Analytical model for MPEG video frame loss rates and playback interruptions on packet networks

**Felix Espina · Daniel Morato · Mikel Izal ·
Eduardo Magaña**

Abstract This paper presents and studies objective video quality evaluation techniques for a network where frame losses can be considered independent, for example a best effort not heavy loaded packet switching network. The total or partial loss of a frame's information affects the quality of video playback, as the frame cannot be decoded and other frames that depend on it cannot be correctly decoded too. Therefore, during some time the video playback has errors in the image and the user will perceive them as interruptions. In this paper, the total number of decoded frames and the video playback interruptions duration will be considered important parameters to quantify the video quality. The analytical formulation for them will be presented and the importance of considering them together will be highlighted.

Keywords Quality of service · Quality of experience · Video communications · Decodable frame rate · Video cut durations

1 Introduction

Video streaming is nowadays one of the hungriest bandwidth consuming applications. Internet video was 40 % of consumer Internet traffic on 2011, and it is expected to reach 62 % by the end of 2015 [5]. Trying to minimize this expense, encoders from MPEG family [18] have been widely adopted for the codification of audio and video. Given the spread of MPEG-2 and MPEG-4/H.264 throughout the industry and reaching millions of user devices, they will most likely remain so in the foreseeable future.

In MPEG standard, the coding process takes advantage of temporal similarities between frames in order to produce smaller compressed frames. The decoding process for most video frames requires previously decoded ones. This hierarchical

F. Espina (✉) · D. Morato · M. Izal · E. Magaña
Public University of Navarre, Campus de Arrosadía s/n, 31006 Pamplona, Spain
e-mail: felix.espina@unavarra.es

structure of MPEG encoding implies a possible error propagation through its frames, and therefore it adds an extra handicap to the transport of MPEG video flows over lossy networks [19]. Small packet loss rates may cause high frame error rates, degrading the video quality perceived by the user.

A video provider needs to quantify the video quality problems, at best before the user perceives them. The concept of *Quality of Experience (QoE)* emerged from this need. ITU-T defines QoE [13] as "The overall acceptability of an application or service, as perceived subjectively by the end-user". The user does not perceive directly the proportion of lost packets, but the proportion of frames that could be or could not be displayed which depends on lost frames and their relationship to the other frames. The proportion of lost packets is a *Quality of Service (QoS)* parameter about the transport network, but according to the definition given above, it cannot be considered a QoE parameter. The proportion of frames that could be decoded, and therefore displayed, is named the *Decodable Frame Rate (Q)* [29] and can be considered a QoE parameter.

User perception is not only affected by the number of non-displayed frames, but also by the grouping in time of these non-displayed frames. This grouping is reflected in the length of the video playback interruptions or cuts.

The main contribution of this paper is to consider the Decodable Frame Rate and the video playback interruptions or cut lengths important parameters to quantify the video quality, as the user perceives both. This paper presents an analytical model to compute the Decodable Frame Rate and the video playback interruption lengths for network scenarios where frame losses can be considered independent. The analytical equations are validated by simulation and used, under certain restrictions, on packet switching networks. These QoE parameters show that different videos can have similar Decodable Frame Rate or cut lengths, but both parameters can only be similar for different videos if the videos transport characteristics are similar.

The rest of the paper is organized as follows. Section 2 presents the related work done on this topic. Sections 3 presents the analytical study of the QoE parameters. Section 4 validates the analytical model for a packet switching network. Section 5 presents results that can be extracted from the model and the validation. Finally, section 6 concludes this paper.

2 Related work

A common strategy for quality evaluation of IP packet transport is the use of QoS measures like packet loss rate or packet jitter. Its analogues for MPEG video transport are QoS measures like packet/frame loss ratio or packet/frame jitter [26]. These measures are not directly related to user perception, therefore they cannot be considered a valid QoE metric.

QoE metrics can be classified based on the availability of the original video signal [7]. In *full reference* (*FR*) metrics the whole original video signal is available. In *reduced reference* (*RR*) metrics only partial information from the original video is available. In *no reference* (*NR*) metrics the original video is not available at all. The signal as received by the user is assumed as available for the three types of metrics.

The most used QoE metrics are FR metrics like the *Peak Signal-to-Noise Ratio (PSNR)* [9] and the *Perceptual Evaluation of Video Quality* (*PEVQ*) [14]. These

metrics can only be carried out by offline computation, or by online computation in highly controlled environments where the original video and the video at user side are available together at some point. In a real video distribution scenario, online QoE computation is preferred, in order to quickly react to quality problems. However, it is possible only if the original video is available at user side, thus making QoE assessment unnecessary. Therefore, RR or NR metrics are more useful.

RR metrics require parameter extraction from the original video and the received one. They measure changes in these parameters, which must be transported by the network and collected at a single site in order to obtain the measured QoE. In practical situations these parameters are hard to obtain. For example, the *Hybrid Image Quality Metric* (*HIQM*) [15] combines five structural parameters that are computed for each frame, both in the original and the received video. There is therefore a clear overhead not only on computation but also on transportation of these meta-data.

The computation of NR metrics does not use the original video or any parameter extracted from it. They can be computed at user side based only on the received video. This paper focuses on these metrics as they are the easiest to implement and use for traffic engineering in a real video distribution network.

The Decodable Frame Rate is the proportion of the video frames that the user will see completely correct, so it is directly related to user perception and it only needs the video signal as received by the user. However, the user perceives non-decoded frames differently if they are temporally contiguous or they are separated, i.e. the user perceives the length of the video cuts too.

The Decodable Frame Rate was introduced in [29]. It was assumed that frame losses were mutually independent and the analytical model for the Decodable Frame Rate depends on the probability of a frame being lost. In [29] this probability is obtained for the case of a packet switching network.

Some papers have used this analytical model for transmission on wireless networks [16], on IPTV networks [1] or in the evaluation of the SCTP protocol [3]. All of them present the Decodable Frame Rate as a valid QoE metric. In fact, the authors from [3, 4] obtain for the same scenarios and parameters the PSNR and the Decodable Frame Rate of some videos. They compare the results and they conclude that the Decodable Frame Rate can reflect the behaviour of the PSNR, and therefore the MOS video quality can be estimated with reasonable accuracy from the Decodable Frame Rate.

However, none of them [1, 3, 4, 16, 29] includes the distribution of the non-decoded frames, which gives more data than the Decodable Frame Rate for the statistical evaluation of video quality. The Decodable Frame Rate can be derived from the distribution of the non-decoded frames, but this distribution cannot be derived from the Decodable Frame Rate.

The distribution of the non-decoded frames is reflected in the video playback interruptions or cuts length. In this paper, the video playback interruption lengths are analytically derived for network scenarios where frame losses can be considered independent.

Both the number of non-decoded frames and the video playback interruption lengths must be taken into account to evaluate video quality. The user perception is affected by the number of consecutive non-displayed frames, as experimental measurements have shown in [22, 23]. For example, video playback interruptions of 200 ms are certainly visible to the user and even a 80 ms long cut may be visible.

In [22, 23], the authors present a mathematical expression to obtain the MOS video quality from the video playback interruption length. They only consider the case of one cut, i.e., the mathematical expression calculates the degradation in quality when the video has a single cut. In [20, 21], a NR metric is proposed to obtain the MOS video quality when the video has multiple playback interruptions. The mathematical expression from [22, 23] is extended to take into account the video playback interruption lengths distribution. The authors obtained experimentally the video playback interruption lengths distribution, thus the metric can only be used after the reception of the video and it can not be used to make predictions about the video quality at the user side. The analytical model of video playback interruption lengths that we propose in this paper can be used to simplify the metric and to make predictions about the video quality at the user side based on current network conditions.

3 Analytical model

Three types of video frames are defined in the MPEG standards [8]: intra-coded frames (I-frames), inter-coded or predicted frames (P-frames) and bidirectional coded frames (B-frames).

