Reliable Transmission of MPEG-2 VBR Video Streams over New Network Adaptation and ATM Adaptation Layers

Xavier Garcia Adanez, Olivier Verscheure and Pascal Frossard
Telecommunication Services Group, TCOM Laboratory
Swiss Federal Institute of Technology, Lausanne
telephone: +41 21 693 5258
fax: +41 21 693 2683
email: {garcia, verscheure, frossard}@tcom.epfl.ch
URL: http://tcomwww.epfl.ch/{~garcia, ~verscheu, ~frossard}

Abstract

This paper adresses the transmission of VBR MPEG-2 video streams on top of both Network Adaptation (NAL) and ATM Adaptation Layers (AAL) [1] for real-time multimedia applications. The NAL, specific to MPEG-2, provides a selective data protection mechanism based on syntactic criteria. The AAL provides a reliable transmission by applying per-cell sequence numbering combined with a selective Forward Error Correction (FEC) mechanism based on Burst Erasure codes. Studies carried out with Constant Bit Rate (CBR) video streams showed improvements in terms of network performance evaluated by the cell loss ratio (CLR) as well as in terms of user perceived quality compared to the performance obtained with AAL5 under the same network conditions. This paper proposes improvements at the NAL and presents the results obtained for the transmission of Variable Bit Rate (VBR) video streams. To evelute the impact of cell losses at the application level, we apply a perceptual quality measure to the decoded MPEG-2 sequences which allows us to evaluate performance at the user level.

1 Introduction

For the past several years, a tremendous number of audiovisual services has been emerging (e.g. Video on Demand (VOD), Interactive Distance Learning (IDL), home shopping, . . .). Today, ATM technology, efficient compression techniques and other developments in telecommunications make it possible to offer such services. However, a lot of work remains to be done to optimize these services so that they can be offered at attractive prices (the user expects an adequate perceived quality at the lowest possible cost).

Source coding of audiovisual information (i.e. the MPEG-2 encoding scheme) naturally generates variable bit rate since the entropy of an audiovisual sequence may vary in time. ATM, the adopted solution for BISDN, has the advantage of handling constant as well as variable bit rate communications. Thouhg CBR connections are easier to manage than VBR ones, VBR connections may use the network bandwith more efficiently through statistical multiplexing. One of the major drawback with such statistically multiplexed connections is that the requested QoS is also statistically guaranteed. It has been shown that cell and frame loss ratios might not be negligibles in such situations [2, 3].

This work studies the transmission of MPEG-2 VBR video streams on top of new Network Adaptation and Multimedia ATM Adaptation Layers. These layers provide a reliable transmission over ATM under cell loss. We study the performance of the new layers from the network perspective via the cell loss ratio (CLR) and packet loss ratio (PLR) metrics. Since the mapping of data loss onto user perceived quality heavily depends on the type and location of the lost information, we also evaluate performance with a video quality metric. This metric has the advantage of giving results correlated with human perception [4].

2   Multimedia Application Requirements

2.1   Perceptual Impact of Data Loss

One of the main characteristic of multimedia data streams is that they are of a continuous nature. Multimedia data embeds a temporal behaviour that naturally fixes the delay bounds within which the information has to be delivered. Beyond these bounds, the information is considered as lost because it is meaningless to the application. In addition to being *time-sensitive*, the audiovisual information is also *loss sensitive* because compression techniques such as MPEG-2 reduce the redundancy in the transmitted data.

The reduction of quality due to data loss heavily depends on the type of the lost information. Losses of syntactic data such as video headers and system information, affect the quality differently than losses of semantic data such as raw video information. Furthermore, the location in the stream of the lost semantic data has also an impact onto quality reduction due to the predictive structure of an MPEG-2 video stream. Figure 1 depicts how network losses map onto decoded visual information in different MPEG-2 pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g. slice headers) mainly due to the use of variable length coding, run length coding and differential coding. This is referred to as spatial propagation. When loss occurs in a reference picture (intra-coded or predictive frame) it will generally remain until the next intra-coded picture is received. This causes the errors to propagate across several non intra-coded pictures which is known as temporal propagation and is mainly due to inter-frame predictions.

