Impact of Packet Loss on H.264 Scalable Video Coding

Siyu Tang
Bell Labs, ALCATEL-LUCENT
Antwerp, Belgium
siyu.tang@alcatel-lucent.com

Patrice Rondao Alface
Bell Labs, ALCATEL-LUCENT
Antwerp, Belgium
Patrice.Rondao_Alface@alcatel-lucent.com

***Abstract*—This paper presents an exact study of the impact of packet loss on H.264 scalable video coding (SVC). A Markov Chain (MC) with 2^N states is developed to describe the error propagation process inside a group of pictures (GOP). The model is extended to estimate the number of frames affected by transmission errors for a video sequence composed of multiple GOPs. By analyzing the inter-frame prediction rules, we examine the performance of different GOP structures against transmission errors. From the exact analysis, several metrics are analytically determined. Based on the proposed metrics, the performance of the SVC hierarchical B-frame structure and the advance video coding (AVC) IPPP structure (compatible base layer in SVC) are evaluated and compared under the assumption of random packet loss with rate p.**

***Keywords*-QoE; SVC; AVC; Markov Chain**

I. Introduction

In order to transport compressed video over a packet-based network (e.g., IP-network), the encoded bit-stream needs to be fragmented according to the *maximum transfer unit* (MTU). In an error-prone environment, packet loss may occur and cause distortion on the perceived video quality at the receiver. The resulting video distortion, however, varies according to the inter-frame predicting rules being used. Hence, it is important to understand the impact of transmission errors on different encoding schemes.

The scalable video coding (SVC) amendment [1] for the H.264/AVC (Advance Video Coding) standard provides scalable video streams in the temporal, spatial and quality dimensions with graceful adaptation between different scalability layers. It is considered to be a promising approach to offer quality adaptation to heterogeneous receivers with varying bandwidth constraints. Video distortion caused by packet loss in SVC is of great importance, as it determines the level of quality degradation, and provides insights on efficient quality adaptation.

Predicting transmission distortion in SVC is challenging, due to the three scalable dimensions and the hierarchical inter-frame prediction structures (the terms of *inter-frame prediction* and *frame dependency* are used interchangeably in this paper.). As we will discuss in Section II, an exact model of the transmission distortion in SVC seems missing. In this paper, we focus on an inter-frame prediction model for the hierarchical temporal prediction structure. In particular, we aim to investigate the error propagation process in SVC hierarchical prediction structure and the robustness of different prediction mechanisms against transmission errors. Encoding schemes studied in this paper are: 1) SVC hierarchical B-frame structure (efficient compression, applicable for non-real-time video application), 2) AVC IPPP mode (SVC base layer compatible, applicable for real-time video delivery with stringent delay requirements).

The contributions of the paper are two-fold. First of all, to our knowledge, it is the first exact analysis that studies error propagation in the SVC hierarchical prediction structure. Instead of relying on extensive simulations, the exact model allows us to evaluate the performance of difference GOP (group of pictures) structures under packet loss accurately. Secondly, results obtained from this work can be used as a guideline in choosing the preferred codec (or GOP structure) that is more robust against transmission errors.

The rest of the paper is organized as follows. In Section II, related work is discussed. Section III presents preliminary definitions and the description of the error propagation problem. Section IV describes the proposed analytic model, along with the performance metrics. In Section V, we present the analytic results and key findings. Section VI concludes the paper.

II. Related Work

There exists a rich number of studies focusing on the transmission distortion problem on the baseline profile of H.264/AVC. To simplify the loss-model, most of the studies assumed an additive model when consecutive packet losses occur, e.g., [6], [7] and [8]. In [9], the non-linearity of transmission distortion was considered. The proposed algorithm has shown superior performance over the linear models.

Existing transmission distortion models in AVC, however, cannot be directly applied to SVC due to the hierarchical prediction structure being employed. Most rate-distortion models regarding SVC aimed to optimize the perceived video quality as a function of different encoding parameters, e.g., [13], [14]. Studies about transmission distortion in SVC are either performed by experimental measures, or via approximations. Monteiro *et al.* [10] quantified the impact of packet loss on the SVC video stream with extensive simulations. Ghareeb *et al.* [11] have shown that the effect of packet loss can be reduced when delivering SVC layers through multiple paths based on experiments conducted in the ns-2 simulator. In [12], a loss-distortion approximation model was developed but did not provide an exact analysis. In order to accurately predict the impact of packet loss on the SVC hierarchical prediction structure, an exact analysis is required, which is the focus of this paper.

III. Preliminary definitions and problem description

We consider the delivery of N encoded frames over a lossy network, where a unique *identification number* (ID) k $(0 \leq k \leq N-1)$ is assigned to each frame. Let d $(0 \leq d \leq n_{gop} - 1)$ be the unique ID of each GOP, where n_{gop} is the total number of GOPs in the video sequence. Let t $(0 \leq t \leq m)$ be the unique identifier of each temporal sublayer (or simply temporal layer) where m is the number of temporal layers in each GOP.

