View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Infoscience - École polytechnique fédérale de Lausanne

Content-based MPEG-2 Structuring and Protection

Pascal Frossard[a] and Olivier Verscheure[b]

[a] Signal Processing Laboratory
[b]Institute for Computer communications and Applications
Swiss Federal Institute of Technology
Lausanne, Switzerland

ABSTRACT

We address a new error-resilient scheme for broadcast-quality interactive MPEG-2 video streams to be transmitted over lossy packet networks. A new scene-complexity adaptive mechanism, namely Adaptive MPEG-2 Information Structuring and Protection (AMISP) is introduced. AMISP lies on an information structuring scheme which modulates the number of resynchronization points (i.e., slice headers and intra-coded macroblocks) in order to maximize the perceived video quality. The video quality the end-user experiences depends both on the quality of the compressed video before transmission and on the degradation due to packet loss. Therefore, the structuring scheme constantly determines the best compromise between the rate allocated to encoding pure video information and the rate aiming at reducing the sensitivity to packet loss. It is then extended with a Forward Error Correction (FEC) based protection algorithm to become AMISP. AMISP triggers the insertion of FEC packets in the MPEG-2 video packet stream. Finally, it is shown that AMISP outperforms usual MPEG-2 transmission schemes, and offers an acceptable video quality even at loss ratios as high as 10^{-2}. Video quality is estimated using the Moving Picture Quality Metric, which proved to behave consistently with human judgement.

keywords: MPEG-2, error resilience, adaptive structuring, content-based protection, perceptual quality.

1. INTRODUCTION

Because of the increasing availability of Internet and ATM networks, packet video is becoming a common support. It is therefore important to fully understand the parameters that may affect the quality of the video delivered to the end-user, and how to cope with the resulting impairments. Both the encoding and the transmission processes may affect the quality of service. The best quality at the lowest bandwidth occupancy can thus only be obtained by optimizing the entire system end-to-end rather than its individual components in isolation.[1,2]

The choice of a compression standard mostly depends on the available transmission or storage capacity as well as the features required by the application. The MPEG-2 standard is an audio-visual standard developed by the International Standards Organization (ISO) together with the International Electrotechnical Commission (IEC).[3] The video part of MPEG-2 permits data rates up to 100 Mbps and also supports interlaced video formats and a number of advanced features, including those supporting HDTV. MPEG-2 is capable of compressing NTSC or PAL TV-resolution video into an average bit rate of 3 to 7 Mbps with a quality comparable to analog broadcast TV.[4]

Like any other compressed data, compressed video is highly sensitive to data loss (see Section 2). Data loss propagates within the sequence and may thus become very annoying for the end-user.[5] The error resilience schemes proposed in the literature[6] could be roughly classified into three categories.[7]

First, the concealment techniques try to estimate missing video data using information available at the receiver. The simplest methods would be to replace the missing block with a similar block. The motion vectors could also

Further author information: (Send correspondence to Pascal Frossard)
Emails: pascal.frossard@epfl.ch, ov1@us.ibm.com
Olivier Verscheure is now with the IBM T.J. Watson Research Center, NY, USA.

be approximated by those of the surrounding video blocks. However, even for the most sophisticated concealment techniques,[8–11] important loss of data may lead to very annoying degradations. Second, the resynchronization or error localization techniques aim at limiting spatially and/or temporally error propagation.[12–15] These techniques however do not take into account the local relevance of video data,[16] which is the only consideration leading to optimal scenarios. Finally, unequal error protection schemes try to efficiently recover the missing video information.[17–20] Similarly to the resynchronization techniques, the best results could though only be obtained with a judicious packet prioritization process.[21] In this category, layered coding[22] and the new Multiple Description Coding schemes[23] could be mentioned as the most promising algorithms. Optimal error resilient schemes should however combine techniques of the three categories. Given bit budget constraints, such a combination is indeed the only way to provide the best video quality.

In this work, we propose an adaptive MPEG-2 information structuring and protection algorithm targeting interactive video applications. This algorithm fully exploits the nonstationary nature of the video signal exactly where it lies. The structuring part of the algorithm (AMIS) adaptively modulates the number of slice headers[24] and intra-coded macroblocks[25] in order to minimize the impact of data loss, and thus to maximize the perceived video quality. To do so, it measures the impact of an hypothetical packet loss. The protection part of the algorithm (AMISP) adaptively triggers the underlying Network Adaptation Layer (NAL) to generate redundancy data (i.e., FEC packets), according to the structuring information. Data packets are thus protected whenever their loss would lead to annoying degradations in the reconstructed video. AMISP produces therefore a very low overhead in comparison to existing techniques.

The paper is organized as follows: Section 2 briefly describes the MPEG-2 standard and the impact of data loss onto the decoded video sequence. Section 3 then states the error resilience problem driven by timing and cost constraints. AMIS, the structuring part of the algorithm is presented in Section 4. The algorithm is then extended by an adaptive protection scheme to become AMISP addressed in Section 5. The performances of AMISP are discussed in Section 6, where AMISP is compared to usual video transmission schemes. Finally, concluding remarks are given in Section 7.