The impact that the loss of syntactic data may have is in general more important and difficult to recover than the loss of semantic information. Indeed, when a header is lost, the entire underlying information cannot be decoded even if received correctly and is then skipped. When this occurs to an intra or predictive frame, it may heavily affect the quality perceived by the user. Clearly, some headers are more important than others because they are important syntactically as well as perceptually.
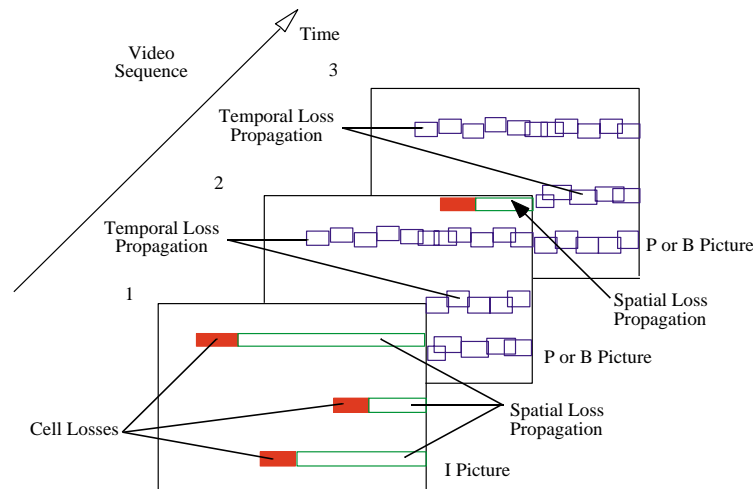

Figure 1: *Data Loss Propagation in a Typical MPEG-2 Decoded Sequence*

Error concealment algorithms have already shown that it is possible to reduce the impact of data loss on the visual information [5, 6]. These error concealment algorithms include, for example, spatial interpolation, temporal interpolation and early resynchronization techniques. Early resynchronization decoding techniques limit the spatial propagation of errors by decoding some semantic information that is normally discarded from the damaged MPEG-2 video streams. In other words, it helps the decoder to resynchronize before the next header. This method is based on the identification of allowed codewords as proposed in [6] and, unfortunately, works only with intra-coded frames. These techniques are thus, by far, not perfect. Data loss may still involve very annoying perturbation in the decoded video, especially

when entire frames are lost.

2.2   Multimedia Protocol Layers for ATM

As already stated in Sec. 2, multimedia applications are delay and loss sensitive. Therefore, they cannot rely on retransmission mechanisms for error correction. The current specifications to transmit MPEG-2 based applications do not make use of any transport protocol albeit they use AAL5. This means that in case of cell loss no action could be taken to correct the error or even to inform the receiver that data loss occurred. In addition, the packet discard mechanism of AAL5 amplifies this problem.

In order to meet the requirements of multimedia applications, some mechanisms need to be included in the AAL. We proposed in [7] a Multimedia AAL (MAAL) mechanism that improves the cell loss detection mechanism giving better results in terms of data loss seen by the application. Basically, we proposed to increase the resolution of the cell loss detection to the cell level. To achieve this, we introduced a per-cell sequence numbering. To make this possible, we use 47-byte cell payloads which frees an octet per ATM cell to insert a Segmentation and Reassembly (SAR) header. The sequence number is contained in this header (see Fig 2). The receiver uses these sequence numbers to detect the number and position of the lost cells in a packet. This information can thus be passed to the upper layers that can take necessary action to conceal data loss. The MAAL uses a dummy cell insertion mechanism, when cell losses are detected.

