemented in end-user equipment and ATM switches while AAL 1 is still catching up. This is why the Video-on-Demand application adopted in this thesis uses AAL 5 services. Other possible schemes to pack MPEG information into AAL-SDU are examined in [GHA93] and [LIN96].

3.3.4 Video traffic and ATM connections

At the ATM level, one or several connections can be used to carry video and control traffic for single direction video delivery. Three scenarios are listed and explained below:

- Single Connection – in this case, only one ATM connection is set up between the sender and the receiver and video information is serialized and transported over this Virtual Channel Connection (VCC) together with control information;

- Dual Connections – in this scenario, two ATM connections are required. One connection is used for video traffic and the other is used to carry control information between the sender and the receiver[13]. These two connections will obviously have very different traffic characteristics and are therefore set up with different QoS attributes;

- Multiple Connections – another alternative is to use multiple connections for the different information components (video, audio, control etc). In MPEG-2 compression, a technique known as layered coding constructs MPEG bit streams into a base layer (which carries important, base resolution information) and an enhancement layer (which carries quality and resolution enhancement information). Video traffic belonging to these two layers has a different priority indicated by the Cell Loss Priority (CLP) bit and carries different QoS attributes. During network congestion, ATM cells carrying the enhancement layer with lower priority are discarded before information belonging to the base layer with higher priority is affected. This decreases the probability of any interruption in the delivery of video streams to a large extent. The synchronization between the various components can be embedded in each connection so that they can be combined at the receiver.

The use of one or more ATM connections to support video traffic is implementation specific, i.e. determined by the architecture of the video application and the video coding scheme employed. Since the aim of the underlying ATM network is to provide the services and resources to satisfy the connection requests from applications, it plays no part in deciding the number of connections used to carry video traffic.

---

[13] The VoD application adopted in the emulated network uses this configuration.

3.4 ATM Network Impairment

This section outlines the network requirements for real-time video delivery. It summarizes the formal definitions of the five types of ATM network impairments considered in this dissertation and the parameters that are used to characterize them. The section also accounts for the occurrences of these impairments from the perspective of end-user applications.

3.4.1 Network requirements for real-time video delivery

In a multimedia communication application such as VoD, the quality of video received at the client depends on many factors. Some of these factors, such as resolution, colour depth, frame rate and the compression scheme employed, determine the original video quality at the source. During the real-time delivery of video stream, certain performance requirements have to be met by the underlying ATM network for the video delivery to be satisfactory from an end-user's perspective. The network performance requirements from such applications that affect the quality of video delivery include the following:

- Bandwidth – it is used to characterize the amount of video information carried by the network;
- Errors – it is the incorrect interpretation of the information bits (0 treated as 1 and vice versa);
- Losses – it is the loss of information bits in the video bit stream;
- Latency – it is the time taken for the video information to flow from the source to the destination;
- Jitter – it is the variation in latency during the delivery of the video information.

It should be obvious that the network needs to provide the amount of network capacity (bandwidth) that is appropriate for a particular video session in order to delivery all the video information to the destination on time. The other four factors are often collectively referred to as ATM network impairments. When video traffic or any other types of traffic is travelling through the ATM network, it is affected by varying degrees of network impairments. Although it appears that the lowest possible level of each type of ATM network impairments is desirable, it is often not necessary for networks to provide video connections with the absolute minimum amount of latency, jitter, errors and losses that it is capable of in order to satisfy the end-users. In many cases, video connections can tolerate a very limited amount of network impairments without causing any dissatisfaction. These limits are dictated by factors such as the encoding scheme, the visual content, the error correction or concealment techniques etc. However, it is important to note that most of the redundancies have been removed from the video bit stream during the compression process. Therefore, when the level of network impairments increases beyond acceptable limits, a gradual increase in the level of dissatisfaction will result from end-users. If impairment level continue to increase, video quality may be regarded as unacceptable.

3.4.2 Types of ATM Network Impairments

ATM network impairments can be classified into five different types: Cell Error, Cell Loss, Cell Transfer Delay, Cell Delay Variation and Cell Mis-insertion. Among them, the first four types of impairments are related to four of the requirements listed above. Although these impairments carry an end-to-end significance, it is important to note that they can actually occur at the Physical, ATM and ATM Adaptation layers, as well as various physical locations within an ATM network. In cases where impairments occur at more than one point in the network, the end-to-end effect is an aggregate of all the occurrences.

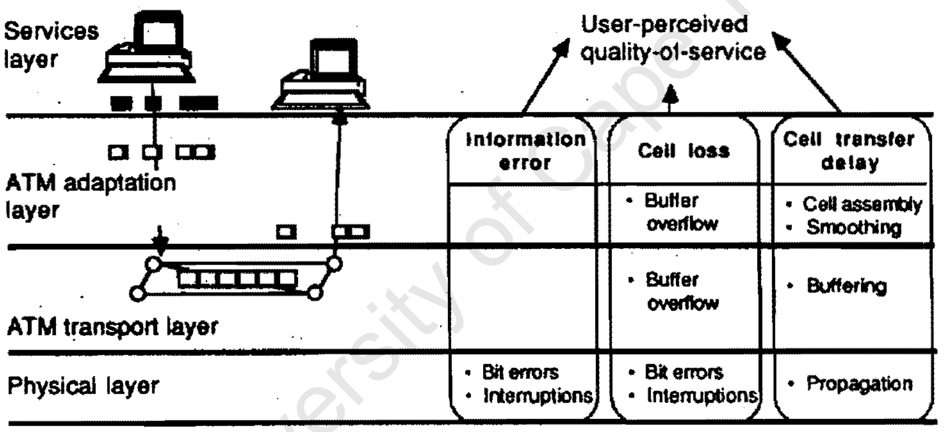


Figure 3-10 Layered view of ATM performance impairments [WOO90]

Figure 3-10 shows a layered view of ATM network impairments [WOO90]. Network traffic originated at the services layer is converted into the ATM format by the ATM adaptation layer. It is then transferred within cells via multiplexing and/or switching by the ATM layer, which is supported by a transmission, or physical layer. Each of these layers may introduce certain information transfer performance impairments perceivable at the service layer.

Statistical multiplexing in an ATM network has the desirable effect of higher bandwidth utilization efficiency among VBR sources as described in section 3.3.1. However, depending on the traffic characterization algorithm, it is possible for statistical multiplexing to cause congestion in a network element when it cannot handle the total load temporarily. In order to avoid or resolve congestion, the network element needs to take actions such as buffering, which causes additional delay and delay variation. In the event of buffer overflow, actions such as cell discarding have to be taken by the network element and cell loss(es) will occur as a result.

The definition of the five types of impairment is based on the ATM cell transfer outcomes defined in [I.356] and is characterized by the five ATM performance parameters defined in sections 3.4.2.1 to 3.4.2.5. ITU-T Recommendation I.353 establishes that a Cell Reference Event (CRE) has occurred when: 1)the VPI/VCI field of a cell after HEC processing corresponds to the VPI/VCI of the monitored connection; and 2)the PT field indicates the cell to be a user information cell. Two cell reference events, CRE_1 and CRE_2, are known as *corresponding* if they are created by the same cell at two different Measurement Points, MP_1 and MP_2, respectively. Since this thesis is concerned with end-to-end ATM network performance, MP_1 and MP_2 are taken as endpoints (source and destination) of the network. When a particular connection is being monitored, a number of possible cell transfer outcomes are defined according to these two corresponding CREs. A transmitted cell can be successfully transferred, errored, tagged, or lost. The ATM cell transfer outcomes are listed below [I.356]:

- Successful cell transfer outcome – a successful cell transfer outcome occurs when a CRE_2 corresponding to CRE_1 happens within a specified time T_{max} of CRE_1 and: 1)the binary content of the received cell information field matches exactly with that of the corresponding transmitted cell; and 2)the cell is received with a valid header field;

- Tagged cell transfer outcome – a tagged cell transfer outcome is the same as the successful cell transfer outcome except for the change of the Cell Loss Priority (CLP) bit from CLP = 0 at MP_1 to CLP = 1 at MP_2 (i.e. from high priority to low priority);

- Errored cell outcome – an errored cell outcome occurs when a CRE_2 corresponding to CRE_1 happens within a specified time T_{max} of CRE_1 but there are one or more bit errors exist in the received cell information field.

- Lost cell outcome – a lost cell outcome occurs when a CRE_2 fails to happen within time T_{max} of the corresponding CRE_1.

- Misinserted cell outcome – a misinserted cell outcome occurs when a CRE_2 happens without a corresponding CRE_1.

- Severely errored cell block outcome – when a certain number of M (or more) Lost, Misinserted, or Errored Cell outcomes are observed in a received cell block of N cells transmitted consecutively on a given connection. This is beyond the scope of the thesis.

The primary causes for the occurrence of each type of ATM network impairments as well as the corresponding ATM performance parameters are described in the sections below.

3.4.2.1 Cell Error

Cell Error corresponds to the Errored Cell Outcome and is described by the Cell Error Ratio parameter (CER). Information error within cells can be caused by physical layer bit errors and interruptions (i.e. errors in the transmission line). Bit errors cause misinterpretation of information at the destination. Their effects on video traffic will be examined in Chapter 6. The Cell Error Ratio parameter for an ATM connection is defined as follows:

$$CER = \frac{\text{Errored Cells}}{\text{Successfully Transferred Cells} + (\text{Tagged Cells}) + \text{Errored Cells}}$$

Equation 3-1 Definition of Cell Error Ratio (CER)

There is some discrepancy between the ITU-T and the ATM Forum with regard to whether Tagged Cells are included in the definition of CER. In the ITU-T Recommendations I.356, CER is defined as above. However, the definition stated in [UNI31] and [TM4.0] from the ATM Forum does not include Tagged Cells. Therefore the tagged cells term is shown in brackets.

The CER depends mainly on the transmission system in the physical layer and the Bit Error Ratio (BER) specified for the transmission medium, therefore CER is described in [TM4.0] as one of the non-negotiated QoS parameters. With the introduction of fibre optics technology in the transmission system, the BER characteristics have improved to the order of 10^{-9} and beyond. The recommended BER values for MPEG-1 and MPEG-2 are 4×10^{-11} and 6×10^{-12} without error handling, and 2.5×10^{-6} and 1.5×10^{-6} with single bit error correction respectively as defined in [ONV94]. If bit errors occur randomly as suggested in [TM4.0], the probability that there is at least one single bit error in the 48-byte payload is given by equation:

$$1-(1-BER)^{384}$$

where 384 is the number of bits in an ATM cell payload. When the physical medium has a high BER = 10^{-6}, the probability of having at least one bit error in the payload is 3.84×10^{-4}. If the BER of the physical medium becomes 10^{-9}, this probability drops to 3.84×10^{-7}.

3.4.2.2 Cell Loss

The definition of Cell Loss is based on the Lost Cell Outcome and is described by the Cell Loss Ratio (CLR) ATM performance parameter. In the physical transmission medium, cell Loss occurs when there is an un-correctable bit error in the header. In the ATM layer, it occurs when buffer or queue overflows at multiplexing or switching nodes. In the ATM adaptation layer, cell Loss occurs when an overflow occurs in the adaptation layer smoothing or send and receive buffer (Figure 3-10). From an application layer's perspective, Cell Loss may also occur at the user-to-network interface due to the usage parameter control (UPC) policing when the source traffic temporarily exceeds the parameters in the connection traffic descriptor specified in the traffic contract. However, since this type of cell loss is caused by the violation of traffic contract by the application, it is not included in the calculation of the CLR parameter, which defines network performance only. The objectives of CLR for MPEG-1 and MPEG-2 are defined in [ONV94] to be 10^{-8} and 2×10^{-9} without error handling, and 9.5×10^{-6} and 4×10^{-6} with cell loss correction respectively. The CLR is one of the negotiated QoS parameters and is calculated as follows:

$$CLR = \frac{\text{Lost Cells}}{\text{Total Transmitted Cells}}$$

Equation 3-2 Definition of Cell Loss Ratio (CLR)

The code used for the HEC function in the ATM cell header is capable of either single-bit error correction or multiple-bit error detection [I.432.1]. The HEC field of the header is used at the end-systems and each network node to check for header integrity. For bit errors in the ATM cell header that can be detected and corrected by the HEC field, the cell is passed, after the error is corrected, to the next node in the case of intermediate nodes and to the AAL for end-systems. Cells with multiple bit errors in the cell header are detected by HEC but cannot be corrected. These cells are dropped by the network and a Lost Cell Outcome results. On the other hand, an ATM cell with a corrupted VPI/VCI field within the cell header that is not detected by HEC will be lost. If this corrupted cell happens to carry a permissible header, it will be mis-directed to another VCC. This results in a Mis-inserted Cell Outcome for the second VCC. Section 3.4.2.5 describes Cell Mis-insertion in more details.

When a conforming cell arrives at a multiplexing or a switching node in the network and finds the buffer of the node to be full due to network congestion, one of the tagged cells already in the buffer gives way to the conforming cell and is dropped. The ATM traffic contract defined in section 2.1.1 includes the cell conformance definition and the QoS requirement for the connection. Cells that are considered to be *non-conforming* by the UPC policing function are either tagged[14] or dropped immediately, resulting in Loss Cell Outcome. Although in theory the CAC procedure should ensure adequate network resource availability when it accepts a new connection and the UPC policing should handle non-conforming cells, unpredictable statistical fluctuation of traffic flows can result in a non-zero probability for buffers to overflow. In fact, even when all the cells of a connection are considered to be conforming, it is still possible for cell losses not exceeding the Cell Loss Ratio QoS parameter to occur for the ATM network to fulfil the traffic contract. QoS of a connection does not guarantee zero cell loss.

3.4.2.3 Information Transfer Latency

Information Transfer Latency between end-user applications is a combination of Cell Transfer Delay introduced by the network and processing delay in end-systems. Cell Transfer Delay is defined as the elapsed time, $t_2 - t_1$, between the occurrence of two corresponding cell transfer events, CRE_1 of MP_1 at time t_1 and CRE_2 of MP_2 at time t_2 ($t_2 > t_1$). In the physical layer, CTD is composed of propagation delay and transmission delay. Propagation delay depends on the distance between the source and the destination and varies slightly with the type of physical medium used[15]. Transmission delay depends on the bit rate supported by the physical link between the source and the destination and becomes negligible as the transmission speed increases.

From the ATM layer perspective, cells in an ATM network traverse through switching and multiplexing nodes. Switching and Multiplexing requires processing time (typically around 10 μs per node), which consists of table lookups for cells and the placement of cells from the input queue to the output queue. In addition, buffering in switches and multiplexers increases cell transfer delay depending on the network load and buffering strategy. In the end systems, the ATM adaptation layer introduces delay during segmentation and reassembly as well as possible smoothing (traffic shaping) of user information. To summarise, the overall delay perceived by an end-user in terms of information transfer can be divided into the following components [I.356]:

- Segmentation and reassembly delay at end systems (subdivided into three components):
  - Segmentation delay in the AAL of the sending side;
  - Possible buffering delay in the AAL of the receiving side to compensate for cell delay variation;

---

[14] A tagged cell is given a low priority indicated by CLP = 1 in the ATM cell header. Tagged cells are the first ones to be dropped when congestion arises.

[15] for optical-fibre cables, the propagation delay is 4μs/km [ONV94]

- Reassembly delay in the AAL of the receiving side;

- Cell Transfer Delay of the network, which is equal to the sum of:
  - Total inter-ATM node transmission and propagation delay;
  - Total ATM node processing (queuing, buffering, switching, multiplexing) delay.

There are two ways of defining CTD parameter in the literature. The ITU-T Recommendations I.356 and the ATM Forum UNI Specification 3.1 both define the Mean Cell Transfer Delay to be the arithmetic average of a specified number of cell transfer delays. On the other hand, the ATM Forum Traffic Management Specification Version 4.0 defines the Maximum Cell Transfer Delay to be a statistical guarantee using a probability parameter α, such that the network has to guarantee the maximum CTD to be met with a probability of $1 - \alpha$ (Figure 3-11). The value of α is generally very small, such as 10^{-6}. It should be noted that for real-time video streaming applications such as VoD, there is a maximum end-to-end delay requirement beyond which ATM cells are treated as lost cells (Figure 3-11). This is because video information contained in late cells cannot be utilized at the destination during decoding and are effectively lost.

3.4.2.4 Cell Delay Variation

Cell Delay Variation specifies the variation in delay incurred in the ATM network. It is caused in the ATM layer by a varying cell buffering or queuing delay in ATM switches and a varying multiplexing delay in ATM traffic multiplexers. As a result, CDV is dependent on the cell scheduling policy in the switches and multiplexers. In the ATM adaptation layer, CDV is caused by a varying SAR sublayer processing delay. There are three performance parameters associated with CDV in the literature, namely 1-point CDV, 2-point CDV and peak-to-peak CDV (p-p CDV).

The 1-point CDV is defined based on the observation of a sequence of consecutive cell arrivals at a single Measurement Point (MP) [I.356], [TM4.0] and [UNI3.1]. The parameter for 1-point CDV describes the variability in the pattern of cell arrival events observed at this MP with reference to the negotiated peak cell rate (1 ÷ minimum inter-arrival of cells) defined in the traffic contract. It can be related to cell conformance at a particular MP, and includes variability that is present at the cell source (customer equipment) as well as the cumulative effects of variability introduced in all connection portions between the cell source and the specified MP. The 1-point CDV (y_k) for cell k at an MP is calculated from the reference arrival time (c_k) and the actual arrival time (a_k) of the cell at the MP as follows:

$$y_k = c_k - a_k$$

Equation 3-3 Definition of 1-point CDV

where the reference arrival time pattern (c_k) is defined as follows:

$$c_0 = a_0 = 0,$$

$$c_{k+1} = \begin{cases} c_k + T & \text{when } c_k \geq a_k ; \\ a_k + T & \text{otherwise.} \end{cases}$$

Equation 3-4 Definition of the reference arrival time c_k for 1-point CDV

1-point CDV may either be positive or negative. While positive values of 1-point CDV correspond to cell clumping (early cell arrivals), negative values of 1-point CDV correspond to gaps in the cell stream (late cell arrivals).

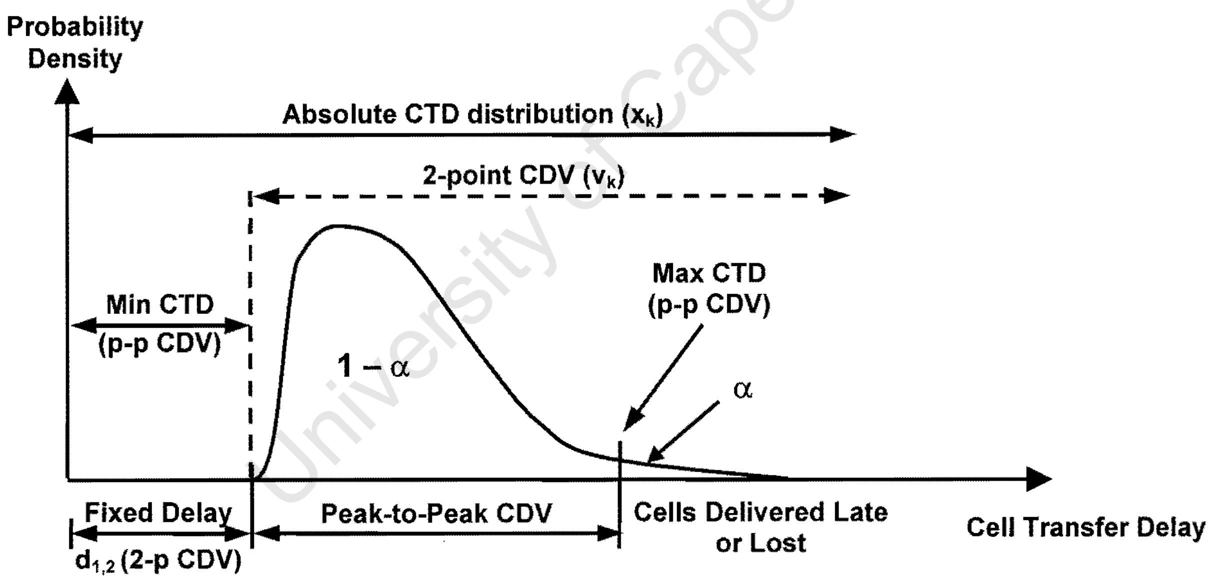


Figure 3-11 Cell Tranfer Delay probability density model – real-time service categories [TM4.0]

2-point cell delay variation is defined based on the observation of *corresponding* cell arrivals at two MPs that delimit a virtual connection portion [I.356] [TM4.0]. The 2-point CDV parameter describes the variability in the pattern of cell arrival events observed at the output of a connection (MP_2) with reference to the pattern of the corresponding events observed at the input to the connection portion (MP_1) [TM4.0]. Therefore, it includes only the delay variability introduced by the connection portion between MP_1 and MP_2.

The 2-point CDV (v_k) for cell k between MP_1 and MP_2 is calculated from the absolute cell transfer delay (x_k) of cell k and a defined reference cell transfer delay ($d_{1,2}$) between the two MPs as follows:

$$v_k = x_k - d_{1,2}$$

Equation 3-5 Definition of 2-point CDV

The absolute cell transfer delay (x_k) of cell k between MP_1 and MP_2 is the actual transfer delay experienced by cell k (as defined in section 3.4.2.3) and is calculated from the cell's actual arrival time at MP_2 (a_{2k}) and the cell's actual arrival time at MP_1 (a_{1k}) as follows[16]:

$$x_k = a_{2k} - a_{1k}.$$

Equation 3-6 Definition of the absolute cell transfer delay x_k for 2-point CDV

The reference cell transfer delay ($d_{1,2}$) between MP_1 and MP_2 is the absolute cell transfer delay experienced by a reference cell between the two MPs.

Positive values of 2-point CDV correspond to cell transfer delays greater than that experienced by the reference cell and negative values of 2-point CDV correspond to cell transfer delays less than that experienced by the reference cell. The distribution of 2-point CDV is identical to the distribution of absolute cell transfer delay displaced by a constant value equal to $d_{1,2}$ (Figure 3-11).

The peak-to-peak CDV is defined in [TM4.0] as one of the QoS parameters that is negotiated using [UNI4.0]. Peak-to-peak CDV refers to the maximum difference in CTD or the difference between the best and the worst case of CTD among all the cells of the connection. The best case (minimum) CTD is the fixed delay and the worst case (maximum) CTD is equal to a value likely to be exceeded with a probability no greater than α. In Figure 3-11, the peak-to-peak CDV value is the $(1 - \alpha)$ portion of the CTD minus the fixed CTD that could be experienced by any delivered cell on a connection. The 2-point CDV and the end-to-end p-p CDV are the same under the following conditions:

- The two MPs for 2-point CDV are taken to be the source and the destination, and;
- The fixed delay for p-p CDV is the same as the reference cell delay $d_{1,2}$ for the 2-p CDV.

---

[16] Variables a_{2k} and a_{1k} are measured with reference to the same reference clock.

The CDV network impairment parameter adopted in the implementation of the Impairment Insertion Modules (section 5.3.2.5) follows the peak-to-peak definition.

3.4.2.5 Cell Mis-insertion

The definition of Cell Mis-insertion is based on the Mis-inserted Cell Outcome and is described by the Cell Mis-insertion Rate (CMR) ATM performance parameter. When bit errors in the cell header are undetected or mis-corrected by HEC and the resulting cell header carries a VPI/VCI that happens to match that of another existing connection, the cell is mis-inserted in a connection to which it does not belong and sent to a wrong destination. On the other hand, if the VPI/VCI fields of a corrupted cell header do not match that of any existing connection, the cell will be discarded immediately. In the case where the cell considered above being a physical layer idle cell or an ATM layer unassigned cell (as defined in [I.321]), only a Mis-inserted Cell Outcome occurs. However, if the cell is a valid cell from an existing connection, a Lost Cell Outcome occurs on this connection in addition to the Mis-inserted Cell Outcome.

The ATM performance parameter for cell mis-insertion is defined as a rate rather than a ratio because the mechanism producing mis-inserted cells is independent of the number of transmitted cells received on the corresponding connection [TM4.0]. The Cell Mis-insertion Rate (CMR) is the total number of mis-inserted cells observed during a specified time interval divided by the length of the time interval as follows:

$$CMR = \frac{Mis-inserted\ Cells}{Time\ Interval}$$

Equation 3-7 Definition of Cell Mis-insertion Rate (CMR)

Bit errors in the cell header that are undetected or mis-corrected are influenced by the transmission error rate in the physical layer. In addition, the likelihood that an undetected or mis-corrected cell header error maps to a valid VPI/VCI is dependent upon the number of VPI/VCI values that are assigned and being actively used [UNI3.1]. Since the number of active ATM sources is likely to be less for a private network than for a public network, the number of sources that can possibly be mapped into another cell address will likely be much larger in a public network than in a private network.

Chapter 4.

Designing the Emulated Network

This chapter is devoted to the design of the emulated network in the realization of the network impairment emulation (NetIE) architecture outlined in Chapter 2. During the design process, two fundamentally different approaches were considered. This chapter first describes and explains the two different approaches considered and then lists the design considerations. It concludes by comparing the two designs.

4.1 Design I

This approach involves the functional simulation of an ATM network as a whole including the network interfaces at end-systems, the intermediate ATM nodes and, most importantly, the network impairments. The input to the simulated environment is encoded video information and the output from it is pre-decoded video bit streams as if they are emerging from an ATM network. Figure 4-1 illustrates this approach to build the NetIE architecture and Figure 4-2 defines the scope of simulation in Design I by considering the functional block diagram of a VoD system.

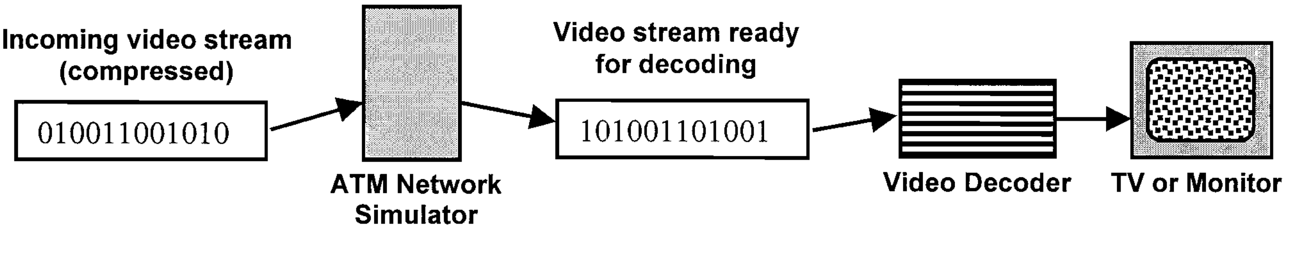


Figure 4-1 The first approach to carry out Network Impairment Emulation

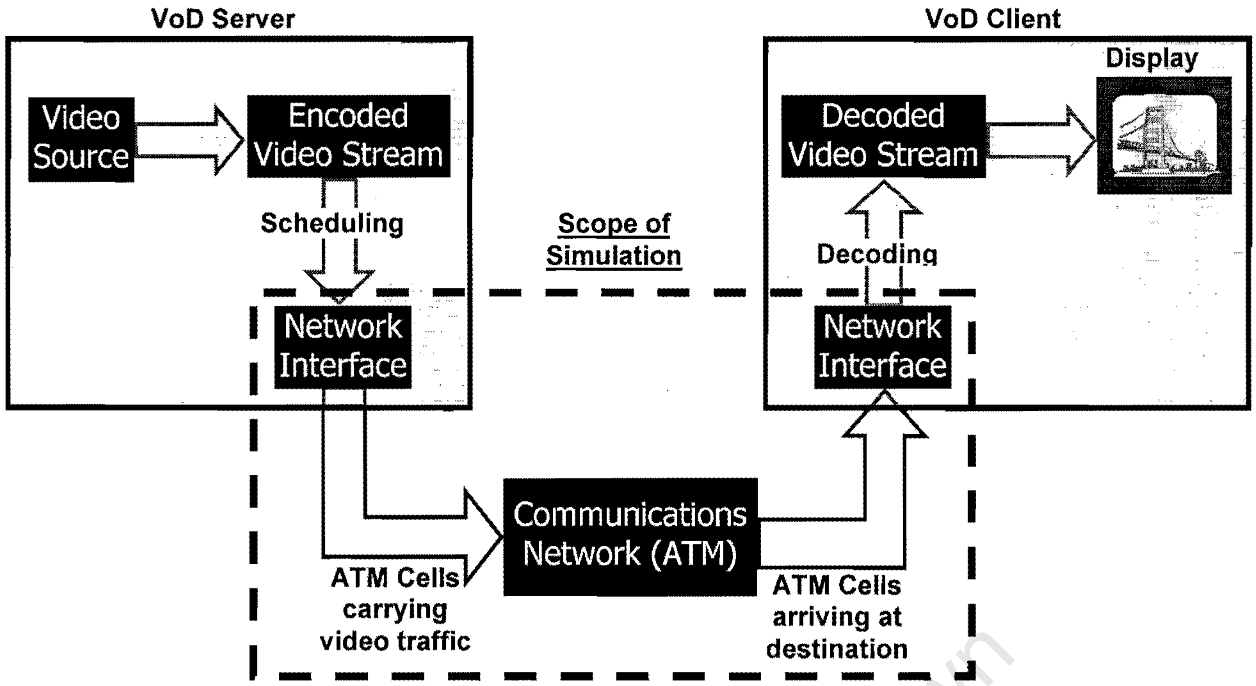


Figure 4-2 The scope of the simulation in Design I

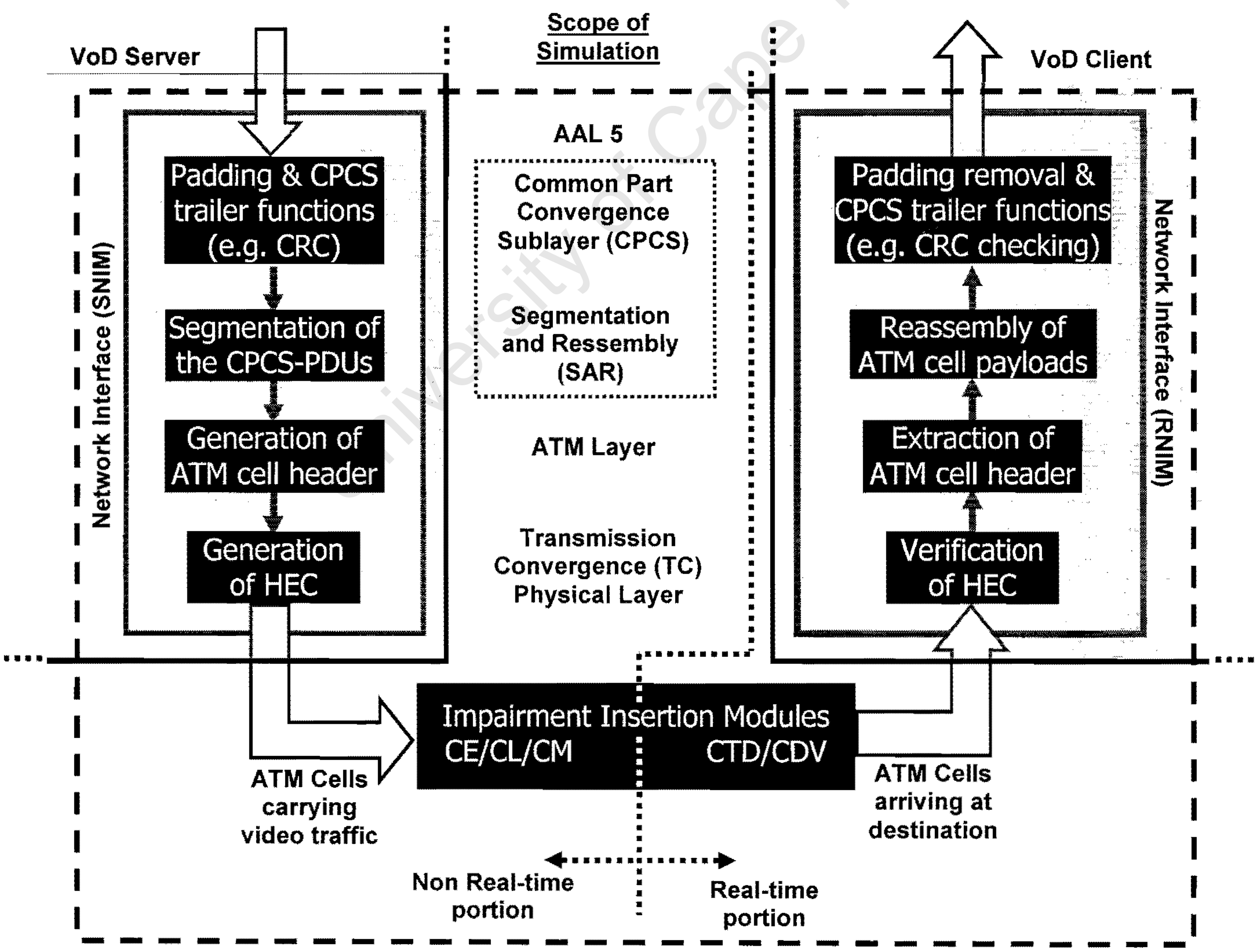


Figure 4-3 Functional block diagram of the simulation in Design I using AAL 5

Figure 4-3 shows the equivalent of Figure 4-2 in the simulated environment. The tasks of the ATM Network Simulator can be broken down to various functional units as follows:

- Sender network interface module (SNIM) – simulate the functions at the network interface of the sender host (e.g. the VoD server);
- Impairment Insertion Modules (IIMs) – impose controllable ATM network impairments as described in section 2.2;
- Receiver network interface module (RNIM) – simulate the functions at the network interface of the receiver host (e.g. the VoD Client).

The simulator will be running a piece of real-time software that is made up of software modules that implement the above tasks. The strategy adopted in the design of each functional unit of the simulator listed above is discussed below.

4.1.1 The Sender Network Interface Module (SNIM)

The sender network interface module handles the incoming video bit stream to be transported over the emulated network in the same way as a VoD server network interface would process video data to be transported over an ATM network. If AAL 5 with a null SSCS is used, the four functions at the network interface of the sender host that need to be simulated include the following (in the order of occurrence):

- Padding and the CPCS-PDU trailer functions (including CRC generation) of the Common Part Convergence Sublayer (CPCS);
- Segmentation of the CPCS-PDU into 48-byte ATM cell payloads at the Segmentation and Reassembly (SAR) sublayer;
- Generation of the ATM cell header at the ATM layer;
- Generation of the HEC field against bit errors in the cell header at the TC sublayer of the Physical Layer.

Depending on the implementation, the four functions can either be integrated into one module or separated into four modules. It is also possible for the sender network interface module to be implemented in a non-real time fashion because there is no timing requirement from the manipulation of the bit streams by the four functions. For example, the input to this module could be a MPEG file containing the MPEG bit stream and the output could be another file called a CELL file containing the simulated ATM cells including the trailer, ATM cell header and HEC, as well as cell boundary information etc. If a simulation on AAL 1 transport of video stream is required, addition real-time software module is required within the sender network interface module because of the ability to handle timing relationship as well as the sequence count operations in the CS sublayer of AAL 1.

Although the ATM cell header has little significance in terms of the end-to-end transportation of video information, it is included in the simulation to cater for cell loss and cell mis-insertion resulting from an undetected VPI/VCI error in the cell header. In addition, the HEC field in the Physical Layer is included in the simulation because in an ATM network, the HEC field offers some protection against bit errors in the cell header.

4.1.2 The Impairment Insertion Modules (IIMs)

The simulated ATM cells resulting from the sender network interface module are forwarded to the IIMs for processing. Each IIM is responsible for the introduction of one type of impairment. It is possible to divide the IIMs into real-time and non real-time categories.

The non real-time impairment insertion modules, i.e. cell error (CE), cell loss (CL) and cell mis-insertion (CM) modules, can be implemented by manipulating the CELL file created by the sender network interface module before the simulation runs. Each of these three IIMs first opens a CELL file, changes the content according to the impairment type and the value of the corresponding network impairment parameter and then saves the content to a different file. Hence there will be the original 'clean' CELL file, and a second 'corrupted' CELL file with the errors and losses having been imposed by the non real-time IIMs.

The real time impairment insertion modules are the ones concerning the timing relationship between the simulated ATM cells at the sender and the receiver network interfaces, i.e. modules that introduce CTD and CDV. These two IIMs are implemented as real-time processes and perform their tasks while the simulation is running. This is known as Cell-by-Cell processing where the simulated ATM cells are handled one after the other from the CELL file generated above. A certain amount of delay is imposed on each simulated cell by the CTD module. On top of a constant amount of delay specified by the CTD parameter, each simulated ATM cell is subjected to a variable amount of transfer delay introduced by the CDV module according to the CDV parameter. These simulated ATM cells are then passed to the receiver network interface module for simulated destination network interface processing.

4.1.3 The Receiver network interface module (RNIM)

Before the decoding of video information is carried out, the simulated ATM cells are passed to the RNIM for destination end-system processing. Besides the CTD and CDV IIMs, the receiver network interface module is also included in the real-time portion of the simulation because of the timing requirements between the simulated ATM cells emerging from the IIMs and the video decoding process (Figure 4-3). If AAL 5 with a null SSCS is used (as in the case of SNIM), the four functions at the network interface of the receiver host that need to be simulated include the following (in the order of occurrence):

- Verification of the HEC field against bit errors in the cell header at the TC sublayer of the Physical Layer;
- Extraction of the ATM cell header at the ATM layer;
- Reassembling of 48-byte ATM cell payloads into CPCS-PDU at the Segmentation and Reassembly (SAR) sublayer;
- Padding removal and the CPCS-PDU trailer functions (including CRC checking) of the Common Part Convergence Sublayer (CPCS).

As in case of the SNIM, the four functions can be integrated into one module or separated into four modules depending on the implementation. The inputs to SNIM are simulated ATM cells emerging from the IIMs and the output from this module is a reassembled video bit stream. The reassembled video bit stream is then forwarded to a receive buffer, which is required to store the bit stream temporarily while it is waiting to be decoded by the video decoder. If the support for a simulation on AAL 1 transport of video stream is required, an addition module is required within the receiver network interface module in order to handle timing relationship as well as the sequence count operations in the CS sublayer of AAL 1.