I-frames can be decoded on their own. P- and B-frames hold only the changes in the image from the reference frames, and thus improve video compression rates. P-frames have only one reference frame, the previous I- or P-frame. In MPEG-2 Part 2 (H.262) [10] and MPEG-4 Part 2 [11], B-frames have two reference frames. These frames are the previous I- or P-frame and the following one of either type. In MPEG-4 Part 10 (MPEG-4 AVC or H.264) [12], B-frames can have up to 16 reference frames, located before or after the B-frame. They can be either I- or P-frames, and even B-frames can be reference frames for other B-frames. In this paper, only videos with "classic" B-frames will be studied, i.e. videos with B-frames with two reference frames of I- or P-frame type.

In MPEG-2 Part 2 and MPEG-4 Part 2, the allowed prediction level is the same for the whole frame. In H.264, the prediction type granularity is reduced to a level lower than the frame, called the slice. A frame can contain multiple slices that are encoded separately from any other slice of the frame and even they can be of different types of prediction level (I-slice, P-slice, B-slice). Therefore, instead of frame losses, we could talk about slice losses.

We have done a brief survey on the presence of slices in videos from present networks [6]. The results show that most of the video sources analyzed use one slice per frame, i.e., the use of more than one slice per frame is infrequent. In the rare event of more than one slice per frame, all the slices are from the same type. Therefore, we have concluded that a model that ignores the presence of slices and simplifies the model of losses is still useful.

The hierarchical structure of MPEG encoding implies a possible error propagation through its frames. Frames that arrive at the destination could be useless if the other frames that they depend on have been dropped by the network. The loss of a frame by the network, or part of it, is named a *direct loss* and it implies that the frame is non-decodable. *Indirect loss* of a frame happens when a frame is considered non-decodable because some frame it depends on is non-decodable. These are common assumptions taken in many other papers [3, 16, 29].

I-, P- and B-frames are grouped into *Groups of Pictures (GoP)*. A GoP is a sequence of frames beginning with an I-frame up to the frame before the next I-frame. The GoP structure is the pattern of I-, P- and B-frames used inside every GoP. A regular GoP structure is usually described by the pattern (N, M) where N is the I-frame to I-frame distance, and M is the I-to-P frame distance (see Table 1 for the notation used in the paper). For example, the GoP structure could be $(12, 3)$ or *IBBPBBPBBPBB*.

$N_{\{I,P,B\}}$ is the number of frames from each type of frames in a single GoP. For any regular GoP, $N_I = 1$ and $N = N_I + N_P + N_B$, but N_P and N_B depend on the GoP structure (N, M). In an open GoP the last B-frames depend on the I-frame from the next GoP, like for example in $(12, 3)$ or *IBBPBBPBBPBB*. In this case N is a multiple of M and the number of P-frames can be computed as $N/M - 1$. In a closed GoP there is no dependence with frames out of the GoP and it ends with a P-frame, like for example in $(10, 3)$ or *IBBPBBPBBP*. In this case $N - 1$ is a multiple of M and the number of P-frames can be computed as $(N - 1)/M$.

Table 1 Notation

Variable	Definition
(N, M)	Regular GoP structure. N is the I-frame to I-frame distance. M is the I-to-P frame distance.
N_I	Number of I-frames in a GoP.
N_P	Number of P-frames in a GoP.
N_B	Number of B-frames in a GoP.
z	Control variable that nullifies terms in the analytical model that only affect open GoPs. For an open GoP $z = 1$ and for a closed GoP $z = 0$.
Q	Decodable Frame Rate
P_I, P_P, P_B	Probability of losing in the network an I-frame, a P-frame or a B-frame.
P_τ	Probability of losing in the network a frame of type $\tau \in \{I, P, B\}$.
T_{if}	Inter-frame time (time between consecutive frames) or the inverse of frames per second.
F	Total number of frames in the video.
N_G	Total number of GoPs in the video.
c	The set of possible cut lengths for a GoP. The values in c depend on the GoP structure (N,M).
$N_{\text{cut}}[c]$	The number of cuts of c frames length when one or more losses happen.
σ	The minimum number of B-frames from the same block that should be available to the decoding process to prevent a cut length larger than $c = 1 \ldots M - 1$, when the first lost B-frame is the rth in the block.
r	Position of the first lost B-frame in the B-frames block to compute the analytical model of the number of $c = 1 \ldots M - 1$ length cuts.
δ	The contribution to the number of cuts of length $c = 1 \ldots M - 1$ frames by the $M - c$ possible loss combinations of a B-frames block that generate a c frames length cut.
P_{cut}	The cut length Probability Mass Function. $P_{\text{cut}}[c]$ is the probability that the cut is c frames long.
L_{cut}	The average cut length.
T_{cut}	Total number of cuts.
p	Packet loss ratio.
d_τ	The average number of packets per frame of type $\tau \in \{I, P, B\}$.

For simplicity, we have defined $N_P = \lfloor (N-1)/M \rfloor$ for any type of GoPs. As $N-1$ is a multiple of M for a closed GoP, then $N_P = \lfloor (N-1)/M \rfloor = (N-1)/M$ for a closed GoP. As N is a multiple of M for an open GoP, then $\lfloor (N-1)/M \rfloor$ is the same as $\lfloor (N-M)/M \rfloor$ and therefore $N_P = \lfloor (N-1)/M \rfloor = \lfloor (N-M)/M \rfloor = \lfloor N/M - 1 \rfloor = N/M - 1$ for an open GoP. Summarizing, the number of frames from each type ($N_{\{I,P,B\}}$) can be obtained for any type of GoP by (1).

$$N_I = 1$$
$$N_P = \left\lfloor \frac{N-1}{M} \right\rfloor$$
$$N_B = N - 1 - N_P = N - 1 - \left\lfloor \frac{N-1}{M} \right\rfloor \tag{1}$$

In an open GoP $N_B = (N_P + 1) * (M-1)$ and in a closed one $N_B = N_P * (M-1)$. We can define the control variable z (2) that nullifies terms in the analytical model that only affect open GoPs. Therefore, when the GoP is an open one $z = 1$ and when the GoP is a closed one $z = 0$. This value can be obtained directly from the GoP structure (2).

$$z = \frac{N_B}{M-1} - N_P = \frac{N - 1 - \left\lfloor \frac{N-1}{M} \right\rfloor}{M-1} - \left\lfloor \frac{N-1}{M} \right\rfloor \tag{2}$$

A GoP structure like $IBBPBBPBBPBB$ or $IBBBPBBBP$ is shown in presentation order, i.e. the order in which the frames will be shown to the user. However, as B-frames require in order to be processed the previous I- or P-frame and the following of either type, the coding/decoding order will be different. The coding/decoding order will be $IbbPBBPBBPBBiBB$ for a (12, 3) GoP structure, where the frames in lower-case correspond to frames from the previous or the next GoP, showing that it is an open GoP.

The transmission order corresponds usually to the coding/decoding order, but the user perceives the cuts as viewed in presentation order. The measurement of cut durations has to take into account this change of frame order.

Parameters presented in this paper like the Decodable Frame Rate and the video playback interruptions or cuts, will be measured in presentation order.

