Joint Source/FEC Rate Selection for Quality-Optimal MPEG-2 Video Delivery

Pascal Frossard, *Member, IEEE,* and Olivier Verscheure, *Member, IEEE*

***Abstract*—This paper deals with the optimal allocation of MPEG-2 encoding and media-independent forward error correction (FEC) rates under a total given bandwidth. The optimality is defined in terms of minimum perceptual distortion given a set of video and network parameters. We first derive the set of equations leading to the residual loss process parameters. That is, the packet loss ratio (PLR) and the average burst length after FEC decoding. We then show that the perceptual source distortion decreases exponentially with the increasing MPEG-2 source rate. We also demonstrate that the perceptual distortion due to data loss is directly proportional to the number of lost macroblocks, and therefore decreases with the amount of channel protection. Finally, we derive the global set of equations that lead to the optimal dynamic rate allocation. The optimal distribution is shown to outperform classical FEC scheme, thanks to its adaptivity to the scene complexity, the available bandwidth and to the network performance. Furthermore, our approach holds for any standard video compression algorithms (i.e., MPEG-x, H.26x).**

***Index Terms*—End-to-end quality, lossy networks, media-independent FEC, perceptual quality, rate selection, video services.**

I. Introduction

STREAMED digital video has already begun its penetration in the market. Transmitting video in digital form is the direct result of the benefits offered by digital compression. The purpose of compression is data rate reduction, which results in lower transmission costs. Although many compression schemes have been studied, a standard was necessary for widespread communications. For digital video, attention is now being focused on the MPEG-2 standard. MPEG-2 aims at diverse applications such as television broadcast over satellite, cable and other broadcast channels (DVB), and digital storage media (DVD).

In general, the distortion the end-user perceives results from compression artifacts, packet losses, delays, and delay jitters. All lossy compression schemes distort and delay the signal. Degradation mainly comes from the quantization, which is the only irreversible process in a coding scheme. Moreover, delays and packet losses are inevitable during transfers across today's networks. The delay is generally due to propagation and queuing. Information loss is mainly caused by multiplexing overloads of high magnitude and duration that lead to buffer overflow in the nodes. Data loss is particularly annoying in video streaming applications due to the predictive structure of MPEG-2 compression, as described in the next section.

Interactive video delivery can significantly be improved by providing sender-side mechanisms [1]. These include 1) structuring techniques [2]–[4] and scalable coding [5], [6] to reduce data loss sensitivity, and 2) forward error correction (FEC) mechanisms to lower the probability of loss at the application layer. FEC 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. This paper explores the media-independent FEC scheme [7]. That is, k video packets are protected by $n-k$ FEC packets within a total of n, regardless of the underlying video sensitivity [8].

Clearly, under a given channel rate, the addition of FEC packets reduces the available rate for source coding. The aim of this paper is to determine the optimal tradeoff between source coding and FEC rates. The optimal distribution changes along the sequence because of varying video and network parameters. A similar problem was addressed in [9] for wavelet-coded images.

The problem is stated in the following manner. Let $R(t)$ denote the channel rate available for transmission at time t for a given time slot. Let $R_S(t)$ and $R_{FEC}(t)$ further denote the MPEG-2 source rate and the rate of FEC packets. Given a set of video and network parameters, the problem is then to *find* the optimal values of $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion[1] *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$.

The paper is organized as follows: Section II briefly describes the transmission of MPEG-2 video streams over packet networks. The source perceptual distortion-rate function is empirically computed in Section III. Section IV then analyzes the FEC efficiency. The residual video loss patterns after FEC reconstruction are computed in the case of a Gilbert-model loss process. Section V studies the error propagation and derives the degradation due to loss from the video loss patterns after FEC recovery. The optimal rate distribution is the topic of Section VI. Section VII shows how our algorithm behaves in different conditions, and compares it to classical FEC schemes in terms of end-to-end distortion. Finally, concluding remarks are given in Section VIII.

II. MPEG-2 Over Packet Networks

In this section, we briefly present the MPEG-2 encoding algorithm. We then explain why data loss is particularly annoying in MPEG-2 streaming applications.

[1]We call "end-to-end video distortion" the distortion as perceived by the end-user.

Manuscript received December 17, 1999; revised May 14, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Henning G. Schulzrinne.

P. Frossard was with the Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland. He is now with the IBM T. J. Watson Research Center, Hawthorne, NY 10532 USA (e-mail: frossard@us.ibm.com).

O. Verscheure is with the IBM T. J. Watson Research Center, Hawthorne, NY 10532 USA (e-mail: ov1@us.ibm.com).

Publisher Item Identifier S 1057-7149(01)10566-X.

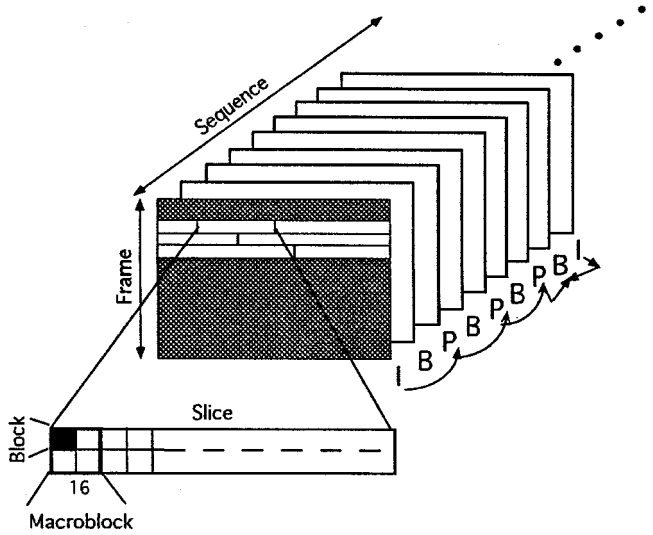

Fig. 1. MPEG-2 video structure.

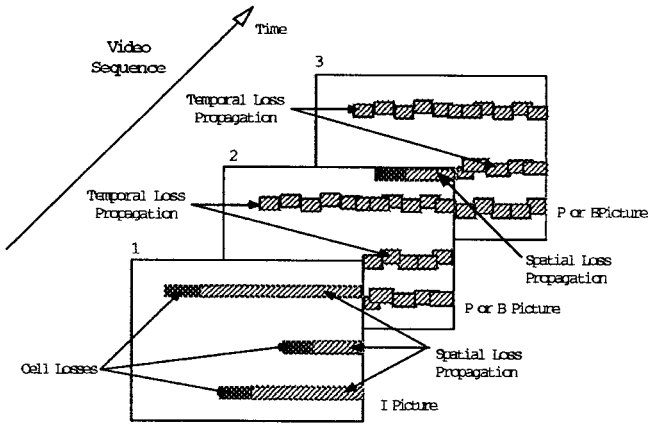

Fig. 2. Data loss propagation in MPEG-2 video streams.

A. MPEG-2 Backgrounder

An MPEG-2 video stream is hierarchically structured, as illustrated in Fig. 1. The stream consists of a sequence composed of several frames. The MPEG-2 video standard defines three different types of frames: intra-coded (I-), predicted (P-) and bidirectional (B-) frames. The use of these three frame types allows MPEG-2 to be robust as I-frames provide error propagation reset points and efficient as B- and P-frames allow a good overall compression ratio. Each frame is composed of slices which are series of macroblocks. Each macroblock (16×16 pixels) contains four blocks (8×8 pixels) of luminance and two, four, or eight blocks of chrominance depending on the chroma format. Motion estimation is performed on macroblocks while the discrete cosine transform (DCT) is calculated on blocks. The resulting DCT coefficients are quantized and variable length coded. The quantizer results from the multiplication of a quantizer scale, MQUANT, and the corresponding element of a quantizer matrix. In general, the higher the MQUANT value, the lower the bit rate but also the lower the quality (well known from the rate-distorsion theory). The resulting video stream finally feeds a packetizer for network transmission (i.e., packet video).