The functional blocks in the RNIM are 'symmetrical' (but in reversed order) to those in the SNIM because of the communications between peer layer pairs in an ATM network. For example, while the SNIM carries out the generation of CRC at CPCS, the RNIM verifies the integrity of the CRC. This simulates the end-to-end error checking in an ATM network at the AAL 5 Common Part Convergence Sublayer.

4.1.4 Implementation considerations

The most obvious way to realize Design I is to implement the entire simulation on a single computer. The non real-time portion of the simulation is carried out first, namely the SNIM and three IIMs. The output of this part of the simulation is a 'corrupted' CELL file. The 'corrupted' CELL file is processed by the real-time portion of the simulation, which includes two IIMs and the RNIM. The video bit stream resulting from the RNIM are decoded and displayed by the video decoder. If the MPEG video encoding technique is employed in the simulation and a software MPEG decoder is used, the processing power requirement for the concurrent processes, i.e. the two IIMs, the RNIM, the software decoder and the video display etc, needs to be considered. This is because the software decoding of MPEG video alone is processor intensive. Therefore, if this design is adopted, the processing power budget for various algorithms of the above real-time simulation modules needs to be evaluated.

4.2 Design II

Instead of a functional simulation of video transportation over ATM, this design approach tackles the problem differently and aims to construct an emulated network for the NetIE architecture. This approach is motivated by the fact that the SNIM and the RNIM in Design I are trying to carry out functions that correspond to the processing of data occurring in the end-systems. Since the SNIM and RNIM implementation aims to imitate the actual implementation of the network interface in end-systems and therefore can only be as good as the latter, a direct alternative is to make use of end-system network interfaces so that SNIM and RNIM are not required.

In this approach, existing end-systems that send and receive digital video are combined to form an emulated ATM network. This emulated network may comprise of a sender, a receiver, an ATM switch and a virtual ATM switch as shown in Figure 4-4. Alternatively, the virtual switch may be connected directly between the end-systems. Figure 4-5 defines the scope of Design II by considering the functional block diagram of a VoD system. In contrast to Design I, this approach aims to emulate the ATM network only. Therefore the scope does not include the network interfaces of end-systems.

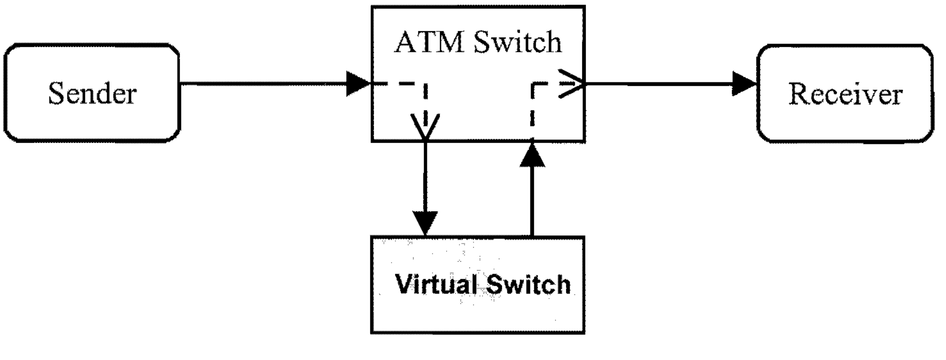


Figure 4-4 The second approach to carry out Network Impairment Emulation

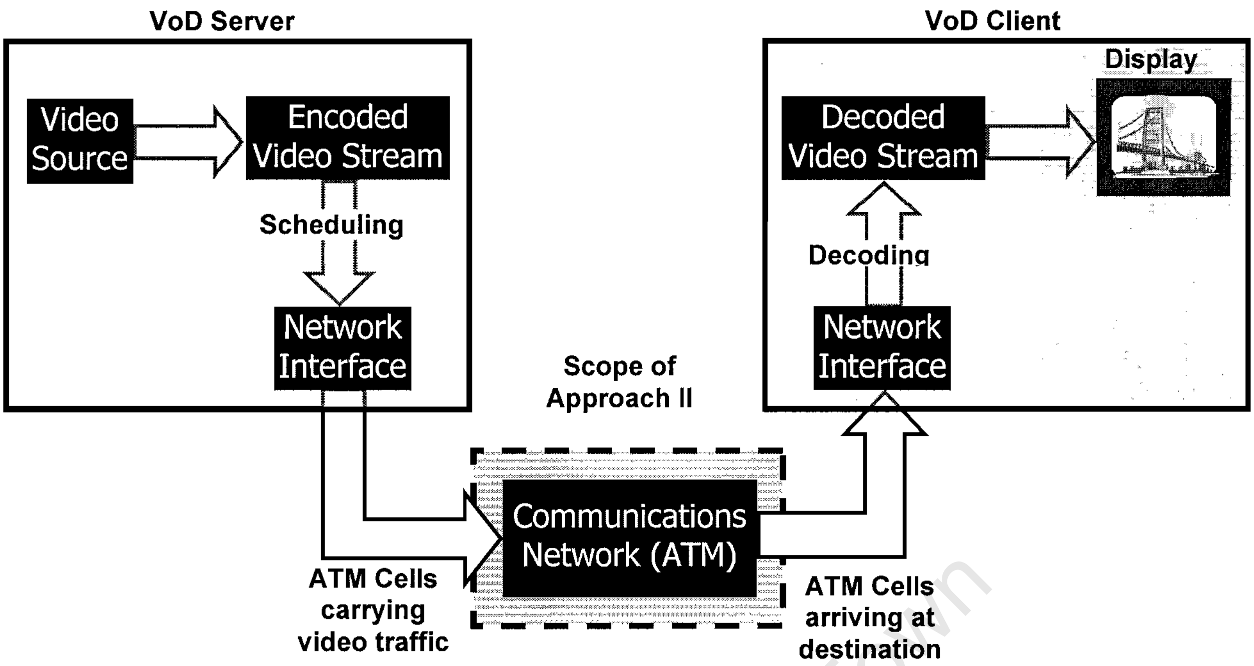


Figure 4-5 The Scope of Design II

Design II can be broken down into the follow components:

- The Virtual ATM Switch
- The interface between the virtual switch and the end-systems
- The Impairment Insertion Modules that reside on the virtual switch.

4.2.1 The Virtual ATM Switch

The idea of a virtual ATM switch originated from the Information and Telecommunications Technology Centre (ITTC) at the University of Kansas [VAS99] and it appears to fit the requirements to accomplish network impairment emulation. The virtual ATM switch distribution releases software for enabling ATM switching capabilities on a Personal Computer running the Linux operating system[17]. The ATM switching functionality is performed in software and uses ATM host adapters as switch ports. Although this may not yield the performance of a real ATM switch, it enables the deployment of an ATM switching platform at a much lower cost for experimentation.

[17] Before a computer running Linux can be used as an ATM virtual switch, ATM support (commonly known as ATM-on-Linux) has to be included in the basic Linux operating system [AOL96]

The virtual ATM switch distribution consists of three components [VAS99]:

- A public-domain user-level tool component to establish associations between ATM adapters and switch ports. A switch port can be established for any type of physical ATM interface (i.e., ATM host adapter) installed on the computer. This component provides support to set up and tear down permanent virtual circuits (PVCs) on the virtual ATM switch and is based on extensions to the set of tools provided in the ATM-on-Linux distribution.

- A public-domain kernel-level component that provides support for ATM cell switching between two or more configured virtual switch ports. The kernel is based on extensions of a standard Linux kernel configured with ATM support.

- A restricted user-level ATM signalling component that provides control for the kernel-based virtual switch fabric. This component allows the setup / teardown of PVCs and SVCs with support for UNI3.0, UNI3.1 and IISP. This component is based on extensions to Q.Port signalling software from Bellcore (Copyright © 1993-1999).

The kernel component and the user-level tool component of the distribution are released as free software under certain copyright restrictions. For the ATM signalling component, the public release of the Q.Port signalling software for the virtual ATM switch fabric is in binary form only.

Before an adapter installed on the virtual switch can be used, a virtual switch port has to be set up and associated with the adapter through the use of the user-level tool. Virtual Channel Connections (VCCs) can then be set up through the virtual switch. If PVCs are required, the user-level tool can be used to specify the input and output ports as well as the VPI/VCI for the connection [SAN99]. Alternatively, the Q.Port signalling software can be used to configure PVCs and SVCs [QPO96]. As an ATM cell arrives at an input port, the physical ATM adapter driver generates an interrupt to the Linux kernel to indicate such an event. A table look-up occurs in the kernel level cell-switching component in order to find an entry that matches the input port number and the VPI/VCI values of the ATM cell. If the entry is found, the cell is updated with outgoing VPI/VCI values and is switched to the output port according to the switching table.

4.2.2 Interfacing with the sender and the receiver

As shown in Figure 4-5, the network interfaces at the end-systems are not within the scope of Approach II. The network interface functions are handled in the end-systems as if the hosts are connected to an ATM network. In fact, it is neither required nor necessary for the end-systems to be aware of the existence of the emulated ATM network. They can be made to believe that they are connected to an ATM network. Depending on the implementation of the application running on the end-systems and that of the virtual ATM switch, it is possible for end-systems to require little or no modification when they changes their connection from an ATM network to the emulated network made up of the virtual switch.

The functions carried out by the network interfaces are dependent on and often limited by the specific implementation of the ATM adapters and the associated driver. In this particular case, the ATM adapters used in the end-systems are the ForeRunner LE series from FORE Systems® because of its availibility. However, according to the technical specifications [LEspec], the LE series adapters only provide hardware support for AAL ¾, AAL 5 and "Raw Cell" formats (AAL 0) but not for AAL 1. Therefore, if the transportation of video streams over AAL 1 is required, an alternative needs to be considered. One of the ways to add support for AAL 1 in the LE series adapters is to use software modules in the end-systems to handle the processing in the AAL 1 and ATM layers so that the ATM adapters are only responsible for the Physical Layer functions. For example, SNIM and RNIM for AAL 1 can be implemented on the sender and receiver end-systems respectively as in Approach I. ATM cells can then be sent and received between the two hosts as raw or AAL 0 cells.

A minimum of two ATM adapters are installed on the virtual switch. One of them connects to the network adapter on the sender host and the other one connects to that on the receiver host. Depending on the implementation of the network components of the video application being connected to the emulated network, a PVC or a SVC can be set up between the two ATM adapters (ports) installed on the virtual switch. The Impairment Insertion Modules are situated between the two switch ports as shown in Figure 4-6.

4.2.3 The Impairment Insertion Modules

Although ATM switches operate up to the ATM layer for user-plane functions, the virtual switch implementation offers optional support for AAL 5 packet switching, i.e. it can be configured to switch at the AAL 5 level or the ATM level. At the AAL 5 level, CPCS-SDUs are reconstructed from the ATM cell payloads before they are being handled by the switching functions. Each CPCS-SDU is switched as a whole to the output port before it is segmented and descends down the protocol stack again. On the other hand, if the virtual switch is operating at the ATM level, ATM cell payloads are not reassembled into CPCS-SDU and switching is performed on a cell-by-cell basis.

Since the virtual switch is implemented in software, all the ATM cells going through the virtual switch are accessible at the virtual switch computer during the switching process. This allows for the Cell-by-Cell processing that is required in network impairment emulation. In this case, the virtual ATM switch are configured to work at the ATM level so that an ATM cell constitutes a switching unit and is not reassembled into CPCS-SDUs.

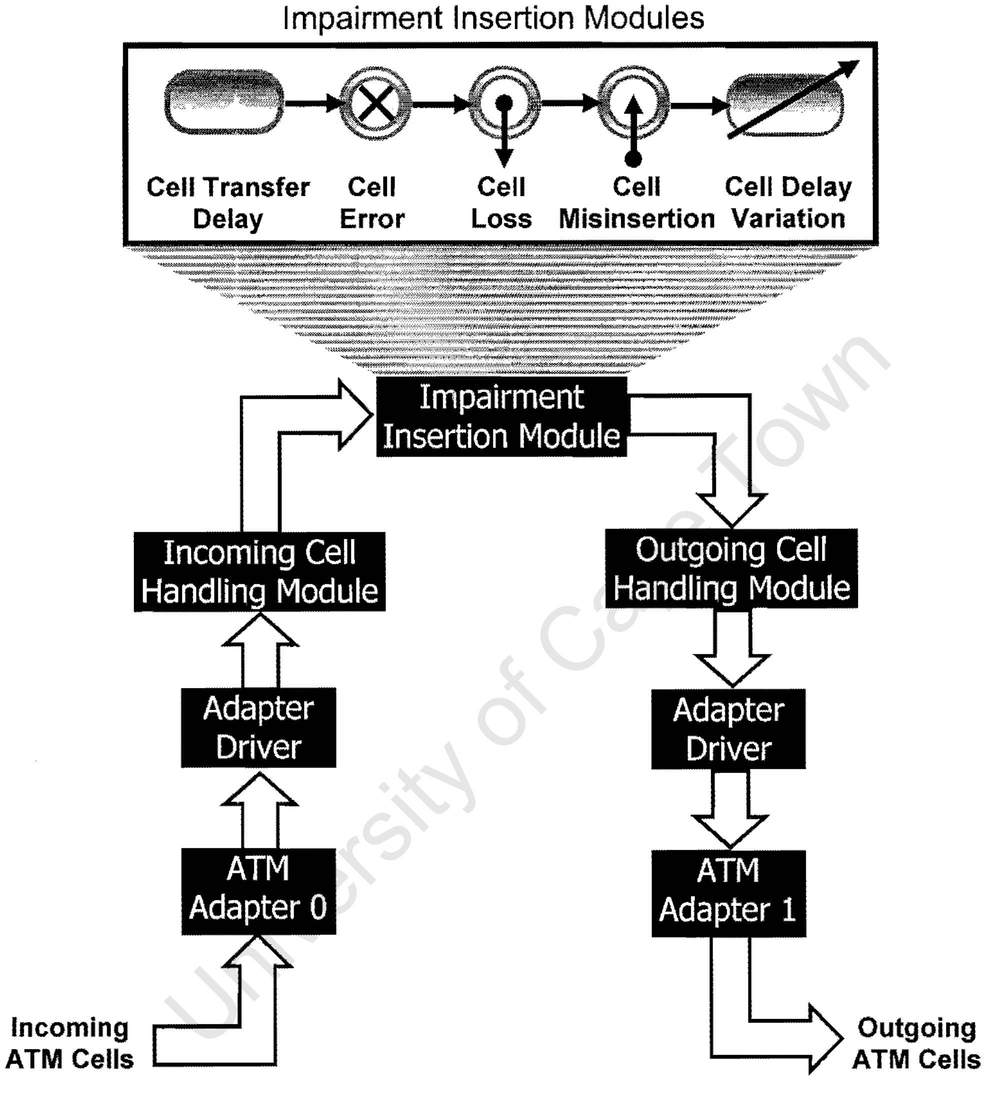


Figure 4-6 Functional block diagram of the virtual ATM switch

In the virtual switch, network impairments at levels specified by the 'knobs' are inserted only on the VCC carrying the video traffic under test. Figure 4-6 shows the functional block diagram of the emulated network comprised of the ATM virtual switch. All ATM cells originated from the video source are handled by the incoming cell-handling module, which checks each incoming ATM cells on all VCCs and ensures that only the ones belonging to the connection under test (i.e. carrying test video traffic) are forwarded to the Impairment Insertion Modules. There is one IIM for each type of impairment as is the case in Design I and they are implemented as software modules running on the virtual switch. ATM cells emerging from the impairment insertion modules are delivered to the outgoing cell handling module, through the ATM adapter and onto the link towards the destination.

In terms of the operating system architecture, it is possible for the Impairment Insertion Modules to be situated in the virtual switch computer at three different levels, namely the kernel-level, the driver-level and the user-level. This section describes the placement of the IIMs in the three levels and the advantages / disadvantages in each case, as well as the fundamental differences between them.

Kernel-level – since the software components responsible for the switching function of the virtual ATM switch is compiled as part of the Linux kernel, it is possible to include the impairment insertion modules as part of the virtual switch software located within the kernel. Incoming ATM cells are handled by the ATM adapter and the adapter driver. If a virtual switch port has been set up for that ATM adapter and VCCs have been configured in the virtual switch, cells that match any of the VPI/VCI entries are forwarded to the switching components in the kernel. Therefore, the impairment insertion software modules can be combined with the switching components so that the required network impairments are added to the cells during the cell-switching process.

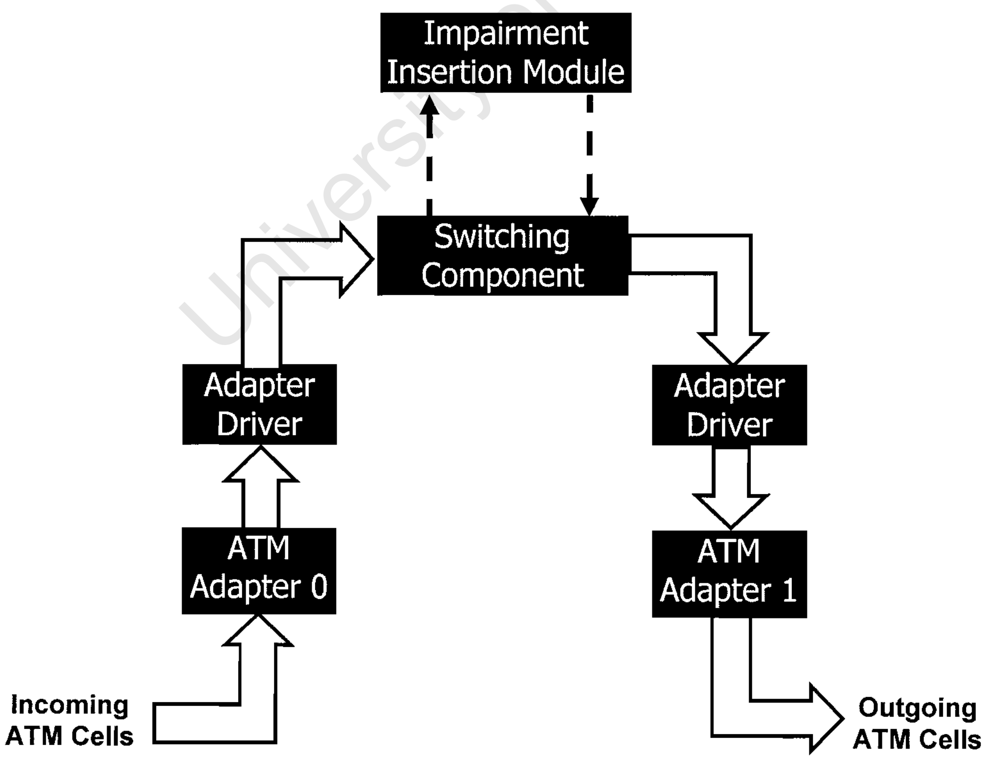


Figure 4-7 Putting the IIMs in the kernel-level

The placement of the IIMs within the switching component in the Linux kernel allows for faster function calls in software because of the lower processing overheads associated with function calls within the kernel than that from, for example, the user-level. In addition, this is an integrated approach for the IIMs to be incorporated in the kernel itself. On the other hand, this method also has a few disadvantages. The presence of the IIM modules within the kernel implies that each time the IIMs is modified[18], the entire Linux kernel needs to be recompiled. The process of recompiling the kernel takes 7 to 20 minutes or more depending on the processor speed on the virtual switch host computer and it becomes very inefficient if frequent modification to the IIM implementation is required, especially during the early stages of development.

One alternative to speed up the development of the IIMs inside the kernel is to compile the IIM as a kernel module. In this case, the kernel only has to be compiled once, during which the kernel becomes aware of the existence of the IIMs but does not include the actual implementation of the modules. When the IIMs are required, they can be inserted into part of the kernel with the '*insmod*' command. Subsequent changes to the IIMs only require the compilation of the IIMs and is therefore more efficient. Whether the IIMs are compiled as part of the kernel or as a kernel module, the placement of IIM at the kernel level results in a kernel of larger size. If the implementation of the IIMs is very large in size, this method may have an adverse effect on the operations of the Linux kernel. Therefore, if this method is adopted, this factor needs to be taken into consideration during the implementation stage.

Driver-level – an alternate method is to modify the ATM adapter driver in order for it to become aware of the existence of the IIMs so that cells belonging to the VCC under test can be processed by the IIMs as required. In the modified ATM adapter driver, the VPI/VCI values of all the incoming ATM cells are monitored in order to find the cells belonging to the test connection carrying video traffic. Instead of passing these cells to the kernel-level switching component, the driver forwards these cells to the IIMs. The IIMs carry out their tasks and redirect the cells back to the driver, which passes these cells to the switching component as usual. Unlike the previous method, the switching component is not aware of the presence of the IIMs in this case.

The advantage of this method is that the incoming cells being handled by the adapter driver are in a native form as they are transported in the network. They have not been subjected to any processing in the kernel-level switching components of the virtual switch. Therefore, the functions carried by the incoming cell-handling module (Figure 4-6) can also be determined by the implementation of the emulated network and not dictated by the implementation of the

[18] a modification refers to any major / minor changes and even the correction of typing errors in the code

virtual ATM switch. On the other hand, the modifications required in the adapter driver imply that this method is not portable across different ATM adapters. If a different adapter is installed on the virtual ATM switch and it requires a different adapter driver, the changes made to the original driver have to be mirrored on the new driver for it to recognize its interaction with the IIMs.

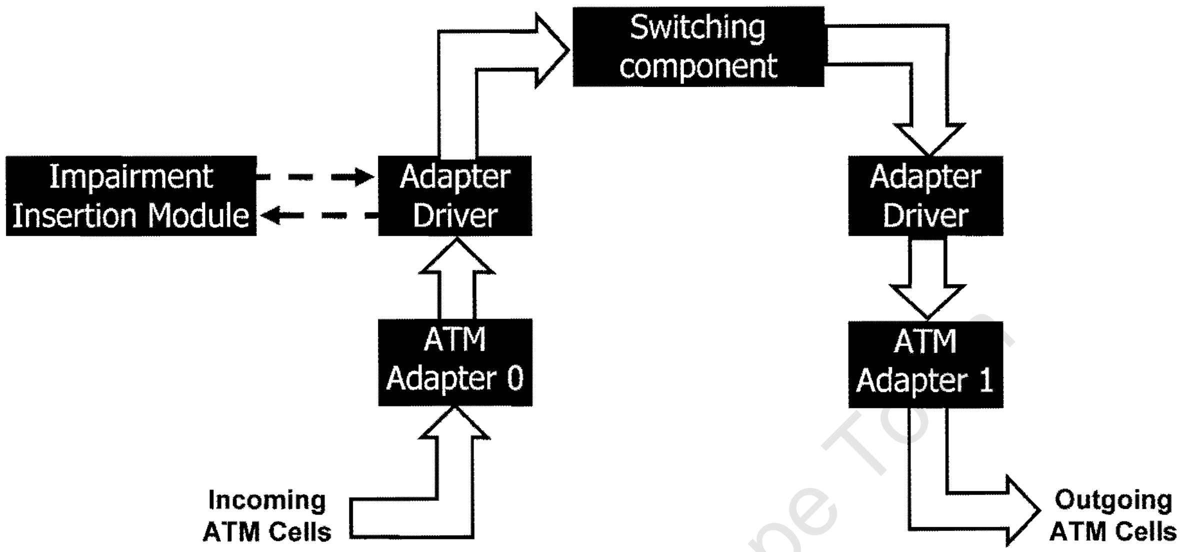


Figure 4-8 Putting the IIMs in the driver-level

User-level – During the investigation into the first two methods, it became apparent that the first two methods requires in-depth knowledge of the Linux kernel as well as the virtual switch software and the adapter driver implementation. The knowledge gained in the studies of the above topics has helped the search of a simpler way of placing the IIMs within the virtual ATM switch. Instead of using the kernel-level switching component in the virtual ATM switch software distribution, this method aims to carry out the functions of the emulated network above the adapter driver (Figure 4-6) at the user-level. This includes the incoming and outgoing cell handling module as well as the impairment insertion modules. These three modules and the sub-modules are incorporated in user-level program(s) running on the virtual switch computer. These user-level programs replace the virtual ATM switch software released by ITTC at the University of Kansas.

The user-level programs send and receive ATM cells through the use of the Application Programming Interface (API) defined for ATM related system services under Linux [API96]. This API is an extension to the 4.3 BSD release of UNIX socket interface with the support for additional functionality required for ATM. The incoming and outgoing cell handling modules create ATM connection sockets on adapter 0 and adapter 1 respectively in order to handle the cells that belong to the connection under test. Just as in the other two methods proposed, the cells that belong to the test connection are processed by the impairment insertion modules. However, in this method, the IIMs are going to be implemented in the user-level and not within the Linux kernel.

One of the advantages of this method is its simplicity. The use of the ATM-API hides the complexity of the Linux kernel and the ATM adapter driver away from the implementation process. The user-level programs are enabled by the ATM-API to gain access to the ATM cell contents and therefore can perform cell-by-cell processing required. The other advantage is that other than the IIMs, this method also enables full control over the functions carried out by the two cell handling modules. As a result, the behaviour of virtually the entire emulated network constructed is controllable during implementation. In addition, the size of the Linux kernel is not increased by the implementation of the IIMs because they are now situated in the user-level.

The disadvantage of the user-level implementation includes the higher inherent overhead associated with user-level function calls. Although the use of a powerful processor could result in an insignificant amount of overhead, the extent of the overhead needs to be considered during the implementation of the software modules.

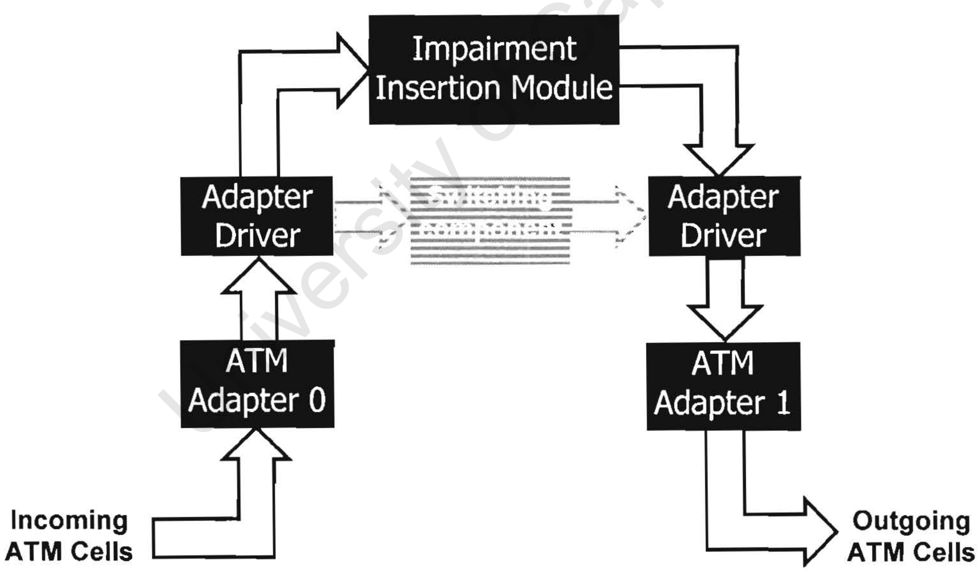


Figure 4-9 Putting the IIMs in the user-level

4.3 A Comparison of the Two Proposals

In this section, the two design approaches are compared against the functional requirements and features described in section 3.4. The two approaches are also compared against each other in terms of their characteristics and design strategies.

First of all, both designs carry out ATM protocol functions as one of their core requirements. Therefore, they are similar functionally. Secondly, the abstraction of real-world network impairments into five types of emulated impairments is handled by the Impairment Insertion Modules (IIM) in a modular fashion. As a result, network impairments can be efficiently introduced to video information. Thirdly, the modularity of both design approaches allows for scalability. New features can be added to the emulated network without having to comprehend the details of the existing modules.

However, there is one key difference between the two designs. Design I is a functional simulation of a VoD application supported by an ATM network. Its scope includes the network functions carried out by the end-systems. As a result, it includes the implementation of SNIM and RNIM, which simulate the tasks carried out by the server and client network interfaces respectively. This implies that Design I is very flexible in terms of end-system processing and that it is possible to experiment with different scheduling and transportation schemes within the simulation.

For this simulation to be a true reflection of a real VoD system, other tasks need to be considered. For example, in a VoD server, the scheduling and synchronization mechanisms are very important. Besides, traffic shaping and flow control schemes are often employed to ensure better performance. In a VoD client, data buffering and flow control mechanisms are often present to enhance video quality. For the simulation to include these functions as well, the implementation would become rather complex.

On the other hand, Design II takes advantage of existing VoD end-systems and focuses on the communications network only. This leaves the complex mechanisms of a VoD system to the implementation of the end-systems. The separation of the multimedia application (VoD) and the communications network (ATM) is crucial to the design because this results in a simpler and more efficient implementation of network impairment emulation.

One of the major advantages of Design II over Design I is its general applicability. The emulated ATM network can be used to study the effects of network impairments on the vast variety of multimedia communication applications described in 1.1. For a particular application, the vulnerability of different video compression formats or different ways to packetize video information towards network impairments can be investigated by modifying the implementation of the application in the end-systems accordingly. In this way, the effects of using different scheduling schemes, synchronization mechanisms and ATM Adaptation Layers in the end-systems can also be analyzed. Moreover, the emulated network can even be used to examine applications transporting information other than digital video. Modification to the emulated network is not required in most cases because it is up to the implementation of the application to decide which video compression technique, AAL, synchronization mechanism etc to use. The virtual ATM switch represents an ATM network only. It does not carry out any processes that are part of the end-systems (Figure 4-5). Therefore, the study of different applications or various mechanisms within applications only requires modifications in the end-systems and not the emulated network. On the other hand, the scope of approach I includes the communications network as well as some end-system processes (Figure 4-2). Despite the modularity of the design, major changes to the implementation are required if it is necessary to apply design I to another application, use another packetization / synchronization scheme, or send traffic over another AAL etc.

Considering the above factors, it should be obvious that Design II should be preferred to Design I not only because of its general applicability, but also for its simpler implementation.

Chapter 5.

Implementation of the Emulated Network

The chapter describes the implementation of the emulated network based on Design II with the user-level impairment insertion architecture. The implementation of the emulated network is very important because it not only validates the NetIE architecture and its design, but also provides the platform over which video artefacts and quality degradation can be examined. This chapter will also address the implementation issues associated with each impairment insertion module as well as the cell handling modules. The mechanisms with which impairment events are generated will be compared to those adopted by the ATM network impairment emulator module (NEM) of the Broadband Series Test Systems (BSTS). This chapter concludes by describing how the emulated network is integrated from its components and the BSTS is utilized in this study.

5.1 The Linux based Virtual ATM Switch

The virtual ATM switch is constructed from a PC Linux machine with two or more physical ATM host adapters. Since the basic Linux operating system is not ATM-aware, ATM support and virtual switch software have to be added before the PC machine can operate as a virtual ATM switch. In the Linux environment, version numbers of various components are important to ensure interoperability. The Virtual ATM Switch assembled in this study runs RedHat Linux 5.2 upgrade to a development kernel version 2.1.126. The virtual switch software and the ATM on Linux distribution used during this study are version 0.5 and 0.52 respectively. The procedures of building a virtual ATM switch (including the installation of Linux and the addition of ATM support) are provided in Appendix C.

The Virtual ATM Switch distribution [VAS99] provides a user-level command line tool '*vsw_ctl*' to create virtual switch ports out of physical ATM host adapters. To establish a PVC through the virtual switch, information regarding the input and output switch port together with the corresponding VPI/VCI fields for the VCCs is provided to *vsw_ctl*. The procedures of creating VCCs are very similar to those for commercial ATM switches.

5.2 The User-Level Impairment Insertion Architecture

Having decided to adopt the user-level impairment insertion architecture in the previous chapter, this section describes how it can be realized. The incoming and outgoing cell handling modules carry ATM cells to and from the virtual switch ports and enable the impairment insertion modules to be implemented in the application-level. The implementation of these two modules is greatly simplified through the use of the Application Programming Interface (API) for ATM on Linux. Each of the cell handling modules creates one communication socket over AAL 0 using the *socket ( )* function. AAL0 is preferred to AAL5 because AAL0 carries no communication overheads and maps user information directly to and from ATM cell payloads (i.e. cell-by-cell processing). If AAL 5 sockets were used, ATM cells will be reassembled by the adapter driver into AAL 5 PDUs and processing can only be performed on a PDU-by-PDU basis. Using the *bind ( )* function, these sockets are bound to the respective ATM network interface and the VPI/VCI used by the VCC carrying video traffic. With the AAL 0 communication sockets, the incoming and outgoing cell handling modules can send and receive ATM cells containing video traffic respectively at the application-level of the Linux operating system architecture (using the *read ( )* and *write ( )* functions). The switching of ATM cells containing control information is performed at the kernel level by the original virtual ATM switch software.

5.3 The Statistical NetIE Model

One of the most important aspects in the implementation of the Statistical NetIE model is the statistical generation of impairment events. An impairment event corresponds to the occurrence of a particular type of network impairment. It can either be monitored (as in the case of a real-world ATM network) or generated (as in the case of this emulated network). The generation of impairment events rely on the values of the impairment parameters and the statistical distribution defined. Two types of statistical distributions described in this section are normal and uniform. Deterministic generation of impairment events will also be explained. Lastly, the application of these distributions on the generated network impairments are explained and compared to the methods adopted by the ATM network impairment emulator in the Broadband Series Test Systems.

5.3.1 Statistical Distributions

It is the task of the Impairment Controller (IC) to determine which ATM cells should be affected by network impairments. Statistical Distributions are used to characterize and generate impairment events. They introduce variability to the generation of events based on the network impairment parameters specified to the IC so that events do not occur in a predictable manner. At this point it is important to note that the use of statistical distributions determines when network impairments should occur and not how they are introduced to the cell stream. The next section will explain how each type of network impairment is introduced to an ATM cell stream.

5.3.1.1 Generating Random Numbers

Statistical Distributions are often characterized by graphs whose shape is determined by the behaviour of a Random Variable governed by a Probability Density Function (PDF). The generation of events that is consistent with a certain statistical distribution requires the generation of random numbers in a specific range. In order to avoid the same sequence of random numbers every time, a random seed, which is usually associated with the time on the host computer, is assigned to the random number generator.

5.3.1.2 Gaussian Distribution

Figure 5-1 shows the probability density function (PDF) of the Gaussian distribution. Two statistical parameters associated with the Gaussian random variable are the mean (μ) and the standard deviation (σ). The mean sets the centre point of the distribution. The standard deviation determines the variability of the distribution. In other words, it controls whether the probabilities of events are clustered around the mean or spread over a wider range. These two statistical parameters correspond to the values of network impairment parameters specified on the 'knobs' of the network impairment emulator. The shape of the Gaussian distribution is governed by the following equation:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \times e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Equation 5-1 Probability Density Function for the Gaussian random variable X

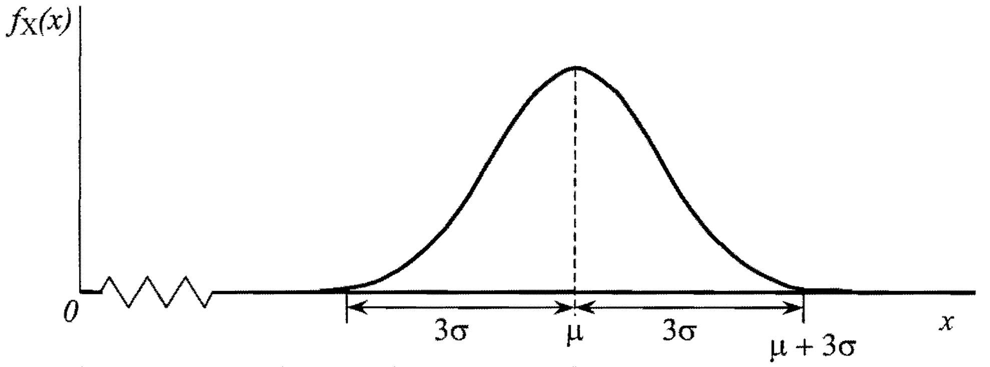


Figure 5-1 The Probability Density Function for the Gaussian Random Variable

The generation of impairment events that conform to the Gaussian Distribution requires the generation of two random numbers. First of all, the two statistical parameters, mean μ and standard deviation σ, are provided to the impairment controller. Secondly, the scope of the generation of impairment events needs to be defined. While 99.74 % of all the possible values within the Gaussian distribution occur between $\mu - 3\sigma$ and $\mu + 3\sigma$, more than 99.99 % of values within the Gaussian distribution are contained between $\mu - 4\sigma$ and $\mu + 4\sigma$. For example, if $\mu = 10^5$, $\sigma = 100$ and $\mu - 4\sigma$ to $\mu + 4\sigma$ are considered, then the impairment events can be generated by the following procedures:

- Generate a random number (called x) in the range from 0 to $\mu + 4\sigma$ (i.e. from 0 to 10,400);
- Substitute x into Equation 5-1 to obtain $f(x)$;
- Generate another random number (called y) in the range from 0 to 1;
- If y is less than or equal to $f(x)$, an impairment event has occurred. If y is large than $f(x)$, the impairment event has not occurred;
- Repeat for the each ATM cell.

Note that the value $1/\mu$ refers to the average occurrence probability of network impairments. It corresponds / relates to parameters such as Cell Loss Ratio or Cell Error Ratio.