3.1 Decodable frame rate Q

The analytical model (3) for the Decodable Frame Rate presented in [29] is valid for open GoPs.

$$Q = \frac{(1 - P_I) + (1 - P_I) \sum_{i=1}^{N_P} (1 - P_P)^i}{N}$$
$$+ \frac{(M-1)(1 - P_I)(1 - P_B) \left[\sum_{i=1}^{N_P} (1 - P_P)^i + (1 - P_I)(1 - P_P)^{N_P} \right]}{N} \tag{3}$$

P_I is the probability of losing an I-frame, P_P is the probability of losing a P-frame and P_B is the probability of losing a B-frame. We use P_τ as the probability of losing

in the network a frame of type $\tau \in \{I, P, B\}$. It is assumed that frame losses are mutually independent. The last term of (3), $(1 - P_I)(1 - P_P)^{N_P}$, corresponds to the last B-frames that depend on the I-frame from the next GoP. A closed GoP does not have these last B-frames, so the analytical model should reflect this difference. Equation (4) presents the Decodable Frame Rate expression, valid for any type of GoP, where the difference between open and closed GoPs is reflected using variable z ($z = 1$ for open GoPs and $z = 0$ for closed GoPs).

$$Q = \frac{(1 - P_I) + (1 - P_I) \sum_{i=1}^{N_P} (1 - P_P)^i}{N} +$$

$$+ \frac{(M - 1)(1 - P_I)(1 - P_B) \left[\sum_{i=1}^{N_P} (1 - P_P)^i + z * (1 - P_I)(1 - P_P)^{N_P} \right]}{N} \qquad (4)$$

3.2 Video playback interruptions or cuts

Video playback interruption lengths or cut lengths can be measured as the number of consecutive non-decoded frames. It is a non-negative integer number c that is related to the cut time duration through the inter-frame time T_{if}. The possible cut lengths depend on the GoP structure of the video and they can be obtained by (5) (see Appendix for more details).

$$c = [1 \ldots M - 1, \ j * N + i * M + z * (M - 1), \ (j + 1) * N + z * (M - 1)]$$
$$i = 1 \ldots N_P$$
$$j = 0 \ldots N_G - 1 \qquad (5)$$

Where $N_G = F/N$ is the total number of GoPs in the video and F is the number of frames.

$N_{\text{cut}}[c]$ (6) is the number of cuts of c frames length when one or more losses happen. It depends on the GoP structure of the video and on the frame loss probabilities, $P_{\{I,P,B\}}$ (see Appendix for more details).

$$N_{\text{cut}}[c] = \begin{cases} N_G * \delta * P_B^c * (1 - P_I) \sum\limits_{m=1}^{N_P} (1 - P_P)^m + \\ + z * N_G * \delta * P_B^c * (1 - P_I)^2 (1 - P_P)^{N_P} & \text{for } c = 1 \ldots M - 1 \\[2ex] N_G * P_I^j * P_P * (1 - P_I)^2 (1 - P_P)^{N_P - i} & \text{for } c = j * N + i * M + z * (M - 1) \\[2ex] N_G * P_I^{(j+1)} * (1 - P_I)^2 (1 - P_P)^{N_P} & \text{for } c = (j + 1) * N + z * (M - 1) \\[2ex] 0 & \text{otherwise} \end{cases}$$

$$(6)$$

Where δ is defined in (7) (see Appendix for more details).

$$\delta = \sum_{r=1}^{M-c} (1 - P_B)^\sigma$$

$$\sigma = \begin{cases} 0 & \text{if } r = 1 \text{ and } r + c = M \\ 1 & \text{if } (r = 1 \text{ and } r + c < M) \text{ or } (r > 1 \text{ and } r + c = M) \\ 2 & \text{if } r > 1 \text{ and } r + c < M \end{cases} \quad (7)$$

The total number of cuts T_{cut} can be computed as:

$$T_{\text{cut}} = \sum_{c=1}^{F} N_{\text{cut}}[c] \quad (8)$$

The proportion of cuts of c frames length ($P_{\text{cut}}[c]$) can be computed dividing the number of cuts of c frames length ($N_{\text{cut}}[c]$) by the total number of cuts (T_{cut}). The cut length Probability Mass Function P_{cut} is the set of all possible values of $P_{\text{cut}}[c]$.

$$P_{\text{cut}}[c] = \frac{N_{\text{cut}}[c]}{T_{\text{cut}}} = \frac{N_{\text{cut}}[c]}{\sum_{i=1}^{F} N_{\text{cut}}[i]} \quad (9)$$

The average cut length L_{cut} can be computed from the cut length Probability Mass Function:

$$L_{\text{cut}} = \sum_{c=1}^{F} c * P_{\text{cut}}[c] = \frac{\sum_{c=1}^{F} c * N_{\text{cut}}[c]}{\sum_{c=1}^{F} N_{\text{cut}}[c]} = \frac{\sum_{c=1}^{F} c * N_{\text{cut}}[c]}{T_{\text{cut}}} \quad (10)$$

The Decodable Frame Rate can be derived from the cut length Probability Mass Function:

$$T_{\text{cut}} = \frac{F * (1 - Q)}{L_{\text{cut}}} \Rightarrow Q = 1 - \frac{L_{\text{cut}} * T_{\text{cut}}}{F} = 1 - \frac{1}{F} \sum_{c=1}^{F} c * N_{\text{cut}}[c] \quad (11)$$

4 Analytical model validation on packet switching networks

In Section 3, we have presented the analytical model for computing the cut length Probability Mass Function P_{cut}, the Average Cut Length L_{cut} and the Decodable Frame Rate Q for network scenarios where frame losses can be considered independent. In this section, we present simulation results that validate the applicability of the analytical model. Statistical results are obtained at 95 % confidence level, but most confidence intervals are too small to be noticed in the figures.

The video source takes a video trace containing the size and timestamps for each frame. The video source generates UDP packets using the frame's size from the video traces. The number of packets per frame depends on the frame length and the selected packet length. Therefore, all packets from the video have the same length, except for the last one from each frame. A summary of the different video traces

from [25, 27] used as video flows is presented in Table 2. Video traces with same and different bit-rate, GoP and relative frame sizes have been used. For example, two versions of *Tokyo* video have been used. Both have the same GoP, but different bit-rate and therefore, as stated on [24], different relative frame sizes.

The analytical formulation requires the frame loss probabilities as an input parameter. If the packet loss ratio can be modelled as an independent rate p, then the probability P_τ of losing a frame of type $\tau \in \{I, P, B\}$ can be approximated using (12) based on the packet loss ratio p and the average number of packets per frame d_τ.

$$P_\tau = \sum_{k=1}^{d_\tau} \binom{d_\tau}{k} p^k * (1-p)^{d_\tau - k} = 1 - (1-p)^{d_\tau} \tag{12}$$

First, we assume an environment where the packet loss ratio that the video flow experiences is an independent rate p. This environment is modelled by a black box network scenario with an i.i.d. packet loss ratio (Subsection 4.1). Afterwards, a more realistic environment will be studied (Subsection 4.2), where the packet loss ratio is the result of output port contention on routers. The analytical model requires that frame losses are independent. In a more realistic environment, when the link is congested, buffer is close to full occupancy and bursty arrivals can result in bursty losses. Therefore, packet losses can present correlation and it can not be asserted that the analytical model is valid. However, as the results will show, the model stays accurate for low to medium link utilizations and only for highly congested

Table 2 Summary of video traces used in the simulations

	The Matrix (1999) (Matrix)	The Lord of the Rings III (LOTRIII)	Tokyo Olympics (TokyoQP4)	Star Wars IV (StarWars IV)	Tokyo Olympics (TokyoQP1)
GoP structure	$G12B2$	$G12B2$	$G16B7$	$G16B7$	$G16B7$
Bitrate (Mbps)	0.403	0.714	0.888	3.143	5.502
I-frames contribution (% of bit-rate)	24.245	21.350	13.312	9.537	7.134
P-frames contribution (% of bit-rate)	29.464	30.161	9.610	8.314	7.934
B-frames contribution (% of bit-rate)	46.291	48.489	77.078	82.149	84.932
Average number of packets per frame	1.899	2.936	3.523	11.179	19.186
Average number of packets per I-frame (d_I)	4.483	6.713	6.924	16.795	21.822
Average number of packets per P-frame (d_P)	2.127	3.431	5.148	14.696	24.222
Average number of packets per B-frame (d_B)	1.491	2.278	3.164	10.527	18.638
Packets per I-frame / Packets per B-frame	3.007	2.946	2.189	1.596	1.171
Packets per P-frame / Packets per B-frame	1.426	1.506	1.627	1.396	1.300

links it deviates from the simulation results. For each of the environments, a specific simulator was developed using OMNeT++ [28].