B. MPEG-2 Sensitivity to Data Loss

Fig. 2 illustrates how network losses map onto visual information losses in different types of pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g., picture or slice headers) mainly due to the use of differential coding, run-length coding and variable length coding. This is referred to as spatial propagation and may damage any type of picture. When loss occurs in a reference picture (intra-coded or predictive frame), the damaged macroblocks will affect the non intra-coded macroblocks in subsequent frame(s), which reference the errored macroblocks. This is due to inter-frame predictions and known as temporal propagation.

However, the error visibility may be dramatically reduced by means of error concealment techniques [10]. These error concealment algorithms include, for example, spatial interpolation, temporal interpolation, and early resynchronization techniques [11]. The MPEG-2 standard proposes an elementary error concealment algorithm based on motion compensation. Basically, it estimates the motion vectors for the lost macroblock by using the motion vectors of neighboring macroblocks in the affected picture (provided that these have not also been lost). This improves the concealment in moving picture areas. However, there is an obvious problem with errors in macroblocks whose neighboring macroblocks are intra-coded, because there are ordinarily no motion vectors associated with them. To circumvent this problem, the encoding process can be extended to include motion vectors for intra macroblocks.[2]

In general, error concealment techniques may efficiently decrease the sensitivity to data loss. However, none of these techniques is perfect. Data loss may still involve annoying degradation in the decoded video.

III. MPEG-2 Perceptual Distortion-Rate Function

Several studies have already been conducted on the analysis of the rate-distortion curve for MPEG codecs [12]–[14]. They lead to the conclusion that the distortion evolves somehow exponentially with the decreasing source rate [15]. However, video distortion was measured by means of pure mathematical metrics such as the MSE.

Experimental results (see Fig. 3) shows that the perceptual video distortion varies exponentially with the quantizer scale factor (i.e., $MQUANT$). The TV-resolution video sequence was encoded in open-loop VBR mode (OL-VBR) using a TM-5 MPEG-2 video coder [16]. The video distortion was measured by means of the perceptual distortion metric (PDM) tool [17], which proved to behave consistently with human judgments [18]. This tool relies on a model of the human visual system (HVS) [19], [20]. We can indeed see that the exponential fitting perfectly matches the experimental data.

Moreover, the average source rate R_S is also evolving exponentially with the $MQUANT$ [21]. Therefore, the source perceptual distortion-rate function can be expressed as (see Fig. 4)

$$D_S(R_S) = \chi_S R_S^{\xi_S} \tag{1}$$

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

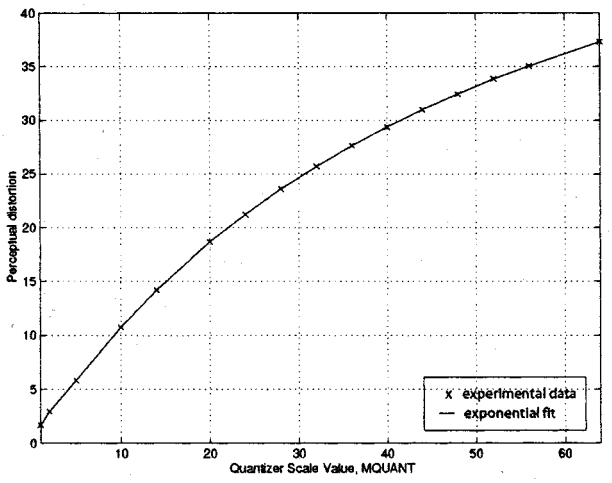

Fig. 3. Perceptual distortion as a function of the quantizer scale MQUANT.

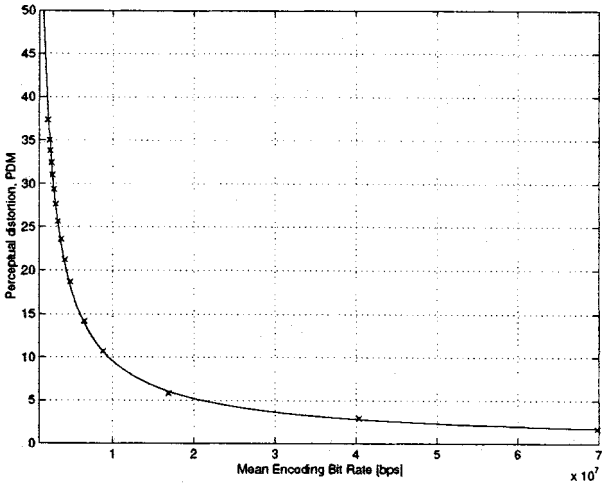

Fig. 4. Perceptual distortion as a function of the mean encoding bit rate. Fitting parameters of (1): $\chi_S = 14.59\ 10^6$ and $\xi_S = -0.883$.

where the parameter ξ_S is related to the encoding complexity of the set of frames under consideration. The parameter χ_S is a constant used for accurate fitting. Indeed, (1) can be rewritten as

$$\log D_S(R_S) = \log \chi_S + \xi_S \log R_S. \tag{2}$$

Extremely low values of R_S must lead to a constant distortion D_S (i.e., independent of the video sequence). Therefore, from (2), χ_S may indeed be represented by a constant while ξ_S can be measured off-line, or picked from a set of values estimated from a catalog of test sequences. Details of the interpretation of (1) can be found in [22]. This relation holds at the sequence, group of pictures, and even frame level.

Finally, it is important to note that Fig. 4 exhibits an important, though trivial, behavior: for MPEG-2 source rates below 10 Mb/s, a small increase in source rate leads to a great decrease in perceptual distortion.

IV. Loss Process Parameters After FEC Recovery

A. Packet-Level FEC

FEC techniques are the preferred error-control scheme for multicast, interactive, or broadcast applications [23]. In this paper, a very simple media-independent FEC mechanism is used. 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. Either the packet is present and correct or it is lost. These losses are mainly caused by network congestion and the resultant buffer overflow and queuing delay. In this case, packet-level FEC schemes [24]–[27] provide an efficient way to fight against losses, although perfect recovery cannot be guaranteed. The description of the FEC algorithms is outside the scope of this paper but can be found in [28]. Recall, however, that common FEC schemes based on Reed–Solomon codes or X-OR functions can generally correct as many losses as the number of redundancy packets.

Assume every block of k video packets are protected by $(n-k)$ FEC packets as represented in Fig. 5. This is referred to as *media-independent FEC*. If at least k out of n packets are correctly received, the underlying video information can be correctly decoded. Otherwise, none of the lost packets can be recovered by the receiver. Hence, the packet loss pattern experienced at the video level is quite different from the loss pattern observed on the lossy channel.

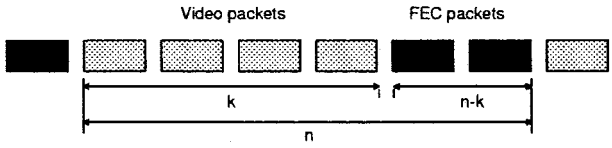

Fig. 5. Media-independent FEC scheme.