If we consider the transport of MPEG-2 Transport Stream (TS) packets, the proposed mechanism has several advantages: first, it always gives an integer number of cells, since the length of the TS packets is 188 bytes which gives exactly 4×47 byte payloads. Second, if dummy cell insertion is applied, it can be used as an error message since sequences of more than 5 zero-octets are not allowed by the MPEG-2 standard [8]. So this mechanism can be exploited by the decoder as an error indicator.
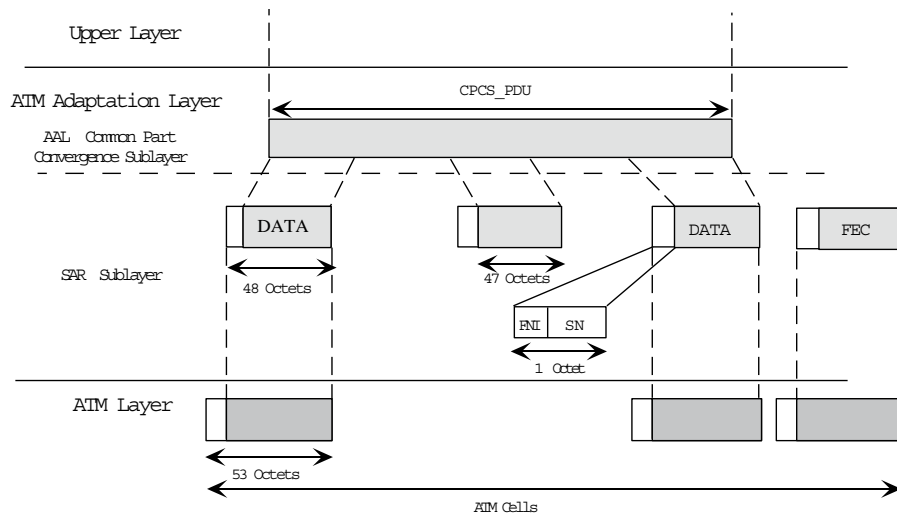

Figure 2: *Multimedia AAL Segmentation Mechanism*

To further improve the reliability of the AAL, a selective FEC mechanism has been proposed in [1]. It is based on burst erasure codes (RSE) [9] which take into account the specifics of ATM to correct cell losses only. RSE codes are simpler than standard Reed-Solomon Correcting codes because they rely on the fact that the position and the size of the erasures is known. This is the case for ATM. The erasures are limited to fixed boundaries (cells). In addition the utilization of a cell sequence number gives the position of the erasure. Since this mechanism does not rely on interleaving, it introduces low delay as shown in Fig 3. The advantage of such a selective method is twofold: first it reduces the overhead, and

second, it adapts to highly structured information such as compressed video.
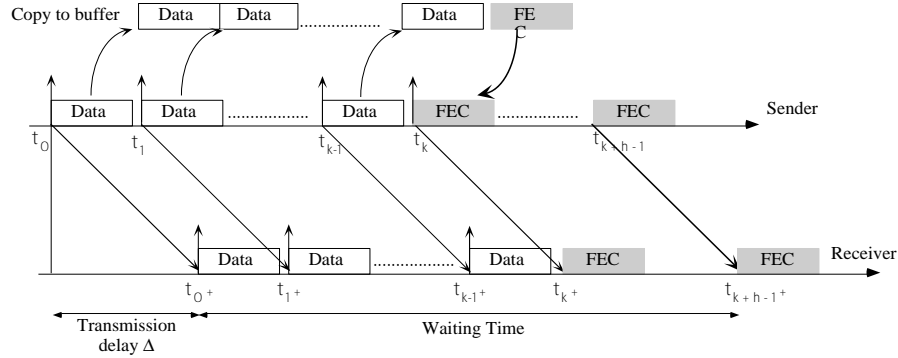

Figure 3: *Delay introduced by the RSE based Forward Error Correction Mechanism*

However, to achieve this selective FEC mechanism, a knowledge of the data to be transmitted is necessary. Since AALs are generic, another layer has to be able to detect which information has to be protected. This is the task of the proposed Network Adaptation Layer (NAL). It is an application specific layer since it has to identify information elements in the bitstream. In this work we develop a NAL specific for the transmission of MPEG-2 video streams. Its main functionalities are the following: delivery of fixed size PDUs (in the case of MPEG-2 it is straightforward) and detection of essential data combined with the generation of FEC_request messages. The latter functionality consists of detecting the most relevant information for the decoder (i.e. syntactic data). Since this selection is based on the most sensitive headers from their perceptual impact viewpoint, we call this mechanism *Perceptual Syntactic Information Protection* (PSIP).

Since the number of FEC cells generated by the AAL may vary, a FEC Number Indication (FNI) field is also included in the SAR header. This field is used by the receiver to know the number of FEC cells appended to a given packet.