4.1 Model validation in a network scenario with i.i.d. packet losses

Figure 1 shows the first four terms of the cut length Probability Mass Function (P_{cut}) versus the independent packet loss ratio p for the different video traces. The analytical results match quite well with the simulations. *LOTRIII* and *Matrix* traces have the same GoP structure, $G12B2$, obtaining a very similar $P_{cut}[c]$. However, this is not always the case. *TokyoQP4*, *TokyoQP1* and *StarWarsIV* have the same GoP structure, $G16B7$, but different $P_{cut}[c]$. The cut length Probability Mass Function does not only depend on the GoP structure, but it also depends on the relation between the average number of packets per each type of frame. Table 2 shows that *TokyoQP4*, *TokyoQP1* and *StarWarsIV* have the same GoP structure, but they differ greatly in the relation between the average number of packets. However, *LOTRIII* and *Matrix* have the same GoP structure and a similar relation between the average number of packets.

As the packet loss ratio p decreases, the cut length Probability Mass Function tends to stabilize. This happens because the probabilities of cut lengths that involve the loss of multiple frames tends to zero. These negligible cut length probabilities depend on the GoP structure, e.g. for $G12B2$ they will be $c \in \{2, 8\}$ and for $G16B7$ they will be $c \in \{2, 3, 4\}$ (see Fig. 1). So, as the packet loss ratio p decreases, the analytical formulation can be simplified assuming that cut lengths coming from the loss of multiple frames can not happen (see Section 5).

Figure 2a shows the simulation results for Average Cut Length (L_{cut}) versus the independent packet loss ratio p. The analytical results match quite well for all video traces. As the packet loss ratio p grows, more frames are lost and it is more probable that these losses interact on greater cut lengths. As *LOTRIII* and *Matrix* have a similar P_{cut} the Average Cut Length tends to the same value when p decreases, something that does not happen with *TokyoQP4*, *TokyoQP1* and *StarWarsIV*.

Figure 2b shows the simulation results for Decodable Frame Rate (Q) versus the independent packet loss ratio. Again, the analytical results match quite well with the simulations. As the packet loss ratio grows, more frames are lost and a lower Q is obtained.

As the bit-rate grows, the number of packets per frame grows, and therefore the frame loss probabilities increase too. So, the same video (e.g. *TokyoQP4*) with a higher bit-rate (*TokyoQP1*) will suffer more losses and will have a worse Decodable Frame Rate, as seen in Fig. 2b. However, the Average Cut Length can be smaller, as seen in Fig. 2a when $p < 0.01$. Basically, *TokyoQP1* suffers more cuts than *TokyoQP4* in the same scenario, but these cuts are shorter. This can be good, for example for video recovery techniques such as interpolations, that work well with little losses. This possible disparity between the Decodable Frame Rate and the Average Cut Length reflects the importance of considering not only the Decodable Frame Rate for a QoE metric, but to complement it with the distribution of the cuts as proposed in this paper.
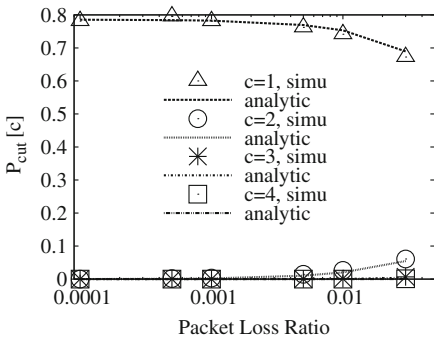
So far, the results have been validated for a scenario with an independent packet loss ratio, where the frame loss probabilities depend only on the frame sizes. On the following, we simulate a network scenario where packet losses can present
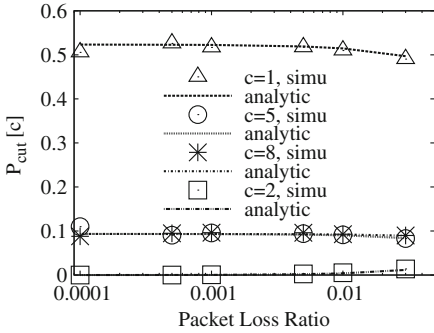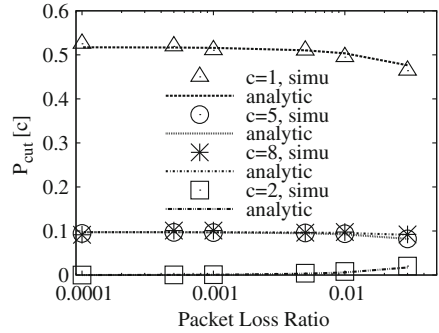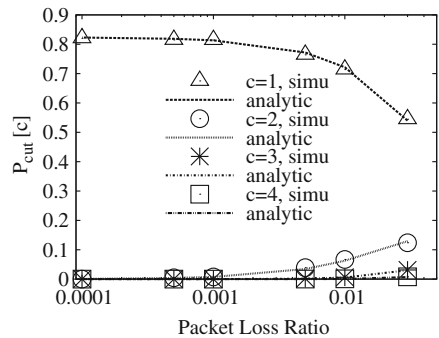

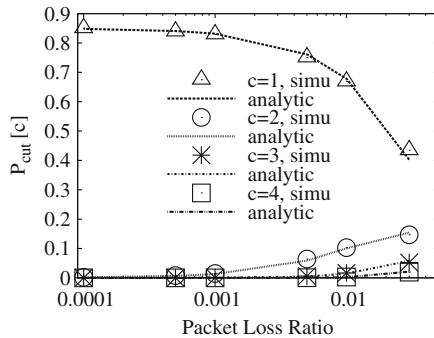**Fig. 1** Cut length probability mass function for different video traces in a network with independent losses
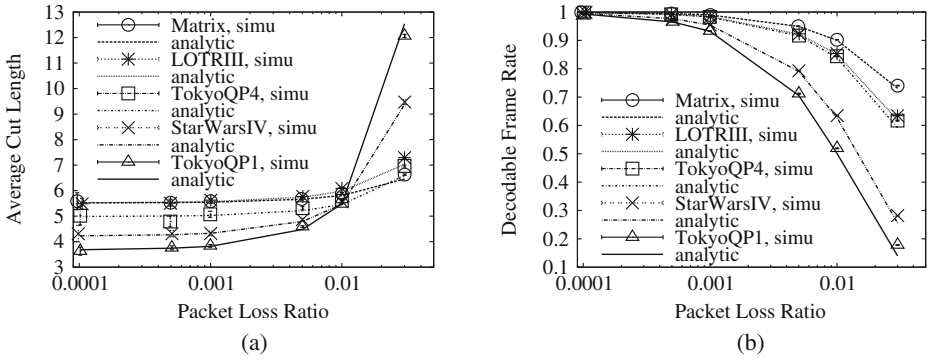
Fig. 2 Average cut length and decodable frame rate for different video traces in a network with independent losses

correlation and hence frame losses too. We check the validity of the formulation when independence of frame loss probabilities is not assured.