The purpose of this section is to compute the video loss process parameters after FEC recovery. These parameters are the packet loss ratio (PLR) π_v and the average burst length α_v. They directly drive the final video quality. The PLR π_v simply represents the ratio between lost and sent video packets, or equivalently the probability for a video packet to be lost. The average burst length α_v is the average length of consecutively lost video packets.

B. FEC Performance in Renewal Error Process

Several studies have been performed to compute the FEC efficiency or the probability for data to be recovered in case of loss [29]–[31]. The probability of recovery is simply given by the probability to have less than $n-k+1$ losses in a block of n packets. However, this parameter does not bring enough information about the loss process after FEC recovery. To correctly model the video quality, at least two parameters (i.e, π_v and α_v) should be computed.

First, assume any packet takes a binary value zero or one, where a zero is for a correctly received packet, and, a one means the packet has been lost or equivalently represents an error. We further assume that the loss process matches a renewal error process. That is, the lengths of consecutive inter-error intervals (also called gaps) are assumed to be independently and uniformly distributed. This assumption appears to be limitative but is extensively used in practice [32]–[35].

Following the development of [29], let $p(i)$ denote the probability that a gap length is $i-1$, i.e., $p(i) = \Pr(0^{i-1}1|1)$, where 0^{i-1} is a shorthand for $i-1$ successive zeros. Similarly, let $P(i)$ denote the probability that at least $i-1$ zeros follow a given error, i.e., $P(i) = \Pr(0^{i-1}|1)$.

Order is irrelevant because of the independence among gap lengths of a renewal process. The events $1\ \ 0^{i-1}$ and $0^{i-1}1$ are therefore equiprobable. From this property, the probability $R(m, n)$ that $m-1$ errors occur in the next $n-1$ packets following an error can be easily computed by recurrence [29]. Thus

$$R(m, n) = \begin{cases} P(n), \text{ for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i) R(m-1, n-i), \text{ for } 2 \leq m \leq n. \end{cases} \tag{3}$$

As mentioned earlier, this probability is not sufficient to model the video loss process. Let $r(m, n)$ denote the probability that $m-1$ errors occur in the $n-1$ packets between two errors

$$r(m, n) = \begin{cases} p(n), \text{ for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} p(i) r(m-1, n-i), \text{ for } 2 \leq m \leq n. \end{cases} \tag{4}$$

Finally, let $\overline{r}(m, n)$ represent the probability that $m-1$ errors occur in the $n-1$ packets following an error and preceding a zero

$$\overline{r}(m, n) = R(m, n) - r(m, n). \tag{5}$$

Let $q(i)$ denote the probability that a burst is of length $i-1$, and $Q(i)$denote the probability that at least $i-1$ ones follow a zero. These probabilities are given by the loss process or can even be deduced from the above variables. The dual of $R(m, n)$, namely, $S(m, n)$, represents the probability to have $m-1$ zeros in the next $n-1$ packets following a zero. This probability is obtained by recurrence from

$$S(m, n) = \begin{cases} Q(n), \text{ for } m = 1 \text{ and } n \geq 1 \\ \sum_{i=1}^{n-m+1} q(i) S(m-1, n-i), \text{ for } 2 \leq m \leq n. \end{cases} \tag{6}$$

The video packet loss rate after FEC recovery is now easy to compute. Two cases are considered with respect to the state of the last video packet of an FEC block. Its loss or its presence directly drives the loss process into the next FEC block. By the renewal process properties, π_v is thus computed by

$$\pi_v = \frac{\pi}{k} \sum_{i=1}^{k} iR(i, k) \sum_{j=\lfloor n-k+1-i \rfloor}^{n-k} R(j+1, n-k+1) + \frac{1-\pi}{k} \sum_{i=1}^{k-1} (k-i)\, S(i, k) \sum_{j=0}^{k-1-i} S(j+1, n-k+1) \tag{7}$$

where the notation $\lfloor x \rfloor$ represents the positive part of x and π represents the global PLR.

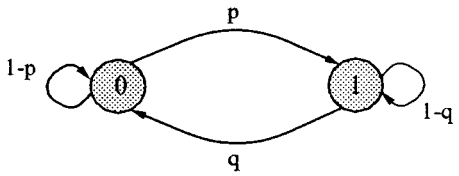

Fig. 6. Two-state Markov chain: Gilbert model.

The average video burst length after FEC recovery, α_v is also computed from the previous development. Since bursts of errors do not have the same probability to start on any packet of the FEC block, each position has to be considered separately. The probability P_j that a burst starts at the jth position of the FEC block is given by

$$P_j = \Pr(x_{j-1} = 0, x_j = 1) \tag{8}$$

where x_i denotes the binary state of packet i (i.e., either lost or not after FEC recovery). The average length L_j of a burst starting at position j in an FEC block is then written as

$$L_j = \frac{\sum_{l=1}^{\infty} l \Pr(x_{j-1} = 0, x_j = 1, \ldots, x_{j+l-1} = 1, x_{j+l} = 0)}{P_j} \tag{9}$$

where only video packets are considered. Indeed FEC packets obviously do not impact the video burst length. The video average burst length α_v is finally given by the following probability-weighted summation:

$$\alpha_v = \frac{\sum_{j=1}^{k} P_j L_j}{\sum_{j=1}^{k} P_j}. \tag{10}$$

Details of the computation appeared in [36]. They are repeated in Appendix I for sake of completeness.

C. FEC Performance in a Gilbert-Model Loss Process

Assume now that the channel loss process can be characterized by the Gilbert model [32]–[35]. Notice that our goal here is not to validate the Gilbert model, but rather to show how, given a channel model, the video loss pattern can be computed. The Gilbert model is a two-state Markovian model [37] with geometrically distributed residence times (see Fig. 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.

The global PLR π corresponds to the stationary probability to be in the loss state: $\pi = p/(p+q)$. The average error burst length α is given by the average residence time in the loss state: $\alpha = 1/q$. These loss process parameters π and α can be sensed through control protocols (e.g., RTCP) or delay measurements [38].

The loss patterns π_v and α_v are easily computed in this case. They obviously depend on both the model parameters p and

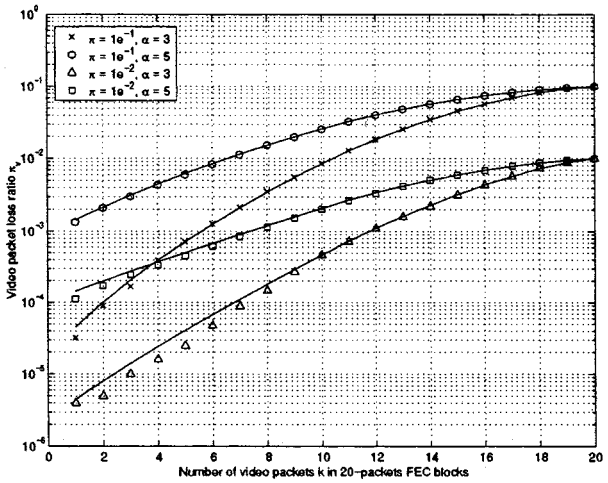

Fig. 7. Evolution of π_v versus the number of video packets k in an FEC block of length $n = 20$ in a Gilbert-model loss process.