2. MPEG-2 SENSITIVITY TO DATA LOSS

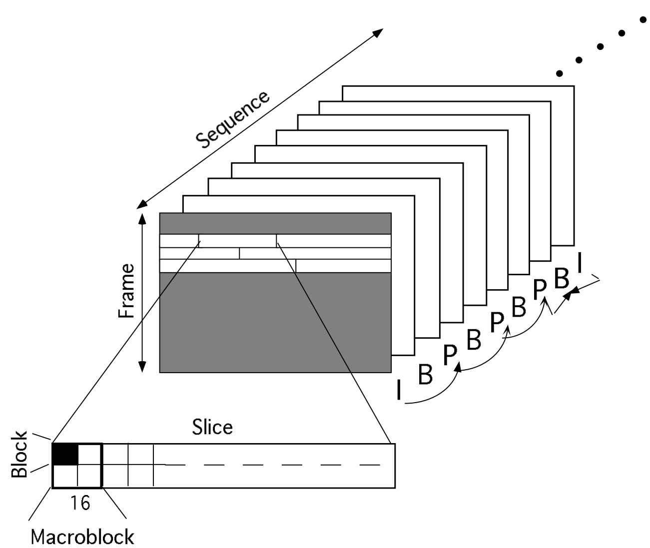

Figure 1. *MPEG-2 video structure.*

An MPEG-2 video stream is hierarchically structured as illustrated in Figure 1. The smallest entity defined by the standard is the *block*, which is an area of 8×8 pixels of luminance or chrominance. A *macroblock* (16×16 pixels) contains four blocks of luminance samples and two, four or eight blocks of chrominance samples, depending on the chrominance format. A variable number of macroblocks is encapsulated in an entity called *slice*. As required by the MPEG standard, each new line of macroblocks begins with a slice header. However, there is no constraint on slice length. To decrease the overhead and hence increase the compression, very often a slice continues all the way to the end of a macroblock line. Slices occur in the bit stream in the order in which they are produced. Thus, each *picture* is composed of a variable number of slices.

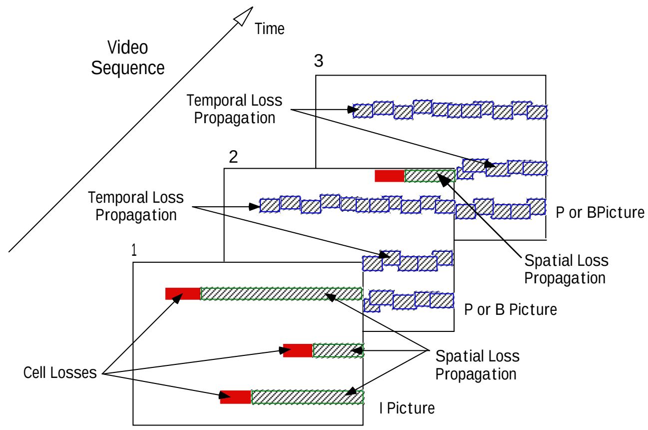

Figure 2. Data loss propagation in MPEG-2 video streams.

Figure 2 shows how network losses map into visual information losses in different types of MPEG frames (I, P or B). Data loss spreads within a single picture up to the next resynchronization point (e.g., picture or slice headers) due to macroblock-to-macroblock differential and variable-length coding. This is referred to as spatial propagation. When loss occurs in a reference picture (I- or P- picture), the lost macroblocks will affect the predicted macroblocks in subsequent frame(s). This is known as temporal propagation.

Error concealment is generally used to reduce the impact of data loss on the visual information. The error concealment algorithms include, for example, spatial interpolation, temporal interpolation and early resynchronization. The MPEG-2 standard[3] proposes an elementary error concealment algorithm based on *motion compensation*. It estimates the motion vectors of the lost macroblock by using the motion vectors of neighbouring macroblocks in the affected picture (provided that these have not also been lost). There is however an obvious problem with lost macroblocks whose neighbours are intra-coded, since usually they do not have associated motion vectors. To get around this problem, the encoding can include motion vectors also for intra macroblocks [*]. Eventhough error concealment may, in general, efficiently decrease the visibility of losses, severe data loss may however still lead to annoying degradations in the decoded video quality.

The robustness of compressed MPEG-2 video may be dramatically increased by judiciously inserting resynchronization points in the bit stream. These can be obtained by extra slice headers to limit the spatial propagation and intra-coded macroblocks to stop the temporal propagation. However, the addition of extra slice headers and/or intra-coded macroblocks is not costless. Under the same video traffic constraints, it reduces indeed the amount of bits available to code pure video information, which affects the quality of the reconstructed video. Similarly, in open-loop VBR transmission (i.e., constant quantizer scale for the whole video sequence), it increases the bit rate to be sent throughout the network without affecting the encoding video quality.

3. PROBLEM STATEMENT

We focus on providing the end-user with the best video quality given some networking QoS parameters (i.e., expected packet loss ratio). The video quality the end-user experiences depends both on the quality of the compressed video before transmission and on the degradation due to packet loss. Therefore, this work addresses the optimal trade-off between pure video information (i.e., source coding) and redundant data (i.e., channel coding) under a given bit budget. This is not a trivial problem as the optimal bandwidth distribution (i.e., source vs. channel coding) strongly depends on the underlying local video content. For example, slow-moving video areas are easy to recover from data loss with appropriate error concealment techniques, and thus do not necessitate any particular channel coding attention.

[*] Some MPEG-2 encoder chips automatically produce concealment motion vectors for all macroblocks.

Therefore, we simplify the problem by making the following assumptions. First, we consider the packet loss process affecting the video stream to be independent.[26] The authors in[5] have shown that this loss pattern leads to the worst video quality under the same packet loss ratio. Also, we consider the sender to be aware of both (i) the packet loss ratio measured at the receiver, and (ii) the error concealment technique implemented at the decoder. The knowledge of the error concealment algorithm leads to the optimal scenario. However, if it is not known a-priori, our algorithm would still behave well.

Furthermore, the portion of the total bandwidth aiming at protecting the stream may be divided into two parts: (i) structuring (i.e., limiting MPEG-2 error propagation) and, (ii) FEC-based protection. In the remainder, we develop AMISP, a content-based FEC technique sitting on top of an adaptive structuring algorithm.

4. ADAPTIVE MPEG-2 STRUCTURING

This section presents the Adaptive MPEG-2 Structuring scheme[25] used in our algorithm. The way macroblock loss probabilities are computed is first presented. The structuring algorithm itself is then briefly described.

4.1. Loss Probability Matrices

It has been noticed that a macroblock may be damaged in any of the three following cases:

- (i) it belongs to a packet that has been lost during transmission
- (ii) it belongs to a slice that has been affected by a packet loss (spatial propagation)
- (iii) it is temporally dependent on a damaged macroblock of a previous reference frame (temporal propagation).

which are discussed below.