3   Performance Evaluation

3.1   Simulation Setup

3.1.1   Video Coding Setup

The traffic under test (TUT) consists of a ski sequence of 1000 frames (720×576) encoded with a constant quantizer scale, with an MPEG-2 software encoder, as interlaced video, with a structure of 12 images per GOP and 2 B pictures between every reference picture, and a single slice per line (i.e. 45 macroblocs per slice).

In our experiments, two different quantizer scales have been considered, 26 and 28 which both produce a CATV quality, in such a way that 10% more bandwidth in average is required by the former. The saved bandwidth will be used for data protection (PSIP). These two quantizers imply different perceptual video encoding qualities (see Fig. 9).

Before being transmitted, the MPEG-2 bitstream is encapsulated into 18800-bytes length Packetized Elementary Stream (PES) packets and divided into fixed length Transport Stream (TS) packets by the MPEG-2 system encoder.

3.1.2   Network Setup

The simulation setup used for our experiments, depicted in Fig. 4, is composed of four multimedia worksta-tions and two ATM switches. Both switching stages are implemented as multiplexers with limited buffer size. The multimedia workstations are connected as two point-to-point communications. One of the
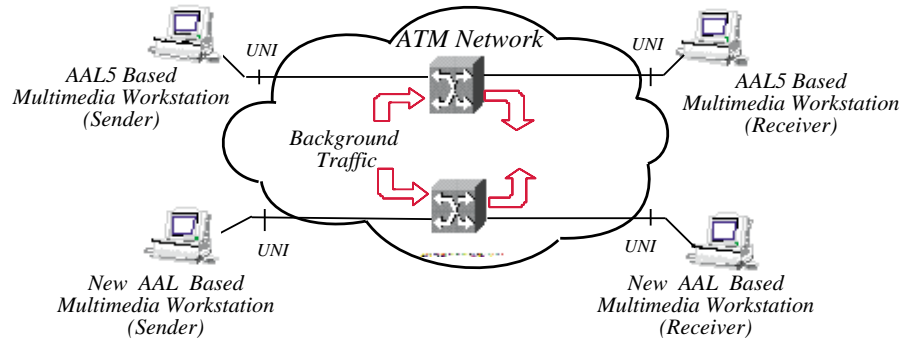
Figure 4: *Simulation Setup*

connections uses AAL5 to transmit the MPEG-2 bit stream while the other uses MAAL. To generate cell losses in the multimedia stream, the multiplexing stages are loaded with background traffic provided by several On-Off sources. This type of source model is widely used to simulate a multiplex of traffic such as the one that could be found at the entrance of an ATM switch. To guarantee the same CLRs to both cell streams, the background traffic is replicated and sent simultaneously to both multiplexing stages. The TUT is then routed to the receiver end system where the data is reassembled prior to decoding.

Currently, there is no specification on how to transmit VBR video over ATM. Originally, AAL2 was defined by the International Telecommunications Union (ITU) as the adaptation layer for VBR time-constrained communications. However, it has never been specified and recently this AAL has been devoted to wireless communications. Therefore, today, the only specification describing the transmission of video over ATM is the approved ATM Forum Video on Demand specification [10]. This document describes the encapsulation of CBR encoded MPEG-2 TS packets into AAL5-SDUs (Service Data Units). This scheme packetizes two single program transport streams (SPTS) packets regardless of their information contents, being of audio, video or timing nature into a single AAL5-SDU. The AAL5 adds its 8 byte trailer and the resulting AAL5-PDU is segmented into 8 ATM cells without any padding. Recently, the ATM Forum has started the specification to carry VBR video over AAL5. We have adopted the same encapsulation scheme for our experiments. Figure 5 depicts the current framework in white. The layers in grey show how the proposed protocol layers seamlessly integrate into the protocol stack.
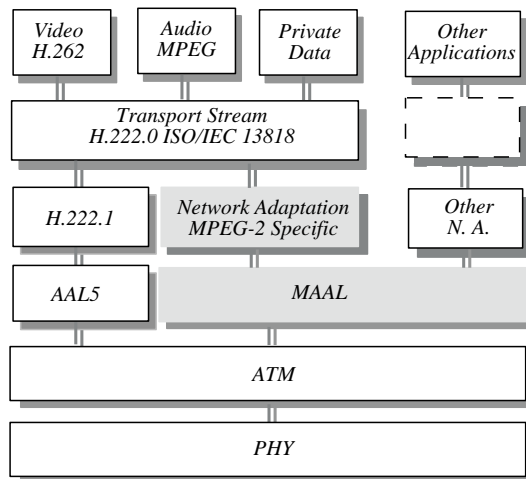