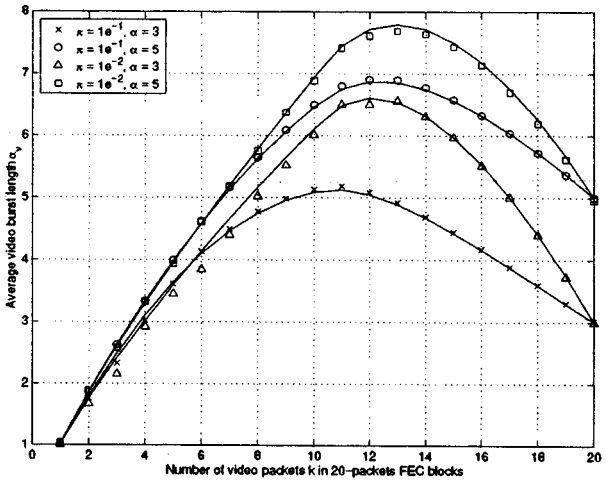

Fig. 8. Evolution of α_v versus the number of video packets k in an FEC block of length $n = 20$ in a Gilbert-model loss process.

q, and the FEC parameters k and n. Indeed, for a Gilbert loss process, the following relations hold:

$$p(i) = \begin{cases} 1-q, & \text{if } i = 1 \\ q(1-p)^{i-2}p, & \text{otherwise} \end{cases}$$

$$P(i) = \begin{cases} 1, & \text{if } i = 1 \\ q(1-p)^{i-2}, & \text{otherwise} \end{cases}$$

$$q(i) = \begin{cases} 1-p, & \text{if } i = 1 \\ p(1-q)^{i-2}q, & \text{otherwise} \end{cases}$$

$$Q(i) = \begin{cases} 1, & \text{if } i = 1 \\ p(1-q)^{i-2}, & \text{otherwise.} \end{cases}$$

The probabilities $R(m, n)$, $r(m, n)$, $\overline{r}(m, n)$, and $S(m, n)$ can be computed by recurrence from (3)–(6), respectively. The video PLR and average burst length are then computed from (7) and (18), respectively. For the remaining of the development, the parameters (π_v, α_v) will be written as $\pi_v(k, n)$ and $\alpha_v(k, n)$ to explicit their dependence on k and n.

Figs. 7 and 8 represent the evolution of the video loss parameters π_v and α_v for different network loss patterns. The analytical values perfectly fit the experimental data. Moreover, π_v obviously increases with k as the amount of protection decreases for a given n. FEC protection becomes also less efficient for bursty loss traffic (i.e., large values of α) for a given π. Moreover, the average length of lost video packets clearly exhibits a maximum. This can be explained as follows. When the amount of protection is very large, α_v stays close to k. When the amount of protection decreases, the video loss pattern gets closer to the channel loss pattern. In between there is a maximum which is less pronounced for bursty process. These behaviors still hold for different FEC block lengths.

V. Perceptual Distortion Under Different Loss Patterns

We now investigate the impact of different packet loss patterns, such as those experienced after FEC recovery, onto video

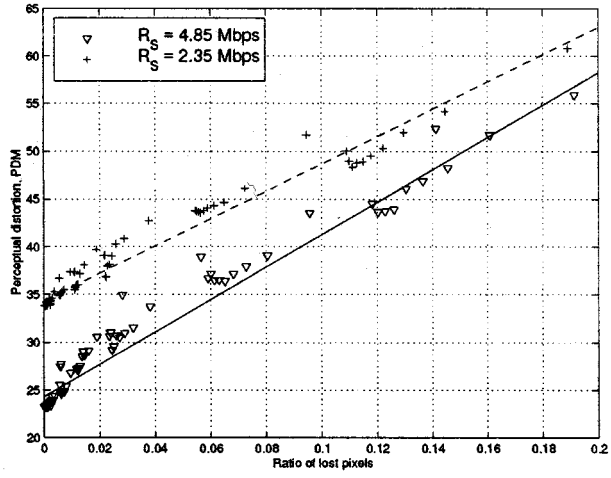

Fig. 9. Perceptual distortion versus percentage of spatially lost pixels for several loss patterns (source rate $R_S \in \{4.85, 2.35\}$ Mb/s).

distortion. Recall that an MPEG-2 macroblock may be damaged in any of the three following cases:

1) it belongs to a video packet that has been lost during transmission;
2) it belongs to a slice that has been affected by a packet loss (*spatial propagation*);
3) it is temporally dependent on a damaged area of a previous reference frame (*temporal propagation*).

Intuitively, the perceptual distortion of the received video is in average proportional to the number of lost macroblocks, hence to the number of lost pixels. To emphasize this, an MPEG-2 transmission system has been simulated using a 400-frame long sequence and a very large set of network loss patterns. Fig. 9 shows the video distortion as a function of the number of lost pixels. The linear proportionality of this result has a correlation factor lying around 0.992. Video distortion is thus indeed directly related to the number of spatially lost macroblocks. Equivalently, the distortion is directly proportional to the number of spatially lost pixels, since macroblocks can only be entirely lost. Hence the perceptual distortion is driven

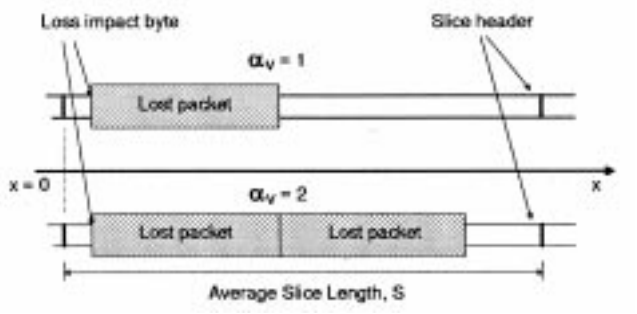

Fig. 10. Spatial error propagation under $\alpha_v = \{1, 2\}$.

by spatial error propagation. Temporal error propagation is indeed also induced by spatially lost macroblocks in reference frames. It is therefore clearly a direct consequence of the lost macroblocks, hence pixels. The number of spatially lost pixels drives the average distortion due to transmission losses.

Let us now further study the spatial error propagation in order to establish a relation between the video loss patterns (i.e., π_v and α_v) and the video distortion. The loss of a single packet can be sufficient to induce the loss of a complete slice due to error propagation.

As an example, Fig. 10 shows how data loss propagates spatially for two different loss patterns (i.e., $\alpha_v = 1$ and $\alpha_v = 2$). Due to differential coding within slices, both cases lead to the same loss length. All bytes between the loss impact byte and the next slice headers are indeed useless. However, in the second case ($\alpha_v = 2$), two lost packets have been absorbed by the spatial error propagation. Therefore, since an entire slice can be fully lost with the loss of a single packet, it is preferable that a burst of lost packets damages the same slice. Accordingly, under a given PLR, bursts of lost packets damage fewer pixels than individual lost packets. Thus the average number of lost pixels, which is directly driven by the rate of lost video bytes $\overline{L}$ and the source rate R_S, decreases inversely proportionally to α_v (see Appendix II). Thus we draw the important conclusion that a uniform and independent loss process corresponds to the worst case with respect to the amount of degradation due to data loss. The probability P_l for pixels to be lost is given by (24) in Appendix II

$$P_l = \frac{\overline{L}}{R_S} = \pi_v + \frac{R_S \pi_v}{2PN_S\alpha_v} \tag{11}$$

where P represents the video packet size (i.e., generally 184 bytes of MPEG-2 video data) and N_S represents the average number of MPEG-2 slices per second.