4.2 Model validation in a network scenario with real traffic

The network scenario, Fig. 3, is similar to the scenario used in other papers [29]. Each switching node is modelled as a router with a finite queue on the output port. There are three background traffic flows in the network. Each background traffic flow only competes with the video traffic for resources (bandwidth and queue space) in one router, i.e., FlowA goes from LAN A1 to LAN A2 through RouterA, FlowB goes from LAN B1 to LAN B2 through RouterB, and FlowC goes from LAN C1 to LAN C2 through RouterC. Each background traffic flow is generated from different Ethernet packet traces from the WIDE project's MAWI Working Group Traffic Archive [2, 17]. The *2010/04/13* set from the *Day in the Life of the Internet project* is used. Chronologically consecutive Ethernet traces are concatenated to obtain at least the duration of the video flow and then packet's inter-arrival time and size are extracted. Each background traffic source uses these inter-arrival time and
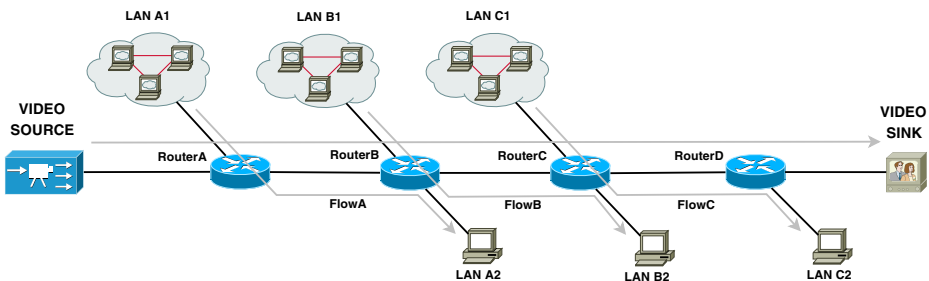

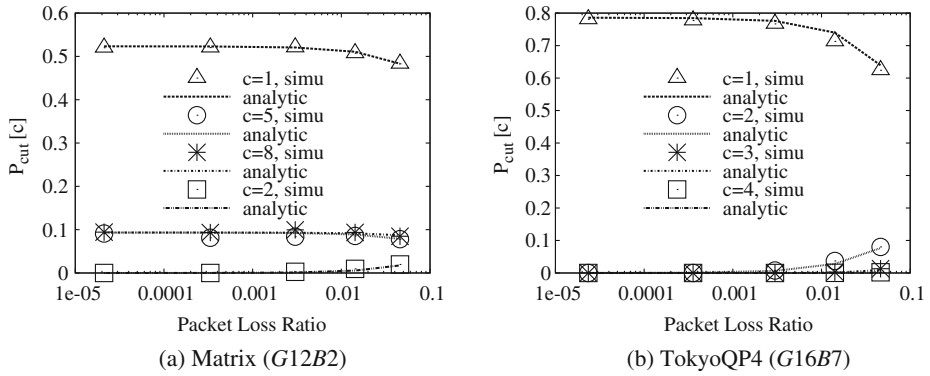**Fig. 3** Network scenario with background traffic

Fig. 4 Cut length Probability Mass Function for different video traces in the network scenario with background traffic

size to generate the corresponding background traffic. The Ethernet traces have an average rate of 200 Mbps, while the simulation links have 2 Gbps bandwidth. Larger background traffic rates are created multiplexing several Ethernet traces before the packet's inter-arrival time and size extraction process. As more background traffic is added on each hop, the packet loss ratio grows.

Figure 4 shows the first four terms of the cut length Probability Mass Function (P_{cut}) versus the experimental packet loss ratio for only *Matrix* and *TokyoQP1*. The analytical results match quite well with the simulation ones for both video traces. Again, as the packet loss ratio decreases, the cut length Probability Mass Function tends to stabilize and the analytical formulation can be simplified assuming that cut lengths that imply the loss of multiple frames will not happen (see Section 5).

Figure 5 shows the simulation results of Average Cut Length (L_{cut}) and Decodable Frame Rate (Q) versus the experimental packet loss ratio. Again, the analytical
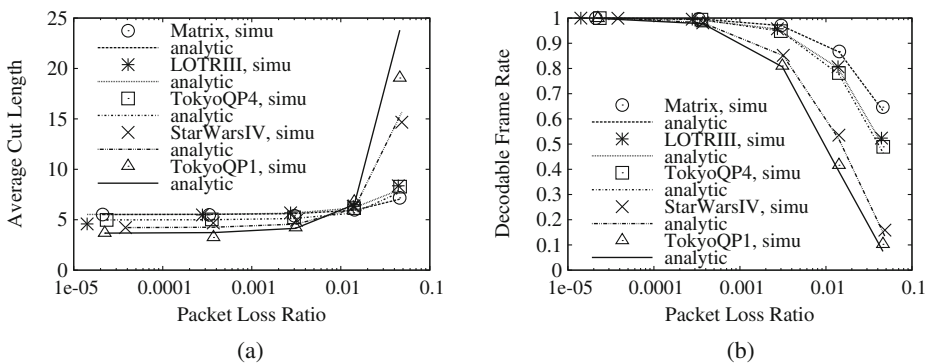

Fig. 5 Average Cut Length and Decodable Frame Rate for different video traces in the network scenario with background traffic

results match quite well for both video traces for small packet loss ratios. For high packet loss ratios ($p \sim 0.04$) the analytical results differ from simulations ones. This happens for high background traffic rates, where the router's queue is saturated most of the time and the losses are bursty. The analytical model needs independent packets losses, so for high background traffic rates the model is not valid. Production networks usually have small packet loss ratios, so the analytical model could be used in realistic packet switching scenarios and the frame loss probabilities can be obtained from the packet loss ratio and the average number of packets per frame.

For moderate or low packet loss ratios ($p \leq 0.01$), the model accuracy is good. The analytical model for the Average Cut Length has an error below 0.5 frames or 20 ms (40 ms inter-frame time) for $p = 0.01$. The authors in [22, 23] show that the MOS does not have significant changes, even in the most sensitive range, for variations in cut lengths below the inter-frame time (Figure 7 in [23]). The analytical model for the Decodable Frame Rate has an error below 3 % for $p = 0.01$. The authors in [3, 4] relate this rate to the PSNR. They show that the PSNR value greatly depends on the content under study, but for the same content the PSNR does not have significant changes for variations in Decodable Frame Rate below that range.

5 Model evaluation results

The analytical model for the Decodable Frame Rate shows that as the bit-rate grows, the Decodable Frame Rate decreases. This happens because the number of packets per frame grows, and therefore the frame loss probabilities for a packet switching network, P_τ, increase too. It can be checked on Figs. 2b and 5b with the extreme cases of *TokyoQP1* and *TokyoQP4*. *TokyoQP1* has a bit-rate six times greater than *TokyoQP4*, and therefore the difference on Decodable Frame Rate is noticeable.

However, the analytical model for video cuts, $P_{\text{cut}}[c]$, has not a so clear behaviour with the bit-rate. The Average Cut Length can be smaller for the higher bit-rate case, as seen in Fig. 5a when $p < 0.01$. *TokyoQP1* has more cuts than *TokyoQP4*, but these cuts are shorter. Video recovery techniques such as interpolations work well in scenarios with few losses.

This disparity between the Decodable Frame Rate and the Average Cut Length reflects the importance of considering both parameters together, not only the Decodable Frame Rate. *TokyoQP1* always will have a worse Decodable Frame Rate than *TokyoQP4*, but for small p this can be compensated by the improvement on the Average Cut Length. In other cases this compensation can be happen for any p.

In the previous section it was shown that the analytical formulation could be simplified when the packet loss ratio p is low. If p is small, it can be assumed that the probability of packet losses in adjacent frames tends to zero, i.e., only cuts resulting one frame losses will be considered. This leads to:

– Only cuts of length $c = 1, i * M + z * (M - 1), N + z * (M - 1)$ can be possible.
– All terms not related to the lost frame can be removed from the analytical formulation, i.e., terms of type $1 - P_{\{I, P, B\}}$ can be removed.