5.3.1.3 Uniform Distribution

Figure 5-2 shows the probability density function (PDF) of the Uniform distribution. The Uniform Distribution contains a uniform probability of impairment occurrence throughout the interval defined by two limits. The upper limit U, and the lower limit L are the two statistical parameters associated with this distribution. They are the network impairment parameters specified on the 'knobs' of the network impairment emulator. The difference between the two limits determines the width of the distribution. The equation that defines the probability density function of the Uniform Distribution is:

$$f(x) = \begin{cases} \dfrac{1}{U - L} & \textit{if } L \le x \le U \\ 0 & \textit{otherwise } (\textit{i.e. } x < L \textit{ or } x > U) \end{cases}$$

Equation 5-2 Probability Density Function for the Uniform Distribution

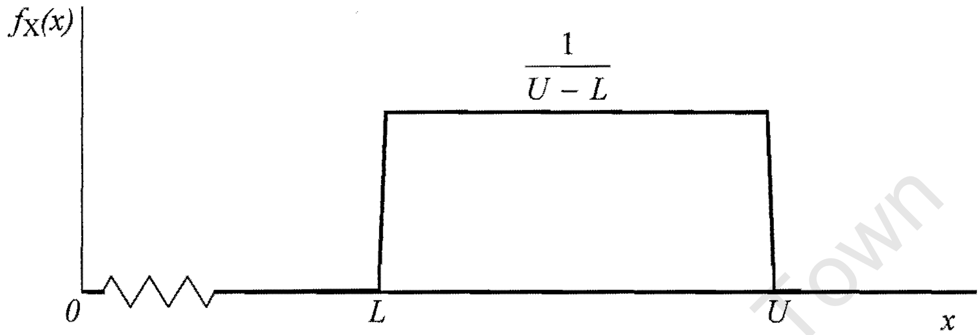


Figure 5-2 The Probability Density Function for the Uniform Distribution

As for the Gaussian Distribution, the generation of impairment events that conform to the Uniform Distribution requires the generation of two random numbers as shown below. The upper limit U and the lower limit L are provided to the Impairment Controller. For example, if U = 9,600 and L = 10,000, then the impairment events can be generated by the following procedures:

- Generate a random number (called x) in the range from 0 to U (i.e. from 0 to 10,000);
- Determine $f(x)$ from Equation 5-2;
- Generate another random number (called y) in the range of 0 to 1;
- If y is less than or equal to $f(x)$, an impairment event has occurred. If y is large than $f(x)$, the impairment event has not occurred;
- Repeat for the each ATM cell.

Note that the value $1/U$ refers to the average occurrence probability of network impairments as shown in Equation 5-3. It corresponds / relates to parameters such as Cell Loss Ratio, Cell Error Ratio, etc.

$$prob\,(impairment\ event) = prob\,(L < x \le U) \times prob\left(y < \frac{1}{U-L}\right)$$
$$= \frac{U-L}{U} \times \frac{1}{U-L} = 1/U$$

Equation 5-3 Calculation of Impairment Event Occurrence Probability

5.3.1.4 Deterministic Distribution

It is possible for network impairments to be generated in a deterministic manner. Equation 5-4 defines the PDF of the Deterministic Distribution. The single statistical parameter associated with this distribution is the mean μ. This is a special type of distribution because the PDF is a single impulse function concentrated at the mean value (i.e. there is no variation in the probability of impairment events). Conceptually, it can be regarded as a special case of the Gaussian Distribution with a very small σ or the Uniform Distribution with $U - L = 1$. The equation that defines the probability density function of the Deterministic Distribution is:

$$f(x) = \begin{cases} 1 & if\ x = \mu \\ 0 & otherwise \end{cases}$$

Equation 5-4 Probability Density Function for the Deterministic Distribution

The generation of impairment events with the deterministic distribution can be carried out as follows:

- Generate a random number (called x) in the range from 0 to μ (i.e. $0 < x \le \mu$);
- If x is equal to μ, then the impairment event has occurred. If x is not equal to μ, then the impairment event has not occurred.
- Repeat for the each ATM cell.

Although the average occurrence probability of network impairments is constant in deterministic distribution, impairment events do not occur in a predictable manner because the random number associated with each ATM cell is not the same every time. Therefore, different ATM cells will be affected by the impairment events each time the network impairment emulation runs.

5.3.2 Network Impairment Insertion

Having determined when impairment events should occur, the section covers the implementation of the Impairment Controller and the Impairment Insertion Modules. It also describes the interaction between them. While the Impairment Controller determines which ATM cells should be affected by network impairments, the Impairment Insertion Modules are responsible for the placement of network impairments within the flow of ATM cells.

5.3.2.1 Cell Error IIM

According to [TM4.0], cell errors occur in an ATM cell stream in a random manner. Therefore, it is suitable for the Deterministic Distribution to be used to generate cell error impairment events in the Impairment Controller (IC), which takes in the value of CER and uses it as the mean (μ.) of the distribution. The generated impairment events are passed to the cell error impairment insertion module. Since the IC only decides which ATM cells should be affected, it is up to the IIM to determine in which bit within a particular ATM cell should the error be present.

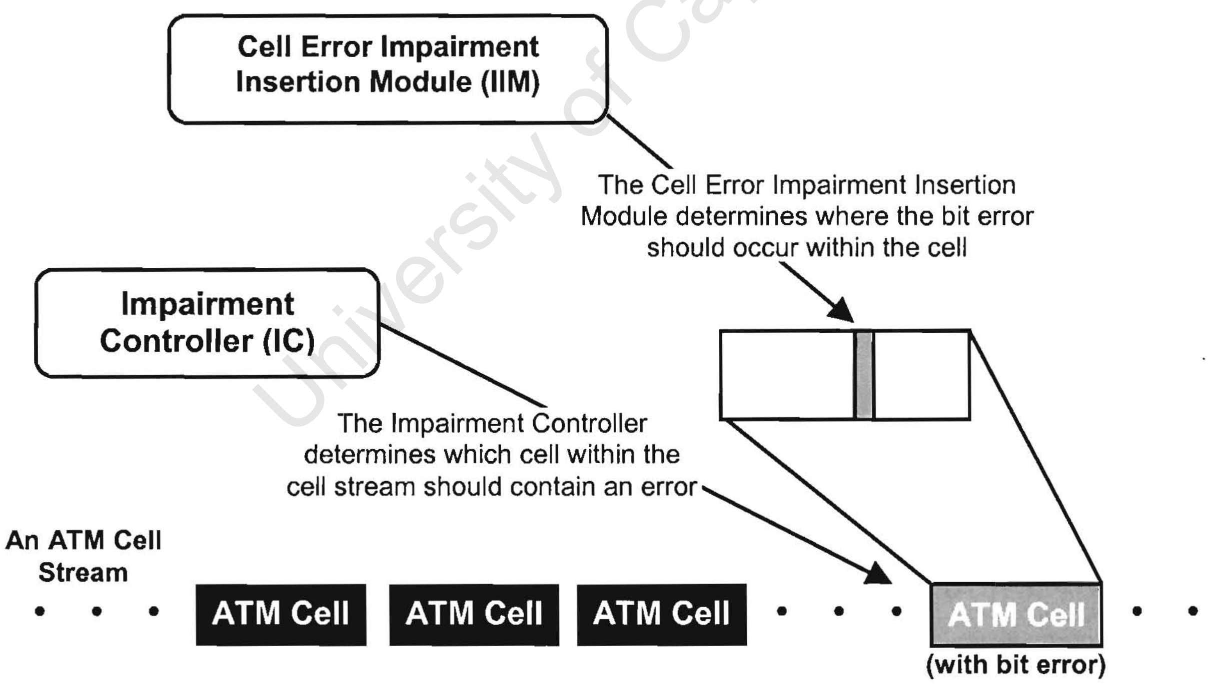


Figure 5-3 Generation Cell Error within an ATM cell stream

The study by Anagnostou et al [ANA92] shows that under most conditions, the occurrence of bit errors within the ATM cell header (resulting in possible cell loss or cell mis-insertion event) is insignificant compared to bit errors within the ATM cell payload. This is because of the difference in size (5 bytes against 48 bytes) and the presence of the Header Error Check.

Therefore, it is assumed that bit errors occurring within the ATM cell header can be neglected without significant effects.

When an ATM cell is indicated by the IC to have a bit error, it is put into an error buffer within the IIM. An error buffer is an array that has space to contain one ATM cell. The Errored cell outcome defined in section 3.4.2 refers to ATM cells with at least one error bit. Therefore, the value of CER is not affected by whether a single or multiple bit error is present in the errored cells. In the implementation of the IIM, each Cell Error impairment event corresponds to a single bit error. Multiple bit errors per errored cell can also be considered with minor modification to the IIM.

To determine where the bit error should occur within the cell, the IIM needs to generate a random variable in the range of 1 to 384. Each number corresponds to one bit in the ATM cell payload and each bit is equally likely to be the error bit as a result. For example, if the random variable is 50, then the value of the 2^{nd} bit (remainder of $50 \div 8$) of the 7^{th} byte ($1 +$ the quotient of $50 \div 8$) is toggled using the XOR operator.

5.3.2.2 Cell Loss IIM

The generation of Cell Loss can follow any one of the three types of statistical distribution according to the 'control panel' of the network impairment emulator. For each type of statistical distribution, the cell loss network impairment parameter (CLR) needs to be provided to the Impairment Controller together with the supplementary parameters such as σ for the Gaussian Distribution and U for the Uniform Distribution. The Impairment Controller then generates Cell Loss impairment events as described above.

Within the IIM, an ATM cell received from the incoming cell-handling module is lost when it is not being forwarded to the outgoing cell handling module. The IIM also has the option to replace lost cells by cells with payload containing 0's only. This can be used to simulate 'dummy cell insertion' in end-systems where 48 bytes of zeros are inserted in place of a lost cell upon detection. Several studies have reported that dummy cell insertion has the ability to conceal video artefacts and therefore reduces video quality degradation caused by cell losses. The effects of dummy cell insertion will be discussed in Chapter 6.

5.3.2.3 Cell Mis-insertion

The generation of Cell Mis-insertion can again follow any one of the three types of statistical distribution as specified on the 'control panel' of the network impairment emulator. For each type of statistical distribution, the cell loss network impairment parameter (CMR) needs to be provided to the Impairment Controller together with any supplementary parameters. It is important to note that the network impairment parameter for Cell Mis-insertion is defined as a rate and not a ratio because the mechanism producing mis-inserted cells is independent of the number of cells within the connection concerned. Therefore, the generation of Mis-insertion impairment events is different from that of Cell Loss and Cell Error. Instead of considering each ATM cell belonging to the video connection, the generation of Cell Mis-insertion is based on the inter-occurrence interval of impairment events (Figure 5-4).

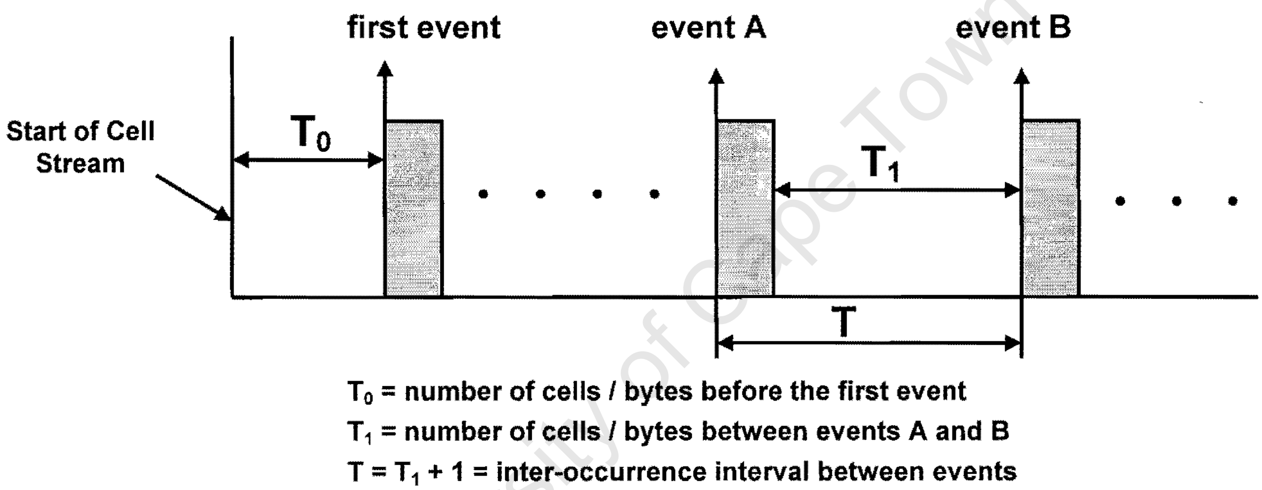


Figure 5-4 Generating Impairments based on Inter-occurrence Interval

When an ATM cell stream arrives at the incoming cell-handling module of the virtual ATM switch, the Impairment Controller decides (using statistical means described above) how many cells in the stream (T_0) should be allowed to pass through the emulated network before the occurrence of the first mis-insertion event. After the desired number of cells have gone passed, the cell mis-insertion IIM is instructed to insert an extra cell to the traffic stream. The IC then generates the next cell mis-insertion event by determining the inter-occurrence interval (T) between the current and the next event. For example, if the first cell mis-insertion event is determined to occur after 5,500 cells and the inter-occurrence interval is 10,500, then the second cell mis-insertion event should occur after the 16,000th cell in the original cell stream[19]. During a cell mis-insertion event, the mis-inserted cell originated from the IIM is passed to the outgoing cell-handling module to be sent to the receiving host.

[19] or after the 16,001st cell in the outgoing cell stream if no other impairment, such as cell loss, is present.

The Cell Mis-insertion Module allows for the content of the mis-inserted cell to be specified. In general, since a mis-inserted cell is either an unassigned cell or a cell from another VCC, the payload bits within the mis-inserted cells are either all set to '0' (i.e. an idle or unassigned [I.321]) or filled with randomly generated bits.

5.3.2.4 Cell Transfer Delay

Cell Transfer Delay refers to the constant delay experienced by all ATM cells and can be implemented using a simple First-in-First-out (FIFO) ring buffer shown in Figure 5-5. This buffer can either be implemented as an array or a singly linked list. As the ATM cells arrive at the CTD IIM, they are issued time stamps and are inserted into the queue. When the cell at the front of the queue has been delayed for the required amount of time, it is removed from the front and forwarded to the next IIM. At the same time, new cells are admitted to the back of the queue. Pointers are required to keep track of the front and the back of the queue so that insert and remove operations can be carried out. They are also used to detect the 'buffer full' condition in which case an error will be signalled.

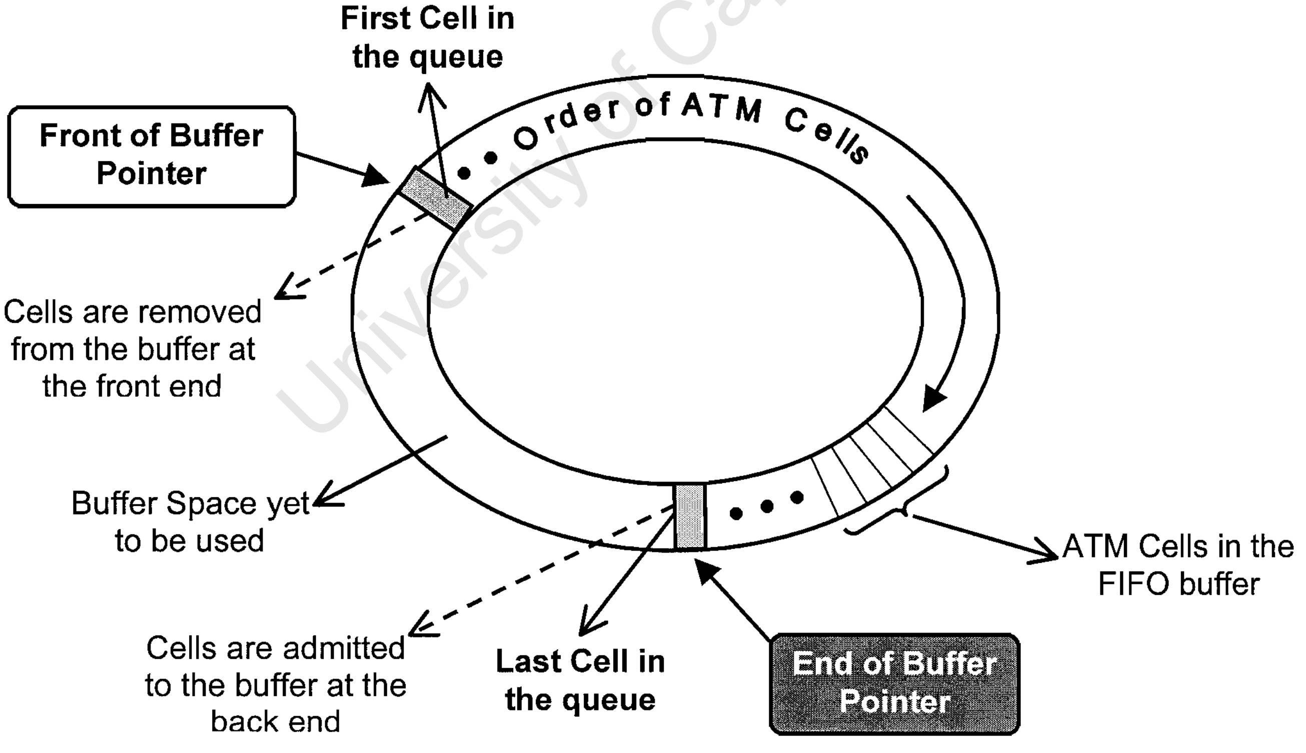


Figure 5-5 FIFO Ring Buffer for the implementation of CTD and CDV

The size of this queue should be related to the desired amount of constant delay introduced to the cells because the amount of delay determines the number of cells that need to be stored in any one instance. Therefore, more buffer space needs to be allocated[20] for the queue when a large transfer delay is required and vice versa.

5.3.2.5 Cell Delay Variation

Variable cell delay can be implemented in the emulated network through the use of a finite state machine. The two states in this state machine correspond to an increase and a decrease in cell delay by one cell time (Figure 5-6). The parameters supplied to the Impairment Controller are the peak-to-peak Cell Delay Variation (defined in section 3.4.2.4) and the maximum desirable amount of cell delay applied to the ATM cells. The IC is responsible for keeping track of the current cell delay value. The initial cell delay value can be set to the average of the maximum and minimum CTD [max CTD – (peak-to-peak CDV ÷ 2)]. Under normal operations, this value is equally likely within each state to stay in the current state or move to the next state. However, when the amount of delay introduced to the cell stream equals the maximum value, it can only be decreased. Similarly, when the amount of delay introduced to the cell stream equals the minimum value (i.e. maximum value – peak-to-peak CDV), it can only be increased.

The implementation of the CDV IIM can be based on that of the CTD IIM. When an ATM cell arrives, the delay value currently stored in the IC is added to the arrival time of the cell to obtain the departure time of the cell. This cell will then enter the FIFO ring-buffer where it will be held until the departure time. In order to preserve the ordering of the cells, the state machine cannot change its state when there is a continuous flow of ATM cells. It can only change state when there is a pause in the arrival of cells.

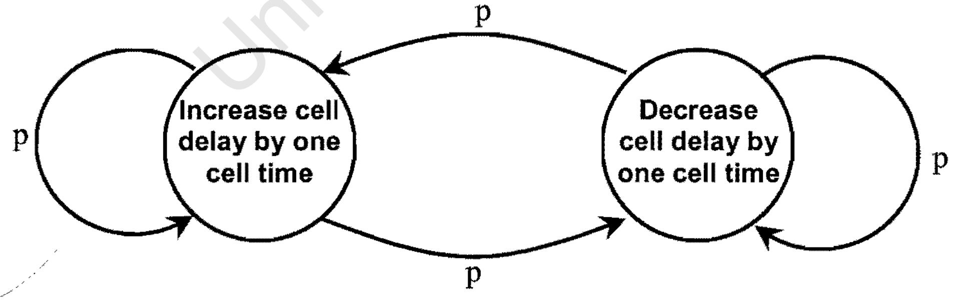


Figure 5-6 Introducing Cell Delay Variation with a Finite State Machine

[20] When the buffer is implemented as an array, buffer space is statically allocated and the maximum amount of buffer space is fixed. When the buffer is implemented as a linked list, buffer space is dynamically allocated and the size of buffer can be modified. However, since the buffer size should correspond to the maximum CTD, static or dynamic allocation of memory is irrelevant and the choice of implementation makes no difference.

5.3.3 Introducing Network Impairments with the BSTS

During the implementation of the NetIE model, the architecture of the Broadband Series Test Systems (BSTS) from Hewlett Packard is taken into consideration. The BSTS is a sophisticated UNIX®-based high speed ATM/B-ISDN test platform. It can perform comprehensive testing of all ATM protocol layers and is used for product development, quality assurance, network operations, type approval and conformance testing. This study is interested in the BSTS because one of its hardware modules, the E4219A ATM Network Impairment Emulator Module (NEM), is functionally similar to the NetIE model in many ways. The E4219A is used with a line interface module (LIF)[21]. The cell stream output from a line interface is routed over the internal cell bus of the BSTS to the E4219A module, which modifies the cell stream to introduce ATM network impairments. The resulting impaired cell stream is then routed back to the output of the line interface over the internal cell bus. This is illustrated Figure 5-7. Comprehensive details of the NEM can be found in [NEM98].

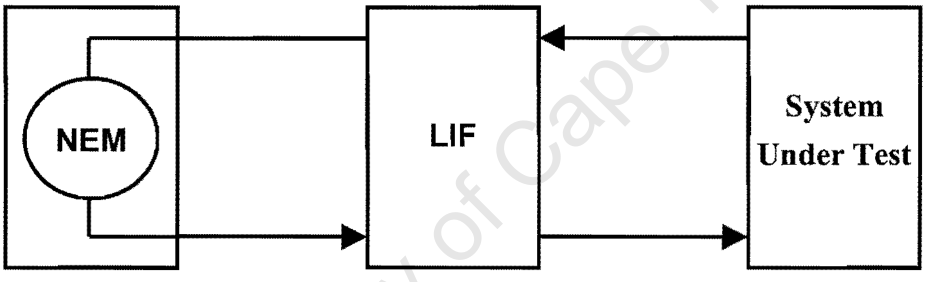


Figure 5-7 The Configuration of NEM to form an Emulated Network [NEM98]

A general comparison between the two test platforms is listed in Table 5-1. This section also aims to examine the mechanisms adopted by E4219A to generate impairment events and compare them to those employed by the NetIE implementation.

Unlike the NetIE implementation, the NEM generates the Cell Loss and Cell Error impairment events based on the inter-occurrence interval of these events as described in Figure 5-4 and section 5.3.2.3. One advantage of this method is lower processing. Instead of processing each ATM cell and determining if an impairment event has occurred, this method processes each occurrence of an impairment event and determines when the next one should occur. On the other hand, the drawback of this method is that impairment events of the same type are bound to be separated by relatively large intervals due to the nature of this scheme. In other words, it is

[21] Optionally, a Cell Protocol Processor (CPP), which generates ATM traffic, can be used in conjunction with the NEM so that generated traffic with known characteristics can be tested against the influence of network impairments.

almost impossible for impairment events of the same type to occur within short intervals under this mechanism. This is a limitation of generating impairment events this way.

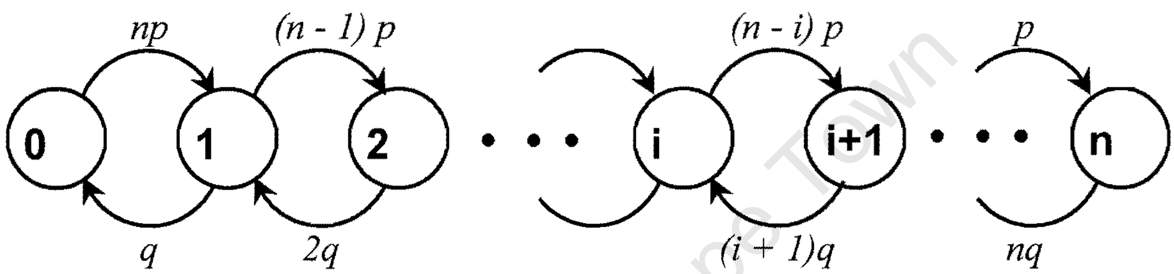
	NEM (BSTS)	NetIE
General application	connects to end systems which send and receive ATM cells, independently introduces five types of network impairments in a controllable manner	
Interfacing with end-systems	Dependant on the line interface module	Dependant on the ATM network adapter installed on the virtual ATM switch
Traffic Flow	Only caters for one-way traffic. For full duplex configuration, use two NEM / NetIE units, one in either direction	
Mechanisms to generate Network Impairments	Using Statistical Distribution only	Using Statistical Distribution and possible to specify the placement of impairment
Range of Statistical Distributions	Normal, Exponential, Deterministic (generate impairments at fixed intervals), Uniform, User-defined	Gaussian, Uniform, Deterministic (Random)
Introduction of Cell Loss and Cell Error	Generation of inter-occurrence intervals of impairment events following statistical distribution	Considering cells individually and generate impairment events according to statistical distribution
Introduction of Cell Mis-insertion	Generation of inter-occurrence intervals according to specific statistical distribution	Similar
Introduction of Cell Delay Variation	Using a 927-state Markov Chain	Using a Finite State Machine
Current situation	Commercially available	Development stage
Applicability	Can be used to test all types of ATM traffic	Can be used to test all types of ATM traffic
Scope for Further extensions	C programs can be run to automate testing	Possible due to modularity of the design
Other features	TTL trigger signal can be created to an external device to measure impairment events	AS model for a comprehensive study on a particular type inform-ation transported over ATM

Table 5-1 Comparisons between NEM from HP and NetIE proposed in this study

The mechanism of introducing Cell Delay Variation is somewhat similar between the NEM and NetIE. The difference is that the NEM uses the Markov process model to implement variable cell delay. Each state in the Markov Chain defines the amount of cell delay (in terms of cell time) to be applied to each matching cell. Therefore, the steady state probabilities of the Markov Chain define the delay distribution. For example, the mean and standard deviation of the Binomial Distribution can be defined as follows[22]:

$$\mu = n \times p$$
$$\sigma^2 = np(1 - p)$$

Equation 5-5 The resulting Binomial Distribution from the Markov Chain



Equation 5-6 The Markov Chain used by NEM to implement Variable Cell Delay [NEM98]

In the NEM implementation, the maximum variable cell delay is limited by the number of states in the Markov Chain. On the other hand, the maximum variable cell delay is limited by the amount of allocated buffer space in the NetIE implementation.

5.4 The Application-Specific NetIE Model

As explained in chapter 2, the application-specific NetIE model applies only to a particular type of traffic, which is MPEG in the context of this study. The generation of impairment events is based on statistical distributions as well as the type of information contained in individual ATM cells. This requires the entire MPEG video stream to be analyzed at the video source before it is sent. The MPEG bit stream is divided into three categories: system information, video information and audio information.

Each of these types of information can be identified by a start code prefix followed by a start code value [M13818]. The start code prefix is a string of twenty-three bits with the value zero

[22] The derivation of them can be found in Appendix A of the NEM User's Guide in the \BSTS directory on the CD

followed by a single bit with the value one, i.e. 0000 0000 0000 0000 0000 0001 in binary or 00 00 01 in hexadecimal notation. The start code value is an eight-bit integer which identifies the type of the start code. System information can be located by the system start codes such as the pack_start_code and the system_header_start_code. The start code values for system start codes ranges from 0xB9 to 0xFF. For example, the start code value for the pack_start_code is 1011 1010 in binary or BA in hexadecimal notation and the entire start code is 0x00 00 01 BA. The start code value for the system_header_start_code is 0xBB.

Video information can be traced by picture_start_code, whose start code value is 0x00 and the entire start code is 0x00 00 01 00. Optionally, video information can be categorized further according to the type of frame by examining the picture_coding_type field. For example, a picture_coding_type field of 001 implies that the current frame is an I frame, while P and B frames have picture_coding_type fields of 010 and 011 respectively. The start of an audio packet can be identified by a 12-bit *syncword* code.

It is important to note that only two types of network impairments, Cell Error and Cell Loss, should be included in the Application-Specific NetIE model. This is because the AS model introduces network impairments to specific type(s) of information only. If Cell Transfer Delay or Cell Delay Variation is introduced to ATM cells containing information in a certain category and not to cells containing other types of information, the order of the ATM cell stream will be altered. This violates the connection-oriented nature of ATM, which provides in-sequence delivery of ATM cells within Virtual Channel Connections as defined in [I.150]. Moreover, Cell Mis-insertion is irrelevant to the AS model because it cannot be applied only to a certain type of MPEG information due to its nature.

5.5 Integration of the Emulated Network

The virtual ATM switch and the Impairment Insertion Modules described above constitute an abstraction of the ATM network core only. For the emulated network to be a realistic representation of a real-world ATM network and a useful platform to experiment with video traffic, it needs to incorporate physical end-systems that send and receive digital video traffic. Throughout this study, a Video-on-Demand application, which is jointly designed and implemented by Mr. M.J. Roux and Mr. P. Vine, is chosen as the application implemented on the end-systems within this emulated network. In this Video-on-Demand application, digital video encoded in the MPEG-1 format is sent uni-directionally over ATM Adaptation Layer type 5 from the server host to the client host. This section highlights some of the important steps taken during the integration of the end-systems and the virtual ATM switch into an emulated network.

The ATM network interface cards (NICs) used in the emulated network are ForeRunner LE series 155 Mbps adapters [LEspec] with RJ-45 connectors. Because the ATM adapters are not attached to an ATM switch but directly connected with one another (commonly known as the back-to-back configuration), it is necessary to use special UTP 5 cross-over cables for the connections. The crossover cables required are different from the standard UTP 5 crossover for Ethernet adapters and their pin-outs are as follows:

RJ-45		RJ-45
1	-------------	7
2	-------------	8
3	-------------	3
4	-------------	4
5	-------------	5
6	-------------	6
7	-------------	1
8	-------------	2

Figure 5-8 Pin-Outs of UTP 5 Cross-Over Cable for ATM NICs [AOL98]

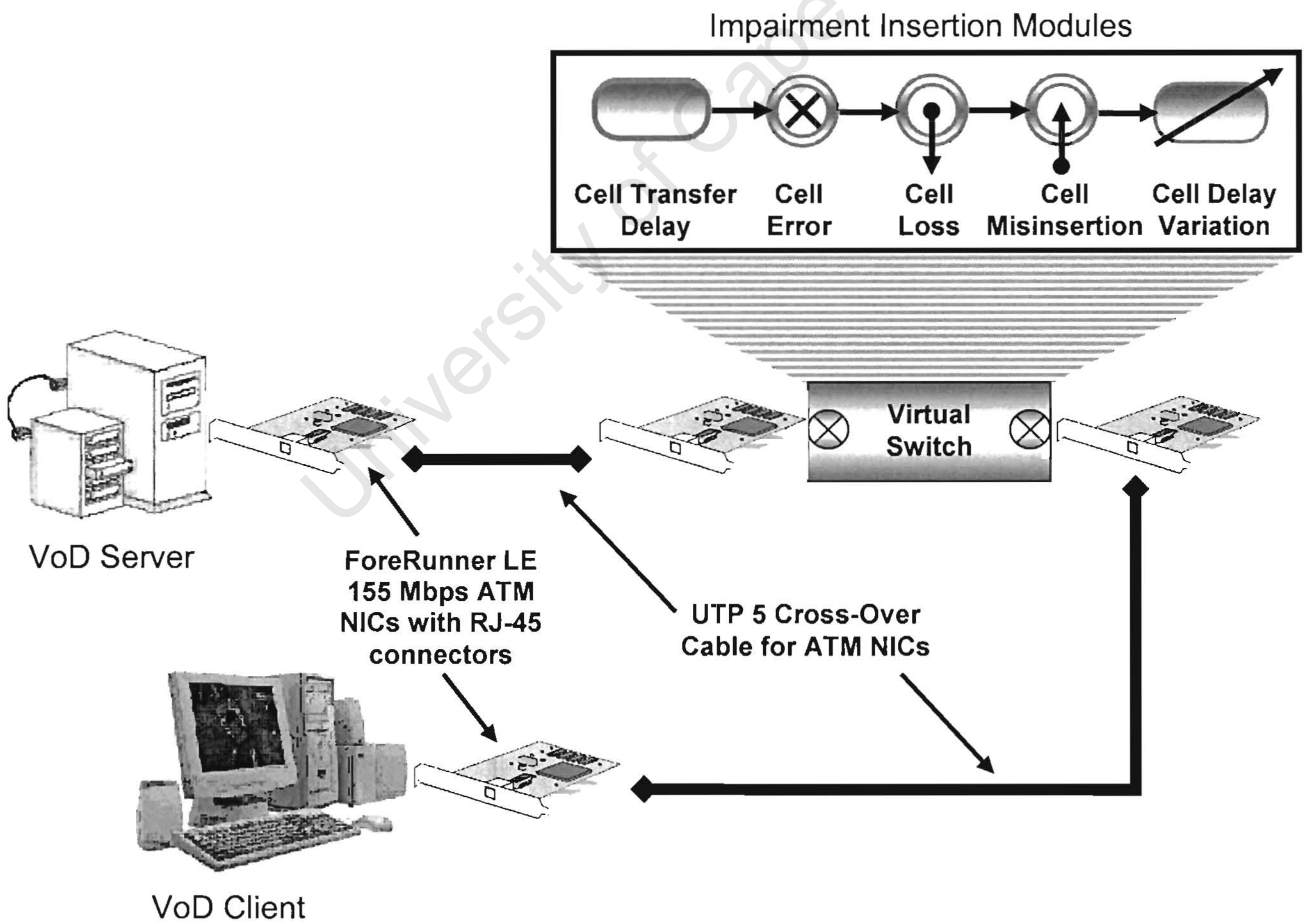


Figure 5-9 Physical Layout of the Emulated Network

Figure 5-9 illustrates the physical layout of the overall emulated network which consists of three computer systems. Since ATM connections between the VoD Server and Client are set up when necessary and they are relatively short-lived, these connections are set up as SVCs. For the VoD system to function properly within the emulated network, ATM signalling function needs to be implemented on the virtual ATM switch in order to handle connection set up / tear down requests. This turns out not to be a straightforward task as the user-level ATM signalling component of the virtual switch software do not operate as expected. Moreover, because this study is focused on ATM network impairments and the resulting video quality degradation, the importance of the ATM signalling function does not justify the inclusion of its development within this work. Therefore, it was decided that the original VoD system should be modified to use PVCs with VPI/VCI values 0/150 and 0/151 for control and video traffic respectively.

Besides using PVCs instead of SVCs, another modification to the VoD system developed by Roux and Vine is required. Since end-users of a VoD application watch video streams that are streamed from the video server in real-time, it is not necessary for the implementation of the VoD client to include the ability to store video received to a local file. However, studying the effects of network impairments on video quality requires video streams that are transported by the emulated network to be stored so that they can be thoroughly examined. Moreover, video and audio artefacts can be comprehensively analyzed by playing back the video streams that are under the influence of network impairments as many times as necessary.

The other important reason for the VoD client to store the video received is that these stored video clips can also be used during user-oriented quality assessments described in section 6.7.2. In order for the responses (in the form of opinion scores or otherwise) provided by different participants to be compared, it is important to ensure that the same video sequences are shown to all participants.

The modifications made to the VoD system can be subdivided into two areas. Firstly, new menu items, *open storage file* and *close storage file*, are added in the VoD client user-interface for file operations. The implementation of the ATM class is then modified in such a way that the content of the video buffer is copied to the storage file whenever segments of video are received.

It is understood that modifying the implementation of the end-systems violates the general applicability principle of Design II. However, it is important to note that the major objectives of the implementation are to test the feasibility (plausibility) of the NetIE architecture and to demonstrate the degradation of video quality under the influence of network impairments. This implementation of the emulated network aims to provide a test platform for further development and not a final product. Therefore, the applicability aspect of the NetIE implementation has been sacrificed in order to ensure that the main objectives are accomplished. Besides, due to the modularity of the design, the general applicability of the NetIE model will be restored when the ATM signalling function is included in the virtual ATM switch.

5.6 Connecting the BSTS to the VoD System

Having introduced the NEM and compared it with NetIE in previous sections, this section suggests how the NEM can contribute to this study by connecting it to the VoD system in place of the virtual ATM switch. A demonstration BSTS unit was kindly supplied by Concilium Technologies for a trial period during which its capabilities were explored and it was used to investigate the effects of ATM network impairments on video quality.

The BSTS demonstration unit consists of the following modules[23]:

- E4200 – 7-slot BSTS base unit;
- E4219A – ATM Network Impairment Emulator Module (NEM);
- E1697A – 155Mb/s (STS-3c/STM-1) Optical Line Interface (LIF);
- E4209B – Cell Protocol Processor (CPP).

Since the demonstration unit comes with an single mode optical line interface module and the ATM NICs installed on the VoD server and client have RJ-45 UTP connectors, the BSTS cannot be directly connected to the VoD system. They are bridged by the ASX-200BX and LE155 Workgroup ATM switches from FORE systems as shown in Figure 5-10. The VoD server and client are connected via normal UTP 5 cables to the LE155 Workgroup switch, which is in turn connected to the ASX-200BX through multi-mode optical fibre. The optical line interface module of the BSTS is then linked to the ASX-200BX using single mode optical fibre. Control traffic of the VoD system is passed from the client to the server directly through the LE155 workgroup switch. In the opposite direction, video traffic moves from the server to the LIF of BSTS through the LE155 and the ASX-200BX respectively, passes through the NEM and back down the LIF, before it is forwarded to the client via the ASX-200BX and the LE155 again. The NEM is configured to only process traffic on the VCC with the VPI/VCI fields equal 0/151 and PVCs are set up on both ATM switches accordingly.

During the use of the NEM to investigate the effects of ATM network impairments on video quality, the load on both ATM switches is kept low in order to prevent them from introducing undesirable network impairments. The results collected using the NetIE model and those obtained with this arrangement will be compared in Chapter 6.