Now the effect of network losses has to be analyzed in terms of distortion. It has been shown here above that, for a given source rate, the perceptual distortion is proportional to the number of spatially lost pixels (see Fig. 9). Showing the dependence of π_v and α_v from k and n, (11) leads to the distortion D_L due to data loss

$$D_L(k, n) = \chi_L \left(\pi_v(k, n) + \frac{R_S \pi_v(k, n)}{2PN_S\alpha_v(k, n)} \right) \tag{12}$$

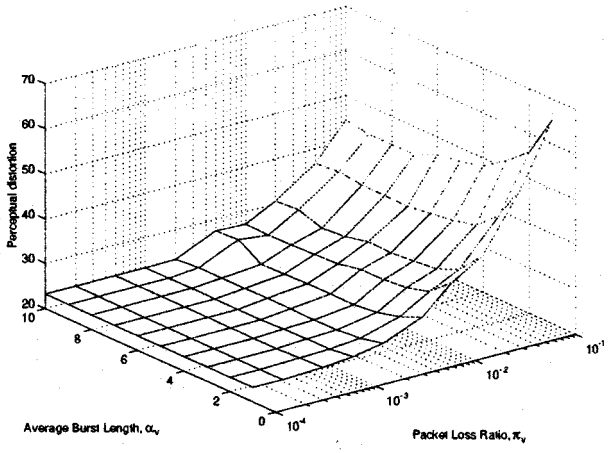

Fig. 11. Perceptual distortion versus the video loss process parameters (source rate $R_S = 4.75$ Mbit/s).

where χ_L is a constant depending on the spatio–temporal complexity of the sequence and the error concealment scheme. It captures the effects of both spatial and temporal error propagation phenomenons whose incidence depends on the nature of the sequence. These parameters can be measured off-line, or picked from a set of values estimated from a catalog of test sequences. The residual loss process parameters π_v and α_v directly depends on the FEC algorithm and the global loss process (see Section IV).

The variation of $D_L(k, n)$ on π_v and α_v as given by (12) is shown in Fig. 11, for a given source rate. It can be seen that the distortion indeed increases linearly with π_v along the logarithmic-scaled x-axis. Also, under a given π_v, the distortion decreases inversely proportionally to α_v. Note that the same behavior has also been observed with the MSE distortion metric. It can be shown that the slope of the perceptual distortion increase varies accordingly to the complexity of the sequence, as expected from (12).

Interestingly, the mapping between the distortion $D_L(k, n)$ and the number of spatially lost macroblocks (i.e., the value of χ_L) is almost independent of the source rate [22]. In other words, the relative distortion of concealed areas is similar in low and high source rates, within the range of MPEG-2 target rates (i.e., 3.5–15 Mb/s). The quality of reconstructed areas through error concealment depends on the accuracy of surrounding video elements, and thus on the source rate. For larger source rate, the reconstructed areas are better approximation of the original data. At the same time the quality of the lossless transmission also increases with the source rate. Hence the relative channel distortion is assumed to be independent of the source rate.

VI. Joint Source/Channel Perceptual Distortion

We now address the initial problem stated in Section I, namely, to find the optimal $R_S(t)$ and $R_{FEC}(t)$ at time t that *minimize* the end-to-end video distortion *under* the constraint $R_S(t) + R_{FEC}(t) \leq R(t)$ *given* a set of video and network parameters. The total bit rate $R(t)$ is one of the inputs to our system. It must be noted that this input may very well be

adjusted to conform to any network policy (e.g., using the flow control algorithm of TCP [39]). Our rate allocation scheme then optimally uses the entire available rate $R(t)$.

From the FEC scheme structure, the source rate is expressed as $R_S(t) = (k/n)R(t)$ and the rate of FEC packets as $R_{FEC}(t) = (n-k)/nR(t)$ (see Fig. 5). Recall that the distortion D_L as given by (12) represents the average distortion between a losslessly and packet-lossily transmitted versions of the same MPEG-2 bit stream. We need however a total end-to-end distortion measure D between the original and received video. Consider now only the video elements (e.g., macroblocks) that are lost but replaced by error concealment at the receiver. Let $\widetilde{D_L}$ be the average distortion between these elements and their original version. The end-to-end average distortion can then be written as

$$D = D_S(1-P_l) + \widetilde{D_L}P_l = D_S + P_l(\widetilde{D_L} - D_S) \quad (13)$$

where P_l represents the average probability for a video element (e.g., macroblock) to be lost. Equation (13) holds for MSE-like distortion metrics. Assume that it is also verified in average for perceptual distortion. In this case, we can interpret the second-term in the right-hand side of (13) as the average distortion due to data loss

$$D_L = P_l(\widetilde{D_L} - D_S) \quad (14)$$

where D_L is directly related to the number of lost pixels from (12). Notice that (14) holds thanks to our definition of the distortion D_L.

Our problem becomes now the following: At time t, find the optimal FEC scheme parameters (k, n) that minimize the end-to-end distortion

$$\min_{k\le n\le N} D(n,k) = \min_{k\le n\le N}\left[\chi_S\left(\frac{k}{n}R(t)\right)^{\xi_S} + \chi_L\pi_v(k,n)\left(1+\frac{kR(t)}{2nPN_S\alpha_v(k,n)}\right)\right] \quad (15)$$

given the total bandwidth $R(t)$, the channel state (i.e., p and q in the Gilbert model) and the scene-dependent parameters (χ_S, ξ_S, χ_L). The constraint $n \le N$ simply imposes a maximum reconstruction time for the FEC decoding.

Since the parameters (k, n) only take integer values, (15) can be solved easily through numerical methods. The optimal values of k and n then lead to the optimal rate distribution between source and FEC rates. They define the media-independent FEC parameters that minimize the end-to-end distortion. These two parameters are adapted along the sequence, according to the transmission parameter changes. Such a dynamic adaptation thus ensures and optimal end-to-end quality.

VII. Experimental Results

The optimal FEC parameters are given in Table I for several network conditions and video scenes of different spatio–temporal complexities. The FEC reconstruction delay is set to approximately 6.5 ms (i.e, $N = 20$ at 4 Mb/s and $N = 30$ at 6 Mb/s). Optimal FEC parameters and hence rate distribution are then numerically computed from (15). It is shown that the FEC rate [i.e., $(n-k)/n$] decreases when the global loss ratio (PLR) decreases. Moreover, the required FEC rate is lower for the Foot sequence (i.e., high spatio–temporal complexity) than for the News video scene (i.e., low spatio–temporal complexity), at least for $PLR \le 0.1$. This intuitive result clearly exhibits the joint role of both the source and FEC rates onto the end-to-end distortion. The evolution of the optimal parameters with the increasing global loss burstiness (ABL) is less straightforward. They indeed result from a tradeoff between the decreasing FEC efficiency and decreasing error propagation effect.

TABLE I
Optimal FEC Parameters for Several Transmissions Conditions and Scenes

		$R = 4$ *Mbps*				$R = 6$ *Mbps*			
		Ski		Foot		Foot		News	
PLR	ABL	k	n	k	n	k	n	k	n
0.1	1	14	18	16	20	23	29	22	29
	2	14	20	15	20	22	30	19	30
0.01	1	19	20	19	20	28	30	28	30
	2	1	1	1	1	29	30	25	30
0.001	1	1	1	1	1	1	1	29	30
	2	1	1	1	1	1	1	1	1

Let us now observe the behavior of our algorithm on an complex video sequence. The sequence is composed of 5 different scenes and is encoded with 12 frame-long GOPs. The available rate R is set to 4 Mb/s for the 204 first frames, and to 6 Mb/s for the 196 last frames. The temporal evolution of the perceptual distortion through a five-scene sequence is given in Figs. 12 and 13. The distortion averaged over sliding windows through Minkowski summation is compared to the one obtained from classical FEC schemes. Thanks to its adaptivity features, our rate distribution algorithm outperforms the common schemes, as reported also by the PSNR evolution (see Figs. 14 and 15). The proposed algorithm indeed adapts to the scene complexity, to the available bandwidth and to the network conditions. It has to be noted that large loss ratio values have been chosen since the length of the sequence is relatively short. However, even if these values seem relatively high compared to usual mean ratios on today's network, they are likely to happen during small time intervals.