Figure 5: *ATM Forum and Proposed Frameworks for Transmission of MPEG-2 Video over ATM*

3.1.3 Video Decoding and Quality Assessment Setups

When the MPEG-2 decoder detects data loss in semantic information, it tries to resynchronize as quick as possible by applying early resynchronization algorithms presented in [5]. For not resynchronized areas, it uses classical spatial and temporal concealment techniques.

When a frame has been lost because of the lack of a frame header, the previous decoded frame is then considered instead. In general, when a system or video header is missing, the decoder tries to recover as much video data as it can.

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of the Human Visual System [11]. The vision model described in [12] has been used to build a computational quality metric for moving pictures which proved to behave consistently with human judgements [4]. Basically, the metric, termed Moving Pictures Quality Metric (MPQM), first decomposes an original sequence and a distorted version of it into perceptual components. A channel-based distortion measure is then computed accounting for contrast sensitivity and masking. Finally, the data is pooled over the channels to compute the quality rating which is then scaled from 1 to 5 (the higher the value the better the quality).

The quality estimation using the MPQM metric is performed on the whole sequence. The mean quality value is then calculated. Further experiments will consider a more accurate quality metric in order to gather the frame-based quality measures into a single value.

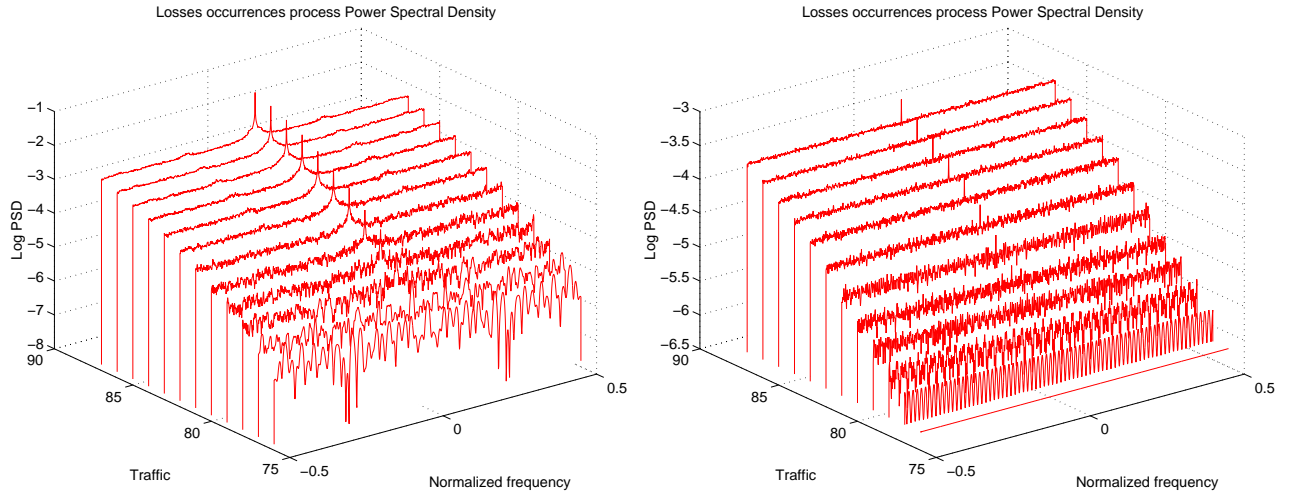## 3.2 Simulation Results and Comparisons

3.2.1 Network Performance

Figure 6: *Power Spectral Density of the TuT Cell Loss Process Correlation*

The utilization of an open-loop VBR encoder leads to a highly bursty traffic profile. It is therefore to be expected that the cell loss process will also show some burtiness. However, it has been shonw in [13] that the recovery efficiency of FEC techniques is reduced when the cell loss process is correlated. We first study the behavior of the cell loss process for the VBR video stream. Fig. 6 depicts the power spectral density (PSD) function of the sampled cell loss processes for both the global and the video streams. The peak at the null frequency $\varphi(0)$ shows that a correlation at the cell level exists for both processes.