The first factor that may affect a macroblock is the transmission error. If we assume a uniform loss pattern, the probability θ for a packet to be lost is given by the PLR experienced on the network. Therefore, without any other information about the packet loss process, every packet has the same average probability to be lost, $\theta = PLR$. Let us now call $B_n(i,j)$, the macroblock at the i^{th} column and the j^{th} row of a given frame n. Under the assumption that a macroblock is lost as soon as part of it is missing, the probability $\lambda_{\mathbf{n}}(\mathbf{i},\mathbf{j})$ for the macroblock $B_n(i,j)$ to be lost is given by:

$$\lambda_{\mathbf{n}}(i,j) = \theta\ \mathbf{N_n}(i,j), \quad \forall\,(i,j)\,|\,1 \leq i \leq B_{row}\ and\ 1 \leq j \leq B_{column}. \tag{1}$$

where $\mathbf{N_n}(i,j)$ is the number of packets containing the macroblock $B_n(i,j)$. B_{row} and B_{column} are respectively the number of macroblocks per frame row and column.

The second factor that may affect a macroblock is *spatial propagation*. In case of transmission error, an MPEG-2 decoder skips all video information up to the next slice header, which acts as a spatial resynchronization point. Consequently, when a macroblock is lost within a slice, all subsequent macroblocks of the same slice are considered as being damaged, even if they do not belong to the lost packet.

Thus, for a given frame n, the probability $\mathbf{P_n}(i,j)$ for a macroblock $B_n(i,j)$ not to be correctly decoded (transmission error + spatial propagation) is given by:

$$\begin{aligned} \mathbf{P_n}(i,j) &= \lambda_{\mathbf{n}}(i,j) + \theta\ \mathbf{M_n}(i,j) \\ &= \theta\ [\mathbf{N_n}(i,j) + \mathbf{M_n}(i,j)], \\ &\quad \forall\,(i,j)\,|\,1 \leq i \leq B_{row}\ and\ 1 \leq j \leq B_{column}. \end{aligned} \tag{2}$$

where, $\mathbf{M_n}(i,j)$ represents the number of packets within the same slice before the first packet related to $B_n(i,j)$ (see Figure 3).

There is an exception to this rule. Indeed, according to the MPEG-2 syntax, every frame is preceded by a header. If a packet containing a frame header is lost, the entire frame is skipped, making Eq. (2) useless. We assume this case to be rare enough to be neglected. This assumption is enforced when these headers are protected via a specific FEC scheme.[27]

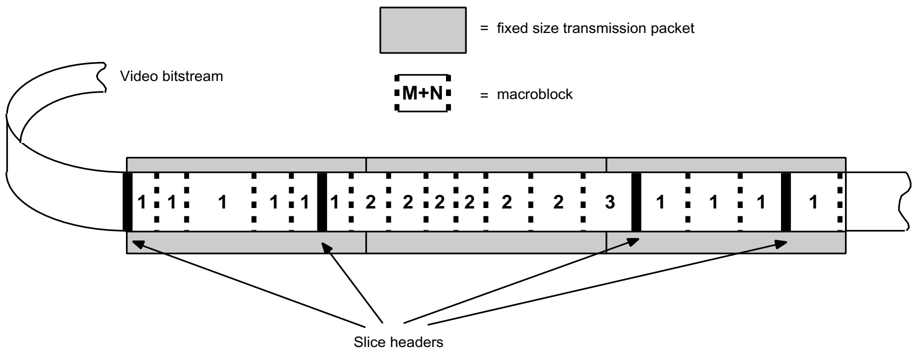

Figure 3. Illustration of $\mathbf{P_n}(i, j)$. The numbers in each macroblock represent $M_n + N_n$ values.

The third factor that may affect a macroblock is *temporal propagation*[19] (see Section 2). This kind of degradation appears when a macroblock motion vector points to an erroneous picture area in the reference frame. Let $\mathcal{E}_{\mathbf{n}}^{\mathbf{n-k}}$ denote the probability for the pixel p to be affected by loss in frame k. Roughly, its computation has to take into account the motion vectors chosen by the encoder, and the loss probability is reset by Intra-coding. Please refer to [25,28] for a detailed formulation of this probability.

4.2. The Structuring Algorithm: AMIS

The Adaptive MPEG-2 Information Structuring (AMIS) algorithm works as follows: an extra resynchronization point is inserted in the bit stream whenever hypothetical data loss would lead to video degradation above a desired threshold, after error concealment.

The *mean luminance difference* (MLD) has been chosen as distortion measure. It corresponds to the simplest metric correlated with human perception[10] (under the assumption that the viewer stands far enough from the monitor). The MLD for $B(i, j)$ is defined as follows:

$$\delta(i,j) = \frac{1}{256}\sum_{p=1}^{256} B^p(i,j) - \frac{1}{256}\sum_{p=1}^{256} \widetilde{B}^p(i,j), \tag{3}$$

where, (i, j) is the macroblock position in the frame and p is the pixel position in the corresponding macroblock. $B(i, j)$ and $\widetilde{B}(i, j)$ represent respectively a correctly (error-free) decoded macroblock and the corresponding damaged macroblock.

AMIS is divided in two distinct parts: (i) the *spatial* part, which deals with slice headers insertion, and (ii) the *temporal* part, which is in charge of deciding when a macroblock should be intra-coded. Indeed, inserting extra slice headers has no effect on temporal error propagation. Also, adding intra-coded macroblock does not help in limiting the spatial error propagation. Therefore, these two parts are considered independently. However, it is clear that the slice structure of reference frames may influence the insertion decision of intra-coded macroblocks.

AMIS-Spatial :

The *spatial* part of AMIS aims at limiting the spatial error propagation, or at least its visible degradation. It introduces an extra slice header as soon as the distortion due to hypothetical loss reaches a given threshold, Δ_s. Clearly a new slice is inserted as soon as:

$$\sum_{B_n(i,j)\in S} \delta_{\mathbf{n}}{}^{\mathbf{s}}(i,j)\ \mathbf{P_n}(i,j) \geq \Delta_s, \tag{4}$$

where, $B_n(i, j)$ is the current macroblock belonging to slice S and, $\delta_{\mathbf{n}}{}^{\mathbf{s}}(i, j)$ corresponds to the expected MLD in case $B_n(i, j)$ was damaged. $\mathbf{P_n}(i, j)$, defined in Eq. (2), represents the probability for $B(i, j)$ to be spatially damaged, by packet loss or spatial propagation.