Taking into account these assumptions, a simplified versions of the number of cuts can be computed:

– Number of cuts of c frames length

$$N_{\text{cut}}[c] \approx \begin{cases} N_G * P_B * N_B & \text{for } c = 1 \\ N_G * P_P & \text{for } c = i * M + z * (M - 1) \\ N_G * P_I & \text{for } c = N + z * (M - 1) \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

– Total number of cuts

$$T_{\text{cut}} \approx N_G * P_B * N_B + N_G * P_P * N_P + N_G * P_I = N_G \sum_{\tau \in \{I, P, B\}} P_\tau * N_\tau \quad (14)$$

– Cut length Probability Mass Function

$$P_{\text{cut}}[c] \approx \begin{cases} \dfrac{P_B * N_B}{\sum\limits_{\tau \in \{I, P, B\}} P_\tau * N_\tau} & \text{for } c = 1 \\ \dfrac{P_P}{\sum\limits_{\tau \in \{I, P, B\}} P_\tau * N_\tau} & \text{for } c = i * M + z * (M - 1) \\ \dfrac{P_I}{\sum\limits_{\tau \in \{I, P, B\}} P_\tau * N_\tau} & \text{for } c = N + z * (M - 1) \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

– Average Cut Length

$$L_{\text{cut}} \approx \frac{P_B * N_B + \sum\limits_{i=1}^{N_P} [i * M + z * (M - 1)] P_P + [N + z * (M - 1)] P_I}{P_B * N_B + \sum\limits_{i=1}^{N_P} P_P + P_I} =$$

$$= \frac{P_B * N_B + \left[M \frac{N_P * (N_P + 1)}{2} + z * (M - 1) N_P \right] P_P + [N + z * (M - 1)] P_I}{P_B * N_B + P_P * N_P + P_I} \quad (16)$$

It can be seen in (15) that the probability of a c frames length cut, $P_{\text{cut}}[c]$, is proportional to the loss probability of the frame that generates that cut length and the number of these frames inside a GoP. So, the more B-frames a GoP contains, the more cuts will be of only one frame length and the Average Cut Length will be smaller. The loss of a B-frame has not a great impact in the average cut length (16) compared to the loss of a P- or I-frame. P-frame losses get amplified their effect on the average cut length by a factor of $M \frac{(N_P + 1)}{2} + z(M - 1)$; I-frames by a factor of $N + z(M - 1)$ and B-frames by a factor of 1. Although the loss of a frame is an sporadic incident when p is small, it is very important the type of lost frame. Therefore, any attempt to improve the transmission of a video must be based on reducing the loss of I- and/or P-frames.

Taken into account the same assumptions, also the Decodable Frame Rate can be simplified. The expression proposed on [29] tends to 1 (17), but the expression

derived from the cut length Probability Mass Function still depends on the GoP structure (and the frame losses) (18).

$$Q = \frac{(1 - P_I) + (1 - P_I) \sum_{i=1}^{N_P} (1 - P_P)^i}{N}$$

$$+ \frac{(M-1)(1-P_I)(1-P_B)\left[\sum_{i=1}^{N_P}(1-P_P)^i + z*(1-P_I)(1-P_P)^{N_P}\right]}{N}$$

$$\approx \frac{1 + \sum_{i=1}^{N_P} 1 + (M-1)\left[\sum_{i=1}^{N_P} 1 + z\right]}{N} = \frac{1 + N_P + (M-1)(N_P + z)}{N} = 1 \quad (17)$$

$$Q = 1 - \frac{L_{\text{cut}} * T_{\text{cut}}}{F} \approx 1 - \frac{L_{\text{cut}}}{N} \sum_{\tau \in \{I,P,B\}} P_\tau * N_\tau$$

$$= 1 - \frac{P_B * N_B + \left[M\frac{N_{P}*(N_P+1)}{2} + z*(M-1)N_P\right]P_P + [N + z*(M-1)]P_I}{N}$$

$$(18)$$

This is a clear advantage of the presented formulation for computing the Decodable Frame Rate as it provides a better approximation for low loss scenarios. It also shows the importance of considering the Decodable Frame Rate together with the video cut lengths.

6 Conclusions

This paper has presented two objective video quality evaluation parameters for a network where losses of video frames can be considered independent: the Decodable Frame Rate and the video cut lengths. The Decodable Frame Rate has been used on previous works, but user perception is not only affected by the number of non-decoded frames, but also by the video playback interruptions caused by the grouping of these non-decoded frames. Therefore, these two parameters have to be considered important parameters to quantify the video quality. The analytical formulation for them has been presented and the importance of considering together the two parameters has been reflected.

The analytical model has shown that as the bit-rate grows, the Decodable Frame Rate decreases. However, the Average Cut Length can be smaller for the higher bit-rate case, because it can have more cuts, but they can be shorter. This reinforces the importance of considering the two parameters together.

The simplified analytical model shows that, as expected, the loss of a B-frame has not a great impact in the average cut length comparing to the loss of a P- or I-frame. Although the loss of a frame is an sporadic incident when the packet loss ratio is small, the simplified analytical model shows that the type of lost frame is very important. Therefore, any attempt at improving the transmission of a video should be

directed at minimizing the Average Cut Length and/or at maximizing the Decodable Frame Rate reducing the number of frame losses. Based on the analytical results on frame losses, depending on the GoP structure the best strategy for this improvement will be the reduction of I-, P- and/or B-frames losses.

Appendix: Cut lengths in a video

In a video flow that suffers frame losses, the cut length can vary from one frame length to the total length of the video. The cut length depends on the grouping of the directly lost frames and of the non-decodable frames due these lost frames. The cuts must be understood as experienced by the user, therefore they must be measured from the presentation order of the videos, not the transmission order where the losses take place.

For example (Fig. 6), in a video with the $IPBBPBB$ GoP structure in transmission order, if the second B-frame and second P-frame are lost, the four last frames of the GoP could not been decoded. But in presentation order, two disconnect cuts of size 1 and 3 will be generated. This makes the analytical model of cuts more complex.

When an I-frame is directly lost the whole GoP cannot be decoded, creating a single cut of at least length N. In the case of an open GoP, the last B-frames from the previous GoP will be non-decodable and they will be part of the same single cut of length $N + (M - 1)$ when considered in presentation order. If we define the control variable $z = \frac{N_B}{M-1} - N_P$, where $z = 1$ if the GoP is an open one, and $z = 0$ if it is a closed GoP, then the cut length will be $N + z * (M - 1)$.

If the I-frame of the next GoP is lost too, all the frames from the next GoP will not be decoded and a single cut of length $2 * N + z * (M - 1)$ will be generated. In general, when j consecutive I-frames are lost, a single cut of length $j * N + z * (M - 1)$ will be generated.

The loss of a P-frame makes impossible to decode all the following frames in the GoP. As all these frames are consecutive, both in transmission and presentation order, only one cut is generated. If the lost P-frame is the last or N_Pth P-frame of the GoP, the cut will be of $M + z * (M - 1)$ frames length. If the lost P-frame is the penultimate or $(N_P - 1)$th P-frame of the GoP, the cut will be of $2 * M + z * (M - 1)$ frames length. If the lost P-frame is the first P-frame of the GoP, the cut will be of
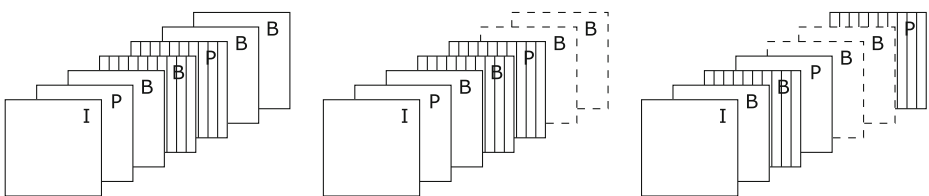

Fig. 6 *Left*: second B-frame and second P-frame in transmission order are lost; *Center*: the last B-frames cannot be decoded; *Right*: in presentation order two cuts of sizes 1 and 3 are generated

$N_P * M + z * (M - 1)$ frames length. In general, when the $(N_P + 1 - i)$th P-frame of the GoP is lost, a single cut of $i * M + z * (M - 1)$ will be generated.

If the I-frame from the next GoP is lost too, then all the frames from the next GoP will not be decoded. Again, all these frames are consecutive and result in only one cut of $N + i * M + z * (M - 1)$ frames. In general, when the $(N_P + 1 - i)$th P-frame of the GoP is lost and the next j I-frames are lost, a single cut of length $j * N + i * M + z * (M - 1)$ will be generated.