[23] Details of the BSTS system and its modules can be found in CD under the */BSTS* directory

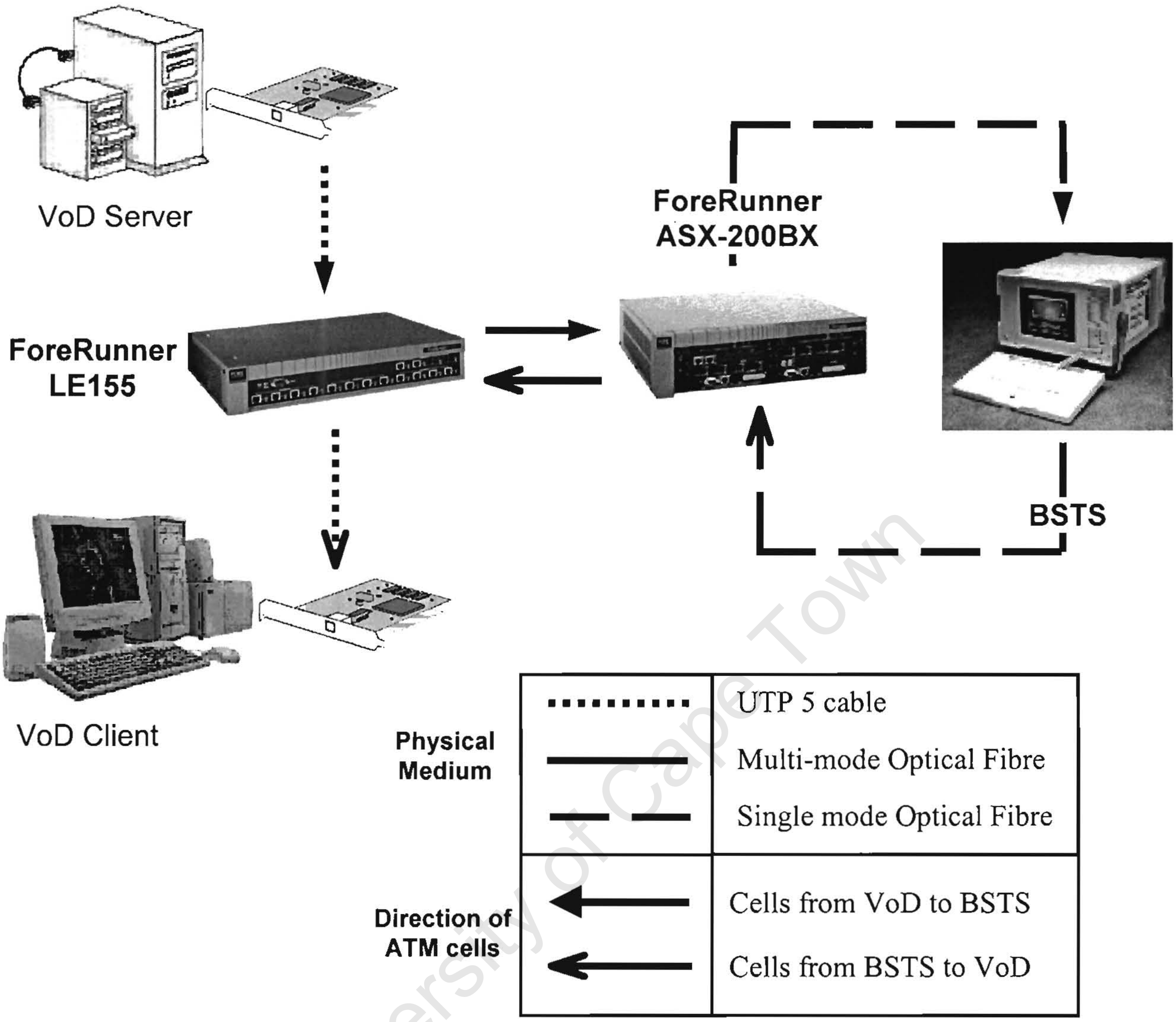


Figure 5-10 Connecting the BSTS to the VoD System

Chapter 6.

Effects of ATM Network Impairments to Video Quality

This chapter presents the effects of ATM network impairments to video quality by introducing controllable network impairments to video traffic in the emulated network. It illustrates video and audio artefacts observed in MPEG video streams resulting from the occurrences of ATM network impairments using both the Statistical and Application-Specific NetIE models. This study demonstrated that these effects are dependent on where impairments event occur and has justified the need for the AS model. The AS model is found to be particularly useful to study specific video distortions identified in this chapter. Preliminary result from the AS model suggests that the impact of network impairments on MPEG system information results in higher distortion than that on video / audio information. Also, audio distortions are generally more noticeable than video distortions so it appears that the human hearing system is less tolerant to audio distortions than the human visual system is to visual distortions. Objective and subjective methods to assess video quality will be examined in this chapter and the procedures of user-oriented quality assessments carried out in this study are outlined. This chapter also describes a method to establish the mapping between network level QoS and user level QoS.

This chapter intends to be interactive when video and audio artefacts caused by network impairments are examined (sections 6.1 and 6.4) and when user-oriented video quality assessment is demonstrated (section 6.7.2). Video clips included in the accompanying CD will be used as examples to demonstrate video and audio distortions identified during this study. Moreover, video sequences that are used as test materials for user-oriented quality assessment are included in the CD. Video frames captured and presented in this chapter are also included in the CD as they appear to be clearer on a computer monitor. Because these images and video clips form part of the materials presented in this chapter, readers are requested to view these video sequences and images as they are discussed. Since 'a picture is worth a thousand words', believed that the understanding of the printed material presented in this chapter will be enhanced as a result of this interactive approach.

6.1 Video and Audio Artefacts

This section demonstrates the video and audio artefacts observed at the VoD Client as a result of the network impairments occurring in the emulated network. Two video sequences, 'Pike Place' and 'UTA Internet Teaching', will be considered. They can be categorized as typical documentary and news reports respectively. Before these sequences are analyzed, it is important to note that since this study is focused on user-level QoS, only artefacts that are perceivable to the human eye will be considered.

'Pike Place' can be categorized as a conventional documentary sequence. Typical characteristics of this type of video clips can be summarized as follows, it contains: 1)narration and background music; 2)zooming in and zooming out; 3)slow to medium speed panning; 4)close ups and distance view; 5)scene changes and medium amount of motion; 6)movement of people and other objects such as cars; 7)both outdoor and indoor shots. The Pike Place video sequence is encoded at 1.5Mbps with a resolution of 352×240 pixels and a frame rate of 30 per second. For better viewing, set the MPEG player to display the video at 200%.

'PikePlace2.mpg'[24] contains the 'Pike Place' sequence as received at the VoD client from the VoD server under the influence of a mixture of network impairments introduced in the virtual ATM switch. After this sequence is being closely examined, the video and audio artefacts and the time at which they occur are identified as follows[25]:

- During the zooming out of the camera starting at *00:05* of the sequence, the zooming speed is suddenly increased at *00:06*. This is closely followed by a severe jump in motion of the camera and a still picture that persisted for about 200ms, before the zooming out action at normal speed is resumed. In terms of audio artefacts, the first syllable in the word 'market' of the narration is truncated at around *00:07*;

- At *00:09* of the sequence, audio distortion occurs in the second syllable of the word 'institution';

[24] Please view the 'PikePlace2.mpg' sequence in the */chapter6.1/* directory of the accompanying CD and compare it with the original sequence 'PikePlace.mpg'.

[25] the 'UTA' sequence can be analyzed in a similar manner.

- A mixture of video distortions occur from *00:38* to *00:43*. As the three persons (man in a grey T-shirt on the right, woman with blonde hair in the middle, and man in a blue shirt on the left) move forward, the walking speed increases at *00:38*. This is closely followed by three jerky movements in quick succession if the attention is focused on the woman in the middle or three severe jumps in motion in quick succession if the attention is focused on the two men either side of her. Each of the jerky movements or severe jumps in motion is immediately followed by a still picture of varying duration ranging from an estimate of 100ms to 300ms. The second and the third still pictures clearly last longer than the first. Normal movement resumes for about 2 seconds before jerky movement or severe jump in motion occurs again depending on where the focus is placed as described above. This is closely followed by two consecutive still pictures, the second of which occurs at a scene change;

- Small audio distortions occur at the same time as the above video distortions from *00:38* to *00:40* firstly between the words 'thousand' and 'people' and then at the first syllable of the word 'market';

- The speed of movement including walking and zooming increases at 00:49. Again, the video artefacts can be rather different depending on where the focus of the viewer lies. If the viewer is looking at the person walking towards the right side of the screen, jerky motion can be observed. If the focus is placed on either the 'POST ALLEY' sign or the person walking towards the left side of the screen, severe jump in movement noticed[26]. A still picture that lasts an estimate of 300ms follows;

- From *00:54* to *00:56*, the jerky movement of the van and the people can be observed. This is followed by a still picture that lasts almost half a second;

- Lastly, the flying fish appears mysteriously with the fisherman's arms at *01:00* of the sequence followed by a still picture of about 300ms. The jerky motion during the throwing action of the fish (at *01:02*) seems to coincide with the audio distortion at the first syllable of the word 'flying'.

Each incident of video artefacts (and the associated audio distortions in some cases) described above is the consequence of a single occurrence of a Cell Loss, Cell Error or Cell Mis-insertion impairment event. An exception is at *00:38* to *00:43* of the sequence, where a Cell Loss event is closely followed by a Cell Error event.

---

[26] This is especially severe for the person walking towards the left hand side. He has jumped at least 1.5 meters!

6.1.1 Observations

- Video artefacts that are identified above can be classified as: *Faster Motion*, *Motion Jerkiness*, *severe jump in movement*, and *frame freezing*. Audio distortions observed can be categorized as *truncation of part or all of a syllable* and *general audio distortion or noise in voice or background music* (another common type of artefact is the lost of lip-synchronization, which can be found in the 'UTA2.mpg' sequence);

- Audio and video artefacts occasionally occur around the same time, so some dependencies seem to exist between them. However, while it appears that audio distortions are always accompanied by video distortions, video distortions can occur without the presence of noticeable audio distortion. The audio distortion identified at *00:09* without a corresponding video distortion is a special case that will be discussed in the next section;

- The type of video artefact observed is partially dependent on the placement of the viewer's focus on the screen at the instance it occurs;

- A general pattern can be identified for the occurrences of video artefacts. When video distortions occur, they occasionally start with an increase in speed of motion (*Faster Motion*). This is closely followed by either *Motion Jerkiness* or a *severe jump in movement*, which is in turn followed by a still picture (*frame freezing*);

- Some instances of video and audio artefacts seem to be more noticeable than others;

- Video and audio distortions seem to be more severe when the occurrences of two impairment events are close to one another. A good example can be found from *00:38* to *00:43* in the sequence.

Before the above observations can be analyzed, it is important to note that the error checking mechanisms of AAL 5 are enabled in the VoD Client. This causes CPCS-PDUs to be discarded when an ATM cell stream is affected by network impairments. The effects of the AAL 5 error checking mechanisms will be explained in section 6.2. At this point, it is sufficient to be aware of the fact that when AAL error checking mechanisms are enabled in end-systems, CPCS-PDUs received with error(s) are dropped during ATM Adaptation Layer processing. As a result, the video decoder skips a certain number of video frames and audio packets during the decoding process. In the PikePlace2 sequence, only 1853 out of the original 2033 are displayed. The 80 missing frames have been discarded by the video decoder due to the presence of network impairments.

6.1.2 Analysis of the identified video and audio distortions

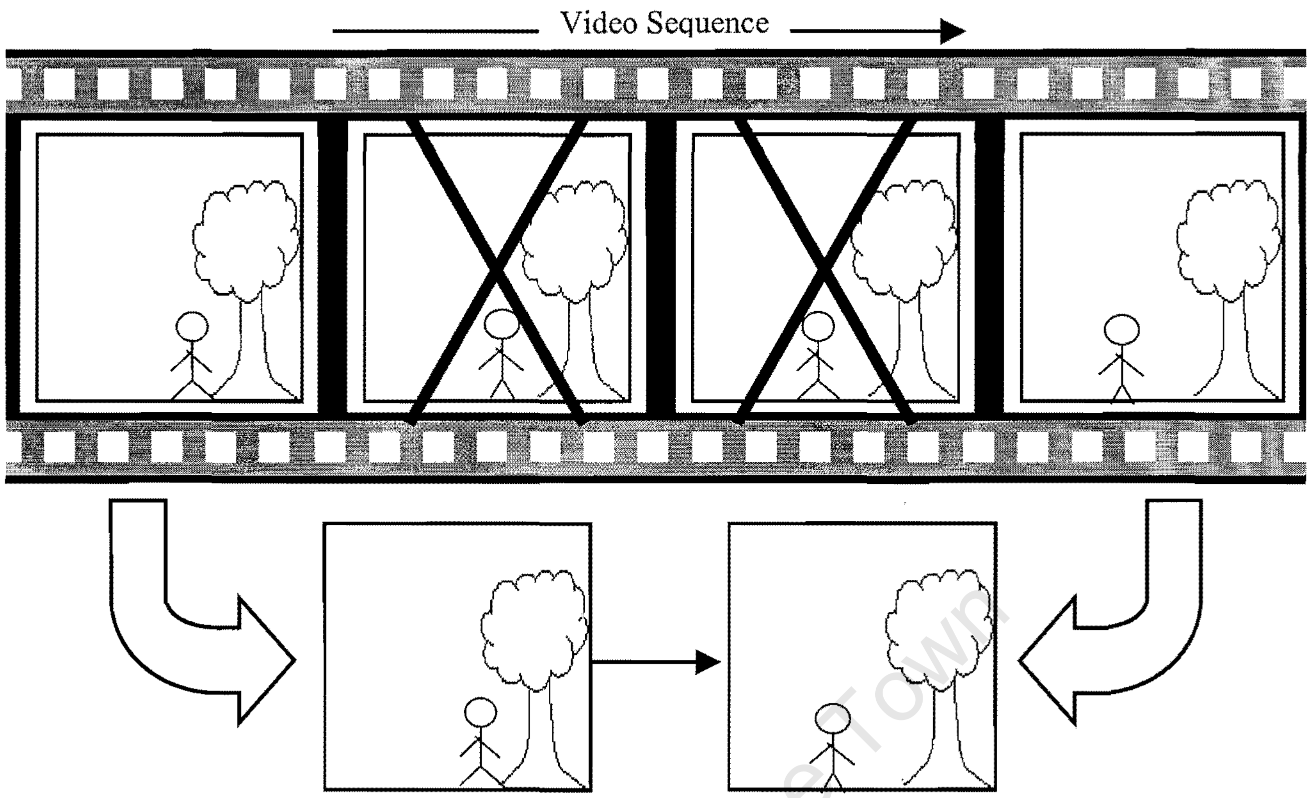
The number of video frames and audio packets being dropped determines the resulting visual effect and is in turn dependent on the importance of the information being affected by the network impairments. For example, an MPEG-1 decoder synchronizes with the received bit stream through the System Clock Reference (SCR) field in the system layer[27]. When the decoder loses sychronization with the MPEG bit stream, the decoding process cannot continue until sychronization is re-established and one or more video frames will be skipped as a result. If important synchronization information is affected by network impairments to a large extent, the time required for synchronization to re-establish will be relatively longer, resulting in more frames being dropped. The general pattern of visual distortions identified above can be seen as a result of the resynchronization process carried out by the MPEG decoder when it losses synchronization with the bit stream due to the occurrence of network impairment(s).

The skipping of a small number of frames can cause *Faster Motion* where people or the camera itself seems to be moving faster than normal[28]. When slightly more frames are skipped, either *Motion Jerkiness* (movement not smooth nor continuous) or *Severe Jump* (larger jump in motion or object) occurs. If yet more frames are skipped, noticeable pause in the sequence of about 100ms or more (commonly known as *frame freezing*) can be observed. In the extreme case, frame skipping can even result in *screen blanking* where a black (blank) screen appears for a considerable period of time.

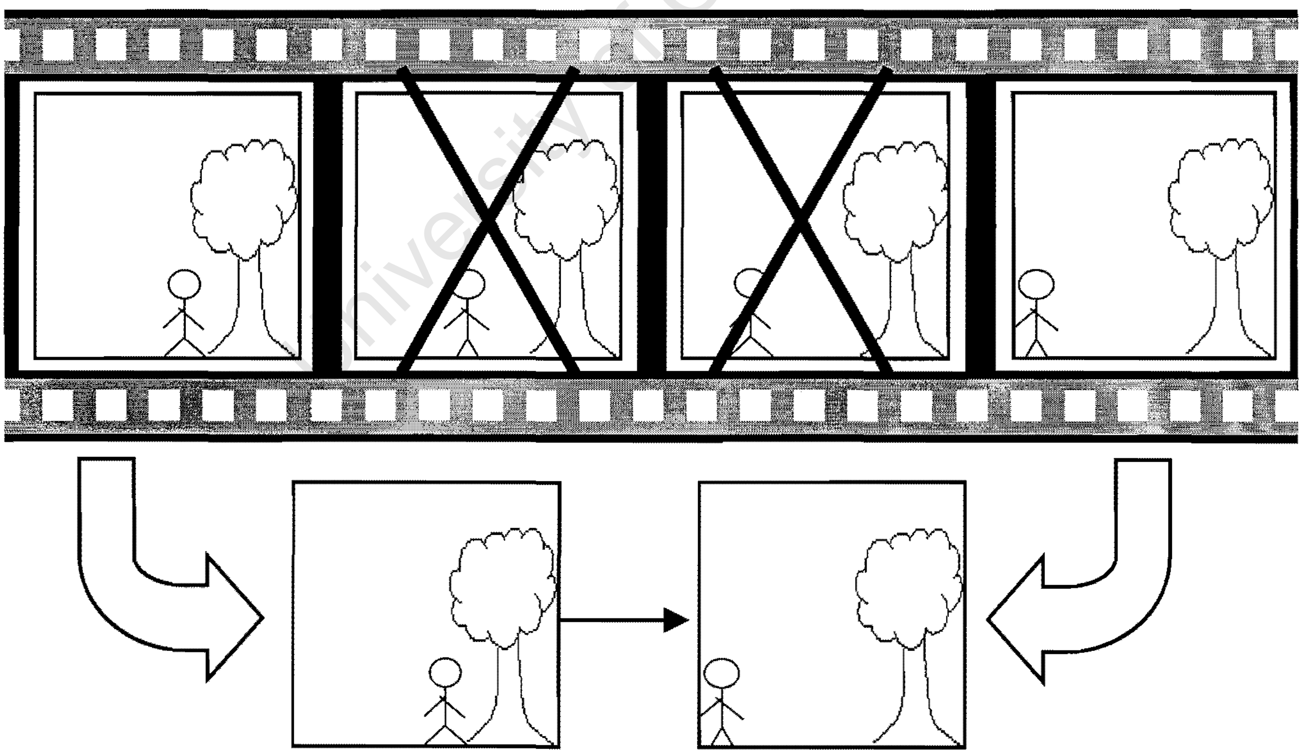
Although both *Motion Jerkiness* and *Severe Jump* result from a comparable amount of frame skipping, they are dissimilar because of a different amount of motion in the video sequence. For example, when the loss of a certain CPCS-PDU results in the drop of three frames, *Motion Jerkiness* occurs if only a small to medium amount of movement occurs within the three lost frames. The net result is a small jump in motion instead of a continuous and smooth motion as shown in Figure 6-1a. On the other hand, if there is a relatively large amount of movement taking place during the time corresponding to the three lost frames, the net result is a large jump in motion that is more obvious (as shown in Figure 6-1b).

[27] For MPEG-2 TS packets, the decoder uses the Sync Byte in the TS header and the Program Clock Reference (PCR) to establish and maintain synchronization

[28] Uneven movement of people and objects occurs in movies from the 60's even when they are played at the correct speed because there are less frames in each second of video than that required to achieve smooth continuous motion. The effect is similar to that of skipping a small number of frames.



a) Motion Jerkiness due to the dropping of two video frames



b) Severe Jump in Motion due to the dropping of two video frames

Figure 6-1 The skipping of video frames and the resulting visual effect

The importance of the information being affected by network impairments also has a direct impact on the resulting video and audio artefacts. Besides the synchronization information described above, header fields such as the sequence header, the group of picture header and the picture header are significantly more important than the headers within a picture. While a corrupted picture header may cause the loss of a frame (or a few frames due to the inter-dependency of frames described below), a corrupted group of picture header may result in the loss of the entire group of picture structure (typically made up of 15 video frames). In the extreme case where the sequence header is damaged, the decoder may even fail to start the decoding process.

In a group of picture, I frames are more important than P frames, which are in turn more important than B frames. For a group of picture with N frames, there are $(N \div 3) - 1$ P frames, ${}^{2}/_{3} \times N$ B frames and only one I frame. The inter-dependencies within a *Group of Picture* structure implies that the decoding of a B or P frame relies on the only I frame and the nearest P frame in the group (Figure 3-2). Therefore, the skipping of an I frame will result in a more severe video distortion than that of a P frame. The skipping of a B frame is relatively least significant.

When two impairment events occur close to each other, they tend to cause significantly more severe video distortions because the resultant video artefacts last for a longer duration (such as from *00:38* to *00:43* in the PikePlace2.mpg sequence). This is especially so when the impairment events occur in such a way that the MPEG decoder is affected by the second impairment before it has recovered from the first one.

Video and audio artefacts resulting from the presence of network impairments are not always perceivable by a human viewer. The response of the human visual system is relatively slow compared to the rate with which video frames are displayed. This is why it is generally accepted that images displayed at 25 to 30 frames per second constitutes continuous, full-motion video. A good example of video distortions that are not so obvious can be found at *00:09* in the PikePlace2.mpg sequence. The *frame freezing* resulting from the skipping of frames is not very noticeable because it happens at a scene change at which time the human visual system (HVS) is adjusting to it. This is why this occurrence of *frame freezing* was not described above.

Since the majority of information within MPEG bit streams is used to represent video, there is a higher probability for network impairments to affect video information than audio or system information. As a result, network impairments that cause video distortions do not always affect the audio signal as observed above. However, audio distortions are almost always accompanied by video distortions because it is quite rare for network impairments to affect only audio information without having some influences on video or system information.

Although this section intended to examine each type of video and audio artefacts identified above and relate them to the type(s) of impairment that cause them so that this analysis is centred around the quality of video perceived by the users, it turns out that such an approach is inappropriate. This is because of the following reasons. Firstly, when the error checking mechanisms of AAL 5 are enabled in the VoD Client, Cell Loss, Cell Error and Cell Mis-insertion cause the same effect to the MPEG bit stream as section 6.2 will explain. As a result, any distinction between these three types of impairments becomes impossible. Secondly, the introduction of Cell Delay Variation does not appear to result in any observable effect in the video displayed at the VoD client. This is possibly due to the presence of a jitter removal buffer within the MPEG decoder which has compensated for the variable cell delay introduced to the video traffic stream. However, as the source code for the MPEG decoder is not available, the presence of such a buffer has not yet been verified. Thirdly, since this study is focused on one-way video traffic, imposing constant Cell Transfer Delay only has very limited effect on the initial response time.

6.2 The Effects of AAL 5 Error Checking Mechanisms

The AAL 5 Common Part Convergence Sublayer Protocol Data Unit (CPCS-PDU) trailer contains a 2-octet length field and a 4-octet CRC field. The length field is used to encode the length of the CPCS-PDU payload and the CRC field contains the value of a CRC-32 calculation performed over the entire content of the CPCS-PDU including the CPCS-payload, the PAD field and the first four octets of the CPCS-PDU trailer. At the Common Part Convergence Sublayer of the receiving host, the length field is responsible for the detection of loss or mis-inserted cells, and the CRC field is able to detect bit errors in the CPCS-PDU.

Errors and losses detected in the received video bit stream are not retransmitted from the source because video information is time-based. For information in a certain video frame to be useful to the decoder, it must arrive at the destination a sufficient amount of time before the corresponding frame is displayed. However, the re-transmission of incorrect or lost video information increases network traffic. In many cases, the re-transmitted video information does not arrive at the destination on time to be used by the decoder and is discarded as a result. Moreover, re-transmission of video frames requires extra capability within the decoder to handle out-of-order video information, resulting in increased implementation complexity. Therefore, it is commonly accepted that re-transmission of video information should not used to deal with network impairments and that the video decoders should compensate for these impairments as far as possible and decode video sequences from the information available.

When a Cell Loss, Cell Error or Cell Mis-insertion occurs, the CPCS-PDU is considered to be corrupted. A corrupted PDU can cause visual artefacts and a certain degree of quality degradation as described above. At this point, it is important to note that the corrupted PDUs can either be discarded by the CPCS or delivered to the Service Specific Convergence Sublayer depending on the architecture of the ATM network adapter. If corrupted PDUs are discarded at the CPCS, then the occurrences of Cell Loss, Cell Error or Cell Mis-insertion will cause the same effect on the video stream (i.e. losing a number of information bits equal to the size of the CPCS-PDU payload). In other words, whether an impairment event is Cell Loss, Cell Error or Cell Mis-insertion will make no difference, the net result is the same to the CPCS-PDU.

Previous research such as [GRI98] suggests that under the influence of network impairments, video of relatively better quality results when corrupted CPCS-PDUs are passed to the video decoder than when they are discarded. The occurrence of a Cell Error event affects one single bit and if a CPCS-PDU (typically containing a few hundred bytes) is discarded when this occurs, over 99% of the information dropped is actually free of errors. Therefore, the importance of the many useful bits that are discarded far exceeds the significance of one error bit among the CPCS-PDU. Similar arguments can be applied to Cell Loss and Cell Mis-insertion whose extend is limited to one ATM cell (containing 48 bytes or 384 bits) within a CPCS-PDU. If the size of the CPCS-PDU is 384 bytes or more as described in section 3.3.3, the amount of valid information lost during the process of discarding corrupted CPCS-PDU is above 87%. This suggests that the dropping of corrupted CPCS-PDUs at the receiving host actually magnifies the effects of network impairments and possibly aggravates the video quality degradation that results. Therefore, the effects of ATM network impairments on video quality need to be investigated with the error checking mechanisms of AAL 5 disabled in order to validate the arguments presented above.

Although the possibility of passing corrupted AAL 5 CPCS-PDUs to the SSCS is outlined in [I.363.5] from the ITU_T, disabling length checking or CRC verification in an end-system is not as easy as it may seem. This is because ATM Adaptation Layer functions are implemented in the ATM NIC and the adapter driver, and the implementation of the ForeRunner LE series adapters does not provide the option to 'turn off' either the length checking or the verification of CRC.

6.3 The 'add-on' module

After further investigation into the subject, it was decided that the illustration of video quality degradation under the influence of network impairments is not complete without comparing the effects of enabling and disabling the error checking functions of AAL 5. Therefore, a technique to compensate for the limitation of the LE adapters needs to be considered. One of the ways is to introduce Cell Loss, Cell Error and Cell Mis-insertion to the video bit stream in an 'offline' manner. In other words, the file containing the video information is manipulated before transmission and video traffic that originates from the VoD Server is already corrupted. As a result, the length checking and the CRC verification at the destination will not cause any errors and the corrupted CPCS-PDU can be passed to the video decoder.

To compensate the effects of AAL 5 error checking using the 'add-on' method described above, end-system processing and the Network Impairment Emulation mechanisms need to be replicated at the video source. This 'add-on' method needs to:

- be aware of the scheduling scheme for the video traffic at the source;
- consider the packetization scheme;
- simulate the segmentation of video information into ATM cells;
- generate Cell Loss, Bit Error and Cell Mis-insertion impairment events according to the Statistical NetIE model;
- insert the generated impairments into the file containing the video information.

Although this method is similar to Design I proposed in Chapter 4 in that the three types of impairments (Cell Losses, Bit Errors and Cell Mis-insertions) are introduced to the video bit streams before they are sent over the emulated ATM network, it is not necessary for SNIM or RNIM functions to be carried out in this case because the network interface processing is carried out on the sender and the receiver by the ATM adapter and its driver. The software program for this 'add-on' module can be found on the CD in the */chapter6.3* directory.

6.4 Analysis of Video Artefacts (with AAL error checking disabled)

Instead of analyzing individual video sequences one by one from start to finish, this section aims to illustrate the different types of video artefacts. These video artefacts can be divided into two main categories:

- Distortions appearing within a video frame (section 6.4.1), and
- Distortions that span over multiple frames (section 6.4.2).

Depending on the appearances of the video distortions, they are further divided into different categories within each main category (sections 6.4.1 and 6.4.2). Although this study attempts to relate the identified video artefacts to the type of network impairments that cause them, this has proved to be very difficult because it is observed that none of the video artefacts is related specifically to a certain type of network impairment. While constant cell delay and variable cell delay do not seem to affect the video quality in any perceivable way, video distortions of different extent result when Cell Loss, Cell Error or Cell Mis-insertion occurs.

As explained in section 6.1, the effects of Cell Losses and Cell Errors on MPEG traffic over ATM are highly dependant on where they occur in the ATM cell stream because information fields in MPEG carry a varying amount of importance to the decoding process. Cell Loss, as well as mis-inserted cells that are passed to the application layer, usually causes the decoder to lose synchronization with the MPEG bit stream. This results in the skipping of one or more video frames and generally produces *Motion Jerkiness* (or *Severe Jump in movement*) and *frame freezing* similar to the case with AAL 5 error checking mechanisms enabled (as demonstrated in section 6.1). A comparison between Cell Loss and Cell Mis-insertion is presented in the */chapter6.4/CM_vs_CL* directory of the CD.

One of the widely used methods to cope with Cell Loss is known as 'dummy cell insertion'. When a cell loss is detected at the end-system receiving MPEG traffic, a 'dummy' cell containing '0's is inserted at where the cell loss has occurred so that the bit count of the MPEG stream is kept intact. Generally, this scheme manages to prevent the MPEG decoder from dropping frames but could cause certain distortions in the decoded video frames. The advantage of dummy cell insertion can be illustrated by comparing two video clips, 'cell_loss.mpg' and 'dummy_cell.mpg' in the */chapter6.4/drop_vs_dummy* directory. These two sequences are affected by the same number of cell losses in the same locations. While the former is affected rather badly by the skipping of video frames due to cell loss, the latter is almost the same as the original sequence (Porsche.mpg) except for some minor distortions. However, dummy cell insertion does not always perform as well as this example. Two more examples are included in the CD to indicate that dummy cell insertion replaces *frame skipping* (caused by dropped cell) by other video artefacts that will be discussed in the next section. The above examples also demonstrate how network impairments affecting different parts of a MPEG bit stream can cause significantly different visual artefacts during video playback.

Cell Error can cause video distortions such as tiling, error blocks and dislocations that are described in the next section. In general, Cell Error causes less severe artefacts to the video playback than Cell Loss does. This can be illustrated by considering two video clips: CL.mpg and CE.mpg in the */chapter6.4/CE_vs_CL* directory. Cell Loss and Cell Error impairment events affect the same cells in these two sequences and dummy cells are inserted in CL.mpg. Video distortions in CE.mpg are generally less noticeable than that in CL.mpg and overall video quality of the former is therefore better. However, it is possible for Cell Loss and Cell Error to cause video distortions that are almost impossible to be distinguished (Figure 6-2). This not only shows that a single bit error is capable of causing video distortion to a large extent, but also demonstrates that the artefacts observed during video playback is largely dependant on the type of information in the MPEG stream influenced by network impairments once again.

6.4.1 Video Artefacts appearing within a single frame

Video artefacts identified within individual video frames during this study are classified into ten categories according to the visual effects they create. All the ones described in this section are caused by Cell Loss with dummy cell insertion and Cell Error. Typically, errors in slice data cause rectangular-shaped objects to appear in the decoded picture. These objects occasionally extend from where the error occurs to the end of the slice. This is known as the spatial propagation of video distortions and will be explained in section 6.5.

In order to provide a clear illustration, video frames are included as examples in each category. These and other video frames are captured in the widely used *Bitmap* format. They can be found in the */chapter6.4.1* directory of the accompanying CD. The video clips from which the frames are captured can also be found in the */chapter6.4.2* directory (refer to section 6.4.2 and Appendix E for information on where these video frames are captured from the video clips). In cases where the original frame is captured for comparisons, the original frame is shown on the right and the distorted frame is shown on the left. The ten categories of video artefacts are listed as follows:

- Tiling (also known as Pixelation) – Tiling is the formation of small blocks with distinct boundaries. The contents within these small blocks are similar to the corresponding areas of the original picture. The effects of tiling can be found between the eyebrows and above the left eye of the person's face in Figure 6-3. In general, tiling occurs due to the presence of bit errors within cell payloads;

- Error Blocks – formation of solid blocks, each with a distinct colour. These blocks are usually green or black in colour, but other colours such as yellow or red are possible. Error blocks are generally created by Cell Loss or Cell Error that cause momentary loss of synchronization between the decoder and the bit stream. Figure 6-3 shows black error blocks close to the bottom of the picture covering the nose, either side of the eyes, and the cheeks of the person. Black error blocks can also be found at the bottom of the picture in Figure 6-4;

- Soft Error Blocks – similar to solid error blocks but each block is not of a distinct colour. These blocks often carry visual contents belonging to another area of the same frame or the same area in another frame close to the current one in the video sequence. In the former case, spatial prediction information is corrupted by Cell Loss or Cell Error. In the latter case, temporal prediction within a Group of Picture structure is inaccurate. Examples of Soft Error Blocks can be found in Figure 6-4a on the bottom left, bottom right and top right corners as well as on tree tops on the right half of the picture;

- Semi-transparent Error Blocks – they are the same as Error Blocks except for their semi-transparency. These are mainly caused by Cell Error or Cell Loss affecting DCT coefficients within the Block layer. Figure 6-6 shows semi-transparent error blocks in green;

- Left-over – when a small portion of the picture remains after movements in the video content or a scene change. This is caused by corrupted motion prediction information. Soft error blocks and left-over are similar in some ways except the contents cannot be identified in the former, and the contents can be identified as left-over from previous frames in the latter. In Figure 6-65, the words "Fred's Swing" that are moving across the screen have left the top half of the character 'g' behind as shown;

- Colour cycling – this refers to the lost of colour stability. This is characterized by the appearing of a range of colours in the picture. Colour cycling shown in Figure 6-8 is caused by a Cell Loss;

- Colour Blocks – this is a combination of soft error blocks and colour cycling (Figure 6-9). They can be attributed to Cell Loss or Cell Error affecting Macroblocks within the Slice Layer or Blocks within the Macroblock Layer in the MPEG bit stream;

- Dislocation – this occurs when two or more frames are super-imposed on one another (Figure 6-10). This is usually resulting from a Cell Loss or Cell Error that corrupts motion prediction information;

- Total distortion – This refers to severe distortion of any kind that affects a large portion of the picture. The frame shown in Figure 6-12 suffers from severe dislocation and is almost beyond recognition;

- Others – Video artefacts that are not observed very often are classified into this category. One of these is 'twisting', which suffers from a combination of colour blocks and dislocation and results in an obscure image (Figure 6-13). Another interesting one is 'beyond range motion', where an object moves beyond the position the same object appears in a normal video sequence (Figure 6-14). This is caused by incorrect motion prediction information.