Actually, the expected distortion is weighted by its likelihood to occur. There is indeed no need to protect an area not likely to be lost, even if the corresponding distortion would be high. The spatial threshold Δ_s regulates the acceptable level of distortion. The smaller the threshold, the higher the number of slices.

AMIS-Spatial also takes the packetization process into account: no more than one slice header is encapsulated in the same network loss entity.[15]

AMIS-Temporal :

The *temporal* part of AMIS is more complex. First, let us assume that losses in different reference frames can be considered independently in regard to their impact on the current frame. Even though not completely correct, this assumption places the encoding process in the worst case from the distortion point of view. It will tend to generate more protection than *effectively* needed, but greatly simplifies the AMIS mechanism.

AMIS-Temporal analyzes every single macroblock and decides whether or not to intra-code it. Again, this decision depends on the macroblock distortion due to temporal propagation of data loss.

The decision may be expressed as follows. The distortion due to temporal error propagation is weighted by the corresponding loss probability matrix and compared to a threshold Δ_t. This weighted distortion is obtained by summing the effects of uniformly-distributed packet losses in every single previous reference frame, up to the last intra-coded picture $(n-I)$. Finally, the condition for a macroblock $B_n(i,j)$ to be intra-coded in frame n is given by:

$$\sum_{k=1}^{I} \left(\frac{1}{256} \sum_{p \in B_n(i,j)} \mathcal{E}_{\mathbf{n}}^{\mathbf{n-k}}(p)\ \delta_{\mathbf{n},\mathbf{k}}{}^{\mathbf{t}}(i,j) \right) \geq \Delta_t, \tag{5}$$

where $\mathcal{E}_{\mathbf{n}}^{\mathbf{n-k}}$ represents the probability for the pixel p to be affected by loss in frame k.[25] The expected MLD between the current MB correctly decoded and its substitute in case of loss in the reference frame k is given by $\delta_{\mathbf{n},\mathbf{k}}{}^{\mathbf{t}}(i,j)$. Again, the temporal threshold Δ_t regulates the acceptable level of distortion. The smaller the threshold, the higher the number of intra-coded macroblocks.

5. ADAPTIVE MPEG-2 STRUCTURING AND PROTECTION

5.1. On-the-fly Adaptive Protection

The structuring scheme presented here above limits loss propagation within the sequence, but does not avoid losses. Its efficiency is limited to low loss ratios. For higher loss rates, there are two major ways to fight against losses: retransmission schemes (e.g., ARQ) and Forward Error Correction (FEC). Considering interactive video applications, FEC is more appropriate than retransmission since it introduces smaller delays.

Forward Error Correction means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender. Due to the low bit error rates associated with the modern communication media, the assumption is made that decoding is mainly impeded by packet loss (i.e., erasures) caused by network congestion and the resultant buffer overflow and queuing delay. In this case, packet-level FEC schemes[29–32] usually protect k_{FEC} video packets by means of $n_{FEC} - k_{FEC}$ redundancy packets (see Figure 4).

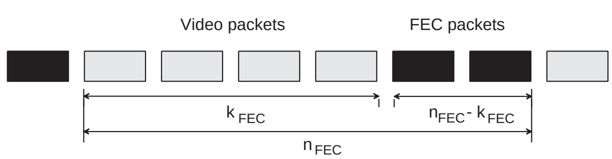

Figure 4. Media-independent FEC scheme.

Such a scheme is able to recover up to $n_{FEC} - k_{FEC}$ lost packets in a block of n_{FEC} packets. The video packet loss process is then modified, and the resulting packet loss ratio for FEC-protected packets becomes θ_{FEC}. Under the assumption of independent losses, θ_{FEC}, is given by[33]:

$$\theta_{FEC} = \theta \left[1 - \sum_{i=0}^{n_{FEC}-k_{FEC}-1} \binom{n_{FEC}-1}{i} \theta^i \ (1-\theta)^{n_{FEC}-i-1} \right], \tag{6}$$

It is clear from the previous equation that FEC algorithms do not guarantee a perfect reconstruction of the video. Usage of FEC thus results from a trade off between delay and reliability. This shows the need of an underlying error resilient encoding scheme to limit remaining losses degradations (i.e., AMIS).

On the other hand, FEC introduces an overhead of $\frac{n_{FEC}-k_{FEC}}{n_{FEC}}$. This overhead could be very important for schemes built to be robust at high loss ratio (i.e., large values of $n_{FEC} - k_{FEC}$). In the same time, network conditions could be highly variable in the today Internet, so that high overhead are not always needed. Optimal FEC overhead should then adapts to network conditions, and the FEC schemes should be able to follow loss ratio fluctuations.[34] However, this kind of evolution towards an adaptive algorithm is based on the assumption that the network PLR is known. If it is not the case, the protection could be totally inefficient. This also arises the need for an underlying video error resilient encoding.

Finally, similarly to the structuring problem, optimal FEC scheme should also take into account the video content. Indeed not all video packets have the same perceptual relevance, following the scene content or the frame type in MPEG-2,[35] for example. The needs for adaptivity to both network conditions and video-content lead to the AMISP algorithm presented in the next section.