VIII. Conclusions

In this paper, we considered a joint source and channel coding problem. More specifically, we proposed a low complexity algorithm for computing the optimal rate distribution between MPEG-2 and media-independent FEC. The optimality has been defined in terms of minimal end-to-end perceptual video distortion. The efficiency of a media-independent FEC algorithm has been studied first. The residual video loss patterns after FEC

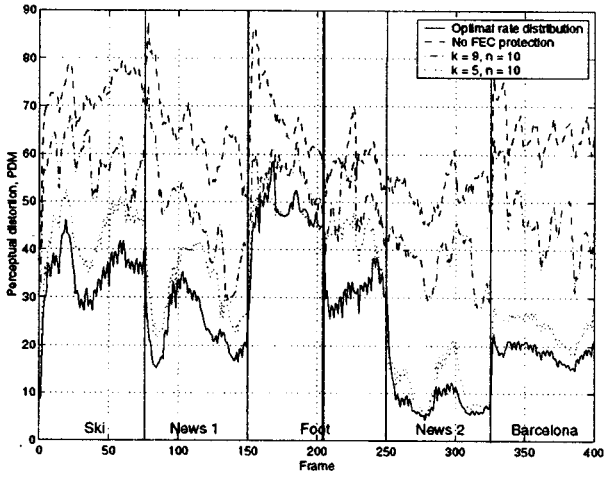

Fig. 12. Perceptual distortion versus frame number for optimal rate distribution scheme and classical FEC schemes. The global loss process parameters are set to $(PLR,\ ABL) = (0.1,\ 1)$.

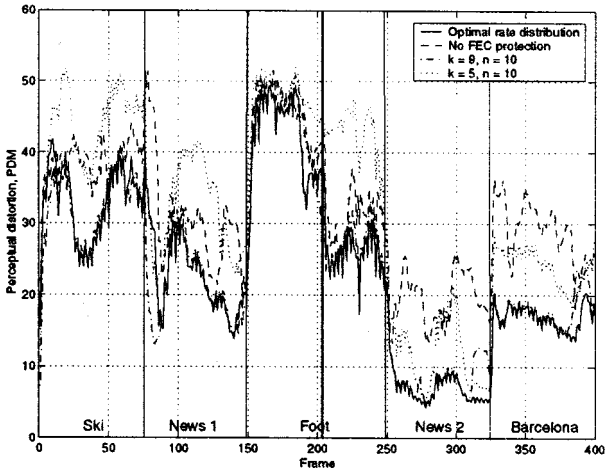

Fig. 13. Perceptual distortion versus frame number for optimal rate distribution scheme and classical FEC schemes. The global loss process parameters are set to $(PLR,\ ABL) = (0.1,\ 1)$.

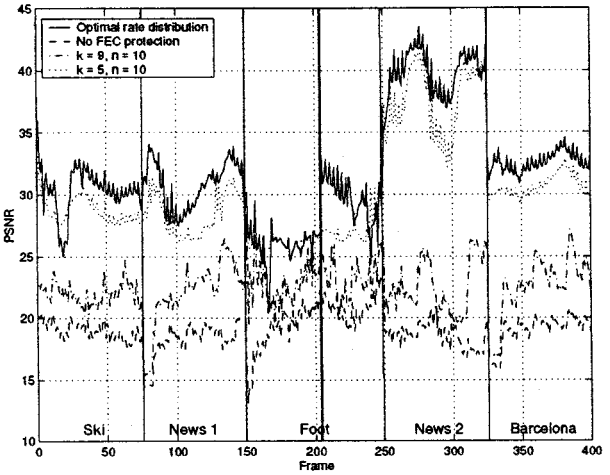

Fig. 14. PSNR versus frame number for optimal rate distribution scheme and classical FEC schemes. The global loss process parameters are set to $(PLR,\ ABL) = (0.1,\ 1)$.

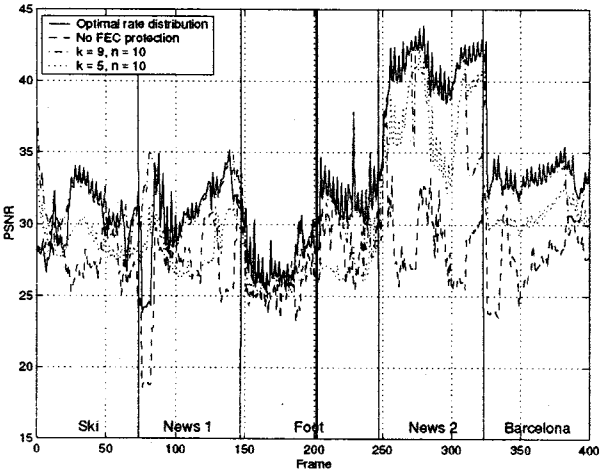

Fig. 15. PSNR versus frame number for optimal rate distribution scheme and classical FEC schemes. The global loss process parameters are set to $(PLR,\ ABL) = (0.01,\ 1)$.

recovery have been computed for a Gilbert-model global loss process. The exponential source perceptual distortion-rate function has then be derived from empirical results. Finally, the distortion due to loss has been shown to be directly proportional to the number of lost pixels. From this set of equations, the optimal rate distribution, as well as the optimal FEC scheme are obtained by solving a simple optimization problem. Finally, the proposed allocation scheme has been shown to outperform classical FEC schemes due to its adaptivity to the video scene, to the available bandwidth and to the network state. Note that the same study directly holds for other video compression schemes (e.g., H.263, MPEG-4).

Appendix I
Average Video Burst Length

In a renewal error process, the probability for a burst to start on the jth packet of an FEC block (i.e., $1 \le j \le n$) is given by

$$P_j = \begin{cases} \pi\left(P(2) + p(1)R_n\right) \displaystyle\sum_{i=n-k}^{n-1} R(i+1,\ n), & \text{if } j = 1 \\ \displaystyle\sum_{i=0}^{n-j} \sum_{x=0}^{k-n+j+i-2} S(x+1,\ j-1)R(i+1,\ n-j+1) \\ \quad \cdot \pi P(2), & \text{if } j > 1 \end{cases} \tag{16}$$

where R_n is the probability that all packets in an FEC block are recovered given that the first packet is missing. It can be written as

$$R_n = \sum_{i=0}^{n-k-1} R(i+1, n).$$

The average length of bursts of lost video packets, α_v, excludes redundancy packets. Let first T denote the number of FEC blocks transitions along the burst of length l

$$T = \text{floor}\left(\frac{l+j-2}{k}\right).$$

Let v denote the position in an FEC block of the first video packet following a burst of length l

$$v = \begin{cases} k, & \text{if } (l+j-1) \bmod k = 0, \\ (l+j-1) \mod k, & \text{otherwise.} \end{cases}$$