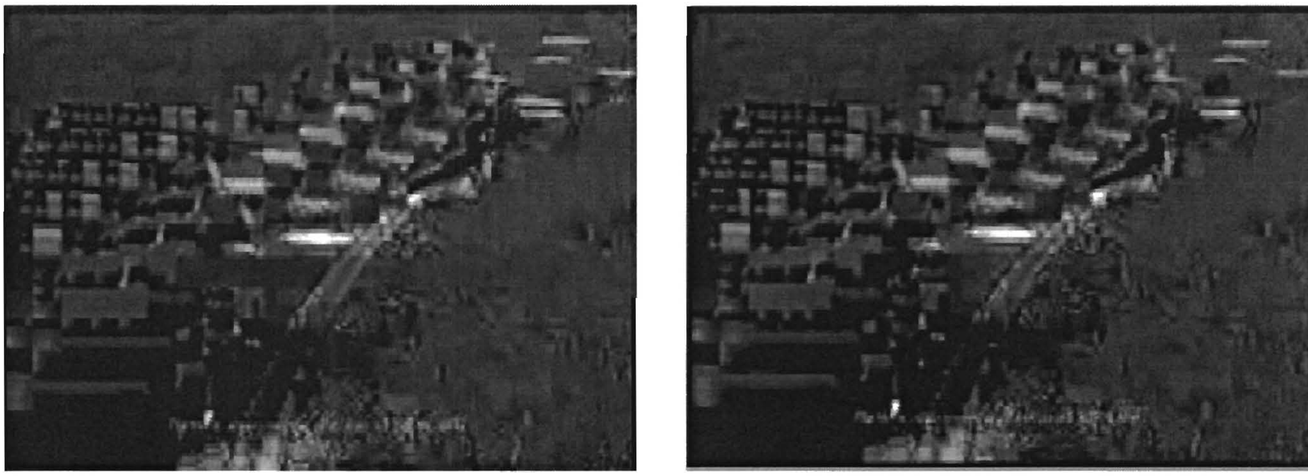


Figure 6-2 Video distorion caused by Cell Loss (left) and Cell Error (right)

Figure 6-3 Video Artefact within a frame - Tiling and Error Blocks

Figure 6-4 Video Artefact within a frame - Soft Error Blocks

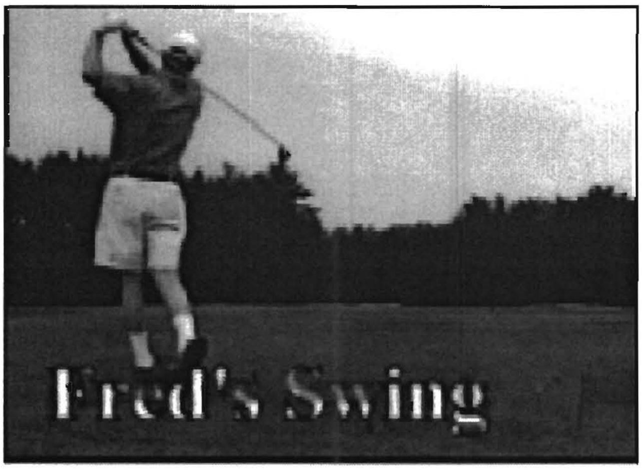


Figure 6-6 Semi-transparent Error Blocks

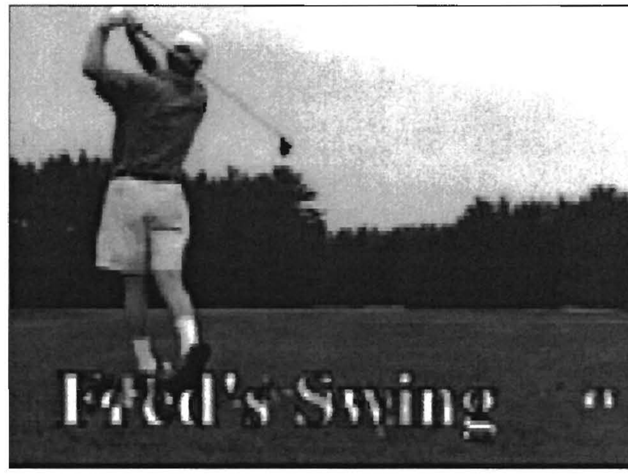


Figure 6-5 Left-over from previous scene

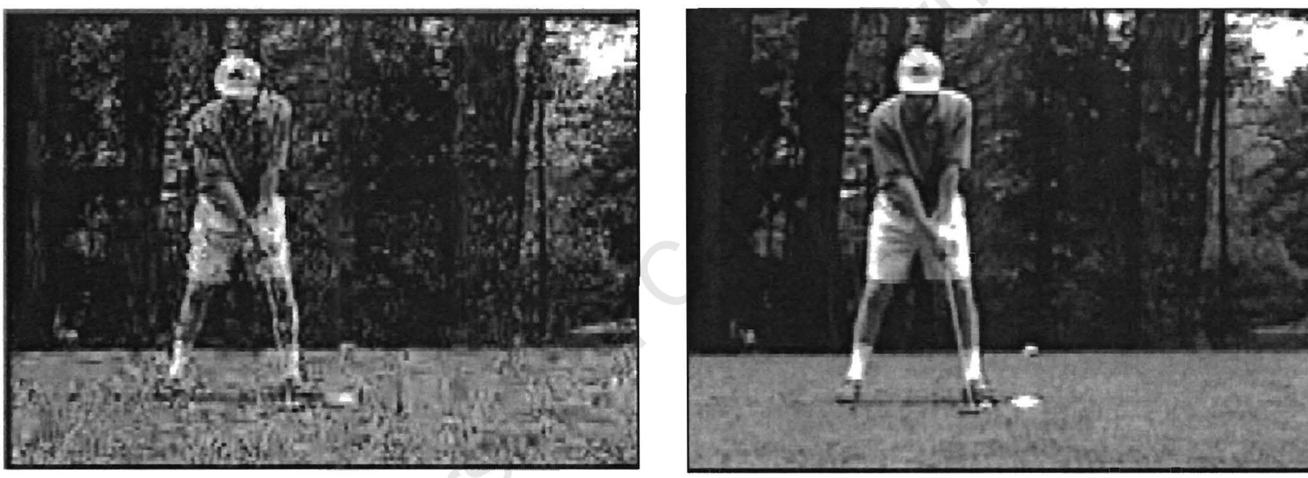


Figure 6-7 Video Artefact within a frame - Colour Cycling

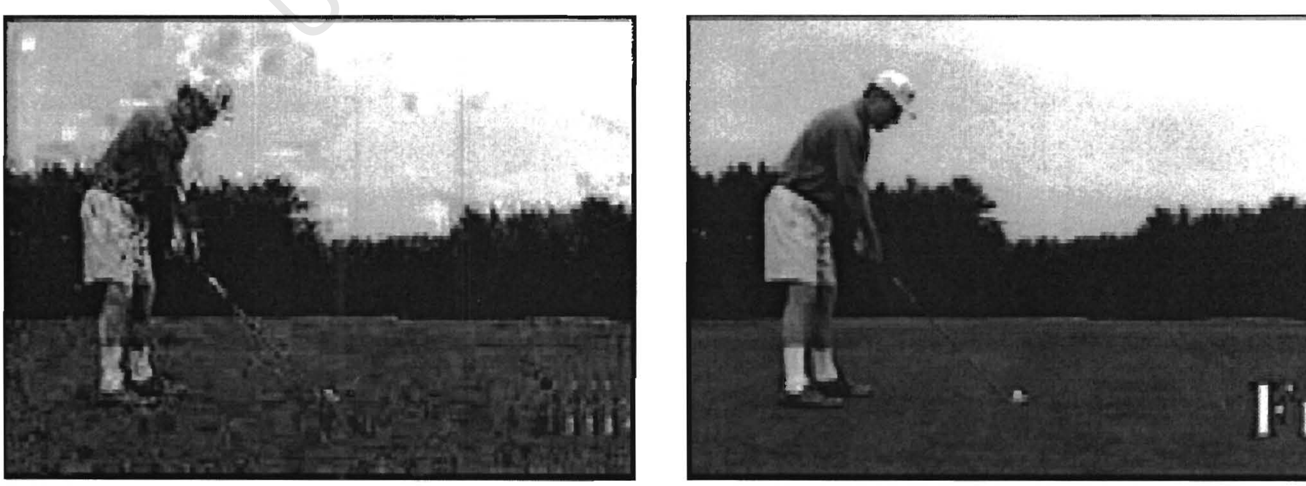


Figure 6-8 Video Artefact within a frame - Colour Blocks

Figure 6-9 Video Artefact within a frame – Dislocation

Figure 6-10 Video Artefact within a frame - Total Distortion

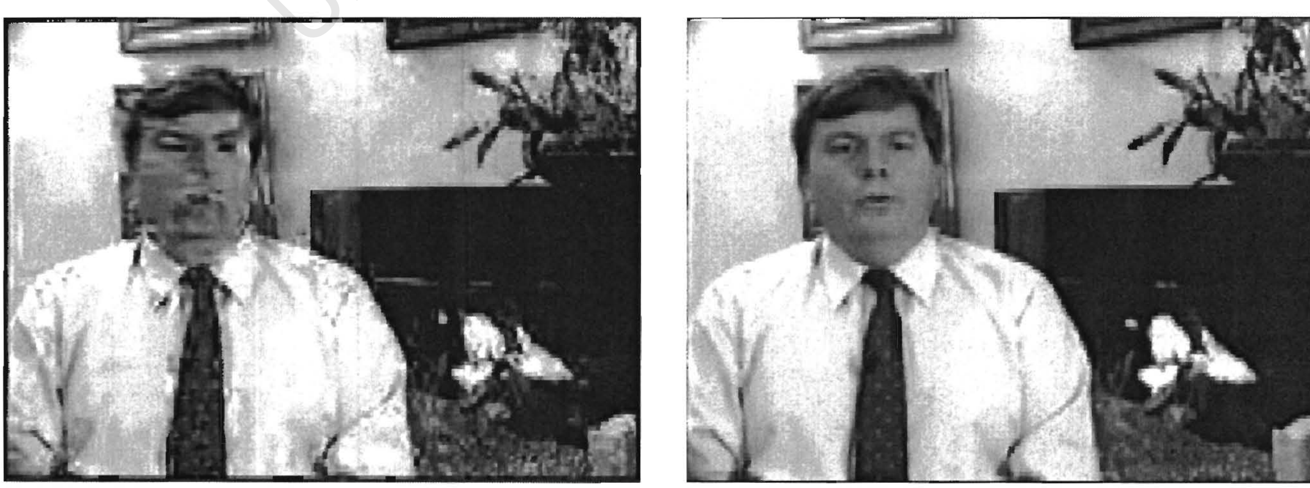


Figure 6-11 Video Artefact within a frame – 'Twisting'

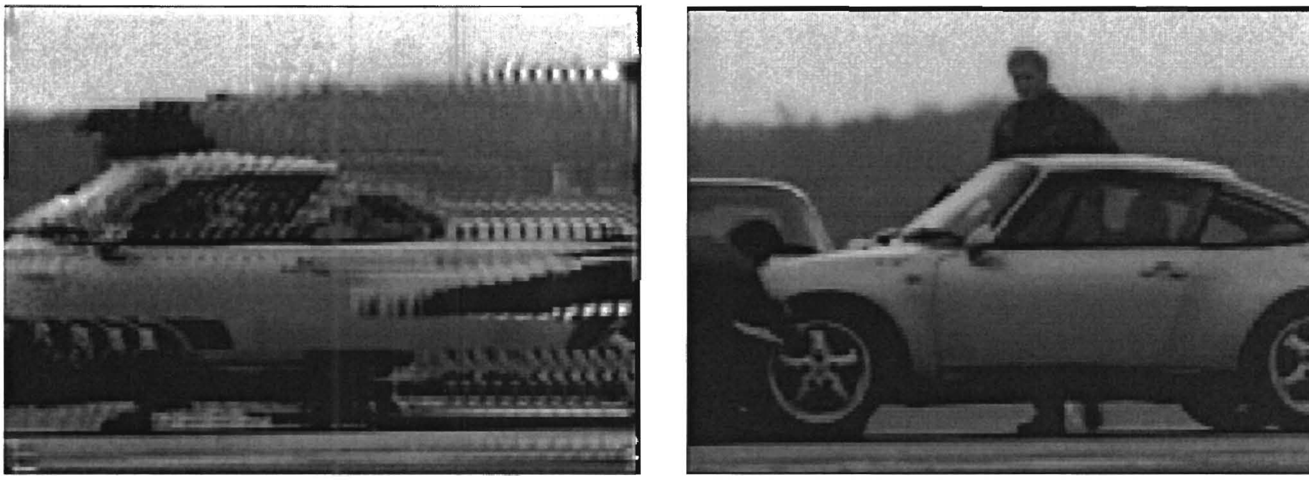


Figure 6-12 Video Artefact within a frame – 'Beyond Range Motion'

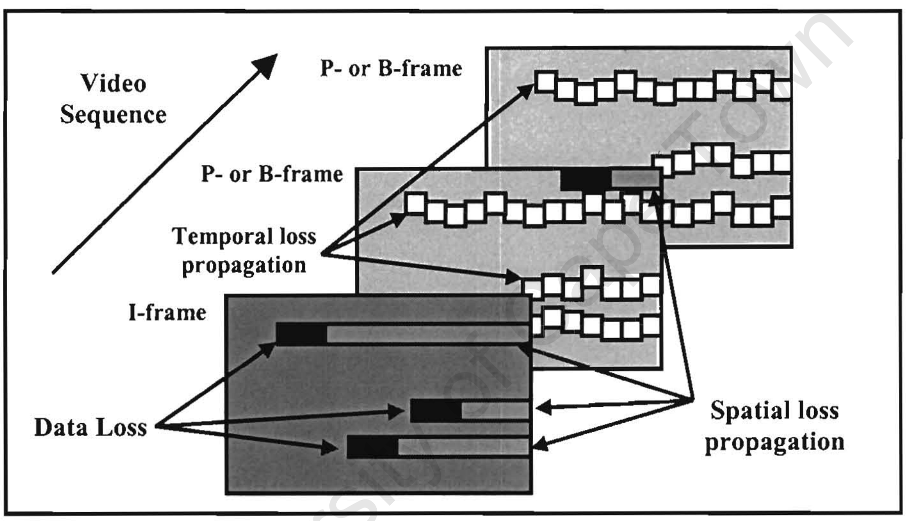


Figure 6-13 Spatial and Temporal Propagation of data loss

Figure 6-14 Spatial Propagation of Error Blocks to a Varying Extent

6.4.2 Video Distortions extending over multiple frames

Video distortions that span over more than one frame can be divided into two categories: those resulting from the skipping of frames and the temporal propagation of video artefacts that appear within a single frame. When AAL error checking is disabled, *Motion Jerkiness*, *Severe Jump* and *Frame Freezing* have been identified when cell loss (without dummy cell insertion) or cell mis-insertion occurs. These network impairments cause the decoder to skip video frame(s). In this case, the resulting visual effects are mainly dependant on the amount of movement in the video sequence and the extent of frame skipping, which is in turn determined by the type of information affected by the impairments as examined in previous sections.

Since MPEG videos are typically displayed at 24 or 30 frames per second. Video artefacts lasting one frame will approximately appear for 33ms to 42ms only and therefore not be very noticeable even if they distort the frame severely. However, video artefacts often propagate temporally into other frames and cause more noticeable effects. In fact, all of the video distortions described in section 6.4.1 are captured from video artefacts that last for multiple frames and the video clips from which the frames are captured can be located in the */chapter6.4.2* directory on the CD.

Figure 6-3 is captured from 'an1.mpg' at *00:00*. Besides tiling and error blocks, this clip also shows video distortions that last multiple frames at two instances: dislocations of the dinosaur and colour blocks at *00:01*, as well as soft error blocks at *00:03*. Among these distortions, the first one is comparatively more visible because of the black error blocks and the proportion of the screen it covers (over 20%). The other two are not as noticeable because they occur on top of moving objects (especially the one at *00:03* which is moving quite fast) and they appear to occupy a small area on the screen compared to the first one (at *00:00*).

Figure 6-4, Figure 6-9 and Figure 6-10 are captured from 'po3.mpg' at *00:06*, *00:08* and *00:09* respectively. Besides these three occurrences of video distortions, this clip also exhibits soft error blocks at *00:02*. The distortion at *00:02* is quite visible because it occupies quite a large portion in the middle of the screen and it lasts for a considerable duration. The various error blocks at *00:06* are not very obvious mainly because they last for a very short time and most of them are located on the side of the screen. Although video artefacts at *00:08* and *00:09* look to be rather severe on Figure 6-9 Figure 6-10 and affect almost the entire screen, they do not appear to be as bad during normal playback. While dislocation at *00:08* occurs on top of a bouncing car and a revolving wheel, beyond range motion and dislocation at *00:09* only flashes pass the screen and it occurs after sideways panning of the camera. In fact, the distortion at *00:09* lasts for such a short time that it disappeared before viewers could have a clear view of what has what has happened. Figure 6-14 was captured from 'po4.mpg'. A comparison between 'po3.mpg' and 'po4.mpg' reveals that the 'beyond range motion' effect found in 'po4.mpg' is more noticeable because it appears on the screen for more than twice as long.

While both Figure 6-5 & Figure 6-6 are captured at *00:03* of 'gs2.mpg', Figure 6-8 & Figure 6-9 are captured at *00:14* and *00:00* from 'gs2.mpg' and 'gs1.mpg' respectively. The transparent error blocks shown in Figure 6-6 do not seem to be visible at all during normal playback. Because the viewers tend to place their attention on the person swinging the golf club and therefore notice the effects of *motion jerkiness* during the swing, the distortions on the phase 'Fred's Swing' and the left-over of the top half of the character 'g' are not as noticeable as they would otherwise. The two occurrences of colour cycling at *00:06* and *00:14* are relatively more serious because the entire screen is affected and they both last long enough to be noticed. In the 'gs1.mpg' sequence, the colour blocks at *00:00* is not very obvious because it only lasts for a brief moment. The slight distortion on the phase 'Fred's Swing' has been overshadowed by the motion jerkiness just before the completion of the swing because of the same reason that applied to 'gs2.mpg' above. Lastly, although the soft and transparent error blocks appearing at *00:10* almost occupy the entire screen, their effects are somewhat reduced because they occur during a scene change.

The effects of 'twisting' and a certain extent of colour cycling (captured from 'tp1.mpg' at *00:05*) shown in Figure 6-13 are quite serious because they not only affect almost the entire screen, but also distort the face of the person, which is likely to be where viewers position their attention. In addition, lip-synchronization is lost (at *00:10*) momentarily due to the dropping of video frames.

Since video distortions caused by ATM network impairments vary significantly in size and shape, it is not possible for this dissertation to include examples for all possible occurrences. Therefore, only a few examples that are illustrative and occur regularly during this study have been described in this chapter. Although the appearances of video artefacts are largely dependant on the implementation of the MPEG decoder and the possible adoption of error concealment techniques, any video sequences that are affected by ATM network impairments can be examined in the same approach adopted in sections 6.1 and 6.4. Note that except for 'po3.mpg' and 'po4.mpg', all the video clips examined in this section ('an1.mpg', 'gs1.mpg', 'gs2.mpg', 'po1.mpg' and 'tp1.mpg) will be used during the user-oriented quality assessment described in section 6.7.2.

6.5 Spatial and Temporal Propagation of Cell Errors and Losses

Since video distortions are more obvious when they occupy a large portion of the screen or when they last for a considerable amount of time as described in the previous section, it is important for this study to investigate why certain video artefacts occupy a larger area or last longer than others. Spatial and temporal propagation refers to the propagation of video distortions within a single video frame and from one frame to the next or more successive frames respectively. Errors and Losses that occur within slices in the picture level of a MPEG encoded bit stream may cause the spatial and temporal propagation of the tiling or error blocks effects. Figure 6-15 illustrates the spatial propagation of video distortions to the end of their respective slice and the temporal propagation of these distortions from an I frame to a P or B frame [VER98a]. In addition, when part of a video frame is derived from another portion of the same frame, the corruption of the latter due to network impairments will cause distortions to propagate spatially to the former. In fact, the effects of temporal and spatial propagation of video distortions have already been encountered in Figure 6-5 and Figure 6-6 respectively.

Video distortions should generally not propagate spatially into the next slice and will definitely not propagate temporally into the next Group of Picture. This is because at the start of each slice, the decoder can re-establish synchronization with the bit stream. At the start of each GOP, the temporal prediction information becomes error free provided that an error free I-frame is present. If the next I frame is corrupted, video distortions will continue to occur. However, this can not be considered as temporal propagation of data loss because the video distortions in this new group of picture are caused by the corrupted I frame.

The number of frames affected by an error propagating temporally depends on the frame type of the corrupted frame. If an I-frame is corrupted, all the frames that use the corrupted I-frame for temporal prediction (including all the frames in the current GOP and the last B-frame in the previous GOP) may be affected. This depends on whether any macroblocks in these frames are predicted from the corrupted macroblocks in the I-frame. If a P-frame is corrupted, all the P-frames after the corrupted P-frame as well as the B-frames that use the corrupted P-frame as prediction reference are likely to be affected. Since no other pictures are predicted from B-frames, video distortions occurring in a B-frame will not propagate temporally.

In general, the effects of Cell Errors and Cell Losses are often intensified by the spatial and temporal propagation of video distortions. The extent of spatial and temporal propagation has a significant effect on video quality. For example, video distortions occurring close to the end of a slice will only propagate to a little extent and occupy small portion of screen as a result. The visual effect is relatively less significant than that caused by video distortions occurring near the beginning of a slice spanning over almost an entire slice. (Figure 6-16). In terms of temporal propagation, video distortions that last for many video frames (or in the extreme case, the duration of a GOP) are more noticeable than one that only lasts very briefly. For example, the 'tp2.mpg' sequence in the */chapter6.5* directory contains, amongst other video artefacts, four occurrences of colour cycling at *00:01*, *00:04*, *00:11* and *00:12* which create approximately the same effects to video frames that are affected. However, the visual effects resulting from these four video distortions are different because each of them lasts for a different duration. While the colour cycling at *00:04* lasts for the longest and is therefore most obvious, the one at *00:11* is the shortest in duration and is therefore least visible.

6.6 Demonstration of Video and Audio Artefacts with BSTS

This section presents the video and audio artefacts that are observed at the VoD client during the streaming of MPEG video when the VoD system is connected to the Broadband Series Test System as shown in Figure 5-10. The approach adopted in this section to identify and analyze video artefacts is similar to that employed in sections 6.1 and 6.4.

In general, the results obtained with the BSTS are similar to that accomplished with the network impairment emulation architecture. Firstly, because the VoD system sends video in one direction only, the effect of constant cell delay (CTD) on MPEG traffic is hardly perceivable except for an extremely small delay in the initial response time (i.e. time elapsed between the video playback request and the start of the actual playback). Secondly, the introduction of variable cell delay (CDV) does not appear to affect the quality of video appearing at the VoD client. Lastly, because the AAL 5 error checking mechanisms are enabled, the occurrence of Cell Loss, Cell Error or Cell Mis-insertion causes the dropping of CPCS-PDUs in the AAL. This results in a combination of *Motion Jerkiness*, *Severe Jump*, *Frame Freezing*, truncation of syllable(s) and general audio twist(s) due to the skipping of video frame(s) and audio packet(s) during the decoding process as explained in section 6.1.

Video playback at the VoD client cannot start at all occasionally when MPEG traffic passes through the BSTS due to the method with which impairment events are generated in the NEM. Instead of cell-by-cell impairment event generation adopted in the NetIE architecture, the NEM relies on inter-occurrence intervals to generate impairment events. As a result, impairment events such as Cell Loss, Cell Error or Cell Mis-insertion often occur shortly after the NEM starts execution. This causes the important Sequence Layer header in the beginning of MPEG bit stream to be corrupted, preventing the video decoder to start the decoding process.

6.7 Measurement of Video Quality

The NetIE architecture has enabled the demonstration of different types of video and audio artefacts that are caused by network impairments during real-time video transfer. The other main objective of this study, i.e. to establish a mapping between network level QoS and application level QoS for video traffic, cannot be achieved without a means to assess and measure video quality quantitatively. This is because only when a satisfactory video quality level can be defined quantitatively can investigations be carried out in an attempt to establish the requirements from the network (network level QoS) that will result in this satisfactory quality level. If, on the other hand, a satisfactory quality level for video cannot be defined, the investigation in network level QoS requirements will not have a standardized quality level as the foundation to base on.

Several assessment schemes have been proposed in the literature [RIL97] [VER98a] [YAM93]. These schemes can be divided into two main categories: Subjective Assessment, and Objective Assessment. The main difference between them is that the former relies on human perception while the latter makes use of analytical methods to determine the video quality. This section will present subjective and objective methods to assess video quality as well as a possible way to establish the mapping between QoS at the network and at the application levels. A summary of all the video and audio artefacts that result in quality degradation examined in this study will be presented.

6.7.1 Subjective Assessment

Methodology for subjective assessment of the quality of television pictures is defined in Recommendation ITU-R BT.500-9 [BT500]. The assessment for the effect of network impairments can be achieved by the *double-stimulus impairment scale method* (also known as the "EBU method"). This method is suitable for the evaluation of the variation in network impairment parameters and the resulting effect. During the assessment session, viewers are first presented with an unimpaired reference video sequence and then the same sequence under the influence of network impairments. The viewers are then asked to vote on the second according to the grading scales in Table 6-1, keeping in mind the first. This procedure can be repeated for a number of video clips during the test session so that different types of video can be studied. The individual ratings of each sequence by all viewers are then combined to form a mean rating. The findings of these test sessions can be used to establish the level of network impairments that is likely to result in satisfactory video quality. The user-oriented video quality assessment performed during this research follows this method and is described in section 6.7.2.

Alternatively, single-stimulus methods can be used where the video sequences under test are presented to the viewers only once in the test session. One of these methods, the adjectival categorical judgement method, requires viewers to assign video clips to one of the categories given in Table 6-2. The categorical scales can either be used to address the level of video quality or visual distortions. The impairment threshold, which represent the level of network impairments that result in unsatisfactory video quality if exceeded, can be determined from the judgement of the viewers.

Grade	**Visual Impairment**
5	**Imperceptible**
4	**Perceptible, but not annoying**
3	**Slightly annoying**
2	**Annoying**
1	**Very annoying**

Table 6-1 The five-grade impairment scale for the EBU method [BT500]

Five-Grade Scale			
Quality		Impairment	
5	Excellent	5	Imperceptible
4	Good	4	Perceptible, but not annoying
3	Fair	3	Slightly annoying
2	Poor	2	Annoying
1	Bad	1	Very Annoying

Table 6-2 ITU-R quality and impairment scales [BT500]

6.7.2 User-oriented Video Quality Assessment

During the course of this research, a user-oriented video quality assessment session was conducted in an attempt to examine the feasibility and implications of trying to map application level QoS and network level QoS. Although the test session is not very comprehensive, it reveals detailed requirements to the running of these assessment sessions and provides valuable insight into how such test sessions can be used to establish the mapping of QoS.

The video quality assessment session aims to follow the double-stimulus impairment scale method proposed by [BT500]. However, some items, such as 'General Viewing Conditions', have not been closely followed due to practical limitations. The items that have been adopted for the test session include:

- Number of observers – while [BT500] recommends to use at least 15 observers for each test material, 50 assessors were used in this test;

- Past-experience of observers – the recommendation suggests that the assessors should be non-expert, i.e. they are not directly concerned with television picture quality as part of their normal work and are not experienced assessors;

- Instructions for the assessment – the recommendation further suggests that the assessors should be carefully introduced to the method of assessment, the types of impairment likely to occur, the grading scale, the test sequences and the timing of the events that will occur during the test session. This is achieved through the use of an explanatory poster and a brief verbal introduction. In addition, figures illustrating video artefacts caused by network impairments are included in the poster and sample sequences are shown to the viewers as proposed by the recommendation;

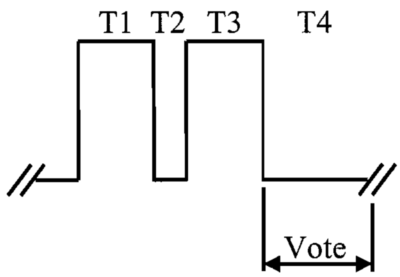
- Training video sequence(s) – a few dummy video clips (such as */chapter6.7.2/dummy sequence.mpg*) are shown to the assessors before the actual assessment starts, after which there is a break to allow time to answer questions from observers as recommended;

- The overall duration test session and the length of video clips – while it is recommended that individual video clips should be around 10 seconds long and the overall test session should not last for longer than 30 minutes, the video clips used range from 10 to 18 seconds and the entire test session lasts three to five minutes. It is important for the test session and the video clips to be short in order to keep the assessors interested and focused throughout;

- The test session – the sequence of events during the test session can be summarized by Figure 6-15;

- Choice of impairment range – it is recommended that the range of impairments chosen for the test material should aim to produce a grand mean score of all test sequences to be close to three;

- Analysis of results – it is proposed in [BT500] that the data collected must be condensed by statistical techniques to yield results in graphical or numerical form as a summary of the outcome of the tests;



Phases of presentation:

T1 = 10 s Reference picture
T2 = 3 s 'blank' screen
T3 = 10 s Test sequence
T4 = 5 – 11 s 'blank' screen

Figure 6-15 Presentation Structure of Test Material [BT500]

Using the *double-stimulus impairment scale method*, a set of network impairment parameter values within a desired range needs to be chosen in a small number of roughly equal steps. Test materials are then prepared with a selection of network impairment parameters from the set chosen above. If all types of network impairments are considered in the test and a few values of each parameter is chosen, a large number of test video clips will be produced. Since it is not advisable to reuse assessors[29], this would require an undesirably large number of participants in the test sessions. Therefore, it was decided that only one type of the network impairment should be introduced to the test materials used for the test session in order to reduce the number of assessors required.

Since the aim of this test session is to investigate the feasibility and implications of QoS mapping and not to provide definite parameters for this mapping, the reduction in participants has more benefits than detriment. The type of network impairment imposed on the test materials is chosen to be cell loss (with dummy cell insertion) because:

- The introduction of Cell Transfer Delay and Cell Delay Variation does not cause any perceivable effects on video quality;
- Although it is shown in sections 6.1 and 6.4 that the effects of Cell Mis-insertion and Cell Loss (without dummy cell insertion) on video quality are comparable, typical values of Cell Mis-insertion Rate are far lower than those of Cell Error Ratio [ANA92]. Therefore, the study of the effects caused by cell loss is more significant than those of cell Mis-insertion;
- It was shown in section 6.4 that dummy cell insertion should be adopted when cell loss occurs because it generally improves video quality;
- It was further observed in section 6.4 that most of the visual effects caused by cell errors are not very obvious and that the effects of cell losses (with dummy cell insertion) on video

[29] Reusing assessors extensively will cause them to become 'experienced', i.e. they become familiar with the video artefacts. Consequently, the test result will be inaccurate.

quality is generally greater than that of cell errors. Therefore, cell loss with dummy cell insertion is chosen in order to produce meaningful results.

The four video clips that are used to generate test materials for the assessment session are: *animation*, *golf swing*, *talking person* and *porsche*. Each of the test clips is created with two different CLR parameters as shown in Table 6-3. The values of CLR vary from a minimum of 6.1×10^{-5} in 'An2' to a maximum of 1.6×10^{-4} in 'Po1'. The CLR values of test sequences created with the same source ranges from a factor of 1.2 (*talking person*) to a factor of 2.5 (*animation*). Cell losses are introduced to the MPEG stream with the Statistical NetIE model. Note that the test materials are created before the test session so that all the assessors are exposed to the same set of sequences with known video distortions instead of different sequences generated during the test. The test materials are shown to viewers randomly during the test session according to the procedures outlined in Figure 6-15. They have been included in the */chapter6.7.2* directory of the CD for reference.

	Animation		*Golf Swing*		*Talking Person*		*Porsche*	
	File name	CLR	File name	CLR	File name	CLR	File name	CLR
Reference	An_ref		Gs_ref		Tp_ref		Po_ref	
First	An1	1.5×10^{-4}	Gs1	1.3×10^{-4}	Tp1	1.3×10^{-4}	Po1	1.6×10^{-4}
Second	An2	6.1×10^{-5}	Gs2	8.5×10^{-5}	Tp2	1.1×10^{-4}	Po2	1.1×10^{-4}
$\frac{CLR_1}{CLR_2}$	2.5		1.5		1.2		1.5	

Table 6-3 Test materials used during the quality assessment session and their CLR

	Animation	*Golf Swing*	*Talking Person*	*Porsche*
Reference	4.8	4.6	4.6	4.8
First	3.8	2.7	3.0	3.1
Second	4.0	2.9	1.9	3.7

Table 6-4 Mean score for each test video sequence

The score assigned to each test video clip by each assessor is recorded and the data collected is then analyzed by calculating the mean of all the scores for each test sequence. The mean scores for each of the test video sequences are recorded in Table 6-4. If mean scores of 3 and 4 are considered to indicate video of 'acceptable' and 'good' quality respectively, then the following can be deduced from the results of the test session:

1. For the *Animation* sequence, a CLR of 6.1×10^{-5} results in 'good' quality. If 'acceptable' video quality is desired, a range of CLR values larger than 1.5×10^{-4} need to be applied to produce test materials for another test session;

2. For the *Golf Swing* sequence, a range of CLR values smaller than 8.5×10^{-5} should be applied to test materials for another test session in order to establish both the 'good' and 'acceptable' quality level;

3. Although it appears that a CLR value of 1.3×10^{-4} would results in 'acceptable' video quality for the *Talking person* sequence, this needs to be verified by further test sessions because of contradictory results obtained from the two test sequences;

4. For the *Porsche* sequence, while a CLR value of around 1.6×10^{-4} should result in 'acceptable' video quality, a range of CLR values above 1.1×10^{-4} should be used for test materials in another test session so that the CLR level corresponding to 'good' quality can be established.

Besides the analysis of CLR on video quality, further analysis of the test results is summarized as follows:

- While a grand mean score of all test sequences should be close to three as recommended in [BT500], the grand mean score for this test is 3.14;

- For test sequences created from *Animation*, *Golf Swing* and *Porsche*, a higher value of CLR results in a lower mean score and vice versa as expected;

- However, for the test materials generated from *Talking Person*, the relationship between CLR and the mean score is reversed. Although the difference between CLR values is relatively small (only a factor of 1.2), the difference in the mean scores is as large as 1.1. There are more video distortions in Tp2 and they are more obvious than those found in Tp1 despite a lower CLR. While there are four occurrences of *colour cycling* of different duration and some *motion jerkiness* in Tp2, there is only one incidence of *twisting* (together with *colour cycling*) and lip mis-synchronization in Tp1. This further illustrates that the locations of cell losses and the number of cell losses are both critical to the occurrences of video distortions and the resulting quality degradation;

- Some viewers are asked to assign scores to the reference sequences. The results show that the reference materials do not always deserve a rating of 5 in the minds of some assessors;

- The mean scores of test materials generated from *Golf Swing* and *Talking Person* are generally lower than those created from *Animation* and *Porsche*. Therefore, it would appear that the difference in scores for the reference sequences has an effect on the scores assigned to the test sequences. This agrees with the understanding that video quality at the destination is partly dependant on the original quality at the source;

- Although CLR_1 and CLR_2 differ by a factor of 2.5 for test materials created from *Animation*, the difference in the mean scores is only 0.2. This is the same as that for test sequences produced from *Golf Swing*, whose CLR_1 and CLR_2 differ by a factor of 1.5 only. This illustrates that the relationship between different CLR values and the resulting change in opinion scores is not necessarily linear. Similar phenomenon can be found with the test video clips generated from *Golf Swing* and *Porsche*: although the CLRs differ by the same factor of 1.5, the differences in mean scores are 0.2 and 0.6 respectively.

It is important to note that since a small-scale test session is conducted, the observations listed above need to be verified with further tests. However, when the same viewers are consulted, their opinions are likely to change as they tend to recognize particular coding artefacts or error patterns. So it may be necessary to use new viewers for each set of tests. In cases where new assessors are difficult to find, Objective Assessment Methods can be used in addition to, or instead of, Subjective Assessment. Further information about the user-oriented video quality assessment conducted in this research (e.g. posters, questionnaire, results from individual assessors etc) can be found in appendix D as well as on the CD.

6.7.3 Objective Assessment

One of the methods to assess video quality objectively is to use a quality metric known as the Peak Signal-to-Noise Ratio (PSNR). It measures the distortions of video frames resulting from network impairments and compares them to the original error-free frames. However, previous studies have shown that PSNR is poorly correlated to the human visual system (HVS) because PSNR does not take the 'masking effect'[30] into consideration. All video frame distortions cause the PSNR to decrease even if they are not perceivable by the HVS. As a result, PSNR provides an unsatisfactory means to measure video quality.

[30] The 'masking effect' occurs when distortions are not noticed by HVS even if they are present.

$$PSNR(dB) = 10\log_{10}\frac{(2^n - 1)^2}{MSE}$$

where,
n is the number of bits required to represent each pixel;
MSE is the mean squared error between the distorted frame relative to the original frame.

Equation 6-1 Definition of Peak Signal-to-Notse Ratio (PSNR)

Another method to measure video quality objectively is to use metrics that rely on a model of HVS. One of these metrics is known as the moving pictures quality metric (MPQM) in which the HVS is modelled as multiple filters that decompose the source and distorted video sequence into multiple channels. The 'masking effect' is simulated to determine which distortions are noticeable to the human eyes. Figure 6-16 shows the block diagram for MPQM and the MPQM tools used to evaluate video quality are presented in [VER98a].

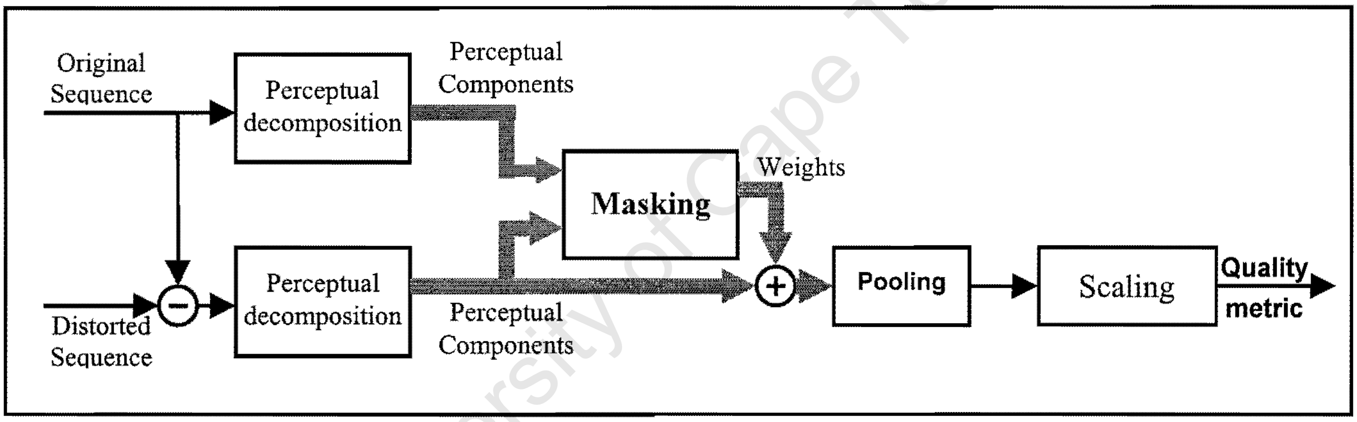


Figure 6-16 Block Diagram for Moving Pictures Quality Metric (MPQM) [VER98a]

The two main advantages of Objective Assessment methods are that they do not require human assessors and that they do not take past experiences of individuals towards video distortions into account. However, a potential disadvantage of these methods is that it might not reflect the responses of the HVS closely. Therefore, during the development of objective assessment techniques, it is advisable to test the results obtained against those provided by user-oriented quality assessment in order to verify its correlation with the Human Visual System.

6.7.4 Quality Degradation caused by Video and Audio Artefacts

This section aims to explain why some of the video artefacts are more obvious than others and summarize the main factors that contribute to quality degradation of video images. It is based upon the video artefacts shown in sections 6.1 and 6.4 and has taken into account the verbal feedback from assessors about the degradation of video during the quality assessment session. Although this research is focused on MPEG video, the analysis technique presented here can be applied to other types of video as well.

The extent of degradation in video quality resulting from distortions within individual frames is significantly dependent on the type, the size and the location of the distortion on the screen. When video playback is affected by distortions that extend over more than one frame, the amount of degradation is largely dependent on the duration of the distortions. The main factors that contribute to quality degradation are summarized as follows:

- **The nature of the video artefact** – certain types of video distortions, such as colour cycling and beyond range motion, generally appear to be more severe than others such as error blocks. This is because the former often affects a large portion of the screen or even the entire screen and the distortion is more obvious as a result;

- **The size of the distortion** – for video artefacts that do not normally affect the entire screen, such as tiling and error blocks, quality degradation is dependant on the size of the distortion;

- **The location of the distortion** – typically, video distortions occurring in the corners of the screen are not as noticeable as those appearing in the centre because of the next factor;

- **The focus of the viewer** – video distortions are generally more obvious if they coincide with the portion of the screen that the focus of the viewers is placed. This is normally the centre of the screen but it also depends on the content of the video. In some cases, the video artefacts observed can even be different when the focus of the viewers is placed on different parts of the screen, such as *motion jerkiness* and *severe jump* illustrated in section 6.1;

- **The duration of the distortion** – for all types of video artefacts identified in this study (frame freezing, motion jerkiness and artefacts identified in section 6.4.1), the resulting visual effect depends significantly on the duration for which they persist on the screen. Some video artefacts just flash pass the screen briefly and are not very obvious unless playback is paused. Therefore, these distortions are not as significant as and do not cause as much quality degradation as those lasting for a few hundred milliseconds because most viewers do not often pause the video playback;

- **The content of the video images** – Moreover, distortions of reasonably fast moving objects or during scene change are less obvious and cause relatively less quality degradation. For example, *frame freezing* at *00:09* of pikeplace2.mpg and at *00:09* of Gs2.mpg. Figure 6-17 shows red error blocks that are not obvious because they occur on the red clothing of the technician.