5.2. The adaptive protection algorithm: AMISP

The proposed adaptive protection algorithm is the following. During the encoding process, a packet p is marked to be protected whenever its hypothetical loss would introduce an unacceptable degradation. Similarly to Eq. (4), the loss probability weighted distortion is compared to a third threshold Δ_{FEC}:

$$\sum_{B_n(i,j)\in p} \delta_s(i,j)\ \theta \geq \Delta_{FEC}, \tag{7}$$

Whenever AMISP decides to protect a packet, it triggers the underlying network adaptation layer (NAL). The NAL starts counting k_{FEC} video packets and then inserts $n_{FEC} - k_{FEC}$ FEC packets in the MPEG-2 bit stream. Of course, if the elected packet already belongs to a FEC block, no additional overhead is inserted. Like in the structuring scheme, the amount of redundancy is driven by the threshold Δ_{FEC} which represents the Quality of Service desired at the receiver. The adaptive FEC algorithm is easily implemented on RTP protocols, thanks to the support for FEC protection.[36]

The structuring part of AMISP still works in the same manner. However, the macroblock loss probability $\mathbf{P_n}$ (see Eq. (2)) becomes $\widetilde{\mathbf{P_n}}$ and is now given by:

$$\widetilde{\mathbf{P_n}}(i,j) = \sum_{p=1}^{N_n(i,j)} \theta_p + \sum_{p=1}^{M_n(i,j)} \theta_p,\ with\ \theta_p\ \in\ \{\theta, \theta_{FEC}\}, \tag{8}$$

where $M_n(i,j)$ still represents the number of RTP packets within the same slice before and excluding $B_n(i,j)$. $N_n(i,j)$ represents the number of packets containing part of the macroblock $B_n(i,j)$. The packet loss probability θ_p is either equal to θ_{FEC} or θ, depending on whether the packet is FEC-protected or not.

It has to be noticed that packet are FEC-protected in regard to their influence onto spatial distortion. These packets very likely contain a slice header due to the similarity between relations (4) and (7). However, the temporal propagation phenomenon is not taken into account by the protection decision process. The reasons of this choice are twofold. First, the temporal propagation of an error in the current frame cannot be predicted. Second, it can be assumed that the most relevant packets (i.e., FEC-protected packets) are the packets that would also cause the highest temporal distortion.

Finally, the video source rate control has also been slightly modified to take into account FEC packets generation. The total source rate (i.e, video and protection information) has indeed to respect bandwidth constraints. Therefore, a FEC packet insertion directly reduces the bit rate available to code video DCT coefficients. Hence, the larger the number of FEC packets, the smaller the bit rate available to code pure video information.

5.3. FEC parameters

One issue was not addressed in the previous section: the choice of the FEC parameters n_{FEC} and k_{FEC}. Several criteria have to be considered. First, the overhead $\frac{n_{FEC}-k_{FEC}}{n_{FEC}}$ as to be kept as small as possible and to be adapted to the expected loss ratio. However, this ratio does not need to be very large to ensure a large recovery probability. It has been shown indeed that, even for a small $\frac{n_{FEC}-k_{FEC}}{n_{FEC}}$ ratio, FEC can be very effective and reduces the loss probability by several orders of magnitude.[37] Figure 5 shows that small overhead can lead to very low loss probabilities after FEC reconstruction, especially for low to medium loss ratios. Moreover, decreasing the overhead (i.e., increasing k_{FEC}) does not lead to a proportional increase in loss probability after FEC reconstruction. The ratio efficiency-overhead is then larger for large k_{FEC} values, assuming losses occur independently.

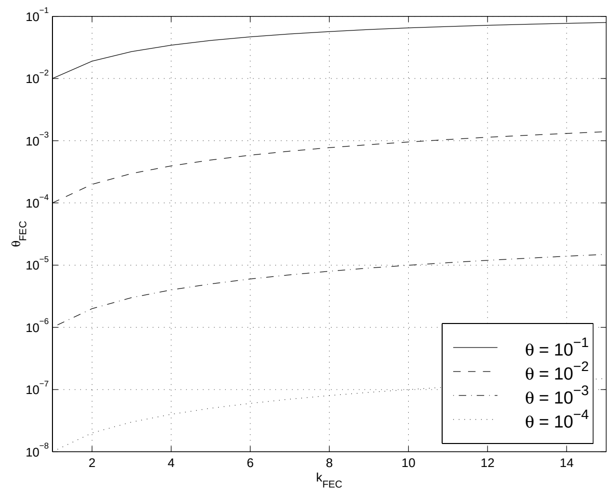

Figure 5. Packet loss probability after FEC reconstruction, versus FEC parameter k_{FEC}, for $n_{FEC} - k_{FEC} = 1$.

Second, the FEC scheme has to meet strict delay constraints in interactive applications. The delay introduced by FEC should not be much larger than one frame, since other delays are also introduced along the transmission path. Since one TS packets represents already a delay of about 5.6 ms in a 6 $Mbit/s$ connection, n_{FEC} should not be larger than 10 to 15 packets. This value is however directly dependent on the bit rate.

Third, it has been shown that for a given overhead, large values of k_{FEC} lead to the best reconstruction probabilities.[17] In the other hand, small k_{FEC} values ensure a more efficient protection of elected packets. All the previous statements suggests that the value of k_{FEC} should be chosen as large as possible, given some delay constraints. Then n_{FEC} value should be computed accordingly to offer a sufficient protection, but also to minimize the overhead.

Finally, FEC parameters could vary dynamically according to loss patterns (i.e., PLR and average length of burst of errors). On-going work is currently trying to optimize these parameters according to network conditions and the degree of protection accuracy. For sake of simplicity, $n_{FEC} - n_{FEC} = 1$ in the following experiments. This allows moreover a very simple and rapid exclusive-OR based FEC scheme.

6. EXPERIMENTAL RESULTS

6.1. Experimental Setup

The setup used to measure the performances of our algorithm is the following.

The input data is a 400-frame long sequence conforming to the ITU-R 601 format (TV-resolution, 720 × 576@25fps). It includes five video scenes that differ in terms of spatial and temporal complexities. An MPEG-2 video encoder based on the Test Model v5[38] has been used to encode the data as interlaced video with a structure of 12 images per GoP and 2 B-pictures between every reference picture. The sequence is encoded in CBR mode, at a total bit rate of 6 Mbps. Motion vectors are produced for each macroblock for the motion compensated error concealment (see Section 2). An MPEG-2 transport stream encoder[39] encapsulates the encoded MPEG-2 bit stream into 18800-byte packetized elementary stream (PES) packets. These are then divided into fixed-length transport stream (TS) packets (i.e., 188 bytes including the TS header).