The loss of c consecutive B-frames of a B-frames block does not affect other frames and it creates only one cut. If the last B-frame of the B-frames block is lost and the next frame (either I-frame or P-frame) is lost too, in presentation order the non-decoded frames will not be consecutive. The B-frames will be on one cut and the frames non-decoded from the lost I- or P-frame will be on the other cut.

The possible cut lengths are summarized in (19), where F is the number of frames in the video.

$$c = [1 \ldots M - 1, \ j * N + i * M + z * (M - 1), \ (j + 1) * N + z * (M - 1)]$$
$$i = 1 \ldots N_P$$
$$j = 0 \ldots N_G - 1 \tag{19}$$

A similar approach was used for the analytical expression of Q in [29]. On the following we obtain the analytical expression of $N_{\text{cut}}[c]$, the number of cuts of c frames length.

$P_{\{I,P,B\}}$ is the probability of a {I, P, B}-frame being directly lost. It is assumed that direct frame losses are mutually independent.

A cut of length $(j + 1) * N + z * (M - 1)$ frames will be generated if $j + 1$ consecutive I-frames are lost and the previous and next (in presentation order) frames are decoded. The next frame will always be the next I-frame, so this I-frame cannot be lost. The previous frame will always be the last P-frame of the previous GoP, so this P-frame cannot be lost. But any of the P-frames and the I-frame of this GoP cannot be lost too, or the last P-frame will not be decoded. Therefore, $N_{\text{cut}}[(j + 1) * N + z * (M - 1)] = N_G * P_I^{(j+1)} * (1 - P_I)^2 (1 - P_P)^{N_P}$, where $N_G = F/N$ is the number of GoPs in the video.

An $i * M + z * (M - 1)$ frames length cut will be generated if the $(N_P + 1 - i)$th P-frame of a GoP is lost, and again if previous and next (in presentation order) frames are decoded. Remember that $i = 1 \ldots N_P$. The next frame will always be the next I-frame, so this I-frame cannot be lost. The previous frame will be a previous P-frame on the GoP or the I-frame of the GoP, so the previous P-frames and the I-frame cannot be lost. Therefore, $N_{\text{cut}}[i * M + z * (M - 1)] = N_G * P_P * (1 - P_I)^2 (1 - P_P)^{N_P - i}$.

A cut of length $j * N + i * M + z * (M - 1)$ frames will be generated if after losing a P-frame the next j I-frames are lost. Then, $N_{\text{cut}}[j * N + i * M + z * (M - 1)] = N_G * P_I^j * P_P * (1 - P_I)^2 (1 - P_P)^{N_P - i}$.

To generate a $M - 1$ frames length cut, the $M - 1$ frames of a B-frames block have to be lost and its neighbouring frames in presentation order have to be decoded. These neighbouring frames will be the two previous P-frames or the previous I- and P-frame in transmission order. As the P-frames depend on the I-frame and its previous P-frames on the GoP, all the previous P-frames and the I-frame on the GoP

cannot be lost. Therefore, for the ith B-frames block the number of cuts of length $M-1$ frames is $N_G * P_B^{M-1} * (1-P_I)(1-P_P)^i$, where $i=1\ldots N_P$.

For an open GoP, the reasoning for B-frames is valid but incomplete, because there are (N_P+1) B-frames blocks on the GoP, not only N_P. The frames in this "extra" block depend not only on the I-frame and the N_P P-frames of the GoP, but on the I-frame of the next GoP too. Then, for the (N_P+1)th B-frames block of an open GoP the number of cuts of length $M-1$ frames is $N_G * P_B^{M-1} * (1-P_I)^2 (1-P_P)^{N_P}$.

In general, the number of cuts of length $M-1$ frames is:

$$N_G * P_B^{M-1} * (1-P_I) \sum_{m=1}^{N_P} (1-P_P)^m + z * N_G * P_B^{M-1} * (1-P_I)^2 (1-P_P)^{N_P}$$

This procedure can be extended to cuts from $c=1$ to $c=M-2$ frames length. But there is an extra difficulty. Now in each B-frames block there are $M-c$ possible loss combinations that generate a c frames length cut. For example, for the case of $c=M-2$ they are two possibilities (see Fig. 7). The first possibility is to lose the first $M-2$ B-frames of the block and to not lose the last B-frame. The second possibility is to not lose the first B-frame of the block and to lose the other $M-2$ B-frames. In both cases, for the ith B-frames block, the number of cuts of length $M-2$ frames is $N_G * (1-P_B) * P_B^{M-2} * (1-P_I)(1-P_P)^i$.

To mathematically express the analytical model of $c=1\ldots M-1$ lengths cut, we define σ and δ. σ is the minimum number of B-frames from the same block that should be available to the decoding process (should not be lost) to prevent a cut length larger than c, when the first lost B-frame is the rth in the block. δ is the contribution to the number of cuts of length c frames by the $M-c$ possible loss combinations of a B-frames block that generate a c frames length cut. (20) presents the expression for σ and (21) presents the expression for δ.
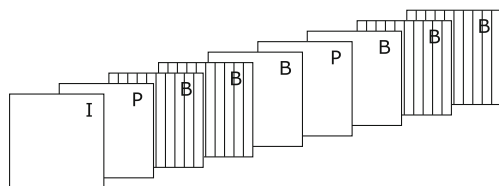
$$\sigma = \begin{cases} 0 & \text{if } r=1 \text{ and } r+c=M \\ 1 & \text{if } (r=1 \text{ and } r+c<M) \text{ or } (r>1 \text{ and } r+c=M) \\ 2 & \text{if } r>1 \text{ and } r+c<M \end{cases} \qquad (20)$$

$$\delta = \sum_{r=1}^{M-c} (1-P_B)^\sigma \qquad (21)$$

In general, the number of cuts of length $c=1\ldots M-1$ frames is:

$$N_G * \delta * P_B^c * (1-P_I) \sum_{m=1}^{N_P} (1-P_P)^m + z * N_G * P_B^c * (1-P_I)^2 (1-P_P)^{N_P}$$

Fig. 7 GoP structure $(9,4)$ with its two possibilities to generate a two frames cut length

Equation (22) presents the general expression for $N_{\text{cut}}[c]$. It depends on the GoP structure of the video and on the loss probabilities of the frames, $P_{\{I,P,B\}}$.

$$N_{\text{cut}}[c] = \begin{cases} N_G * \delta * P_B^c(1-P_I) \sum\limits_{m=1}^{N_P} (1-P_P)^m + \\ + z * N_G * \delta * P_B^c * (1-P_I)^2 (1-P_P)^{N_P} \quad \text{for } c = 1 \ldots M-1 \\[2ex] N_G * P_I^j * P_P * (1-P_I)^2 (1-P_P)^{N_P-i} \quad \text{for } c = j*N+i*M+z*(M-1) \\[2ex] N_G * P_I^j * (1-P_I)^2 (1-P_P)^{N_P} \quad\quad\quad \text{for } c = (j+1)*N+z*(M-1) \\[2ex] 0 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{otherwise} \end{cases} \tag{22}$$

The total number of cuts T_{cut} can be computed as:

$$T_{\text{cut}} = \sum_{c=1}^{F} N_{\text{cut}}[c] \tag{23}$$

The proportion of cuts of c frames length ($P_{\text{cut}}[c]$) can be computed dividing the number of cuts of c frames length ($N_{\text{cut}}[c]$) by the total number of cuts (T_{cut}). The cut length Probability Mass Function P_{cut} is the set of all possible values of $P_{\text{cut}}[c]$.

$$P_{\text{cut}}[c] = \frac{N_{\text{cut}}[c]}{T_{\text{cut}}} = \frac{N_{\text{cut}}[c]}{\sum\limits_{i=1}^{F} N_{\text{cut}}[i]} \tag{24}$$