Moreover, let r_{n-k} and $\overline{r}_{n-k}$ denote the probability to loose all the video packets of an FEC block and that the first video packet of the next FEC block is respectively erased or not (before FEC recovery). These probabilities can be written as

$$r_{n-k} = \sum_{\lfloor n-2k+1 \rfloor}^{n-k} r(i+1, n-k+1)$$

and

$$\overline{r}_{n-k} = \sum_{\lfloor n-2k+1 \rfloor}^{n-k} \overline{r}(i+1, n-k+1).$$

Similarly, r^j_{n-k} and $\overline{r}^j_{n-k}$ denote the probability that the last video packet (i.e., the packet k) of the first FEC block is lost and the first video packet of the second block is respectively erased or not (before FEC recovery). It is assumed that all video packets between packets j and k are lost. These probabilities are expressed as

$$r^j_{n-k} = \sum_{i=0}^{n-k} r(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1)$$

and

$$\overline{r}^j_{n-k} = \sum_{i=0}^{n-k} \overline{r}(i+1, n-k+1) \sum_{x=0}^{2k-n+i-2} S(x+1, j-1).$$

Finally, the average length of a burst of lost video packets starting on the jth video packet of an FEC block is given by

$$L_j = \sum_{l=0}^{\infty} l p(1)^{l-T-1} \lambda_j(l) \tag{17}$$

where $\lambda_j(l)$ is the probability to have a video packet burst of length l starting at j. The conditional probability $\lambda_1(l)$ for bursts starting with the first video packet of the FEC block (i.e., $j = 1$) can be written as

$$\lambda_1(l) = \begin{cases} \dfrac{\pi(P(2) + p(1) R_n)}{P_1} \\ \quad (r_{n-k} R_n + \overline{r}_{n-k}) {r_{n-k}}^T, & \text{if } v = n, \\ \dfrac{\pi(P(2) + p(1)\, R_n)}{P_1} \\ \quad P(2) \displaystyle\sum_{i=0}^{k-2} S(i+1,\, n-v) {r_{n-k}}^T, & \text{otherwise.} \end{cases}$$

For $2 \le j \le k$, λ_j is expressed as

$$\lambda_j(l) = \begin{cases} \dfrac{\pi p(1)}{P_j} \displaystyle\sum_{i=0}^{n-v-1} \sum_{x=0}^{k-3-i} \\ \quad S(x+1,\, j-1) \\ \quad p(1) S(i+1,\, n-v), & \text{if } l+j-1 < k, \\ \dfrac{\pi p(1)}{P_j} (r^j_{n-k} R_n + \overline{r}^j_{n-k}), & \text{if } v = k \text{ and } T = 0, \\ \dfrac{\pi p(1)}{P_j} (r_{n-k} R_n + \overline{r}_{n-k}) \\ \quad r^j_{n-k} {r_{n-k}}^T, & \text{if } v = k \text{ and } T \ne 0, \\ \dfrac{\pi p(1)}{P_j} p(1) \displaystyle\sum_{i=0}^{k-2} \\ \quad S(i+1, n-v) \\ \quad r^j_{n-k} {r_{n-k}}^{\lfloor T-1 \rfloor}, & \text{otherwise.} \end{cases}$$

Finally, α_v is given from (16) and (17) by

$$\alpha_v = \frac{\displaystyle\sum_{j=1}^{k} P_j L_j}{\displaystyle\sum_{j=1}^{k} P_j}. \tag{18}$$

Appendix II
Number of Spatially Lost Pixels

The aim of this section is to compute the number of lost pixels when the video loss patterns (i.e., π_v and α_v) are known. Let x denote the byte position within a video slice where the loss starts (see Fig. 10). The distribution of x is assumed uniform in the interval $[0,\ S]$. Let then P and S, respectively, represent the size of the transmission packet and the average slice length, in bytes. Recall that α_v represents the average number of consecutively lost packets. We can now compute the maximal number of slices σ that can be impaired by a loss of α_v consecutive packets

$$\sigma = \left\lceil \frac{\alpha_v P}{S} \right\rceil. \tag{19}$$

As shown in Section II-B, bytes within a slices are useless for the decoder if previous bytes of the slices have been lost. Therefore, scanning the complete range of x, the number B of lost video bytes per lost packet burst is given by the following piecewise linear relation:

$$B = \begin{cases} \sigma S - x, & \text{if } 0 \le x < \sigma S - \alpha_v P, \\ (\sigma+1) S - x, & \text{if } \sigma\, S - \alpha_v\, P \le x < S. \end{cases} \tag{20}$$

This behavior is illustrated in Fig. 16. Assuming a uniform distribution of x, the average number of lost video bytes per burst is given by

$$E[B] = \alpha_v P + \frac{S}{2} = \alpha_v P + \frac{R_S}{2 N_S} \tag{21}$$

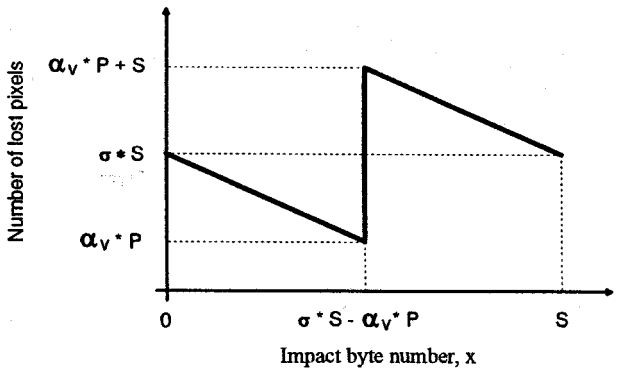

Fig. 16. Number of damaged pixels versus the impact byte in the stream.

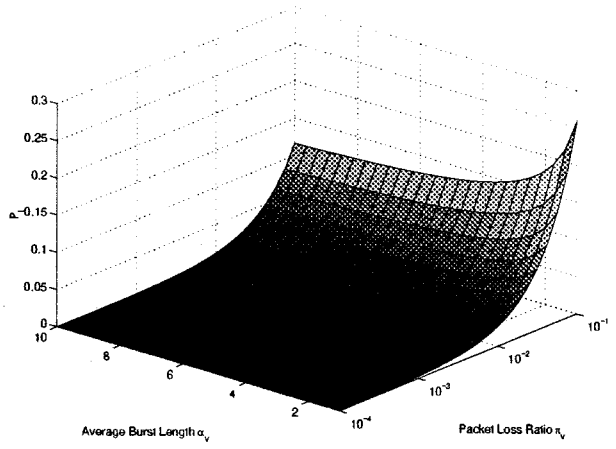

Fig. 17. Number of lost pixels versus the Average Burst Length α_v and the PLR π_v (source rate $R_S = 4.75$ Mbit/s).

where R_S represents the source rate and N_S denotes the number of slices per time unit and is used here to explicit R_S in the formulation of $E[B]$.