A network adaptation layer (NAL) provides selective forward error correction (FEC) and error detection capabilities. FEC packets decrease the bit rate of MPEG-2 sequence so that the total bit rate remains unchanged. The real-time Internet protocol stack including the Real-Time Protocol (RTP), UDP and IP is considered for the bit stream transmission. The RTP/UDP/IP protocol stack has now been widely accepted for the delivery of delay-sensitive and loss-sensitive services over packet networks. In our simulations, every RTP packet contains one MPEG-2 TS packet as in most current implementations.[40]

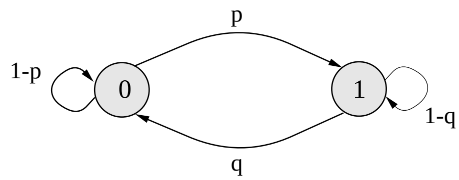

Figure 6. The Gilbert model.

A model-based data loss generator reproduces transmission losses. For this purpose, we used a two-state Markovian model (Gilbert model,[41] see Figure 6). States 0 and 1 correspond respectively to the correct reception and loss of a packet. The transition rates p and q between the states control the lengths of the error bursts. Hence, there are three parameters to be controlled: the fixed packet size (PS), the packet loss ratio (PLR) and the average number of packets lost in a burst of errors (ABL). In our experiments, losses are independent, so that $p = 1 - q$.

The final video quality after error concealment is evaluated by means of the MPQM tool,[42] which proved to behave consistently with human judgments. Indeed, many studies have shown that the usual PSNR metric is poorly correlated with human perception as it does not take visual masking into consideration. The per-frame quality values given by the MPQM tool are then gathered together by means of a Minkowski summation[43] (i.e., weighted average).

6.2. Experimental results and comparisons

Figure 7 first shows the influence of the k_{FEC} parameter onto the final video quality. At low loss ratios, all schemes have the same behavior thanks to the adaptivity feature of AMISP. The protection scheme generates almost the same number of redundancy packets in each case. However, at high loss rates, smaller k_{FEC} values provide the best quality, since a very efficient protection is mandatory for a good decoding. At medium loss rates, larger k_{FEC} values offer the best compromise between efficiency and overhead, and thus lead to the best quality. As stated before, $k_{FEC} = 10$ seems to fit both delay and robustness requirements, at least in the most common θ range (i.e., between 10^{-4} and 5×10^{-2}).

Figure 8 compares AMISP with a basic TM-5 video encoding protected by a regular (by opposition to adaptive) FEC scheme, which generates one redundancy packet every ten video packets. The total bit rate (i.e., video information and FEC overhead) is fixed to 6 Mbps for each transmission. It is clearly visible that AMISP provides a better end-to-end quality over the complete packet loss ratio range. At low θ values, the improvement in quality is mainly due to the adaptivity feature of AMISP: it generates less redundancy, and thus provides more accurate video information. At high loss rates, both schemes perform similarly in terms of protection. However, the quality offered by AMISP is much higher thanks to the underlying structuring scheme (i.e., AMIS). This scheme indeed limits the error propagation within the decoded sequence. These two phenomenon are described further in the following figures.

Figure 9 emphasizes the usefull adaptivity feature to network conditions. The end-to-end quality of AMISP is compared to the same video bit rate but protected by a regular FEC scheme. At low loss ratio AMISP provides a better quality since it does not generate useless overhead. Meanwhile, both algorithms are equivalent at high loss ratios. The number of packets AMISP has to protect becomes very large in these conditions. Hence, the adaptive protection becomes very close to a regular protection scheme.

Finally, Figure 10 demonstrates the advantages of the underlying structuring scheme.[28] AMISP is compared to a TM-5 encoding protected by the same adaptive FEC scheme used in AMISP. It is clear that both algorithms leads to the same quality at low loss rates. Indeed, losses that would cause important degradations are recovered by the FEC algorithm. However, the gap between both schemes grows rapidly with the loss ratio. Indeed, the protection

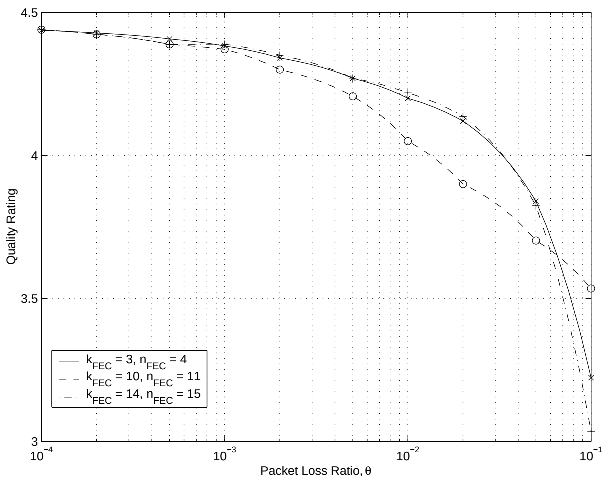

Figure 7. AMISP end-to-end quality versus the packet loss ratio. Comparison of schemes with different k_{FEC} values in terms of quality. The total bit rate is about 6 Mbps.

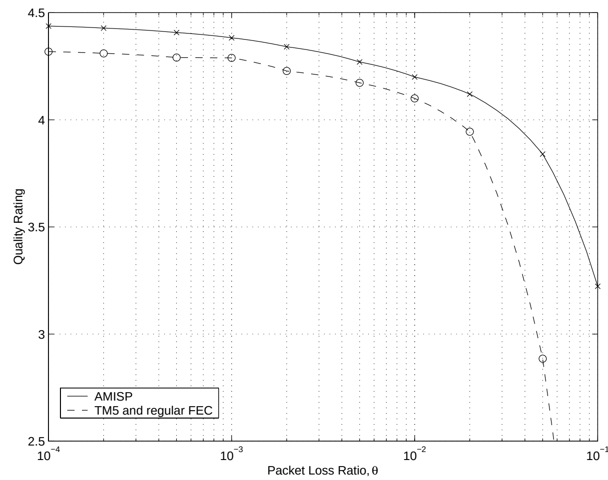

Figure 8. End-to-end quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding and a regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.