Figure 6-17 Solid Error Blocks that are not obvious

- For cell losses (without dummy cell insertion), cell mis-insertion or cases where AAL error checking mechanisms are enabled, video quality degradation is significantly dependent on the number of video frames and audio packets being skipped. This is in turn dependent on the importance of the information being affected by the network impairments.

6.8 Mapping Application and Network Level Quality-of-Service

A possible way to establish the mapping between application level QoS and network level QoS is by conducting user-oriented video quality assessment outlined in section 6.7.2. Although only cell loss with dummy cell insertion has been applied to the test materials used in the quality assessment session, the degradation in video quality caused by other types of network impairments can be investigated in a similar manner. Alternatively, the quality degradation resulting from specific video artefact(s) can be tested for. This can be achieved by careful manipulation of the MPEG bit stream so that it exhibits video artefacts according to the need for the test.

However, the establishment of one set of QoS parameters for MPEG video traffic is almost impossible because of the high diversity in the requirements of individual video clips as demonstrated in section 6.7.2. The QoS requirements of video sequences depends on the following factors:

- The purpose of the video – video clips can be used as news article, documentary, movie etc. Because of the inherent differences among these video types, this determines the following factors to a large extent;

- The quality of the video source – high quality video sequences usually have a more stringent QoS requirement than video clips captured as lower quality;

- The amount of details – video scenes containing many detailed objects are less tolerant to network impairments and therefore requires strict QoS support;

- The amount of motion – video scenes with fast movement of objects are more prone to network impairments and requires a more strenuous QoS in order to achieve good video quality;

- Other attributes of the video sequence – this includes frame rate, resolution etc.

Because an ATM network needs to support a wide variety of videos, it is advisable to divide video sequences into a number of categories. Each category contains videos of similar quality and other attributes. The mapping of application level QoS and network level QoS for each category of video can be achieved by establishing the level of network impairments that would result in 'good' or 'acceptable' video quality using either subjective or objective assessment schemes outlined in section 6.7.

The mapping between network QoS and user QoS is not a straightforward task because it depends significantly on particular end-users and implementation factors of the application and involves extensive experimentation. Firstly of all, not all video distortions are noticeable to all users. Secondly, the level of acceptable quality is highly dependent on the purpose of the video stream. Thirdly, this mapping depends on the nature of the video sequence. Other factors include the decoding algorithm and its vulnerability towards network impairments, the error concealment schemes implemented etc. However, this mapping is necessary if ATM networks are to provide video services that are satisfactory to end-users.

6.9 Solutions against ATM Network Impairments

This section outlines a few techniques that are capable of reducing the effects of ATM network impairments on video quality. These techniques include dummy cell insertion, error correction, error concealment, robust coding and slice size variation. Other techniques, such as those proposed in [SHA90], [KIT90] and [RAY96], have also been found in the literature.

6.9.1 Dummy Cell Insertion

This technique has been briefly described in section 6.4 and is implemented in the end-systems receiving video traffic. When the error checking mechanisms in the AAL detect an error in the length field that indicates a cell loss, a dummy cell that contains only 0's is inserted in the CPCS-PDU to replace the lost cell. The location of the lost cell can be discovered by checking the sequence counter of the cells that are received. Although AAL 5 does not provide a sequence counter in the way that AAL 1 does, a similar approach can be adopted.

Dummy cell insertion maintains the bit count integrity of MPEG encoded streams and can normally prevent the skipping of video frames that is caused by the lost of synchronization between the decoder and the bit stream due to the loss of information. In fact, dummy cell insertion replaces cell loss with a number of bit errors. Although frame skipping is replaced by various video artefacts illustrated in section 6.4.1, dummy cell insertion is generally considered to improve the quality of video sequences that are under the influence of cell loss.

6.9.2 Error Correction

Feedback error correction schemes such as automatic repeat request (ARQ) are considered to be inappropriate for real-time video transmission. This is because of the extra delay introduced by the acknowledgement and retransmission or errored cells and the resulting out of sequence delivery of video information. Forward Error Correction (FEC) adds redundancy to the video information in order to provide protection against errors in transmission. FEC schemes are specified by the number of original information symbols and the number of symbols after the addition of redundancy. For example, the Reed-Solomon (128,124) code used in AAL 1 adds 4 redundant symbols to 124 original information symbols to produce 128 output symbols.

FEC schemes increases the bandwidth requirements of video traffic due to the addition of redundant information. They have different error protection capabilities depending on the amount of redundancy added to the original information. When the amount of errors exceed the protection a FEC scheme can provide, errors will still occur.

6.9.3 Error Concealment

Error concealment techniques are implemented in the video decoder with the aim of hiding the video distortions caused by errors in video information and improves video quality as a result. The video decoder must first detect the error before these mechanisms can be activated. Error concealment can be classified into three main types:

- Temporal concealment – this replaces a lost or corrupted area with pixels in the same location of previous video frames. This scheme is only effective when there are little changes between the corrupted frame and the previous frame. However, if the corrupted frame is situated just after a scene change, the left-over effect may occur.

- Spatial concealment – this technique replaces a distorted block by an estimation made from adjacent error-free blocks. This scheme is effective when the details in the scene is not too high.

- Motion-compensated concealment – this is an improvement of the temporal concealment scheme by estimating the motion vectors for the lost blocks instead of using pixels from the same location. This scheme is generally considered to be more effective than the other two.

6.9.4 De-jitter Buffering

The variable cell delay experience by video traffic can be compensated by the introduction of a de-jitter buffer in the video decoder. At the start of the decoding process, such a buffer is filled up to half its capacity. Variable cell delay introduced to video cells by the ATM network can then be 'absorbed' by the de-jitter buffer. The output of this buffer is fed to the decoding mechanisms, which receives video bit stream that is jitter-free.

6.9.5 Robust Coding

Another technique to reduce the effect of ATM network impairments on video quality is to encode the video information differently so that the encoded bit stream is more robust to errors. Besides improving video quality, this technique must ensure that the encoded bit stream to be compatible with standard video decoders. Two such techniques proposed in the literature are scalable or layered coding and temporal localization.

Layered video coding techniques encode video into two or more layers, each contains different component(s) of the original video information. The video sequence is encoded at a low quality to form the base layer. The difference between the base quality and the original quality of the source is encoded as one or more enhancement layers. While decoding the base layer results in video of low quality, the decoding of the base layer with the enhancement layer(s) results in video of improved quality. The base and the enhancement layers are sent over multiple VCCs with different QoS requirements. Since the base layer is more significant to the decoding process than the enhancement layer(s), it is handled with a higher priority and a stricter QoS compared to the enhancement layer(s) during transmission. As a result, the base layer is affected by network impairments to a lesser extent and maintains basic video quality when the enhancement layer(s) are affected by network impairments. Figure 6-18 shows the functional blocks of the layered coding technique.

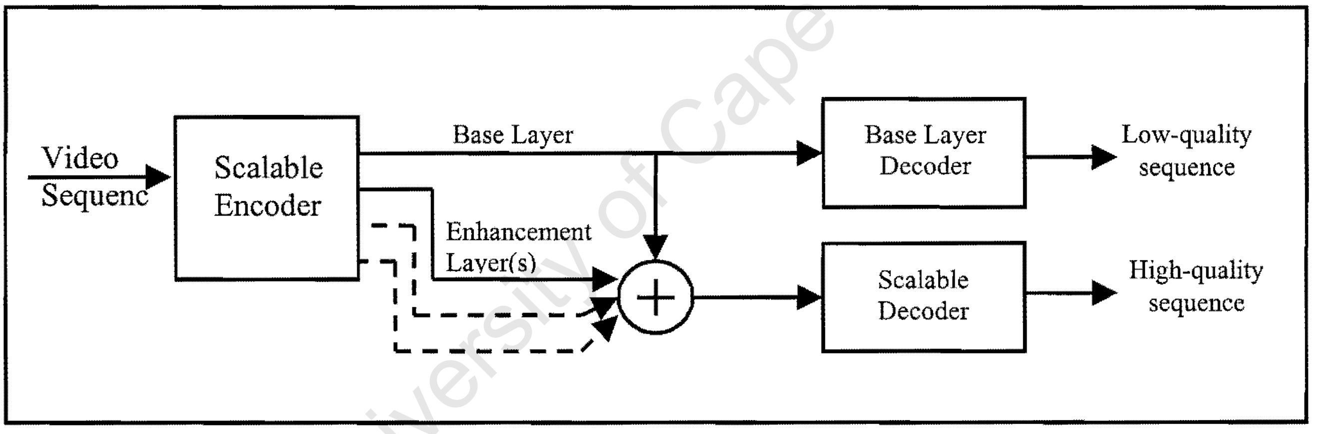


Figure 6-18 Block Diagram of scalable or layered coding system [RIL97]

Since temporal propagation of video distortions does not proceed beyond an error free I frame in a MPEG sequence, reducing the size of the group of picture, i.e. reduce the number of frames between consecutive I frames, should reduce the extent of such error propagation. Temporal localization techniques reduce the temporal spacing of I frames and hence increase the proportion of I frames in the coded sequence. Although the extent of temporal propagation of video distortions is reduced, this method has the disadvantage of decreasing the coding efficiency.

Chapter 7.

Conclusions

The main goal of this research is to investigate the effects of ATM network impairments to video quality. To study the distortions and quality degradation affecting video traffic over ATM, many issues need to be considered such as the characteristics of the video encoding technique, the transportation schemes for video over ATM, the types of network impairments in ATM and how they affect video traffic etc. This goal has been achieved through the development of the network impairment emulation (NetIE) architecture.

An emulated network was designed and implemented to introduce controllable network impairments to video traffic in ATM. Two NetIE models, Statistical and Application-Specific, are presented in this dissertation. These two models complement the limitations of each other and can be used to study the effects of network impairments on video traffic in general and on video traffic of a particular type respectively. The principle of network emulation with controllable impairment testing has also been adopted in the Broadband Series Test Systems from Hewlett-Packard.

The video distortions and quality degradations presented in this thesis are based on the MPEG encoding format because it achieves a high degree of compression and is widely used. However, it is important to note that to study ATM traffic containing video encoded in another format, modification is not required in the emulated network as far as the Statistical NetIE model is concerned. Because the sole difference between the testing of MPEG and video in another format is the type of information transferred from the source and the destination, the only modification required on the overall video testbed is the applications running on the two end-systems.

Since the study of video quality is based on the MPEG video standard, conclusions specific to the transportation of MPEG over ATM and the resulting video quality can be drawn:

- More than ten categories of video artefacts have been identified and illustrated during this research. The spatial and temporal propagation of distortions and their effects on video quality have been examined;

- If the implementation of AAL error correction discard the entire CPCS-PDU upon the detection of a CRC or length field error, a large proportion of the discarded bits is free of errors. Amongst other consequences, this causes Cell Loss, Cell Error and Cell Mis-insertion to produce the same effect to the decoder, which loses synchronization with the MPEG bit stream due to the loss of information bits equal to the size of the CPCS-PDU payload;

- The insertion of dummy cells when cell losses occur generally replaces *motion jerkiness* and *frame freezing* with less severe video distortions and therefore often improves overall video quality;

- Despite dummy cell insertion, cell losses generally cause more severe video distortion than cell error because more information bits are affected by the former;

- The occurrences of video distortion and the resulting quality degradation are highly dependant on where the impairment occurs with respect to the information fields in the MPEG bit stream. This is because of the varying importance carried by different information fields;

- The MPEG bit stream could be encoded in base and enhancement layers. The use of multiple connections each with a suitable level of ATM QoS support to transfer the MPEG video layers should result in an improvement of video quality;

- The measurement of video quality using objective and subjective assessment techniques has been summarized and the detailed procedures of user-oriented video quality assessment has been presented.

This research addresses the mapping of application level QoS and network level QoS. It also demonstrates why it is impossible to have a single set of QoS for all types of video traffic and presents a possible method to establish this mapping.

The design strategy of the emulated network adopted by this study has enabled the study of ATM network impairments on other types of traffic such as data or voice. The wide variety of application for the emulated network, besides acting as the core of the video testbed, has justified the choice in favour of Design II over Design I.

This research has opened many new doors and the NetIE architecture has much scope for expansion. It provides the platform for a variety of investigations in the diverse field of supporting video traffic over ATM. The programmability made available by the virtual ATM switch may also prove to be useful in other areas of ATM research.

Chapter 8.

Recommendations for Future Work

During the course of this research, several areas of interest have surfaced but could not be included in the scope of this project. The following list outlines possible further work based on this study:

- investigate other possible methods to generate impairment events within the emulated network;
- investigate how the AAL error checking mechanisms at the destination can be disabled;
- investigate the effects of sending MPEG video over multiple ATM connections with different QoS and compares video quality under the influence of network impairments with that obtained in this study;
- develop the Application-Specific NetIE model further to study the more specific aspects of MPEG encoded video;
- upgrade to version 0.6 of the virtual switch software which provides SVC signalling support so that the network emulation architecture can be applied to applications that requires switched virtual circuits;
- pursue QoS mapping further with the platform provided by this research;
- test other ATM applications with the emulated network developed in this study.

References

[ANA92] M.E. Anagnostou, M.E. Theologou, E.N. Protonotarios, "Cell Insertion Ratio analysis in Asynchronous Transfer Mode networks", Computer Networks and ISDN Systems, Vol.24, No.4, 1992, pp. 335 – 344

[AOL00] ATM on Linux Home Page
http://lrcwww.epfl.ch/linux-atm/

[AOL98] Usage Instructions – ATM on Linux, release 0.52, December 1998.

[API96] ATM on Linux Application Programming Interface (API)
http://icawww1.epfl.ch/linux-atm/doc.html#api

[AUR98] Cristina Aurrecoechea, Andrew T. Campbell, Linda Hauw, "A Survey of QoS Architectures", ACM Multimedia Systems, Vol.6, No.3, 1998, pp.138-151

[BEC96] Michael Beck, Linux Kernel Internals, Addison-Wesley, Inc, 1996

[BT500] Recommendation ITU-R BT.500-9, "Methodology for the subjective assessment of the quality of television pictures", 1992.

[CHA96] Sze Keong Chan, "End to End Performance of MPEG over ATM Networks", PhD Thesis, University of Toronto, 1996

[CHA98] SK Chang, E. Hassanein, CY Hsieh, "A Multimedia Micro-University", IEEE Multimedia, Vol.5, No.3, July-September 1998, pp.60-68
Knowledge System Institute is on the World Wide Web at http://www.ksi.edu

[CIS98] Video over ATM and existing networks, Cisco Systems 1998

[CIS99] Designing Internetworks for Multimedia – Cisco Systems Inc. 1999
http://www.cisco.com/cpress/cc/td/cpress/ccie/ndcs798/nd2013.htm

[CLA92] William J. Clark, "Multipoint Multimedia Conferencing", IEEE Communications Magazine, Vol.30, No.5, May 1992, pp.44-50

[COF97] Video over ATM networks – Gregory A. Coffey
http://www.cis.ohio-state.edu/~jain/cis788-97/video_over_atm/index.htm

[FORLE] ForeRunnerLE[tm] ATM Adapters - Series of ATM Adapters for Desktops
http://www.fore.com/products/le-adpt/le-adpt_hp.html

[GAR96] Xavier Garcia Adanez, Andrea Basso, Jean-Pierre Hubaux, "Study of AAL5 and a New AAL Segmentation Mechanism for MPEG-2 Video over ATM", TCOM Laboratory, Telecommunications Services Group, Swiss Federal Institute of Technology, Lausanne Switzerland

[GEL97] E. Gelenbe, X. Mang and R. Onvural, "Bandwidth Allocation and Call Admission Control in High-Speed Networks", IEEE Communications Magazine, Vol.35, No.5, May 1997, pp. 122 – 129

[GHA93] Mohammad Ghanbari, Charles J. Hughes, "Packing Coded Video Signals into ATM cells", IEEE/ACM Transactions on Netowrking, Vol.1, No.5, October 1993, pp.505-508

[GRI98] Steven Gringeri, Bhumip Kansnabish, Arianne Lewis, Khaled Shuaib, Roman Egorov and Bert Basch, "Transmission of MPEG-2 Video Streams over ATM", IEEE Multimedia, January-March 1998, pp. 58 – 71.

[GUH95] Aloke Guha, Allalaghatta Pavan, Jonathan Liu, Ajay Rastogi, Todd Steeves, "Supporting Real-Time and Multimedia Applications on the Mercuri Testbed", IEEE JSAC, Vol.13, No.4, May 1995, pp. 749 – 763

[H.222.1] ITU-T Recommendation H.222.1, "Multimedia multiplex and synchronization for audiovisual communication in ATM environments", March 1996

[H.261] ITU-T Recommendation H.261, "Video codec for audiovisual services at p x 64 kbit/s", March 1993

[HES98] Christopher K. Hess, Daniel Lin, Klara Nahrstedt, "VistaMail: An integrated Multimedia Mailing System", IEEE Multimedia, Vol.5, No.4, October 1998, pp.13 – 23

[HOD95] Winston Hodge, Interactive Television, McGraw-Hill, Inc. 1995

[HP-1] Broadband Series Test System, Hewlett Packard http://www.tm.agilent.com/tmo/Products/English/BroadbandSeriesTestSystem.html

[HP-2] HP E4219A ATM Network Impairment Emulator http://www.tm.agilent.com/tmo/datasheets/English/HPE4219A.html

[HP-3] HP E4200B Form 7 Transportable Base Platform http://www.tm.agilent.com/tmo/datasheets/English/HPE4200B.html

[HP-4] HP E1697A 155 Mb/s Optical Line Interface http://www.tm.agilent.com/tmo/datasheets/English/HPE1697A.html

[HP95] Hewlett-Packard, "The ATM Impairment Testing Handbook", 1995

[HUA98] Chung-Ming Huang and Chian Wang, "Synchronization for Interactive Multimedia Presentations", IEEE Multimedia, Vol.5, No.4, October 1998, pp.44-62

[I.112] ITU-T Recommendation I.112, "Vocabulary of Terms for ISDNs", March 1993

[I.121] ITU-T Recommendation I.121, "Broadband Aspects of ISDN", April 1991

[I.150] ITU-T Recommendation I.150 "B-ISDN ATM functional characteristics", November 1995.

[I.211] ITU-T Recommendation I.211, "B-ISDN Service Aspects", March 1993

[I.321] ITU-T Recommendation I.321, "B-ISDN protocol reference model and its application", April 1991

[I.327] ITU-T Recommendation I.327, "B-ISDN functional architecture", March 1993

[I.350] ITU-T Recommendation I.350, "General aspects of quality of service and network performance in digital networks, including ISDNs", March 1993

[I.353] ITU-T Recommendation I.353, “Reference events for defining ISDN and B-ISDN performance parameters”, August 1996

[I.356] ITU-T Recommendation I.356, “B-ISDN ATM layer cell transfer performance”, October 1996

[I.361] ITU-T Recommendation I.361, “B-ISDN ATM layer specification”, November 1995

[I.363.1] ITU-T Recommendation I.363.1, “B-ISDN ATM Adaptation Layer specification: Type 1 AAL”, August 1996

[I.363.5] ITU-T Recommendation I.363.5, “B-ISDN ATM Adaptation Layer specification: Type 5 AAL”, August 1996

[I.371] ITU-T Recommendation I.371, “Traffic control and congestion control in B-ISDN”, March 1993 (08/96)

[I.432.1] ITU-T Recommendation I.432.1, “B-ISDN user-network interface - Physical layer specification: General characteristics”, February 1999

[JUN96] J. I. Jung, “Translation of user’s QoS requirements into ATM performance parameters in B-ISDN”, Computer Networks and ISDN Systems, Vol. 28, 1996, pp. 1753 – 1767.

[KER97] Brigitte Kervella, Valerie Gay, "MHEGAM - A Multimedia Messaging System", IEEE Multimedia, Vol.4, No.4, October 1997, pp.22-29

[KIT90] T. Kitami, I. Tokizawa, “Cell Loss Compensationn Schemes Employing error Correction Coding for Asynchronous Broadband ISDN”, Proceedings of IEEE INFOCOM’90, 1990.

[KIT97] Frederick Lee Kitson, Tom Malzbender, Vasudev Bhaskaran, “Opportunities for Visual Computing in Healthcare”, IEEE Multimedia, , Vol.4, No.2, April-June 1997, pp.46 – 57

[KWO98] Timothy Kwok, ATM: The New Paradigm for Internet, Intranet, and Residential Broadband Services and Applications, Prentice-Hall PTR, 1998.

[LAP95] Video Communications in B-ISDN, Yael Lapid February 1995 http://bbs-mac.uniinc.msk.ru/tech1/1994/atm/vid-atm.htm

[LEB92] Jean-Yves Le Boudec, “The Asynchronous Transfer Mode: A Tutorial”, Computer Networks and ISDN Systems, Vol.24, No.4, 1992, pp. 279 – 309

[LEL98] A. Lele, S.K. Nandy, “Can QoS Guarantees be supported for Live Video over ATM networks?”, Proceedings of GlobeCom’98, 1998, pp.2315 – 2323

[LEO94] Alberto Leon-Garcia, Probability and Random Processes for Electrical Engineering, Addision-Wesley Publishing Company, Inc, 1994

[LEspec] ForeRunnerLE[tm] ATM Adapters – Technical Specification http://www.fore.com/products/le-adpt/le-adpt_ts.html

[LIA98] Wanjiun Liao, Victor O.K. Li, "Synchronization of distributed multimedia systems with user interactions" ACM Multimedia Systems, Vol.6, No.3, May 1998, pp.196 – 205

[LIN96] Mengjou Lin, David Singer, Alagu Periyannan, "Supporting constant-bit-rate encoded MPEG-2 transport over local ATM networks", Multimedia Systems, Vol.4, No.2, 1996, pp.87 – 98.

[M13818] ISO/IEC 13818-1 Information technology – Generic coding of moving pictures and associated audio information: Systems 1996
ISO/IEC 13818-2 Information technology – Generic coding of moving pictures and associated audio information: Video 1996
ISO/IEC 13818-3 Information technology – Generic coding of moving pictures and associated audio information: Audio 1995

[MAG00] Magellan University – America's First Online and Video University http://magellan.edu/

[MCD95] David E. McDysan, Darren L. Spohn, ATM: Theory and Application, McGraw-Hill, Inc. 1995.

[MIN94] Dan Minoli, Bob Keinath, Distributed Multimedia Through Broadband Communications, Artech House, 1994

[MUR96] Shree Murthy, Hari Lalgudi, "Impact of QOS requirements on video coding for ATM networks", Multimedia Systems, Vol.4, No.6, 1996, pp.316-327

[NEM98] ATM Network Impairment Emulator User's Guide, Broadband Series Test System, Hewlett-Packard Australia Limited, 1998.

[ONV94] Raif, O. Onvural, Asynchronous Transfer Mode Networks: Performance Issues, Artech House, Inc. 1994

[ORZ98] Michael Orzessek, Peter Sommer, ATM & MPEG-2 – Integrating Digital Video into Broadband Networks, Prentice-Hall Inc, 1998

[OZE95] Jan Ozer, Video Compression for Multimedia, Academic Press, Inc. 1995

[PAN94] Pramod Pancha and Magda El Zarki, "MPEG Coding for Variable Bit Rate Video Transmission", IEEE Communications Magazine,Vol.32,No.5, May 1994, pp.54–66

[PNNI-1] The ATM Forum Technical Committee, "Private Network-Network Interface Specification Version 1.0 (PNNI 1.0) March 1996 (code: af-pnni-0055.000)

[Q2931] ITU-T Recommendation Q.2931, "Digital Subscriber Signalling System No. 2 - User-Network Interface (UNI) layer 3 specification for basic call/connection control", February 1995

[QPO96] Bellcore, "Q.Port Portable ATM Signaling Software – Tcl Command Reference", Release 2.0, September, 1996

[RAS95] Y. Rasheed ,"ATM Adaptation Layers for Video Applications", M.A.Sc. Thesis, University of Toronto, January 1995

[RAY96] D. Raychaudhuri, D. Reininger, R. Siracusa, "Video transport in ATM networks: a systems view" Multimedia Systems, Vol.4, No.6, 1996, pp.305-315

[RIL97] Martyn J.Riley, Iain E.G. Richardson, Digital Video Communications, Artech House, 1997

[ROS92] Jonathan Rosenberg, Robert E. Kraut, Louis Gomez and C. Alan Buzzard, "Multimedia Communications for Users" IEEE Communications Magazine, Vol.30, No.5, May 1992, pp.20-36

[ROU99] M.J. Roux, "ATOM: A Distributed System for Video Retrieval via ATM Networks", MSc Thesis, University of Cape Town, September 1999

[RUB98] Alessandro Rubini, Linux Device Drivers, O'Reilly & Associates, Inc. 1998

[SAN99] R.Sanchez, "Virtual ATM switch for ATM on Linux (Release 0.5)" March 1999, http://www.ittc.ukans.edu/~rsanchez/software/vswitch/README-0.5

[SHA90] N. Shacham, P. McKenney, "Packet recovery in high-speed networks using coding and buffer management", Proceedings of IEEE INFOCOM'90, 1990

[TAT98] Mallikarjun Tatipamula, Bhumip Khasnabish, Multimedia Communications Networks – Technologies and Services, Artech House, Inc, 1998

[TM4.0] The ATM Forum, "Traffic Management Specification, version 4.0", April 1996 (code: af-tm-0056.000)

[TRY96] Christos Tryfonas, "MPEG-2 Transport over ATM Networks", MSc Thesis, University of California, 1996

[TSA96] Rose P. Tsang, David H.C. Du, Allalaghatta Pavan, "Experiments with video transmission over an Asynchronous Transfer Mode (ATM) network", Multimedia Systems, Vol.4, No.4, 1996, pp.157-171

[UNI3.1] The ATM Forum, "ATM User-Network Interface Specification Version 3.1", September 1994

[UNI4.0] The ATM Forum Technical Committee, "ATM User-Network Interface (UNI) Signalling Specification, version 4.0", July 1996 (code: af-sig-0061.000)

[UTA00] The Centre for Distance Education, University of Texas at Arlington http://distance.uta.edu

[VAS99] Virtual ATM switch driver for ATM on Linux – Home Page http://www.ittc.ukans.edu/~rsanchez/software/vswitch.html

[VER88] W. Verbiest, L. Pinnoo, B. Voeten, "The Impact of the ATM Concept on Video Coding", IEEE JSAC, Vol.6, No.9, December 1988, pp. 1623 – 1632

[VER96] O. Verscheure and X. Garcia Adanez, "Perceptual Quality Metric as a Performance Tool for ATM Adaptation of MPEG-2 based Multimedia Applications", Swiss Federal Institute of Technology, Switzerland.

[VER98a] O.Verscheure, Xavier Garcia, Gunnar Karlsson and Jean-Pierre Hubaux, "User-Oriented QoS in Packet Video Delivery", IEEE Network, Vol. 12, No. 6, November/December 1998, pp. 12 – 21.

[VER98b] Olivier Verscheure, Pascal Frossard and Maher Hamdi, "Joint Impact of MPEG-2 Encoding Rate and ATM Cell Losses on Video Quality", Proceedings of GlobeCom'98, 1998, pp.188-193

[VOD97] The ATM Forum, "Technical Committee Audiovisual Multimedia Services: Video on Demand Specification 1.1, March, 1997 (code: af-saa-0049.001)

[WEL98] Matt Welsh, Phil Hughes, David Bandel et al, Linux Installagion and Getting Started, Specialized Systems Consultants, Inc. 1998
available for download at http://metalab.unc.edu/mdw/docs.html#guide

[WON97] Stephen TC. Wong, HK. Huang, "Networked Multimedia for Medical Imaging", IEEE Multimedia, Vol.4, No.2, April-June 1997, pp.24-35
abstract available on the World Wide Web
http://computer.org/multimedia/mu1997/u2024abs.htm

[WOO90] Gillian M. Woodruff and Rungroj Kositpaiboon, "Multimedia Traffic Management Principles for Guaranteed ATM Network Performance", IEEE JSAC, Vol.8, No.3, 1990, pp. 437 – 446

[XIE99] Yong Xie, Changdong Liu, Myung J. Lee, Tarek N. Saadawi, "Adaptive multimedia synchronization in a teleconference system" ACM Multimedia Systems, Vol.7, No.4, July 1999, pp.326-337

[YAM93] K. Yamazaki, M. Wada, Y. Takishima and Y.Wakahara, "ATM Networking and Video-Coding Techniques for QoS Control in B-ISDN", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 3, No. 3, June 1993, pp. 175 – 181

[ZAH97] T. Zahariadis, C. Rosa, M Pellegrinato, A.B. Lund, G. Stassinopoulos, "Interactive Multimedia Services to Residential Users", IEEE Communications Magazine, Vol.35 No.6 June 1997, pp.61-68

[ZAR96] Panagiotis N. Zarros, Myung J. Lee, Tarek N. Saadawi, "A synchroniation algorithm for distributed multimedia environments", Multimedia Systems, Vol.4, No.1, 1996, pp.1-11

Appendix A

Overview of the B-ISDN and ATM standards

This section aims to provide an overview of the B-ISDN and ATM standards but does not intend to be a comprehensive introduction. Therefore, only concepts that are fundamental and relevant to the thesis project will be introduced.

A.1 B-ISDN and ATM

The concept of ISDN (Integrated Service Digital Network) was adopted by the International Telegraph and Telephone Consultative Committee[31] (CCITT) in 1984. ISDN is defined to be a network that provides digital connections between user-network interfaces[32] and supports a range of different telecommunication services [I.112]. The main feature of the ISDN concept is the support of a wide range of audio, video and data applications in the same network [I.121]. The significant higher bit rate requirement for broadband video applications than data applications has led to the development of Broadband-ISDN. According to ITU-T Recommendation I.113, broadband refers to "a service or system requiring transmission channels capable of supporting rates greater than the primary rate". The ISDN primary rate channel referred to above is a 2 megabits per second (Mbps) circuit switched channel [RIL97].

The availability of high-speed transmission, switching and signal processing technologies has made the development of B-ISDN possible. Some of the other drivers behind the development of

[31] CCITT is now known as the ITU-T, the Telecommunication Standardization Sector of the International Telecommunication Union

[32] This is the interface between the terminal equipment and a network termination at which interface the access protocol (i.e. a defined set of procedures) apply

B-ISDN include: 1)the need to integrate both interactive and distribution services as well as both circuit ad packet transfer mode into one universal broadband network, and 2)the need to provide flexibility in satisfying the requirements of both user and operator. The ITU defines B-ISDN services into four classes [I.211]:

- Conversational services – bi-directional communication with real-time (no store-and-forward) end-to-end information transfer from user to user, for example video telephony and video conferencing;
- Messaging services – with user-to-user communication between individual users via storage units with store-and-forward, such as e-mail incorporating video and audio;
- Retrieval services – where users can retrieve information stored in information centres on demand such as video library or Video on Demand (VoD);
- Distribution services – where a continuous flow of information is distributed from a central source to an unlimited number of authorized receivers connected to the network, for example digital television broadcast or High Definition Television (HDTV).

Asynchronous transfer mode (ATM) is the transfer mode for implementing B-ISDN and is independent of the means of transport at the Physical Layer [I.121]. ATM is a cell-based switching and multiplexing technology designed to be a general-purpose transfer mode for a wide range of services [MCD95]. It allows for the transportation of data, voice and video traffic simultaneously over high bandwidth networks such as B-ISDN. An ATM-based B-ISDN network appears to be suitable to handle real-time video traffic because it offers the following features:

- High bandwidth capacity with typical bit rates being 155Mbps and 622Mbps
- The flexibility to handle continuous as well as bursty traffic (constant or variable bit rate services)
- The 'bandwidth on demand' nature (dynamic bandwidth management)
- Quality-of-Service (QoS) guarantees for connections accepted by the network
- Its ability to handle delay sensitive traffic

A.2 Principles of the Asynchronous Transfer Mode

ATM is a connection-oriented technique and an ATM cell consists of a 5-byte header and a 48-byte payload as shown in Figure 8-1. ATM cells from different network connections are multiplexed onto shared links. Cells on a shared link are identified as belonging to a particular virtual path (VP) or virtual channel (VC) (Figure 8-2) by the VPI/VCI fields within the cell header. The physical link on which the ATM cells flow can be viewed as carrying multiple Virtual Path Connections (VPC). Each VPC in turn carries multiple Virtual Channel Connections (VCC) across the ATM network.

Bit

8	7	6	5	4	3	2	1	Octet
GFC				VPI				1
VPI				VCI				2
VCI								3
VCI				PT		RES	CLP	4
HEC								5
PAYLOAD (48 octets)								6...53

GFC	Generic flow control	PT	Payload type
VPI	Virtual path identifier	CLP	Cell loss priority bit (0 = higher priority)
VCI	Virtual circuit identifier	HEC	Header error check

Figure 8-1 ATM cell format (User to Network Interface)

The difference between the User-Network Interface (UNI) and the Network-Node Interface (NNI) is that the GFC field does not exist in the NNI cell format and the VPI field takes up the 4 bits and becomes a 12-bit field. The Payload Type Indicator (PT) is primarily used to distinguish between user cells (those carrying user information) and non-user cells (those carrying operations, administration and management OAM and traffic management information). The PT is also used by AAL 5 as described in section 3.3.3. The Cell Loss Priority (CLP) bit is assigned by ATM endpoints to indicate the relative priority of the cell for a particular VCC. A CLP value of 0 for a particular cell indicates that it is a high priority cell, while CLP=1 indicates a low priority cell. When an ATM network experiences congestion at one or more switches, cells with CLP=1 are subjected to discard by these congested switches in order to ensure the arrival of high priority cells at their respective destinations. The Header Error Check (HEC) is used to protect only the ATM cell header by detecting and optionally correcting a single bit error in the header fields. Cell header protection is important because it contains vital switching information in the VPI/VCI fields.

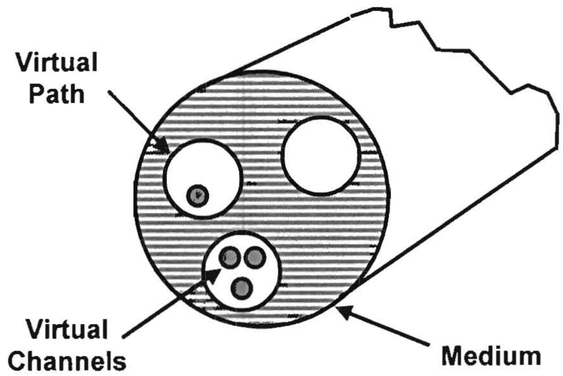


Figure 8-2 Virtual paths and virtual channels

When two hosts on an ATM network needs to communicate with one another, they first have to set up a virtual connection between them. This connection is called a virtual circuit and is identified at the two hosts by its VPI/VCI values, which may or may not be the same at the two hosts for this connection. This virtual connection consists of the concatenation of virtual circuits on each of the separate links throughout the network (Figure 8-3). There are two types of virtual circuits: permanent virtual circuit (PVC) and switched virtual circuit (SVC). Permanent virtual circuits are generally pre-configured manually by network administrators. They remain in service for a considerable duration and are only occasionally reconfigured. On the other hand, SVCs are established by end users on demand using the ATM signalling protocol and are often torn down when the virtual circuit is no longer required.

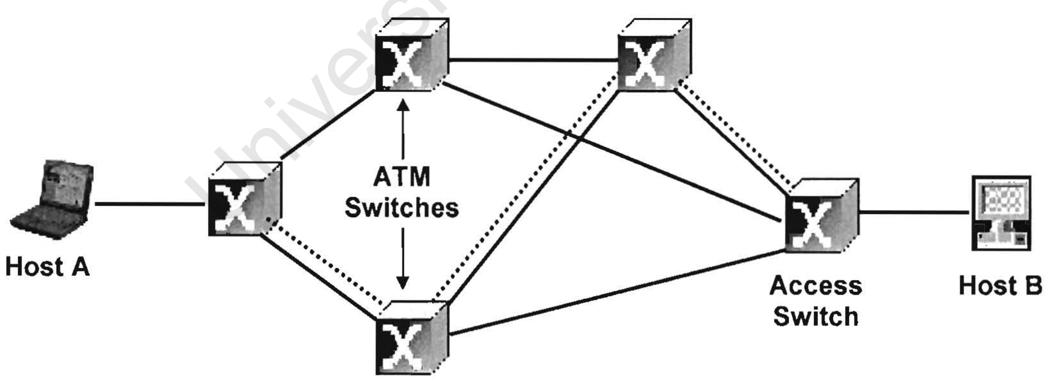


Figure 8-3 A virtual circuit in an ATM network

Depending the type of virtual circuit required by the communication session, one of the hosts sends a request for a SVC to the network by means of the ATM signalling protocol. Traffic descriptors and the chosen QoS class are provided by the host and the connection admission control (CAC) mechanism at the access network handles the request. If there is enough resource in the ATM network to handle the connection without compromising the QoS for the existing connections, the virtual circuit is established and a traffic contract is set up between the hosts and the network. The traffic contract is an agreement between a user and a network regarding the Quality of Service (QoS) that a cell flow is guaranteed if the cell flow conforms to a set of traffic parameters [MCD95].

The traffic contract consists of three main sections [TM4.0]: 1)the Connection Traffic Descriptor; 2)the requested QoS class, and 3)the definition of a compliant connection. The Connection Traffic Descriptor includes the source traffic descriptor, the cell delay variation tolerance (CDVT)[33] and the cell-by-cell conformance definition based on the Generic Cell Rate Algorithm (GCRA). The source traffic descriptor is a set of traffic parameters that describe the traffic characteristics of a source and define at least the Peak Cell Rate (PCR) (as in [I.371] and [UNI3.1]). It may optionally define a Sustainable Cell Rate (SCR), a Maximum Burst Size (MBS) and a Minimum Cell Rate (MCR) [TM4.0].