However, if we now compare the Packet Loss Ratio (PLR) observed for AAL5 and we plot it as a function of the CLR (Fig 7), it turns out that the simulated figure follows very closely the theoretical figure obtained by assuming a random loss process. In fact, if we assume independent and identically distributed cell losses we can derive that the probability of losing a packet follows a binomial distribution given that the loss of at least one and at most all the cells in a packet entails the loss of a packet. We can
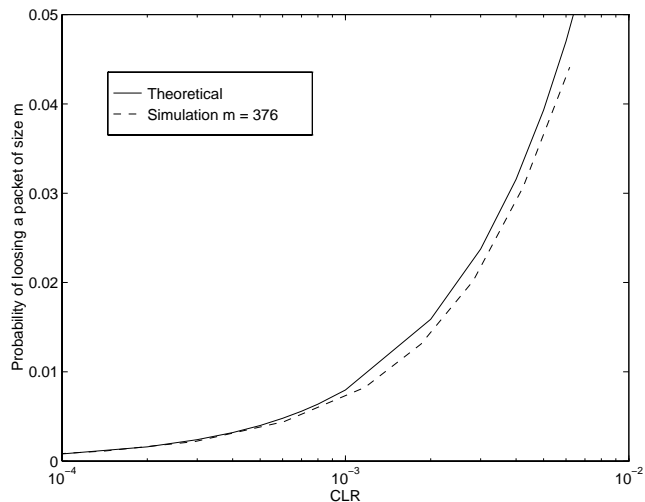
Figure 7: *Cell Loss Ratio vs Packet Loss Ratio for Open Loop VBR Video Traffic*

conclude that from a packet scale the loss process behaves close to a random loss process. This suggests that the correlation at the cell level is relatively short. This is especially true when the TuT uses a small fraction of the link rate which is the case in our experiments. Indeed the correlation effects will increase if better video quality is required which implies the usage of a smaller quantization factor. Therefore, the recovery efficiency of the FEC mechanism would slightly suffer from the burstiness of the VBR video.
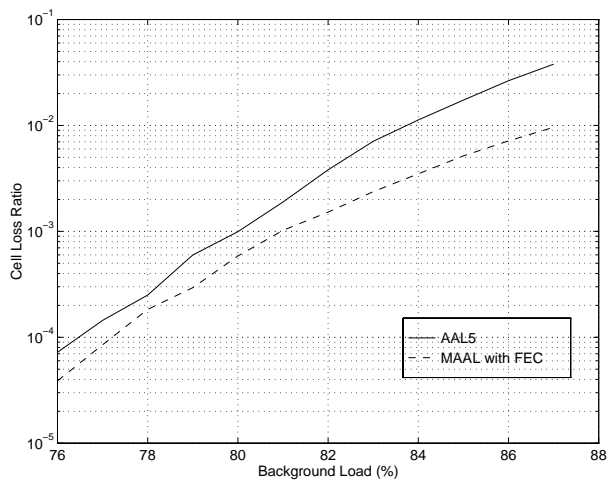

Figure 8: *Cell Loss Ratios observed by both AALs*

Figure 8 shows the CLRs measured at the receiver, which for AAL5 includes all the data lost due to the packet discard mechanism.

The CLR measured for the MAAL is actually equal to the network CLR since there is no extra discard of information. Comparatively, the CLRs achieved by AAL5 are 4 times higher. This is due to the packet discard mechanism which clearly amplifies the CLR seen by the application. The size of the packets used in the simulation is of 8 ATM cells. The fact that the CLR difference between both AALs is not closer to the packet size is due to the clustering effect observed in the cell loss process which reduces the negative impact of the packet discard mechanism. It is interesting to note that changing the size of the packets to be transmitted does not affect the performance of the MAAL. This is not the case for AAL5 whose performance depends on the packet size.