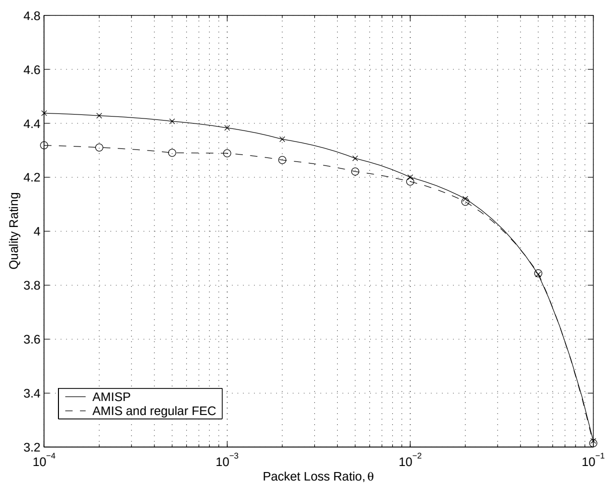

Figure 9. End-to-end quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with an AMIS encoding and a regular FEC scheme ($k = 10$ and $n = 11$). The total bit rate is about 6 Mbps.

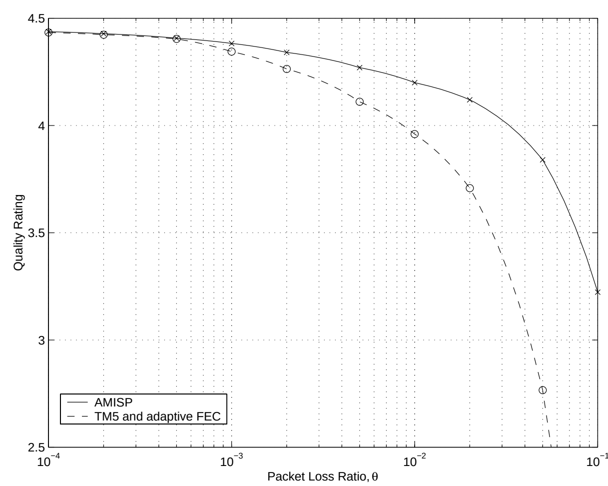

Figure 10. End-to-end quality versus the packet loss ratio. Comparison of the AMISP algorithm ($k_{FEC} = 10$ and $n_{FEC} = 11$) with a TM-5 encoding and the same adaptive FEC scheme. The total bit rate is about 6 Mbps.

algorithm looses some of its efficiency. The errors propagate within the TM-5 sequence, while they are kept to an acceptable level in AMIS.

7. CONCLUSIONS

We have presented a new adaptive error resilient scheme for TV-resolution MPEG-2 video streams interactive delivery, namely AMISP. It includes a media-dependent FEC algorithm relying on an MPEG-2 syntactic structuring technique. A judicious combination of protection redundancy, MPEG syntactic data and pure video information showed to greatly improve the final quality under a given bit budget. Experimental results have shown that AMISP dramatically outperformed existing techniques, thanks to its efficient adaptivity to the network conditions. Moreover, it must be noted that AMISP does not significantly increase the MPEG-2 encoding complexity. Finally, AMISP could also be applied to other video standards at the cost of a few modifications.

REFERENCES

1. Karlsson G., "Asynchronous Transfer of Video," *IEEE Communications Magazine* **34**, pp. 118–126, August 1996.
2. Verscheure O., Garcia Adanez X., Karlsson G. and Hubaux J.-P., "User-Oriented QoS in Packet Video Delivery," *IEEE Network Magazine* **12**, pp. 12–21, November/December 1998.
3. ISO/IEC JTC 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1, 2 and 3.* ISO/IEC 13818, 1996.
4. Haskell B.G., Puri A. and Netravali A.N., *Digital Video: an Introduction to MPEG-2*, Digital Multimedia Standards Series, Chapman and Hall, 1997.
5. Zhang Y.-Q. and Lee X., "Performance of MPEG Codecs in the Presence of Errors," *Journal of Visual Communication and Image Representation* **5**, pp. 378–387, December 1994.
6. Wang Y. and Zhu Q.-F., "Error Control and Concealment for Video Communication: A review," *Proceedings of the IEEE* **86**, pp. 974–997, May 1998.
7. Lee W.S., Pickering M.R., Frater M.R. and Arnold J.F., "Error Resilience in Video and Multiplexing Layers for Very Low Bit-Rate Video Coding Systems," *IEEE Journal on Selected Areas in Communications* **15**, pp. 1764–1774, December 1997.
8. Lopez Fernandez C., Basso A. and Hubaux J.-P., "Error concealment and early resynchronization techniques for MPEG-2 video streams damaged by transmission over ATM networks," in *Proceedings of the SPIE*, vol. 2668, pp. 372–383, 1996.
9. Chen C.-T., "Error Detection and Concealment with an Unsupervised MPEG2 Video Decoder," *Journal of Visual Communication and Image Representation* **6**, pp. 265–279, September 1995.
10. Chen M.-J., Chen L.-G. and Weng R.-M., "Error Concealment of Lost Motion Vectors with Overlapped Motion Compensation," *IEEE Transactions on Circuits and Systems for Video Technology* **7**, pp. 560–563, June 1997.
11. Ghanbari M. and Seferidis V., "Cell-Loss Concealment in ATM Video Codecs," *IEEE Transactions on Circuits and Systems for Video Technology* **3**, pp. 238–247, June 1993.
12. Richardson I.E.G. and Riley M.J., "Varying slice size to improve error tolerance of MPEG video," in *Proceedings of the SPIE*, vol. 2668, pp. 365–371, 1996.
13. McCanne S and Jacobson V., "vic: A Flexible Framework for Packet Video," in *Proceedings of the ACM Multimedia Conference*, pp. 511–522, (San Francisco, CA), November 1995.
14. Kawashima M., Chen C.-T., Jeng F.-C. and Singhal S., "Adaptation of the MPEG Video-Coding Algorithm to Network Applications," *IEEE Transactions on Circuits and Systems for Video Technology* **3**, pp. 261–269, August 1993.
15. Zhang J., Frater M.R., Arnold J.F. and Percival T.M., "MPEG 2 Video Services for Wireless ATM Networks," *Journal on Selected Areas in Communications* **15**, pp. 119–128, January 1997.
16. Batra P. and Chang S.-F., "Effective algorithms for video transmission over wireless channels," *Signal Processing: Image Communication* **12**, pp. 147–166, April 1998.
17. Carle G. and Biersack E.W., "Survey of Error Recovery Techniques for IP-Based Audio-Visual Multicast Applications," *IEEE Network* **11**, pp. 24–36, November/December 1997.
18. K. G., "Layered Error-Control Coding for IP Multicast," in *Proceedings of the Third International Workshop on High Performance Protocol Architectures*, (Uppsala, Sweden), June 1997.
19. Wada M., "Selective Recovery of Video Packet Loss Using Error Concealment," *IEEE Journal on Selected Areas in Communications* **7**, pp. 807–814, June 1989.
20. Parthasarathy V., Modestino J. W. and Vastola K. S., "Design of a Transport Coding Scheme for High-Quality Video over ATM Netwroks," *IEEE Transactions on Circuits and Systems for Video Technology* **7**, pp. 358–376, April 1997.
21. Albanese A., Blomer J., Edmonds J., Luby M. and Sudan M., "Priority Encoding Transmission," *IEEE Transactions on Information Theory* **42**, pp. 1737–1744, November 1996.