The average cut length L_{cut} can be computed from the cut length Probability Mass Function:

$$L_{\text{cut}} = \sum_{c=1}^{F} c * P_{\text{cut}}[c] = \frac{\sum\limits_{c=1}^{F} c * N_{\text{cut}}[c]}{\sum\limits_{c=1}^{F} N_{\text{cut}}[c]} = \frac{\sum\limits_{c=1}^{F} c * N_{\text{cut}}[c]}{T_{\text{cut}}} \tag{25}$$

References

1. Bikfalvi A, García-Reinoso J, Vidal I, Valera F, Azcorra A (2011) P2P vs. IP multicast: Comparing approaches to IPTV streaming based on TV channel popularity. Comput Networks 55(6):1310–1325
2. Borgnat P, Dewaele G, Fukuda K, Abry P, Cho K (2009) Seven years and one day: sketching the evolution of internet traffic. In: INFOCOM 2009. IEEE, pp 711–719
3. Cheng RS, Lin CH, Chen JL, Chao HC (2012) Improving transmission quality of MPEG video stream by SCTP multi-streaming and differential RED mechanisms. J Supercomputing 62:68–83. doi:10.1007/s11227-011-0624-2
4. Chih-Heng Ke CKS (2008) An evaluation framework for more realistic simulations of MPEG video transmission. J Inf Sci Eng 24(2):425–440

5. Cisco Visual Networking Index (2011) Forecast and methodology, 2010–2015. Tech Rep Cisco Systems Inc http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf. Accessed 7 May 2012
6. Espina F, Morato D (2012) Survey on the current uses of H.264. Tech Rep Public University of Navarre (Spain). https://www.tlm.unavarra.es/~felix/publicaciones/TR/H.264.pdf
7. Fiedler M, Hossfeld T, Tran-Gia P (2010) A generic quantitative relationship between quality of experience and quality of service. IEEE Netw 24(2):36–41
8. Heyman D, Lakshman T (1996) Source models for VBR broadcast-video traffic. IEEE/ACM Trans Netw 4(1):40–48
9. Huynh-Thu Q, Ghanbari M (2008) Scope of validity of PSNR in image/video quality assessment. Electron Lett 44(13):800–801
10. ISO JTC 1/SC 29 (2000) ISO/IEC 13818-2:2000: information technology – generic coding of moving pictures and associated audio information: video. ISO
11. ISO JTC 1/SC 29 (2009) ISO/IEC 14496-2:2004: information technology – coding of audio-visual objects – Part 2: visual. ISO
12. ISO JTC 1/SC 29 (2012) ISO/IEC 14496-10:2010: information technology – coding of audio-visual objects – Part 10: advanced video coding. ISO
13. ITU-T Study Group 12 (2008) Recommendation P.10/G.100 (2006) Amendment 2: vocabulary for performance and quality of service – new definitions for inclusion in recommendation ITU-T P.10/G.100. ITU-T
14. ITU-T Study Group 9 (2008) Recommendation J.247 (08/2008): objective perceptual multimedia video quality measurement in the presence of a full reference. ITU-T
15. Kusuma T, Zepernick HJ (2003) A reduced-reference perceptual quality metric for in-service image quality assessment. In: Mobile future and symposium on trends in communications, 2003. Joint First Workshop on SympoTIC '03, pp 71–74
16. Lin CH, Ke CH, Shieh CK, Chilamkurti N (2006) The packet loss effect on MPEG video transmission in wireless networks. In: Advanced information networking and applications, 2006. In: 20th International Conference on AINA 2006, vol 1, pp 565–572
17. MAWI (Measurement and Analysis on the WIDE Internet) Working Group (2012) 150 megabit ethernet anonymized packet traces without payload: WIDE-TRANSIT link @ Tokyo, Japan. http://mawi.wide.ad.jp/mawi/ditl/ditl2010/. Accessed 7 May 2012
18. Moving Picture Experts Group (2012) The moving picture experts group (MPEG) home page. http://mpeg.chiariglione.org/. Accessed 7 May 2012
19. Osama A, Lotfallah MR, Panchanathan S (2006) A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks. EURASIP J Adv Signal Process 2006:042083. doi:10.1155/ASP/2006/42083
20. Pastrana-Vidal R, Gicquel J (2006) Automatic quality assessment of video fluidity impairments using a no-reference metric. In: Proc. of int. workshop on video processing and quality metrics for consumer electronics
21. Pastrana-Vidal R, Gicquel J (2007) A no-reference video quality metric based on a human assessment model. In: Third international workshop on video processing and quality metrics for consumer electronics VPQM, vol 7, pp 25–26
22. Pastrana-Vidal R, Gicquel J, Colomes C, Cherifi H (2004) Frame dropping effects on user quality perception. In: 5th international workshop on image analysis for multimedia interactive services
23. Pastrana-Vidal RR, Gicquel JC, Colomes C, Cherifi H (2004) Sporadic frame dropping impact on quality perception. In: Rogowitz BE, Pappas TN (eds. Human vision and electronic imaging IX, vol 5292. SPIE, pp 182–193
24. Reisslein M, Lassetter J, Ratnam S, Lotfallah O, Fitzek FH, Panchanathan S (2002) Traffic and quality characterization of scalable encoded video: a large-scale trace-based study, part 1: overview and definitions. Arizona State Univ., Dept. of Electrical Eng., Tech. Rep. http://trace.eas.asu.edu/publications/p1.pdf. Accessed 7 May 2012
25. Seeling P, Reisslein M, Kulapala B (2004) Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial. IEEE Commun Surv Tutor 6(3): 58–78
26. Tionardi L, Hartanto F (2003) The use of cumulative inter-frame jitter for adapting video transmission rate. In: TENCON 2003. conference on convergent technologies for Asia-Pacific region, vol 1, pp 364–368
27. Van der Auwera G, David PT, Reisslein M (2008) Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension. IEEE Trans Broadcast 54(3):698–718

28. Varga A, Hornig R (2008) An overview of the OMNeT++ simulation environment. In: Simu-tools '08: Proceedings of the 1st international conference on simulation tools and techniques for communications, networks and systems & workshops. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, pp 1–10
29. Ziviani A, Wolfinger BE, Rezende JF, Duarte OC, Fdida S (2005) Joint adoption of QoS schemes for MPEG streams. Multimedia Tools Appl 26(1):59–80

Felix Espina received his M.Sc. in telecommunications engineering from the Public University of Navarre in 2004. Since 2009 he has been with the Department of Automatics and Computer Science of the Public University of Navarre as an assistant lecturer. He is currectly pursuing his Ph.D. in the field of the multimedia content on next-generation optical Internet.

Daniel Morato received the M.Sc. degree in telecommunication engineering and the Ph.D. degree from the Public University of Navarre, Spain. During 2002 he was a visting postdoctoral fellow at the Electrical Engineering and Computer Sciences Department, University of California, Berkeley. Since 2006 he has been working at the Department of Automatics and Computer Sciences, Public University of Navarre, as an associate professor. His reseach interests include high-speed networks, performance and traffic analysis of Internet services and network monitoring.

Mikel Izal received his M.Sc. and Ph.D. degrees in telecommunication engineering in 1997 and 2002 respectively. In 2003 he worked as a scientific visitant at Institute Eurecom, Sophia-Antipolis, France, performing measures in network tomography and peer-to-peer systems. Since then, he has been with the Department of Automatics and Computer Sciences of the Universidad Pública de Navarra where he is an Associate Professor. His research interests include traffic analysis, network tomography, high speed next generation networks and peer to peer systems.

Eduardo Magaña received his M.Sc. and Ph.D. degrees in Telecommunications Engineering from Public University of Navarra, Pamplona, Spain, in 1998 and 2001, respectively. He is an associate professor at Public University of Navarra. During 2002 he was a postdoctoral visiting research fellow at the Department of Electrical Engineering and Computer Science, University of California, Berkeley. His main research interests are network monitoring, traffic analysis and performance evaluation of communication networks.