Now that we have define the average number of lost pixels per burst, we want finally to compute the pixel loss probability. To this aim, let $\overline{R_B}$ define the mean rate of packet loss bursts

$$\overline{R_B} = \frac{R_S \pi_v}{P \alpha_v}. \tag{22}$$

Therefore, from (21) and (22), the rate of lost video bytes $\overline{L}$ is simply given by

$$\overline{L} = E[B]\overline{R_B} = R_S \pi_v + \frac{{R_S}^2 \pi_v}{2PN_S \alpha_v}. \tag{23}$$

Finally, since in average the number of lost pixels is proportional to the number of lost video bytes, the ratio of lost pixels is equivalent to

$$P_l = \frac{\overline{L}}{R_S} = \pi_v + \frac{R_S \pi_v}{2PN_S \alpha_v}. \tag{24}$$

Equivalently, P_l represents also the probability for a pixel to be lost. The expected behavior of the number of lost pixels versus the average burst length is illustrated in Fig. 17. It represents the number of spatially damaged pixels versus the video average burst length and the PLRs π_v. The number of lost pixels is clearly decreasing in $1/\alpha_v$. Moreover, the phenomenon is more visible for low α_v values (i.e., typical values). Finally, it has to be noted that the influence of the Average Burst Length decreases when the encoding bit rate decreases, since the slice size decreases.

Finally, it has to be noticed that the computed value of P_l is very close to experimental data. The correlation factor computed between simulation results and analytic value from (24) lies around 0.995 for a large set of source rates.

ACKNOWLEDGMENT

The authors are grateful to Prof. M. Kunt from the Signal Processing Laboratory and to anonymous reviewers for their relevant comments and remarks on the paper.

REFERENCES

[1] O. Verscheure, X. Garcia Adanez, G. Karlsson, and J.-P. Hubaux, "User-oriented QoS in packet video delivery," *IEEE Network*, vol. 12, no. 6, pp. 12–21, Nov./Dec. 1998.

[2] I. E. G. Richardson and M. J. Riley, "Improving the error tolerance of MPEG video by varying slice size," *Signal Process.*, vol. 46, pp. 369–372, 1995.

[3] P. Frossard and O. Verscheure, "AMISP: A complete content-based MPEG-2 error-resilient scheme," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 989–998, Sept. 2001.

[4] J. Zhang, M. R. Frater, J. F. Arnold, and T. M. Percival, "MPEG 2 video services for wireless ATM networks," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 119–128, Jan. 1997.

[5] U. Horn, K. Stuhlmuller, M. Link, and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *Signal Process. Image Commun.*, vol. 15, no. 1–2, pp. 77–94, 1999.

[6] S. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 983–1001, Aug. 1997.

[7] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet-loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, pp. 40–48, Sept. 1998.

[8] A. J. McAuley, "Reliable broadband communications using a burst erasure correcting code," in *ACM SIGCOMM*, Philadelphia, PA, Sept. 1990, pp. 297–306.

[9] M. J. Ruf and J. W. Modestino, "Operational rate-distortion performance for joint source and channel coding of images," *IEEE Trans. Image Processing*, vol. 8, pp. 305–320, Mar. 1999.

[10] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: A review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.

[11] C. Lopez Fernandez, A. Basso, and J.-P. Hubaux, "Error concealment and early resynchronization techniques for MPEG-2 video streams damaged by transmission over ATM networks," *Proc. SPIE*, vol. 2668, pp. 372–383, 1996.

[12] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Processing*, vol. 3, pp. 533–545, Sept. 1994.

[13] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 12–20, Feb. 1996.

[14] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 246–250, Feb. 1997.

[15] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 446–459, Aug. 1998.

[16] C. Fogg, "mpeg2encode/mpeg2decode," in *MPEG Software Simulation Group*, 1996.

[17] S. Winkler, "A perceptual distortion metric for digital color video," in *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3644, San Jose, CA, Jan. 23–29, 1999.

[18] C. J. van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio–temporal model of the human visual system," *Proc. SPIE*, pp. 450–461, Feb. 1996.

[19] C. J. van den Branden Lambrecht, "Perceptual models and architectures for video coding applications," Ph.D. dissertation, Swiss Federal Inst. Technol., Lausanne, Switzerland, 1996.
[20] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Process.*, vol. 78, pp. 231–252, Oct. 1999.
[21] S. Sakazawa, Y. Takishima, M. Wada, and Y. Hatori, "Coding control scheme for a multi-encoder system," in *Proc. 7th Int. Workshop Packet Video*, Mar. 1996, pp. 83–88.
[22] O. Verscheure, P. Frossard, and M. Hamdi, "User-oriented QoS analysis in MPEG-2 video delivery," *J. Real-Time Imag.*, vol. 5, pp. 305–314, Oct. 1999.
[23] U. Reimers, *Digital Video Broadcasting—The International Standard for Digital Television*. New York: Springer-Verlag, 2001.
[24] M. O. Rabin, "Efficient dispersal of information for security, load balancing and fault tolerance," *J. ACM*, vol. 36, pp. 335–348, Apr. 1989.
[25] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," *ACM Comput. Commun. Rev.*, vol. 27, pp. 24–36, Apr. 1997.
[26] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[27] N. Shacham and P. McKenney, "Packet recovery in high-speed networks using coding and buffer management," in *Proc. IEEE INFOCOM Conf.*, vol. 1, Los Alamitos, CA, 1990, pp. 124–131.
[28] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.
[29] E. O. Elliott, "A model of the switched telephone network for data communications," *Bell Syst. Tech. J.*, vol. 44, pp. 89–109, Jan. 1965.
[30] J.-P. A. Adoul, "Error intervals and cluster density in channel modeling," *IEEE Trans. Inform. Theory*, vol. 20, pp. 125–129, Jan. 1974.
[31] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.
[32] J. Bolot, S. Fosse-Parisis, and D. Towsley, "Adaptive FEC-based error control for interactive audio in the internet," in *Proc. IEEE INFOCOM*, 1999.
[33] M. Yajnik, J. Kurose, and D. Towsley, "Packet loss correlation in the MB one multicast network: Experimental measurements and Markov chain models," Dept. Comput. Sci., University of Massachusetts, Amherst, Tech. Rep. 95-115, 1995.
[34] V. Paxson, "End-to-end Internet packet dynamics," *ACM Comput. Commun. Rev.*, vol. 27, pp. 139–152, Oct. 1997.
[35] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*. New York: Plenum, 1992.
[36] P. Frossard, "FEC performances in multimedia streaming," *IEEE Commun. Lett.*, vol. 5, pp. 122–124, Mar. 2001.
[37] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, pp. 1253–1265, Sept. 1960.
[38] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis, "Framework for IP performance metrics," IETF, Tech. Rep. RFC 2330, May 1998.
[39] V. Jacobson, "Congestion avoidance and control," in *Proc. ACM SIGCOMM*, 1988.

Pascal Frossard (M'96) received the M.S. and Ph.D. degrees, both in electrical engineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, in 1997 and 2000, respectively.

From 1997 to 1998, he was a Research and Teaching Assistant with the Institute for Computer Communications and Their Applications (ICA), EPFL. From 1998 to 2000, he was with the the Signal Processing Laboratory, EPFL, as Reasearch and Teaching Assistant under a grant from Hewlett-Packard. Since April 2001, he has been a Research Staff Member with the IBM T. J. Watson Research Center, Hawthore, NY. His main research interests concern signal and video compression, video streaming, error control, and video distribution.

Olivier Verscheure (S'97–A'98–M'00) received the B.S. degree in electrical engineering in 1995 from Ecole Polytechnique de Mons, Belgium. He received the Ph.D. degree in 1999 from the Swiss Federal Institute of Technology, Lausanne.

He is a Research Staff Member with the IBM T. J. Watson Research Center, Hawthorne, NY. He was a Visiting Researcher with the Hewlett-Packard Laboratories, Palo Alto, CA, during the Summer of 1997. He joined IBM Research in July 1999. His research lies within the areas of scalable multimedia servers, packet video, and vision science.