22. McCanne S., Vetterli M. and Jacobson V., "Low-complexity Video Coding for Receiver-driven Layered Multicast," *IEEE Journal on Selected Areas in Communications* **16**(7), pp. 983–1001, 1997.
23. Jiang W. and Ortega A., "Multiple Description Coding via Polyphase Transform and Selective Quantization," in *Proceedings of Visual Communications and Image Processing, VCIP'99*, (San Jose, USA), January 1999.
24. Verscheure O. and Frossard P., "Perceptual MPEG-2 Syntactic Information Coding: A Performance Study based on Psychophysics," in *Picture Coding Symposium*, pp. 297–302, (Berlin), September 1997.
25. Frossard P. and Verscheure O., "AMIS: Adaptive MPEG-2 Information Structuring," in *SPIE International Symposium on Voice, Video, and Data Communications*, vol. 3528, pp. 113–123, (Boston, USA), November 1998.
26. Bolot J.-C., "End-to-End Packet Delay and Loss Behavior in the Internet," *ACM SIGCOMM Computer Communication Review* **23**, pp. 289–298, October 1993.
27. Garcia Adanez X., *Designing New Network Adaptation and ATM Adaptation Layers for Interactive Multimedia Applications.* PhD thesis, Swiss Federal Institute of Technology, Lausanne, Switerland, 1998.
28. Frossard P. and Verscheure O., "AMISP: A Complete Content-Based MPEG-2 Error-Resilient Scheme," *IEEE Journal on Selected Areas in Communications* , 1999. Submitted paper to Special Issue on Error-Resilient Image and Video Transmission.
29. Rabin M.O., "Efficient Dispersal of Information for Security, Load Balancing and Fault Tolerance," *Journal of the ACM* **36**, pp. 335–348, April 1989.
30. Rizzo L., "Effective Erasure Codes for Reliable Computer Communication Protocols," *ACM Computer Communication Review* **27**, pp. 24–36, April 1997.
31. Lin S. and Costello D.J.jr, *Error control coding: fundamentals and applications*, Prentice-Hall, 1983.
32. Shacham N. and McKenney P., "Packet recovery in High-Speed Networks Using Coding and Buffer Management," in *Proceedings of the IEEE INFOCOM Conference*, vol. 1, pp. 124–131, (Los Alamitos, CA), 1990.
33. Nonnenmacher J., Biersack E.W. and Towsley D., "Parity-Based Loss Recovery for Reliable Multicast Transmission," *IEEE/ACM Transactions on Networking* **6**, pp. 349–361, August 1998.
34. Park K. and Wang W., "AFEC; An Adaptive Forward Erroc Correction Protocol for End-to-End Transport of Real-Time Traffic," in *Proceedings of the IEEE IC3N*, pp. 196–205, (Los Alamitos, CA), 1998.
35. Park K. and Wang W., "QoS-Sensitive Transport of Real-Time MPEG Video using Adaptive Forward Error Correction," in *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pp. 426–432, (Florence, Italy), June 1999.
36. Rosenberg J. and Schulzrinne H., "An RTP Payload Format for Generic Forward Error Correction," Internet Draft, IETF, 1999.
37. Biersack E., "Performance Evaluation of Forward Error Correction in an ATM Environment," *IEEE Journal on Selected Areas in Communications* **11**, pp. 631–640, May 1993.
38. Fogg C., "mpeg2encode/mpeg2decode," in *MPEG Software Simulation Group*, 1996.
39. ISO International Standard, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1: Systems Specification.* ISO/IEC 13818-1, 1996.
40. Hoffman D., Fernando G., Goyal V. and Civanlar R., "RTP Payload Format for MPEG1/MPEG2 Video," Tech. Rep. RFC 2250, Internet Engineering Task Force, January 1998.
41. Gilbert E.N., "Capacity of a Burst-Noise Channel," *The Bell System Technical Journal* , pp. 1253–1265, September 1960.
42. van den Branden Lambrecht C. J. and Verscheure O., "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," in *Proceedings of the SPIE*, vol. 2668, pp. 450–461, January 1996.
43. van den Branden Lambrecht C.J., "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), May 1996.