The QoS of a connection is described by the following Quality of Service performance parameters [TM4.0]:

- Cell Loss Ratio (CLR)
- Cell Error Ratio (CER)
- Severely Errored Cell Block Ratio
- Cell Mis-insertion Rate (CMR)
- Cell Transfer Delay (CTD)
- Cell Delay Variation (CDV)

The user of a virtual path connection VPC or a virtual channel connection VCC is provided with one of a number of QoS classes supported by the network [I.150]. Each QoS class defines particular values of the QoS parameters. Users can request and receive different QoS classes on a connection-by-connection basis so that distinct performance needs of different services and applications can be met [I.356]. In place of QoS classes, [TM4.0] states that QoS may be specified in terms of individual numeric parameter values using the procedures defined in [UNI4.0] and [PNNI-1].

The precise definition of a compliant connection is network specific and does not necessary require all cells to be conforming. In other words, while a connection for which all cells are conforming shall be identified as compliant, it is possible for non-conforming cells to exist in a

[33] CDVT is a mandatory parameter in any connection traffic descriptor and can either be explicitly specified at subscription time or implicitly specified. It specifies the tolerance of an ATM network for incoming traffic that exceeds the PCR value. CDVT is not the same as CDV, which is a QoS parameter.

compliant connection if the network operator finds them acceptable. According to the Conformance Definition, the agreed QoS objectives should be met for at least the number of cells equal to the conforming cells in a compliant connection. On the other hand, if a connection is determined to be non-compliant by the network, the agreed QoS objectives need not be respected [UNI3.1].

Once the connection is set up, the two hosts can commence communication. For most networked applications, a duplex communication is carried out by the two hosts (ATM cells are sent and received by both hosts) and one or more virtual circuits are set up in each direction. The intermediate nodes perform switching functions on the ATM cells according to their VPI/VCI values. The traffic is monitored and controlled at the user access by the Usage Parameter Control (UPC) function according to the Conformance Definition specified as part of the traffic contract. UPC protects network resources from malicious as well as unintentional misbehaviour which can affect the QoS of other already established connections.

In order to achieve high-speed operation, error recovery of user data is not performed by the network and is left to the end hosts. Since all the ATM cells in a communication session follow the same pre-configured virtual circuit, the cell sequence integrity is preserved within virtual path connections (VPCs) and virtual channel connections (VCCs) [I.150] (i.e. the order in which cells arrive at the destination is the same as the order in which cells are transmitted).

A.3 ATM Protocol Architecture

The B-ISDN protocol reference model is a layered architecture and is described in ITU-T Recommendation I.321 by the following diagram. It is composed of three planes [I.321]:

- The user plane, which has a layered structure, provides for user information flow transfer;
- The control plane, which has a layered structure, performs the call and connection control functions. It deals with the signalling necessary to set up, supervise and release calls and connections;
- The management plane, which provides two types of functions. Layer Management is concerned with the management of protocol entities in each of the various protocol layers. Plane Management has no layered structure and performs management functions related to system as a whole and provides co-ordination between all the planes.

This thesis is mainly focused on the user plane, which is divided into three layers: the physical layer, the ATM layer and the ATM adaptation layer. Each layer utilizes the services provided by the layer below and provides services to the layer above through Service Access Points (SAP).

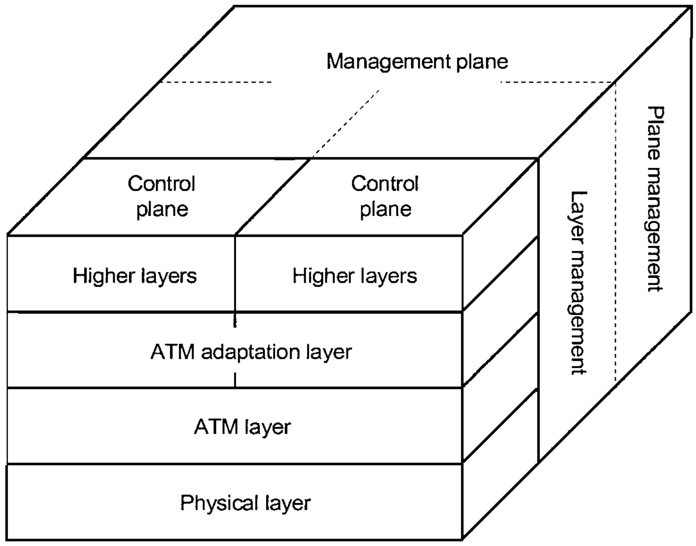


Figure 8-4 B-ISDN protocol reference model [I.321]

A.3.1 Physical Layer

The physical layer is split into two sub-layers:

- The Physical Medium (PM) sub-layer includes only physical medium dependant functions such as electrical or optical interfacing, bit transmission, bit timing and line coding.

- The Transmission Convergence (TC) sub-layer performs all functions required to transform a flow of cells into a flow of data units (e.g. bits) which can be transmitted and received over a physical medium. These functions include cell rate decoupling, HEC generation and verification, cell delineation, transmission frame adaptation and transmission frame generation and recovery.

The service data units (SDUs) between the physical layers and ATM layer above are valid ATM cells.

A.3.2 ATM Layer

The ATM layer is independent of the physical medium and provides cell transfer for all services and carries out cell multiplexing and demultiplexing, cell VPI/VCI translation and cell header generation / extraction. The service data units (SDUs) that are passed between the ATM layer and the AAL above are 48-octet payloads.

A.3.3 ATM Adaptation Layer (AAL)

The ATM adaptation layer (AAL) enhances the service provided by the ATM layer to support functions required by the next higher layer. The AAL only exists in end systems and is responsible for the mapping the service-specific higher layer protocol to the service-independent ATM protocol. One of the key functions of the AAL is to perform error control mechanisms against bit errors as well as lost or mis-insert cells. The AAL is divided into the Convergence Sub-layer and the Segmentation and Reassembly Sub-layer.

The Convergence Sub-layer (CS) is further divided into Service Specific (SS) and Common Part (CP) components and is responsible for handling cell delay variation, end-to-end synchronization as well as lost and mis-inserted cells. Its functions is dependant upon the higher layer requirements and may be empty in some cases.

The Segmentation and Reassembly Sub-layer (SAR) performs the segmentation of protocol data units (PDUs) into a size suitable for the information field of an ATM cell and reassemble the contents of ATM cell information fields into higher layer information.

Three types of AAL are relevant to this thesis report: AAL0, AAL1 and AAL5. They will be described in more details from A.4.1 to A.4.3. The other types of AAL are omitted as they are beyond the scope of this thesis.

A.4 Service Classification for the AAL

Since the CS functions of the AAL depend on the type of services supported in the higher layers, a classification of services into application classes is defined in order to minimize the number of AAL protocols. Four AAL service classes are defined based on the following parameters:

- Timing relation between source and destination (required or not required);
- Bit rate (constant or variable);

- Connection mode (connection-oriented or connectionless)

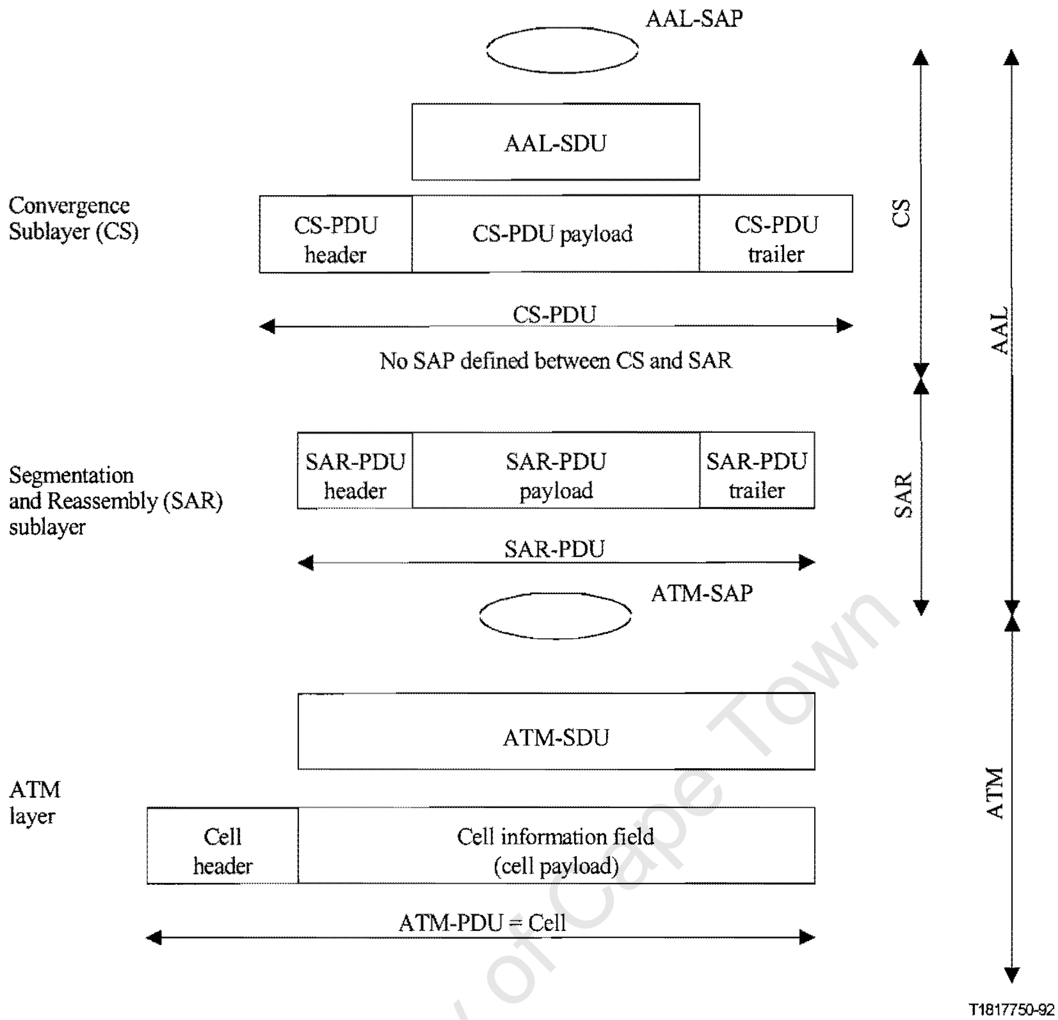
For example, depending on the type of encoding used, video traffic can either be constant bit rate (CBR) or variable bit rate (VBR) and can be classified as class A and B respectively.

	Service Class			
Attribute	Class A	Class B	Class C	Class D
Timing relation	Required		Not required	
Bit Rate	Constant	Variable		
Connection mode	Connection-oriented			Connectionless
Examples	DS1, E1 Circuit Emulation	Variable bit rate video and audio	Connection-oriented data transfer	Connectionless data transfer

Table 8-1 ATM/B-ISDN Service Classes [MCD95]

A.4.1 AAL 1

AAL 1 was designed to provide class A service, which is to support CBR applications. The AAL 1 SDU size can be one bit long (for circuit emulation) or one octet long (for video and voice transport). At the sending ATM endpoint, the AAL 1 service must accept data continuously at a constant rate. The payload bits / octets from the AAL user are assembled into the SAR-PDU (one ATM cell). At the receiving ATM endpoint, the bits / octets of AAL user information in the cell payload are disassembled and the received bit stream are presented at a constant rate to the higher layers by AAL 1. The transportation of MPEG video information over AAL 1 is described in section 3.3.2. Figure 8-5 illustrates the data unit naming convention for AAL 1.



NOTE – ATM Adaptation Layer-Protocol Control Information (AAL-PCI) consists of the SAR-PDU header, CS-PDU header, CS-PDU trailer, and SAR-PDU trailer.

Figure 8-5 Data unit naming conventions for AAL type 1 [I.363.1]

A.4.2 AAL 5

AAL 5 can be used to provide both class A and class B services. It was designed to be a simple and efficient AAL type that carries relatively low transmission and processing overhead. Because a Service Specific Convergence Sublayer (SSCS) has been defined in AAL 5 to support ATM signalling protocols, it has become widely implemented in all ATM switches and end-stations that implement SVCs. The transportation of MPEG video over AAL 5 is described in section 3.3.3. The data unit naming convention for AAL 5 is illustrated in Figure 8-6.

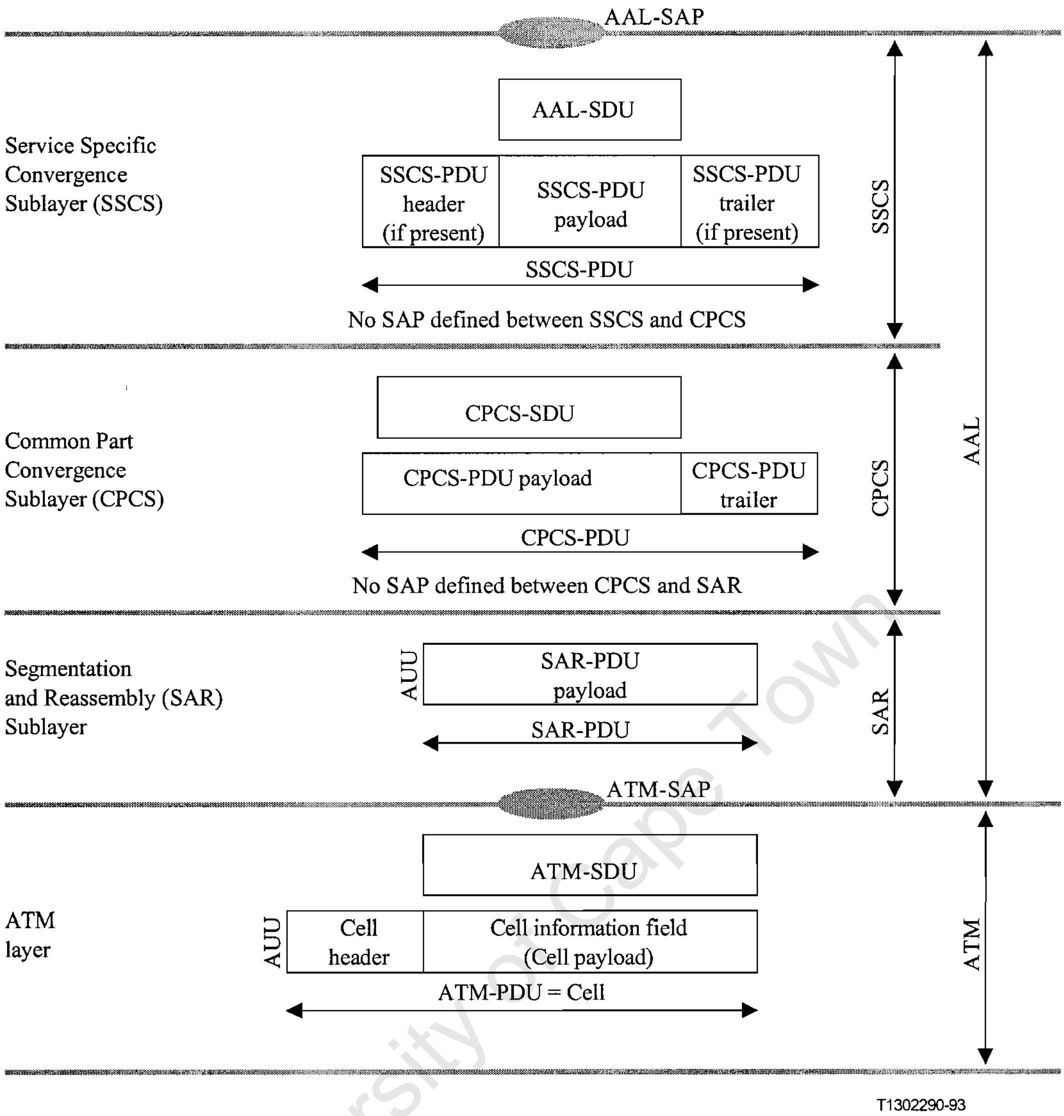


Figure 8-6 Data unit naming conventions for AAL type 5 [I.363.5]

A.4.3 AAL 0

AAL 0 is also known as a null or a raw AAL. There is no processing associated with this type of AAL and user data is directly mapped into the 48-octet payload of an ATM cell. The relevance of AAL 0 in this thesis is demonstrated in Chapters 4 and 5.

A.5 ATM Layer service categories

Another service model that is based on the ATM layer has been proposed in [TM4.0]. This new model is a parameter-based approach and classifies services for all applications into five ATM layer service categories. Each service category can be specified by a particular set of traffic and

QoS parameters. Applications supported within a service category differ only by the values of the corresponding set of traffic and QoS parameters. While [I.356] only defines QoS classes, individual QoS parameters can be specified in [TM4.0][34].

			CBR	rt-VBR	nrt-VBR	UBR	ABR
Attributes	Traffic Parameter	PCR, CDVT [1&2]	specified			specified	Specified
		SCR, MBS, CDVT	n/a	specified		n/a	
		MCR[1]	n/a			n/a	Specified
	QoS Parameter	Peak-to-peak CDV	specified		Unspecified		
		Max CTD	specified		Unspecified		
		CLR[1]	specified			unspecified	
	Other	Feedback	unspecified				Specified

[1] These parameters are either explicitly or implicitly specified for PVCs or SVCs.

[2] CDVT refers to the Cell Delay Variation Tolerance and is not signalled. In general CDVT need not have a unique value for a connection. Different values may apply at each interface along the path of a connection.

Table 8-2 ATM Layer Service Categories and Attributes [TM4.0]

Constant Bit Rate (CBR) service class provides a fix bandwidth transmission circuit that is equal to PCR. Real-time streaming applications such as video on demand can use this service category. Variable bit rate (VBR) service class is intended for bursty traffic that may be real-time or non-real-time. The real-time Variable Bit Rate (rt-VBR) service category is intended for real-time applications, i.e., those requiring tightly constrained delay and delay variation. rt-VBR connections are characterized in terms of a Peak Cell Rate (PCR), Sustainable Cell Rate(SCR) and Maximum Burst Size(MBS). Real-time VBR service expects traffic rate to varies with time and supports statistical multiplexing. Video and voice applications would also fit into this category. The differences between supporting video applications as CBR and VBR service classes are presented in section 3.3.1.

The other service categories are not significant to this thesis but a brief description is included for completeness sake. The nrt-VBR service category is intended for non-real-time applications which have bursty traffic characteristics and no required delay bound. Unspecified bit rate (UBR) service is designed to provide a best effort service with the network providing no guarantees on bandwidth, delay or cell loss. Available bit rate (ABR) service provides a minimum bandwidth guarantee as well as a flow control mechanism.

[34] QoS classes are retained in [TM4.0] for backward compatibility.

Appendix B

Overview of Digital Video and Video Compression

This section serves to be a brief introduction to digital video and digital video compression in general. A few digital video compression standards will also be introduced.

B.1 Digital Video

Video is a stream of data composed of discrete frames, usually including both sound and pictures. Analog video is a naturally continuous signal with breath and depth. Each image frame in an analog video has virtually infinite resolution, (i.e. is infinitely magnifiable without any loss of quality.) On the other hand, digital video is not continuous and colour values with finite precision (represented by zeros and ones) are assigned to fixed points called pixels. Each picture frame is a collection of pixels; therefore digital video only has a finite resolution [OZE95]. Since video streams being transmitted over ATM are in digital format, video in analog format needs to be translated (digitized) into digital format before transmission. Digitizing video involves taking an analog video signal and converting it to a digital video stream using a video capture board as shown in Figure B-1. The resulting picture quality is mainly dependent upon three factors:

- Resolution – the horizontal and vertical dimensions of the video pictures;
- Colour depth – the number of bits that are used to express colour;
- Frame rate – the number of frames that are displayed per second.

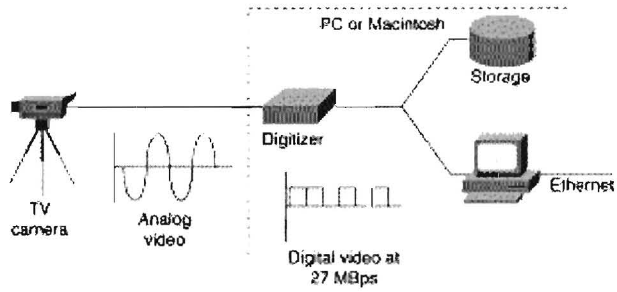


Figure B-1 A video capturing and digitizing example [CIS99]

Instead of the RGB (red, green, blue) colour component system, simplified luminance and chrominance information for each colour, represented by the Y Cr Cb colour system, can equivalently be used to represent digital video information. When bit rate is not a concern, digital video can be used for transmission immediately after the digitization process. However, the bit rate requirement of video in various formats is too high for network bandwidth to be utilize efficiently. The solution is to compress (or encode) the digital video information before transmission in order to reduce the bandwidth requirements for video streams to levels acceptable for existing networks.

Table B-1 shows the bandwidth requirements for various full-screen display standards used by computer monitors (from CGA to SVGA), and for analog video signals used by televisions and VCRs in the U.S., Europe, and France (NTSC, PAL and SECAM respectively) [HOD95]. It can be seen that the image size on the screen, resolution, image (or colour) depth, and frame rate all affect the bit rate (Mbps) requirement, which can be calculated by the following equation:

$$BitRate(bps) = {}^{Bits}\!/_{Pixel} \times {}^{Pixel}\!/_{Line} \times {}^{Line}\!/_{Frame} \times {}^{Frame}\!/_{Second}$$

Equation B-1 Relationship between video attributes and bit rate requirements

Format	Pixels per line	Line per frame	Pixel per frame	Bits per pixel	Bits per frame	Frames per second	Megabits per second (Mbps)
CGA	640	200	128,000	4	512,000	60	30.7
EGA	640	350	224,000	6	1,344,000	60	80.6
VGA	640	480	307,200	6	1,843,200	60	110.6
SVGA	800	600	480,000	8	3,840,000	72	276.5
NTSC	600	485	291,000	24	6,984,000	30	209.5
PAL	580	575	333,500	24	8,004,000	50	400.2
SECAM	580	575	333,500	24	8,004,000	50	400.2

Table B-1 Bandwidth Requirements without Compression [HOD95]

B.2 Video Compression

Video compression is a process where a collection of algorithms and techniques replace the original pixel-related information with more compact mathematical descriptions [OZE95]. Decompression is the reverse process of decoding the mathematical descriptions back to pixels for ultimate display. One of the most fundamental concepts in compression techniques is the difference between lossless and lossy compression. Lossless compression techniques create compressed information that decompresses into exactly the same material as the original. This type of compression is widely used for computer data files where any loss of information is generally considered unacceptable. The compression ratio achieved by lossless techniques is typically around 2:1 to 4:1, which is not enough for most digital video streams. The alternate solution is to use lossy compression techniques, which achieves compression ratio ranging from 7:1 to as high as 100:1 and beyond due to the lossy nature. These compression techniques are primarily used on digital videos and still images, and create compressed files that decompress into frames and images that look similar to the original. In general, the change in quality, resulting from the losses incurred during lossy compression, is hardly noticeable by the human eye. For example, the chrominance signals of a video frame can be represented with a lower resolution than the luminance signals (resulting in a irreversible loss of information) without significantly affecting the visual quality presented to the human viewers. This is acceptable because the human visual system (HVS) is less sensitive to chrominance (colour) information than it is to luminance (brightness) information.

Video compression includes compression within a frame (also known as spatial or intraframe compression) and compression between frames (temporal or interframe compression). Intraframe compression removes the significant amount of spatial redundancy in a still image or a single frame within a video sequence using techniques such as differential pulse code modulation (DPCM) and discrete cosine transform (DCT). The resulting coefficients representing the original image are encoded into variable-length codes (VLCs) in an entropy coding scheme such as Huffman encoding. Other techniques such as Vector Quantization and Wavelets can also be used. On the other hand, interframe compression removes the temporal redundancy among similar successive video frames in a video sequence by assigning a 'reference' frame on a regular basis and represent subsequent frames by the difference between the current frame and the 'reference' frame (differential prediction). Further interframe compression can be achieved by motion prediction when changes in the current frame from the 'reference' frame are due to movement in the scene that can be approximated as a linear motion.

Video Compression dramatically reduces bandwidth requirements as shown in Table B-2. A brief introduction of these compression standards follows in the next few sub-sections.

Standard / Format	Approximate Range of Data Rate	Compression*
Motion JPEG	10 – 20 Mbps	7 – 27 times
MPEG-1	1.2 – 2.0 Mbps$^{\alpha}$	100 times
H.261	64 Kbps – 2 Mbps	24 times
DVI (digital video interactive)	1.2 – 1.5 Mbps	160 times
CD-I (compact disc interactive)	1.2 – 1.5 Mbps	100 times
MPEG-2	4 – 80 Mbps$^{\delta}$	30 – 100 times

Mbps – megabits per second
Kbps – kilobits per second
* Compared to broadcast quality
$^{\alpha}$ Baseline standard; other rates are also possible
$^{\delta}$ Image quality becomes asymptotic above 8 to 10 Mbps

Table B-2 Bandwidth requirements for different compression scheme [MIN97]

B.2.1 Motion JPEG

The Motion JPEG standard is an extension to the joint ITU and ISO JPEG standard for still images. It uses only spatial compression and compresses individual video frames independently as a still image. Because motion JPEG does not reduce the redundancies that exist between frames, it achieves less compression than other techniques (such as MPEG) that removes both intra-frame and inter-frame redundancies. In addition, audio is not integrated into the compression method because JPEG was not originally designed for full-motion video. On the other hand, the lack of inter-frame dependencies in motion JPEG allows faster random access to any frame of the video material than schemes that provides inter-frame compression.

B.2.2 H.261

The H.261 protocol suite developed by ITU is an encoding standard for two way audio and video transmission and is often used in videoconference device for connecting to 64 kbps or 128 kbps ISDN lines. It defines two fixed resolutions 352x288 and 176x144 and employs motion compensated prediction among video frames. H.261 encoders using a feedback mechanism from the output buffer can produce a near constant bit rate output.

B.2.3 Digital Video Interactive (DVI)

DVI is an Intel de facto compression standard. It increases the storage capacity of CD-ROM disc and has brought video to the PC. For example, DVI encoding has enabled the storage of a one hour full-screen, full-motion video with audio or four hours of ¼ screen, full-motion video with audio on a CD-ROM disc.

B.2.4 CD-I

CD-I video compression also provides full-screen, full-motion video. Typically, the throughput achieved by a CD-I player is approximately 1.4Mbps, of which 1.2Mbps must be used for video and 0.2 Mbps can be used for the associated audio.

B.2.5 MPEG

The MPEG standard make use of both temporal and spatial compression techniques and is a widely used format for coding digital video and associated audio information. This section describes the MPEG video compression standard in more details because it is used as an example to illustrate the effects of ATM network impairment has on compressed digital video streams in a Video-on-Demand (VoD) application.

The first of the MPEG standards is known as ISO11172 or MPEG-1. The second MPEG standard is known as ISO13818 or MPEG-2, which builds on the completed MPEG-1 video standard as a compatible extension. While MPEG-1 supports the coding of video and associated audio into a single data stream at a bit rate of about 1.5 Mbps, MPEG-2 extends the functions provided by MPEG-1 in terms of providing a wide range of resolutions, bit rates and encoding options. Some of the differences between MPEG-1 and MPEG-2 are listed below:

- MPEG-2 supports the 4:2:2 and 4:4:4 chrominance formats on top of the 4:2:0 format supported in MPEG-1. A 4:2:0 macroblock consists of 6 blocks (4Y 1Cb 1Cr). A 4:2:2 macroblock consists of 8 blocks (4Y 2Cb 2Cr) and a 4:4:4 macroblock consists of 12 blocks (4Y 4Cb 4Cr) [M13818];

- MPEG-2 supports video of higher resolution and frame rate than MPEG-1 (Table B-3);

Description	Approx. Resolution	Approx. bit rate
Home entertainment video – MPEG-1 quality	352x240, 30 fps	1.5 Mbps
Typical digital television	720x576, 30 fps	10—15 Mbps
Highest supported resolution	1920x1152, 60 fps	60—80 Mbps

Table B-3 Resolution and bit rate for MPEG-2

- Slices do not have to start and end on the same horizontal row of macroblocks in MPEG-1 but in MPEG-2, slices always start and end on the same horizontal row of macroblocks;

- In MPEG-2, different profiles and levels are defined for different types of applications (Table B-4 and Table B-5). Five 'profiles' are defined as subsets of the entire bitstream syntax and within these profiles, four 'levels' are defined on the constraints imposed on parameters in the bitstream[35].

Profile	Features
Simple	4:2:0 sampling, I/P pictures only, no scalable coding
Main	As above, plus B pictures
SNR	As above, plus SNR scalability
Spatial	As above, plus spatial scalability
High	As above, plus 4:2:2 sampling

Table B-4 MPEG-2 Profiles

Level	Maximum Resolution
Low	352 x 288 luminance samples, 30 frames per seconds
Main	720 x 576 luminance samples, 30 frames per seconds
High 1440	1,440 x 1,152 luminance samples, 60 frames per seconds
High	1,920 x 1,152 luminance samples, 60 frames per seconds

Table B-5 MPEG-2 Levels

- MPEG-2 defines four scalable coding modes: spatial scalability, Signal-to-Noise ratio (SNR) scalability, temporal scalability and data partitioning.

[35] Note that not all profile-level combinations are defined in the MPEG-2 specification

Figure B-2 shows the functional block diagram of the MPEG-2 encoding process. At discrete time intervals the video encoder receives uncoded digitized pictures called video presentation units (VPUs) and the audio encoder receives uncoded digitized blocks of audio samples called audio presentation units (APUs). The times of arrival of VPUs are not necessarily aligned with those of APUs. The video and audio encoders produce coded pictures called video access units (VAUs) and coded autio called audio access units (AAUs).

A collection of VAUs and AAUs are combined and packetized separately to form video and audio packetized elementary streams (PES) respectively. They are then converted into either Program Stream (PS) or Transport Stream (TS). The Program Stream is analogous to MPEG-1 streams and may be either fixed or variable rate. It is tailored for communication or storing one program of coded data in environments where there is a low probability of bit errors or data losses. The Transport Stream consists of TS packets of 188 bytes long with 4 bytes of header information and is intended for transportation over media where bit value errors or loss of packets are more likely.

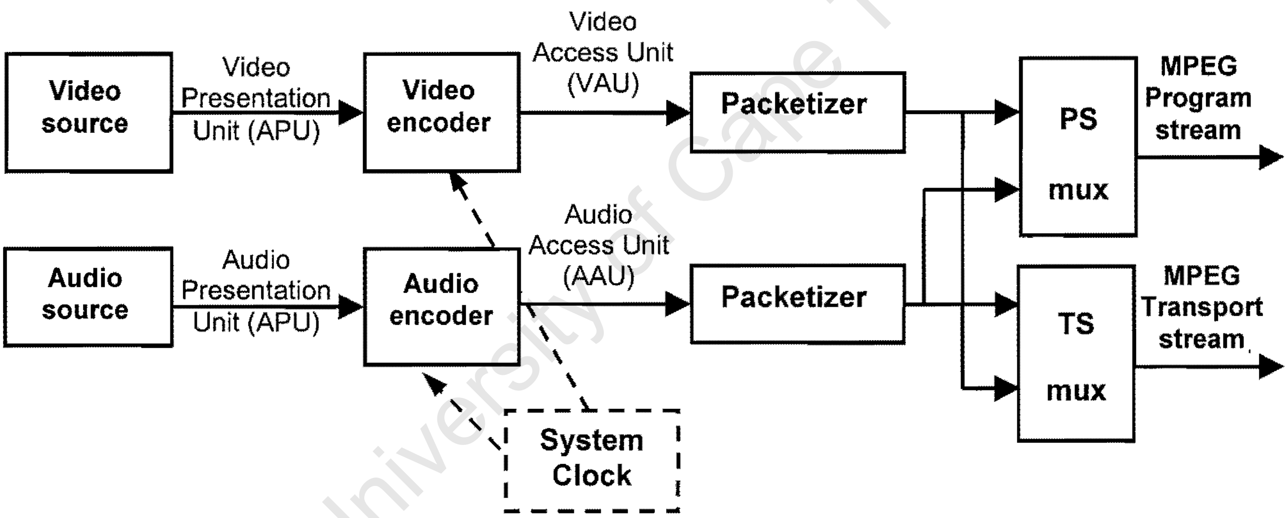


Figure B-2 Functional block diagram of the MPEG-2 encoding process

An important aspect of MPEG is synchronization. Video and audio signals must be stored, retrieved, transmitted, received and decoded with strict timing relationship in a networked video application. Video and audio multiplexing and synchronization features are provided by the systems layer. Synchronization is achieved through the use of time stamps and reference clock signals. Two time stamps: presentation time stamps (PTS) and decoding time stamp (DTS) are used to inform the decoder when to display decoded information to the end user and when to decode information in the decoder buffers respectively. The synchronization between the encoder and the decoder clock is accomplished through the use of System Time Clock (STC) of the same frequency with the help of System Clock Reference (SCR) in Program Streams and Program Clock Reference (PCR) in Transport Streams.

Appendix C

ATM-on-Linux and the Virtual ATM Switch

This section presents an informal guide to the main steps taken during this study to learn about ATM-on-Linux and to construct the virtual ATM switch in the NetIE architecture. It is not a comprehensive step-by-step 'fool-proof' instruction manual by any means but intends to serve as guidelines to make the process less painful. For basic commands and new concepts, refer to information presented in web sites, READMEs, manual pages etc. Note that a lot of this information (e.g. IP address, equipment etc) is specific to the Communications Research Group laboratory environment and to RedHat Linux, adjust information accordingly if necessary.

The Construction of the Virtual ATM Switch involves the following steps:

1. Install RedHat Linux on the Physical Machine intended to be used as the switch;
2. Upgrade the Linux kernel;
3. Include ATM on Linux support;
4. Load the Virtual ATM Switch driver;

Before we get started, let me state that we will be running the following:

- Linux Kernel version 2.1.126,
- Linux-ATM version 0.52, and
- Virtual ATM Switch for ATM on Linux (release 0.5).

Before you start, make sure you are aware of the following:

1. **IP number** – contact system administrator to obtain one (e.g. for ethernet connectivity, 10.128.0.xx, where xx is the last octet allocated to you)
2. **RPM packages** – this is for RedHat Linux only. To install a RPM package, use 'rpm -i package-xx.yy.zz.i386.rpm'.
3. **RPM requirement** – for kernel 2.1.126, RPM 2.2.7 or later is required. To check the version of RPM currently installed, use 'rpm –version'.
4. **The 'node number'** – this is the network point to which your UTP cable connects. You also need to know the difference between the ethernet hub and the ATM switch.
5. **Adapter driver** – If you are using the ForeRunner LE155 ATM adapter, be aware that you will be using a driver called 'nicstar'.
6. **Version Matching** – as with many other things in Linux, packages work with some kernel distributions but not with others. Make sure the interdependencies and pre-requisites among all the components are checked.

Detailed description of each step follows.

C.1 Installation of Linux

The basic Linux operating system can either be installed from a FTP site or a distribution CD. Refer to the 'Linux Installation and Getting Started' book (LIGS) [WEL98] or the University of Cape Town LEG (Linux Enthusiasts Group) site http://www.leg.uct.ac.za/starting_out.html for information to get started.

C.1.1 Components to install

When prompted to choose what components to install, select 'choose individual packages' and choose the following components amongst other things (they can also be installed later):

- kernel-ibcs
- kernel-source
- ncpfs
- autofs
- procinfo
- ipxutils

C.1.2 Installation Hints

For RedHat Linux, all the steps during installation can be found at:
http://www.leg.uct.ac.za/mirrors/RH_Manual/doc000.htm.

A complete log of the installation process can be found in /tmp/install.log.

To change which services automatically start on reboot, you can run /usr/sbin/ntsysv or /sbin/chkconfig after the installation.

Make boot disk. If you boot your system with the boot diskette (instead of LILO), make sure you create a new boot diskette if you make any changes to your kernel. For more info, see the mkbootdisk man page.

Install LILO in /dev/hda (Master Boot Record) and not /dev/hda1 (First sector of boot partition).

Listing and explanations of packages installed during installation can be found in Appendix C of the LEG documentation.

In terms of Linux documentation, here's what you should look for (optional, but may prove to be useful later on):

- A brief history of Linux – Many aspects of Linux are the way they are because of historical precedent. There is also a Linux culture that, again, is based to a great deal on past history. A bit of knowledge about the history of Linux will serve you well, particularly as you interact with more experienced Linux users on the Internet.

- An explanation of how Linux works – While it's not necessary to delve into the most arcane aspects of the Linux kernel, it's a good idea to know something about how Linux is put together. This is particularly important if you've been working with other operating systems; some of the assumptions you hold about how computers work may not transfer from that operating system to Linux. A few paragraphs that discuss how Linux works (and particularly how it differs from the operating system you're used to), can be invaluable in getting off to a good start with your Linux system

- An introductory command overview (with examples) -- This is probably the most important thing to look for in Linux documentation. The design philosophy behind Linux is that it's better to use many small commands connected together in different ways than it is to have a few large (and complex) commands that do the whole job themselves. Without some examples that illustrate the Linux approach to doing things, you will find yourself intimidated by the sheer number of commands available on a Linux system.

At this point you need to copy the following files from the accompanying CD into your home directory:

modutils-2.1.121.tar.gz
ppp-2.3.7-1.i386.rpm
linux-2.1.126.tar.gz
atm-0.52_tar.gz
vswitch-0.5.tar.gz
vswitch_qport-0.5.tar.gz

C.2 Upgrade the Linux Kernel

The Unix kernel acts as a mediator for your programs and your hardware. First, it does (or arranges for) the memory management for all of the running programs (processes), and makes sure that they all get a fair (or unfair, if you please) share of the processor's cycles. In addition, it provides a nice, fairly portable interface for programs to talk to your hardware.