As already explained in Sec. 2.1, one of the major difficulties with video applications is the mapping of data loss impact onto the quality perceived by the user. Indeed, network performance measurements do not give enough information concerning the final quality of the displayed video. Therefore in our experiments, we evaluated the performance of the proposed protocol layers by means of a perceptual quality metric applied to the received video streams. The results are discussed in the next section.

3.2.2 User Perceived QoS

Figure 9 shows the perceptual quality assessment as a function of the CLR for the transmission of MPEG-2 VBR streams over AAL5 and MAAL. Fig 9 (a) compares the quality of the same video stream transmitted with both AALs. The MAAL shows a better behaviour under medium to high CLRs without PSIP. Note that both the AAL5 and MAAL figures begin to diverge for a CLR value close to 10^{-5}. This is due to the better performance of MAAL. However, under very high CLRs both figures converge again. This is due to the limited burstiness of the cell loss process which groups the losses and therefore reduces the negative effect that the AAL5 packet discard mechanism produces. However, the MAAL is still more reliable without any extra overhead which still can be explained by the fact that the cell loss clustering effect due to the correlated traffic is actually very limited which clearly penalizes the AAL5.
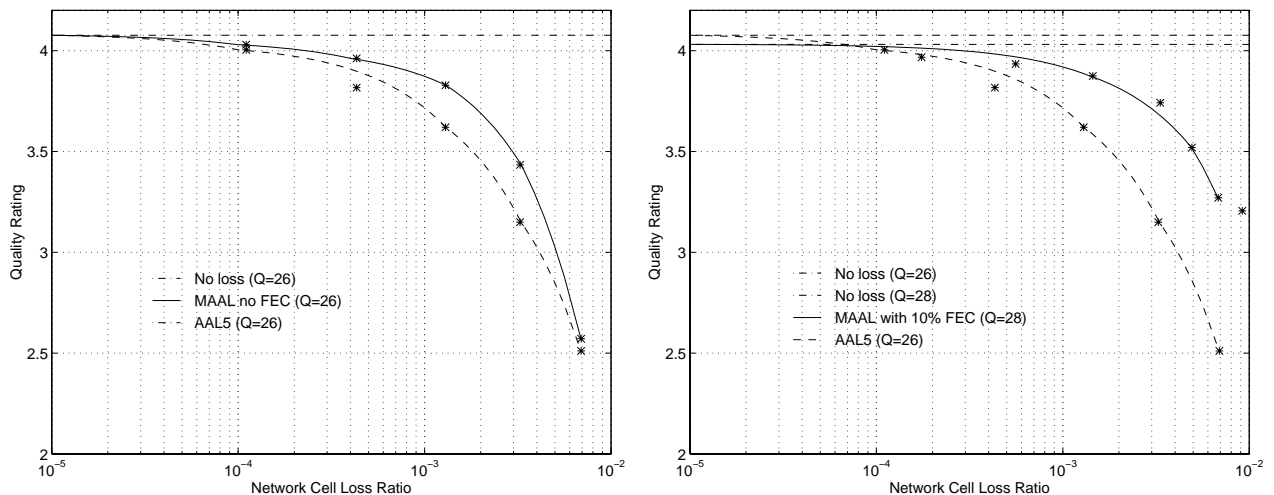

Figure 9: *Quality Rating vs Network CLR for VBR Video Transmission over AAL5 and MAAL*

Fig 9 (b) depicts the results obtained for the transmission of both an unprotected video stream over AAL5 and a PSIP protected one over MAAL. The figure which gives the best quality is the transmission of the lowest video encoding quality stream with MAAL and PSIP. It is interesting to note that even if the quality of the encoded flow is lower, the quality at the receiver is better for a CLR beyond 10^{-4}. For higher CLR values both figures diverge increasing the difference in quality achieved by the MAAL. Even if we have shown that the burstiness of the cell loss process is limited it is true that the FEC mechanism is penalized under such conditions. Bearing in mind that the protection mechanism adds a single redundancy cell, configuration which is not optimized for such bursty loss processes, we can conclude that further improvements on the PSIP mechanism could lead to better results. We can then conclude that a tradeoff may be found which optimizes the perceived quality (the bitrate) and the protection efficiency (the overhead).

The probability of observing such high CLRs in ATM networks due only to congestion is relatively reduced. Nevertheless, our results show that under high CLRs the same or even better perceived quality could be achieved with the combined NAL-MAAL layers. This leads to two possible utilizations: the first is the possibility of applying the proposed mechanisms to a wireless environment where such high CLRs are expected to occur. The second one is related to the cost. If a network operator applies statistical multiplexing the way the resources are allocated depends on the cell loss ratio an application tolerates.

Since the MAAL with the PSIP mechanism achieves the same perceptual quality for higher CLRs than AAL5 it is possible to multiplex more connections per link thus obtaining a better network resource utilization.

4 Conclusion

Based on recently proposed Network Adaptation and ATM Adaptation Layers, we studied the transmission of VBR encoded video over ATM under different network conditions. We have shown that the utilization of a simple FEC mechanism not totally adapted to VBR traffic, still improves the user perceived quality under fairly high cell loss ratios. Even if further investigations are necessary to better tune the PSIP to bursty loss conditions, these results suggest that the proposed layers could be used in wireless environments were such high CLRs are to be expected. Also, they introduce the possibility of reducing costs of multimedia applications by applying aggressive statistical multiplexing strategies.

References

[1] Xavier Garcia Adanez, Olivier Verscheure and Jean-Pierre Hubaux, "New Network and ATM Adaptation Layers for Real-Time Multimedia Applications: A Performance Study Based on Psychophysics", *in Multimedia Transports and Teleservices, International COST 237 Workshop*, Barcelona, Spain, November 1996, available on `http://tcomwww.epfl.ch/~garcia`.

[2] Thomas Stock and Reto Gruenenfelder, "Frame Loss vs. Cell Loss in ATM Concentrators and Multiplexing Units", *in EFOC&N'94*, 1994.

[3] Israel Cidon, Roch Guerin, Asad Khamisy, and Kumar N. Sivarajan, "Cell versus Message Level Performances in ATM Networks", *in First International ATM Traffic Expert Symposium*, Basel, Switzerland, Apr. 1995.

[4] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *in Proceedings of the SPIE*, vol. 2668, pp. 450–461, San Jose, CA, January 28 - February 2 1996, available on `http://tcomwww.epfl.ch/~verscheu`.

[5] Carlos Lopez Fernandez, Andrea Basso and Jean-Pierre Hubaux, "Error Concealment and Early Resynchronization Techniques for MPEG-2 Video Streams Damaged by Transmission over ATM Networks", *in SPIE*, San Jose, USA, February 1996.

[6] S. Lee, J. Youn and S. Jang, "Transmission error detection, resynchronization, and error concealment for mpeg-2 video decoder", *in SPIE*, 1993.

[7] Xavier Garcia Adanez, Andrea Basso, and Jean-Pierre Hubaux, "Study of AAL5 and a New AAL Segmentation Mechanism for MPEG-2 Video over ATM", *in 7th Workshop in Packet Video*, Brisbane, Australia, Mar. 1996, available on `http://tcomwww.epfl.ch/~garcia`.

[8] ISO/IEC JTC-1, editor, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video*, ISO/IEC JTC-1, 1994.

[9] A.J. McAuley, "Reliable Broadband Communications Using a Burst Erasure Correcting Code", *in ACM SIGCOMM'90*, pp. 297–306, Philadelphia, PA, 1990. ACM, Academic Press.

[10] The ATM Forum Tecnical Committee, "Video on Demand Specification 1.0", Jan. 1996.

[11] Serge Comes, *Les traitements perceptifs d'images numerisées*, PhD thesis, Université Catholique de Louvain, 1995.

[12] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", *in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2293–2296, Atlanta, GA, May 7-10 1996, available on `http://ltswww.epfl.ch/~vdb`.

[13] Ernst W. Biersack, "Performance Evaluation of Forward Error Correction in an ATM Environment", *IEEE Journal on Selected Areas in Communications*, vol. 11, 1994.