Information on upgrade the Linux kernel can be found in the LIGS book (p.217) and at http://www.redhat.com/support/docs/rhl/kernel-upgrade/kernel-upgrade-2.html

C.2.1 Preparing for the Upgrade

1) Find out what version of the kernel RPM's you have installed by the following command:

rpm –q kernel kernel-headers kernel-ibcs kernel-pcmcia-cs kernel-source

Kernel-2.0.36-0.7
Kernel-headers-2.0.36-0.7
Package kernel-ibcs is not installed
Kernel-pcmcia-cs-2.0.36-0.7
Package kernel-source is not installed

and

rpm -q mkinitrd SysVinit initscripts

mkinitrd-1.8-3
SysVinit-2.74-5
initscripts-3.78-1

2) If 'kernel-ibcs' and 'kernel-source' have not been installed already, install from the RHCD by:

mount /mnt/cdrom
cd /mnt/cdrom/Redhat/RPMs
rpm –ivh kernel-ibcs kernel-source

check that you have installed them by repeating the above.

3) Make emergency boot floppy (as root) – in case the upgrade fails, this saves the trouble of having to start from the beginning!

Mkbootdisk - -device /dev/fd0 2.0.36-0.7

4) BEFORE upgrading the kernel, it is VERY important to check out the following file: *<u>/usr/src/linux/Documentation/Changes</u>*. It tells you the Minimal Software Revision Requirement for the upgrade. When I installed RedHat 5.2 from the CD (kernel version 2.0.36), a few of those software packages are not up to date. In my case, the following updates are necessary:

   a) Kernel modules upgraded from 2.1.85 to 2.1.121 (modutils-2.1.121.tar.gz).

      i) First copy it into /usr/src
      ii) Untar/gzip it by the command 'gzip –cd modutils-2.1.121.tar.gz | tar xsfv –` '
      iii) You should now get a new directory with the same name, change into this new directory and either check out the README file or do the following:
         (1) ./configure (./configure –help)
         (2) make install
      iv) Check by insmod -V

   b) ppp is installed (ppp-2.3.7-1.i386.rpm). I am not sure if this is absolutely necessary, but no harm doing it. (check by 'pppd –v')

   c) procinfo version 14 was installed (not upgraded) from the RedHat CD. The path is /mnt/cdrom/Redhat/RPMS/ and the file is procinfo-14.2.i386.rpm (check by 'proinfo -v')

d) autofs version 3.1.1 was installed (not upgraded) from the RHCD. The path is /mnt/cdrom/Redhat/RPMS/ (check by 'automount –version')

e) ncpfs version 2.2.0 was installed (not upgraded) from the RHCD. This package (ncpfs-2.2.0-1.i386.rpm) depends on ipxutils, which also needs to be installed (check for ncpfs by 'ncpmount –V')

f) PCI Utilities (lspci) was omitted, but it doesn't seem to affect the new kernel.

C.2.2 The Upgrade Process

Upgrade to kernel version 2.1.126

1) Obtain linux-2.1.126.tar.gz and store it under /usr/src

2) gzip -cd *filename*.tar.gz | tar xfv –

3) Make these symbolic links as follows:
   cd /usr/include
   rm -rf asm linux scsi
   ln -s /usr/src/linux/include/asm-i386 asm
   ln -s /usr/src/linux/include/linux linux
   ln -s /usr/src/linux/include/scsi scsi

4) cd /usr/src/linux

5) make mrproper

6) Ensure you are in /usr/src/linux. We must now edit the kernel:

   make menuconfig

   a) remove SCSI support and remove sound support
   b) add PnP support and (in my case) NE2000 network device support
   c) add MSDOS fs support (in order to mount dos stiffies)

7) make dep; make clean

8) make bzImage – this takes 10 to 20 minutes depending on the specification of you computer and should finish with a message like this: Boot Sector 512 bytes, Setup is 1288 bytes, system is 495kB, make[1]: Leaving directory '/usr/src/linux/arch/i386/boot'

9) make modules

10) make modules_install

11) The build has created a kernel image which we must move to the /boot directory by:
mv /usr/src/linux/arch/i386/boot/bzImage /boot/bzImage

12) Edit /etc/lilo.conf and add the following:

```
image=/boot/bzImage
label = atmlinux
root=/dev/hda1
read-only
```

13) Rerun lilo by:
/etc/lilo.conf

14) REBOOT and the new kernel is there

C.3 Add ATM on Linux Support

In order to install ATM support on Linux, you need the following components (refer to /usr/src/atm/USAGE for details):

- The package itself → atm-0.52.tar.gz
- Linux kernel version 2.1.126 (which you are now running hopefully)
- Perl, version 4 or 5 (use *perl - -version* to check, you should have 5.004)
- If you need memory debugging (which you probably do), get MPR version 1.6

Here we go ☺

1) Obtain 'atm-0_52_tar.gz' and put it in /usr/src (you might have already guessed that this is the place where we put all the source files)

2) Unpack the packages in /usr/src by:
   tar xvfz atm-0_52_tar.gz

3) Patch the Linux Kernel with the Linux-ATM distribution as follows:

   /usr/src/linux/patch –s –p1 < /usr/src/atm/atm.patch

or if this doesn't work, do the following.

   cd /usr/src/linux
   patch –s -p1 <../atm/atm.patch

4) cd /usr/src/linux

5) make menuconfig

   a) Under 'Code maturity level options', enable 'Prompt for development and/or incomplete code/drivers'. And BINGO!! All the ATM stuff appears!

   b) Under 'Networking Options', select:
      i) Asynchronous Transfer Mode (ATM, Experimental)
      ii) Classical IP over ATM
      iii) LAN Emulation (LANE) support (optional)

   c) Under 'Network device support', select:
      i) IDT 77201 (NICSTAR)

6) After doing make menuconfig, you need to rebuild the kernel again

   a) cd /usr/src/linux
   b) make dep; make clean
   c) make bzImage
   d) make modules
   e) make modules_install
   f) The build has created a kernel image which we must move to the /boot directory
      mv /usr/src/linux/arch/i386/boot/bzImage /boot/bzImage

7) Re-run lilo and then reboot

8) If you want to install MPR, you MUST do so BEFORE configuring and building the ATM tools. Copy the 'mpr-1.9.tar.gz' file to /usr/src and do the following:

    a) cd /usr/src/mpr-1.9
    b) gzip –cd mpr-1.9.tar.gz | tar xfv -
    c) ./configure x86-linux
    d) make
    e) make install
    f) install –c –m 0644 libmpr.a /usr/lib
    g) install –c –m 0755 mpr mprcc mprhi mprlk mprsz /usr/local/bin
    h) install –c –m 0755 mprfl mprnm mprpc mprdem mprmap /usr/local/bin
    i) refer to the DOC file for usage instruction.

9) Now the ATM tools are ready to be built and configured

    a) cd /usr/src/atm
    b) make depend
    c) make
    d) make install

   * make install will install executables in the directory /usr/local/bin and /usr/local/sbin, respectively. Libraries and header files are installed in /usr/lib and /usr/include, respectively. Man pages are installed in /usr/local/man.

10) Now you can configure your system to use the ATM adapter and after you have done that, you can switch over to ATM only.

    a) Goto /etc/sysconfig
    b) To play safe, cp network network.eth.old
    c) Edit the file 'network' to look as follows:

```
NETWORKING=yes
FORWARD_IPV4=false
HOSTNAME=testsw.uct.atm
DOMAINNAME=uct.atm
GATEWAY=10.192.255.254
GATEWAYDEV=atm0
#NISDOMAIN=crg
```

You may or may not have the last line, but if you do have it, make sure there is a # in front of it. You will see why in a minute.

d) Goto /etc/sysconfig/network-scripts and create the file 'ifcfg-atm0' to look as follows:

```
DEVICE=atm0
IPADDR=10.192.0.20
NETMASK=255.255.0.0
NETWORK=10.192.0.0
BROADCAST=10.192.255.255
ONBOOT=yes
```

e) Bring down the old ethernet interfact with: ifdown eth0

f) Now we need to reboot the machine with ATM connection, i.e. unplug the cable from the ethernet card and connect it to the ATM card (WAIT!! Don't do it yet, there are a few things you have to do before you can do so)

   i) Goto /etc/sysconfig and edit the file 'network'. If none of the lines says NISDOMAIN=crg, then quit, else add a # in front of that line.

   ii) Goto /etc and edit the file 'fstab'. If any of the lines begins with 'crglnx', add a # in front of it.

   The above 2 steps disable NIS/NFS (if you don't know them, ignore it for now). Since we are about to boot up the machine with only ATM connectivity and the ATM adapter is not running yet, the machine will end up looking for things on the network that it will never find if you don't STOP it from trying to find them.

   iii) Now shutdown your machine

   iv) Go to the Ethernet hub and unplug the cable which comes from the correct network node number (still remember which network point you are on?) [**N.B. make sure you are unplugging your own connection, if you unplug the wrong cable, you will see an angry face coming towards you at high as you walk out of that room ☹.] Then, plug the cable into the ATM workgroup switch. **If you are in doubt, ASK!!!

   v) Now plug the network cable into the ATM adapter and start the machine.

g) (optional) Goto /usr/local/sbin and run these daemons:

./atmsigd –b
./ilmid –b → …hmm.. this one talks to the ATM switch, returns an address and freezes, this is normal, just Ctrl-C it. I think if you see an length string of numbers and chars, you can go straight to vii).
./atmarpd –b
./atmarp-c
./ifconfig atm0 10.192.0.20 up
./atmarp –s 10.192.255.254 47 . arpsrv

Bring the atm interface down and then up: ifdown atm0 ifup atm0

11) Now we need to make the kernel boot up correctly. In Linux, services are all scripts which get started by the symbolic links in /etc/rc.d/rc?.d, where ? is your startup run level. For example, if ? is set to 5, X-window should start automatically (WAIT check that X-window starts up properly before you change the startup run level, otherwise the computer might boot up in a way that you can't see the display at all!!).

For example, to start the inet service at run level 3, the following link exists in /etc/rc.d/rc3.d:

S50inet → ../init.d/inet

What this means is that when the kernel is in run-level 3, the inet service is started because the kernel sees the symbolic link, goes to /etc/rc.d/init.d and runs the inet script.

12) Create a new symbolic link in /etc/rc.d/rc5.d and rc3.d:

cd /etc/rc.d/rc5.d
ln-s /etc/rc.d/init.d/atm-daemons S09network
cd /etc/rc.d/rc3.d
ln-s /etc/rc.d/init.d/atm-daemons S09network

13) Goto /etc/rc.d/init.d/and create the file 'atm-daemons' to look as follows:

#!/bin/sh
echo "ATM: Starting Daemons"
/usr/local/sbin/atmsigd –b

/usr/local/sbin/ilmid –b
/usr/local/sbin/atmarpd –b
/usr/local/sbin/atmarp –c atm0
ifdown atm0
ifup atm0
/usr/local/sbin/atmarp –s 10.192.255.254 47000580ffe1000000f21a015d0020481a015d00
arpsrv

*** the last line 47000.... is actually NOT a new line, it only appears on a new line because the whole line just wouldn't fit.

14) Change the 'atm-daemons' file to executable with: chmod a+x atm-daemons

15) (optional) Now we remove sendmail and smb so that we can unplug the ethernet card. We keep them here incase we might need it in the future.

mv S80sendmail sendmailS80
mv S91smb smbS91

16) Edit rc.local in /etc/rc.d/ and add the following at the end:

ifdown eth0

17) Reboot your machine and you now have ATM on Linux. If you don't believe me, you can always 'ping' around or browse a few web pages and hopefully you will be convinced.

C.4 ATM on Linux Basics

C.4.1 Useful Commands

Some useful commands for ATM on Linux are listed below:

1) ATM tools

   a) Executables in the directory /usr/local/bin and /usr/local/sbin
   b) Libraries and header files are installed in /usr/lib/and /usr/include
   c) Man pages are installed in /usr/local/man

2) Files in /proc/atm

   a) *arp* information specific to CLIP (table of currently active IP over ATM VCs).
   b) *devices* lists all active ATM devices and statistics
   c) *pvc* list all PVC sockets → VPI, VCI, AAL, traffic class and PCR for receive and transmit direction
   d) *svc* list all SVC sockets → VPI, VCI, SVC state and the address of the remote party. SPECIAL: SVC with interface number 999

3) atmdiag → in /usr/local/bin, queries various counters of the ATM device drivers

4) System files located in /usr/local/sbin

   a) atmaddr → Address Maintenance. It lists & maintains (appends) local ATM address.
      i) No flag – lists the local ATM address
      ii) -a – append the address at the end of the list (used for setting up local ATM addresses in <u>back-to-back configuration</u>)
      iii) -d – delete specified address from the list
      iv) -n – numeric address output only, no address to name translation
      v) -r – reset (clear) the local address list of the specified interface

   b) atmarp → Administer CLIP. It maintains ATMARP table of the ATMARP daemon.
      i) -a – list the current ATMARP table
      ii) -c – create a specified IP interface (perhaps more than one??)
      iii) -s – setup a PVC or create a SVC. Options such as QoS / temp are allowed. (It can be used to inform hosts of a defined PVC at the switch and such PVC will show up in /proc/atm/pvc)
      iv) -q – sets the default QOS for all VC in IP network
      v) -d – delete an ARP entry

   c) atmarpd → ATMARP daemon. It is configured from the command line with atmarp. The ATMARP table is written to file /var/run/atmarpd.table after every change.

   d) atmsigd → ATM signalling daemon. It implements the ATM UNI signalling protocol. The kernel sends requests to establish, accept, or close ATM SVCs to the signalling daemon, which then performs the dialog with the network.
      i) -m – mode of operation (user/network). For <u>back-to-back setup</u>, one host will run <u>network</u> mode and the other <u>user</u> mode.

ii) -q – sets QOS for the signalling VC other than the default UBR at link speed.
iii) /var/tmp/atmsigd.pid.status.version → default location of status dumps
iv) /var/tmp/atmsigd.pid.trace.version → default location of signalling trace dumps

e) ilmid → talks to the ATM switch and gets the address of it

5) Other useful files:

a) ifconfig → (under /sbin) configures a network interface.
b) atmtcp, aread, awrite, aping, br, bw etc
c) /var/run/atmarpd.table → contains the ATMARP table

C.4.2 Classical IP over ATM (CLIP)

1) test1 (002048EC32B) can ping test2 (0020480EC2D8) and vice versa. For example, when test1 (IP address 10.192.0.20) pings test2 (IP address 10.192.0.21), this is what happens:

On test1.
Atmsigd:UNI: Active open succeeded (CR 0x000002, ID 0xc373bc00, to 47xxx0020480EC2D800)

On test2.
Atmsigd:UNI: Passive open succeeded (CR 0x800800, ID 0xc3738600, from 47xxx0020480EC32B00)

2) To see that an SVC is established between test1 and test2, 'cat /proc/atm/svc', and you will see the VPI/VCI values for the connection.

3) To see all the connections, run '/usr/local/sbin/atmarp –a' or cat '/var/run/atmarpd.table. What happened above was that during *ping*, a SVC is automatically set up for the IP traffic between test1 and test2, as shown. The SVC disconnects itself when a timeout occurs or it can be disconnected by using 'atmarp –d [hostname]'.

4) To set up PVCs (as opposed to SVCs) for CLIP, not native ATM: (http://www.multikron.nist.gov/scalable/publications/ascii/Configure_Linux-ATM.txt)

a) Create a PVC entry on the LE155 switch.

i) Finding out what node number each of your host is connected to (hopefully you have done that by now) and the PORT number each node is connected to (i.e. the name of the socket on the LE155 switch, D3, A1, etc)
ii) Telnet into the switch (ask the ATM administrator for details).
iii) Manually set up a PVC entry (one way or two ways) between the two hosts
iv) Type '?' for a list of command in the current level
v) To create a new PVC you must do the following:

new [input port] [*vpi*] [*vci*] [output port] [*vpi*] [*vci*] [optional name]

E.g. to set up a 2-way PVC between port C1 and port D3 on the switch using vci 150:

new C1 0 150 D3 0 150 (this pvc goes from C1 to D3)
new D3 0 150 C1 0 150 (this pvc goes the other way)

Note that we normally use the same VPI/VCI numbers for both ways. Also note that you can also name a PVC to ease the pain of management.

b) In order to use host name rather than IP numbers, edit the file '/etc/hosts'. Change the IP number and domain for the current machine to the ATM IP number and domain (since that entry was created when ethernet was still being used). Add a new entry for the other machine according to the format of the other entries. If you don't do this, you can only use IP numbers to specify a machine.

c) A PVC is identified by three numbers, separated by a period. The numbering scheme is [interface.vpi.vci]. Interface is 0 for the first ATM adapter, 1 for the next, etc.

d) Tell both hosts that a PVC exists between them for IP traffic
i) On test1, issue the command:

atmarp –s test2 0.0.150

ii) On test2, issue the command:

atmarp –s test1 0.0.150

e) The PVC setup is complete. Do a 'cat /proc/atm/pvc' now. Notice that when you ping one machine from the other, the response times are ALL less than 1ms. If you didn't have a PVC setup, then the 1^{st} response time would be a lot greater, because of the time it takes to setup a connection.

5) I think that at this point it is quite important to realize the difference between IP traffic over ATM and native ATM. Up to this point, we have been doing ONLY IP over ATM. Now we will use the PVC for native ATM traffic.

C.4.3 Native ATM utilities

There are a few utilities in the ATM on Linux distribution: aread/awrite, br/bw and aping. They all take [interface.vpi.vci] as their arguments.

Let's say we have two hosts: test1 and test2 and a 2-way PVC 0.150 via an ATM switch. We want to send a text string 'hello' from test1 to test2. These are the two steps required:

Test2> aread 0.0.150 &	(so that test2 'listens')
Test1> awrite 0.0.150 hello	(to send the string from test1 to test2).

However, issuing the command 'awrite 0.0.150 hello' on test1 result in an error. The prompt comes back on test1, but it freezes (with no way out other than a hard reset). On testt2, the following line shows up at the prompt.

5: 68 65 6C 6C 6F

which is the correct hex representation of the word 'hello'!!

It turns out that there is a bug in atm-0.52, the fix patch can be found at http://www.execpc.com/~mitch/linux/atm/. The fix was integrated in 0.53 and newer versions.

Another useful utility is *aping*. It just sends AAL5 PDUs back and forth. You can see the AAL5 cell count increases by using atmdiag.

C.4.4 Connecting two ATM NICs back-to-back

1) First you need two computers running Linux with ATM support (let's call them test1 and test2 for example). You also need an ATM cross over cable, which is different from the standard UTP-5 cross over cable (only works for ethernet). Refer to /usr/src/atm/USAGE for details.

2) Then find out the ATM addresses of the two NICs. This can be done by the atmaddr tool.

3) *** Once you have got that, you need to prevent the atm-daemons (that we created earlier) from starting when the machines boot up. (/etc/rc.d/rc3.d/S09network)

4) Start the ATM signalling daemon in 'network' mode for test1 and in 'user' mode for 'test2' (as root on both machines!).

    Test1> atmsigd –b –m network
    Test2> atmsigd –b

5) Create a /etc/hosts.atm file containing the ATM address of the two NICs.

    47.0005.80FFE1000000F20F4546.0020480EC2D8.00   test1
    47.0005.80FFE1000000F20F4546.0020480EC32B.00   test2

   I think you could use spurious addresses as well since we are not going to be connected to the LE switch. However, you may as well use the real addresses.

6) Then we need to set the addresses correctly in the driver.

    Test1> atmaddr –a test1
    Test2> atmaddr –a test2

7) Start atmarpd on both machines to use CLIP (not required for aread/awrite etc)

8) To verify connectivity, do one of the following:

    Test1> aping 0.150
    Test2> aping 0.150

   which sends endless AAL5 cells over the cross-over cable, or:

    Test1> br 0.150
    Test2> bw 0.150

9) You can now type anything on test2 and the same text will appear on test1. You can even use VPIs other than 0. However, the machine running bw might crash when you exit from it, so it is better to pause the process rather than stopping it. Also note that awrite doesn't crash the machine in a back-to-back environment.

C.5 Set up the Virtual ATM Switch

These are the steps to set up the virtual ATM Switch:

1) NB! Read the README for version 0.5 of the Virtual ATM Switch Distribution. In fact, most of the steps covered in this section is directly from the README, which is quite self-explanatory.

2) Insert 2nd adapter into the virtual switch computer. Connect the 1st one to the LE switch.

    a) Copy vswitch-0.5.tar.gz to /usr/src
    b) Tar xfz vswitch-0.5.tar.gz → this will create 2 subdirectories, DSKI/ and KURT/ and 2 patch files under vswitch-0.5/

        Upon restart,

            [rt.c:register_rtmod:342] Registered process module.
            [rt.c:register_rtmod:342] Registered rt_switch module.
            Calibrating cpu ….. 167048083 cycles per second.

3) (For nicstar driver only) In /usr/src/linux/drivers/atm/nicstar.h, change the constant NS_MAX CARDS from 1 to 2. At bootup, both nicstar cards will be recognised.

4) Disable atm-daemons from starting up: comment out the daemons in /etc/rc.d/init.d/atm-daemons on the virtual switch and the two test hosts.

5) Now we need to build the ATM tools again on the vswitch machine becasuse the vswitch patch has provided new tool (follow the vswitch README-0.5 section 3.3)

6) Correct the minor data type problem in the atm.h header file that results in a compilation error: http://www.ittc.ukans.edu/~rsanchez/software/vswitch/vswitch-0.5-tools-bugfix.

7) Create virtual switch ports and VCCs using the 'vsw_ctl' command as describe in the README. You have now got a fully functional virtual ATM switch

Happy Switching ☺

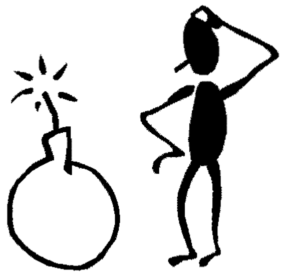
Further information can be found in the following web sites:

Web Sites	Description
www.leg.uct.ac.za	The University of Cape Town Linux Enthusiasts Group (LEG) – a very good starting point
www.linux.org	The Linux Home Page – useful general information and many interesting articles
www.linuxdoc.org or www.leg.uct.ac.za/mirrors/LDP/	The Linux Documentation Project – information to get started, 'how-to's, as well as useful books available for download. This site is mirrored at LEG on a daily basis
www.kernel.org	For Linux Distribution, kernel related information and download different releases of the Linux kernel
www.redhat.com	For information specific to the RedHat distribution of the Linux operating sytem
http://lrcwww.epfl.ch/linux-atm/	The ATM-on-Linux home page – research and development is actively going on with regard to ATM on Linux support
http://www.ittc.ukans.edu/~rsanchez/software/vswitch.html	The Virtual ATM Switch home page – the virtual ATM switch software release is currently being refined. This is the site with the latest updates

Appendix D

User-Oriented Video Quality Assessment

This section presents the materials related to the user-oriented video quality assessment that was conducted on 4th November 1999 including the posters, questionnaire, results from individual assessors etc. Photographs taken during the test session are included in the *\appendixD* directory of the CD.



What is going on?!?

Well, as the banner suggests, this demonstration has got something to do with 1) Digital Video, 2) ATM and 3) Quality Assessment. But how are these three seemingly unrelated terms related to one another? To answer this question, let me start off by explaining these terms.

Digital Video, as opposed to Analog Video(what you see on SABC TV everyday), is represented digitally (i.e. 1's and 0's are used to represent the information). Examples of it includes pictures from Satellite TV (DSTV) and Video CD (VCD).

ATM stands for Asynchronous Transfer Mode and NOT Automatic Teller Machine (or Another Terrible Mistake etc). For those of you who are not comfortable with the long name, just use ATM.

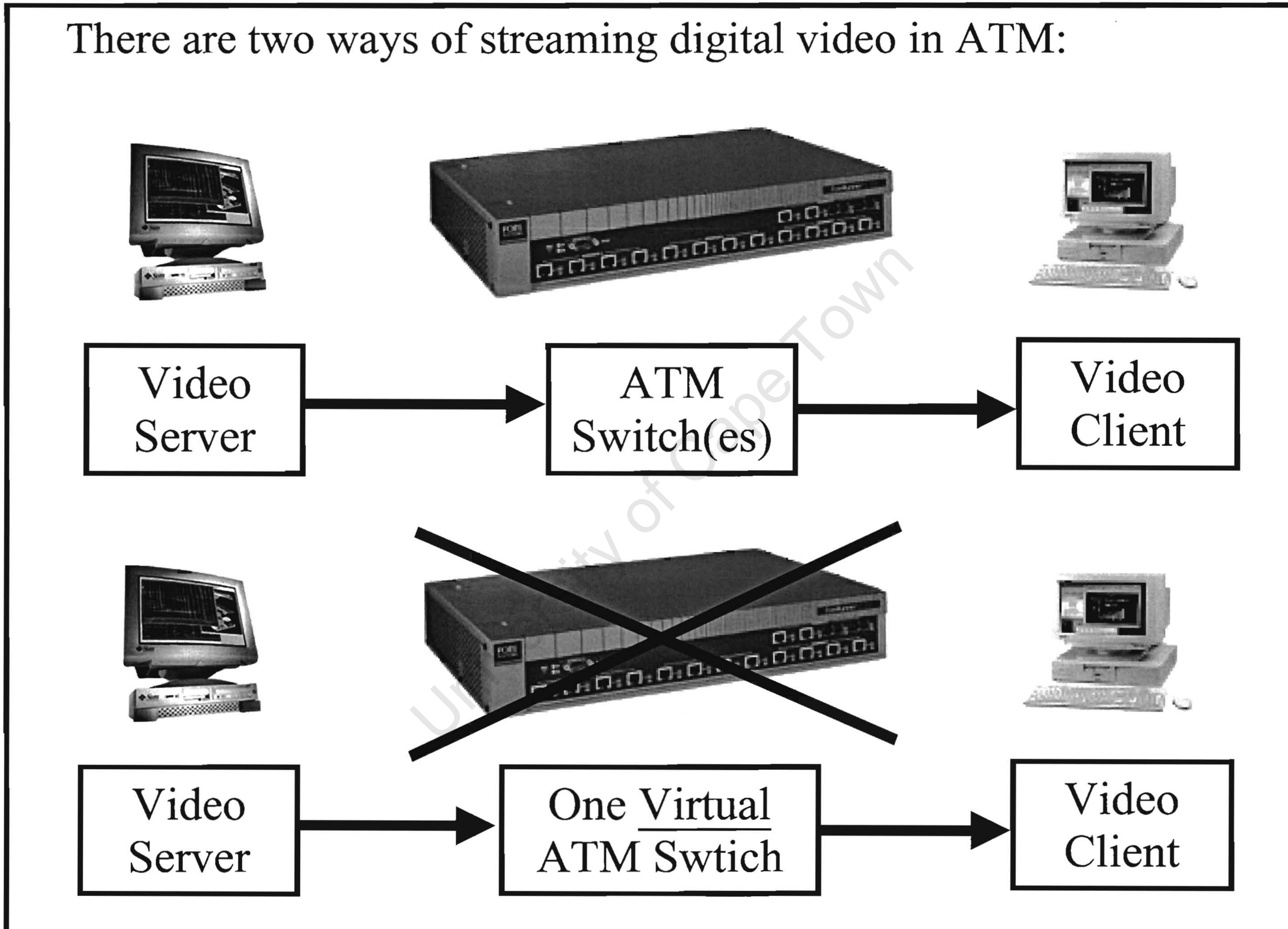
(I am sure when Automatic Teller Machine first came out, people found it as mysterious as Asynchronous Transfer Mode)

In short, ATM is a network protocol for transferring information between hosts on computer networks. Some advantages of ATM:

- it allows a single network to be used to support all types of information, data, voice and video
- it has high bandwidth capability (155Mbps/622Mbps) (Mbps=mega bits per second)
- it provides QoS (Quality-of-Service) support for network applications
- it is connection-oriented (i.e. a connection is set up before the transfer of data).
- it is distance independent (can be deployed as LAN or WAN)

When I tell people about video over ATM, they ask me "in the future, am I going to watch video while I wait for my cash withdrawal to be authorized at an ATM?" Seriously, the transportation of Digital Video over ATM can be done in 'offline' or 'online' mode. In 'offline' mode, video information is sent as if it is data (just like downloading a file from the web) with little time-constraint. 'Online' mode is the real-time streaming of video from a video server to the client. This is what you are about to see.

When digital video is streamed over an ATM network, it is subjected to certain network impairments. These impairments causes the video playback at the client to degrade. So a means to establish what is 'acceptable' quality and what is not is required.


There are two ways of streaming digital video in ATM:
Video Server
ATM Switch(es)
Video Client
Video Server
One Virtual ATM Swtich
Video Client

You can join the fun too?

Since video quality is user-oriented and is hightly subjective, we need to get feedback from 'users' and this is how you can join the fun too. To participate in this quality assessment takes three to five minutes. Four short digital video clips will be play three times each. For each video clip, the first run is 'perfect' (also called the reference clip) while the other two are subjected to different degree of network impairments. What you then do is to grade the video clips according to the table show in your form. Thank you for your participation and have a nice day ☺

The Effect of Cell Loss

Cell losses occurs in the presence of network congestion, errors in cell headers and bufer overflow. The phenomenon of cell loss results into lost blocks of information after the decoding process. This, in turn, may result in a degraded video quality. The effect of cell losses is illustrated below in Figure 1.A and Figure 1.B

Figure 1.A Original Frame

Figure 1.B Impaired Frame (with cell loss inserted)

The Effect of Bit Errors

Bit errors can be attributed to the inefficiencies of the physical transmission medium. The mean rate at which errors occur is described by the bit error ratio (BER) and depends on the characteristics of the physical transmission media. Bit errors may occur in bursts. These burst errors can lead to cell loss. The effect of bit errors is illustrated below in Figure 2.A and Figure 2.B

Figure 2.A Original Frame

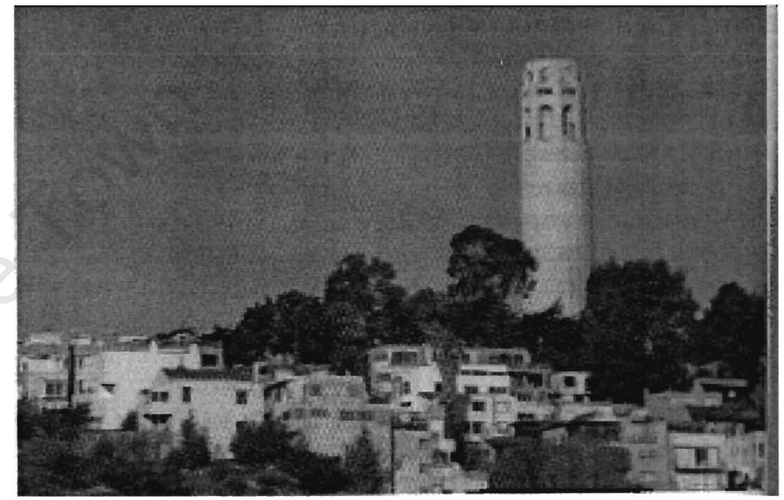
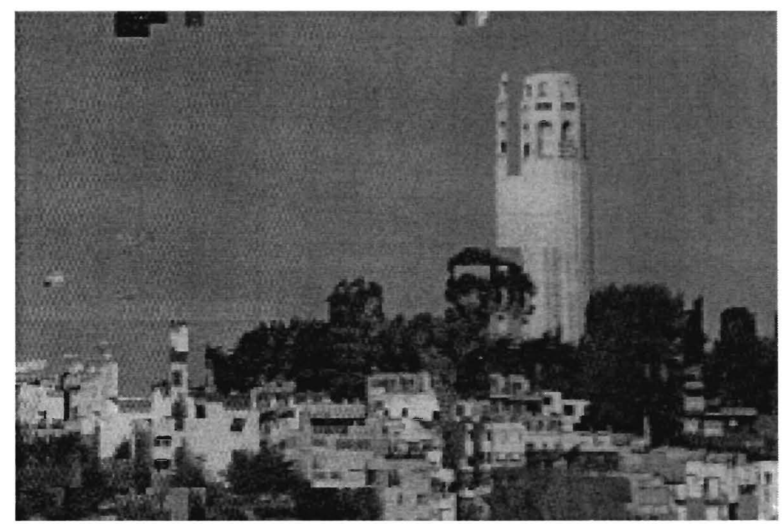


Figure 2.B Impaired Frame (with bit errors inserted)



Source of Images: [GRI98]

Open Day Demonstration: 04 October 1999

Subjective Assessment of the Effect of Impairments to Video Quality

Questionnaire

Abstract

The intrinsic nature of practical communications networks causes errors to be introduced in the transmitted video sequence. The resultant loss or misrepresentation of information causes impairments in the quality of the video that may be visually perceptible and sometimes annoying to the viewer.

Objective and Procedure

Please take a few minutes to partake in this assessment to evaluate the effect of impairments to video quality. The test material consists of four video clips. You will first be presented with the unimpaired version of the video clip, followed by the same video impaired. Use this sheet to compare the two clips by judging the overall impression given by the video sequence.

Please score all four video clips and return it to your contact person.

Assessment

Grading Scale:

For this assessment the following five-grading scale should be used:

5. Imperceptible
4. Perceptible, but not annoying
3. Slightly annoying
2. Annoying
1. Very annoying

Video Clip	Impairment Grading Score - Tick appropriate box														
	Reference					Impairment 1					Impairment 2				
Fatman	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
Porsche	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
Animation	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
Golf	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1

Comments:

Thank you.

Individual Scores for each test video sequence

Scores	Animation			Golf Swing			Talking Person			Porsche		
	Cell Loss Ratio											
	Ref	1.5×10^{-4}	6.1×10^{-5}	Ref	1.3×10^{-4}	8.5×10^{-5}	Ref	1.3×10^{-4}	1.1×10^{-4}	Ref	1.6×10^{-4}	1.1×10^{-4}
1	5	3	4	5	4	4	5	3	2	5	4	4
2	4	3	3	5	1	2	5	2	1	4	2	3
3	4	4	4	4	3	3	5	3	3	5	4	4
4	5	4	4	5	2	4	5	3	2	5	5	4
5	5	4	4	5	1	2	5	3	1	5	2	3
6	5	5	4	5	4	3	5	4	3	5	4	4
7	4	4	4	5	3	4	4	2	1	4	3	3
8	5	4	4	5	3	3	4	3	2	5	3	4
9	5	4	5	5	2	2	5	4	2	5	3	4
10	5	4	3	4	2	2	4	3	1	5	3	4
11	5	3	3	5	2	3	4	3	2	4	2	4
12	5	3	4	5	4	3	4	2	1	5	3	3
13	5	4	4	4	3	3	5	2	1	5	3	4
14	5	5	5	5	4	4	5	3	3	5	4	4
15	5	4	3	5	3	4	4	4	2	5	4	5
16	3	3	4	4	2	3	4	3	2	4	3	3
17	4	4	5	5	2	2	4	2	1	4	2	3
18	5	3	5	3	2	3	4	3	1	4	2	4
19	5	4	2	5	1	2	4	3	3	5	4	4
20	2	4	5	5	3	3	5	4	2	5	3	2
21	5	4	5	5	3	3	5	3	3	5	3	4
22	5	4	4	4	4	2	5	3	2	5	3	4
23	4	4	4	5	3	3	4	2	2	5	3	4
24	4	4	4	5	3	2	5	3	2	5	3	4
25	4				4	4	5	4		5	3	4
AVG	4.8	3.8	4.0	4.6	2.7	2.9	4.6	3.0	1.9	4.8	3.1	3.7

Appendix E

Captured Video Frames

This section presents the files and the locations within the files from which video frames have been captured for reference purposes.

	TIME	From which file was it captured?	
		Left Hand Side	Right Hand Side
Figure 6-2	00:01	/chapter6.4/CE_vs_CL/CL.mpg	/chapter6.4/CE_vs_CL/CE.mpg
Figure 6-3	00:00	/chapter6.2/an1.mpg	/chapter6.2/animation.mpg
Figure 6-4	00:06	/chapter6.2/po3.mpg	/chapter6.2/porsche.mpg
Figure 6-5	00:03	/chapter6.2/gs2.mpg	/chapter6.2/golf_swing.mpg
Figure 6-6	00:03	/chapter6.2/gs2.mpg	/chapter6.2/golf_swing.mpg
Figure 6-7	00:14	/chapter6.2/gs2.mpg	/chapter6.2/golf_swing.mpg
Figure 6-8	00:00	/chapter6.2/gs1.mpg	/chapter6.2/golf_swing.mpg
Figure 6-9	00:08	/chapter6.2/po3.mpg	/chapter6.2/porsche.mpg
Figure 6-10	00:09	/chapter6.2/po3.mpg	/chapter6.2/porsche.mpg
Figure 6-11	00:05	/chapter6.2/tp1.mpg	/chapter6.2/talking_person.mpg
Figure 6-12	00:09	/chapter6.2/po4.mpg	/chapter6.2/porsche.mpg
Figure 6-14		/chapter6.5/tp2.mpg at 00:10	/chapter6.5/po3.mpg at 00:08
Figure 6-17	00:03	/chapter6.7.4/po3.mpg	/chapter6.7.4/porsche.mpg

Appendix F

What is on the CD?

Directory	Description
ATM Forum Spec	ATM Forum Specifications: [TM4.0], [UNI3.1], [UNI4.0], [VOD97]
BSTS	Documents such as user-guide, programming guide, release notes etc. for the Network Impairment Emulator Module, Line Interface, Cell Protocol Processor
Chapter1.1	Contains the UTA Internet Teaching news report
Chapter1.1.2	Contains a range of video clips with different quality
Chapter1.1.3	Quality of Service for Internet Application over ATM
Chapter6.1	Video and Audio Artefacts with AAL error checking mechanisms enabled
Chapter6.3	Software Program for the 'add-on' module
Chapter6.4	Comparisons between 1)Cell Loss and Cell Error, 2)Cell Mis-insertion and Cell Loss, 3)Cell Drop and Dummy Cell Insertion
Chapter6.4.1	Captured video frames (much clearer then those shown in this thesis)
Chapter6.4.2	Video clips from which the pictures in section 6.4.2 were captured
Chapter6.5	Figure 6-14 and the video clips from which they were captured
Chapter6.7.2	Test materials used during the user-oriented video quality assessment conducted in this study. Photographs and Appendix D are also included
Chapter6.7.4	Figure 6-17 in this dissertation
Dissertation	Electronic version of this document
MPEG Players	Software MPEG decoders for Windows and Linux
VATMSwitch	Packages required for the Virtual ATM Switch and Appendix C