Employing variable bit-rate (VBR) encoding at the encoder leads to variable frame sizes after encoding. Let s_k denote such a random variable (r.v.), where k is the frame ID. In IP-networking, the *maximum service unit* (MST) is fixed to 1460 bytes excluding the 40-byte header. Hence, the number of packets n_k consisting of frame k after IP fragmentation is also a random variable. The total number of fragmented packets of the N frames is therefore obtained by $M = \sum_{k=0}^{N} n_k$.

Losing a packet in frame k will not only affect the current frame, but also propagates the initial error to subsequent frames due to the hierarchical inter-frame coding. In this paper, we identify the reason for a frame becoming erroneous as follows. If at least one packet is lost in a frame, we say that the frame is *erroneous due to packet loss*. Error propagation within a GOP takes place step by step, which is defined as *error propagation steps* r. We consider a worst case where a corrupted Macroblock will be used for inter-frame prediction and further propagates the error to successive frames. Any frame that is predicted from an erroneous frame due to the inter-prediction structure is considered as *being erroneous due to frame dependency*. Note that a frame impacted by packet loss can be again affected by frame dependency. We do not apply any advanced error concealment technique to the video sequence, so that error propagation is only evaluated under the influence of GOP structure. Pixels containing errors are simply concealed with zeros.

We measure the error propagation process within a GOP by the total number of erroneous frames, $Y[r]$, after each propagation step r. Let $\{Y[r], r \geq 0\}$ describe such a stochastic error propagation process. In the successive steps, all erroneous frames in the previous step keep disseminating the error to their neighbouring frames according to the inter-prediction rule, resulting in $Y[r]$ erroneous frames after step r. The total number of erroneous frames, $Y[r]$, is non-decreasing with r. The process of $Y[r]$ varies with respect to the prediction structure being used.

State index i	$f_{N-1}f_{N-2}...f_3f_2f_1f_0$
0	00......0000
1	00......0001
2	00......0010
......	......
$2^N - 1$	11......1111

Table I
State Space of the Error Propagation Process with N frames.

Given the above definitions, we formulate our problem as follows: given 1) different GOP structures; 2) M encoded frames with variable frame size and number of fragmented IP packets; and 3) random packet loss over the N fragmented packets with probability p, we want to find out: The *probability density function* (pdf) that there are exactly y frames affected by packet loss and by frame dependency respectively.

IV. Modeling error propagation in SVC

A. A Markov Chain with 2^N states

In this section, we develop an exact analysis of the error propagation process in a GOP. The notion N is confined as the total number of frames in a GOP. The analysis is extended to predict the number of erroneous frames in a video sequence with multiple GOPs in Section IV-C2. At each discrete propagation step r, an arbitrary frame k can enter two states: 1) affected by errors, denoted by $F_k[r] = 1$; and 2) not affected by errors, denoted by $F_k[r] = 0$.

The state $Y[r]$ of the GOP at step r is defined by all possible combinations of the states, in which the N frames can be at step r

$$Y[r] = [Y_0[r], Y_1[r], ..., Y_{2^N-1}[r]]^T \tag{1}$$

where $Y_i[r] = 1$ if $i = \sum_{k=0}^{N-1} F_k[r] \cdot 2^k$, and $Y_i[r] = 0$ otherwise. The total number of states is $\sum_{k=0}^{N} \binom{N}{k} = 2^N$, and the state space of the error propagation process is organized with $f_k \in \{0, 1\}$ as shown in Table I.

The error propagation process can be described exactly as a discrete Markov Chain (MC) since the current erroneous frames $Y_j[r]$ at step r only depends on those from the previous step $Y_i[r-1]$. The number of states with i erroneous frames is $\binom{N}{i}$ out of the 2^N ones.

Let P be an $(2^N) \times (2^N)$ transition probability matrix. Each entry in P, $P_{ij} = \Pr\,[Y_{r+1} = j | Y_r = i]$, denotes the probability that the MC moves from state i to state j in one step. The *probability state vector* $s[r]$ in step r is denoted by $s[r] = [s_0[r], s_1[r], ..., s_{2^N-1}[r]]$, with $\sum_{i=0}^{2^N-1} s_i[r] = 1$ and $s_i[r] = \Pr\,[Y_r = i] = \Pr\,[F_0[r] = f_0, F_1[r] = f_1, ..., F_{N-1}[r] = f_{N-1}]$.

The probability state vector can be calculated in terms of the initial state vector $s[0]$ and the transition probability matrix P from

$$s[r] = s[0] \cdot P^r \tag{2}$$

In the above proposed discrete MC, the maximal number of steps until entering an *absorbing state* (defined by $P_{ii} = 1$) is bounded by $r_{max} = 2^N - 1$. In other words, we take $2^N - 1$ as the upper bound of r for numerical calculations. Consequently, the steady-state vector is given by

$$\pi = s[r_{max}] = s[0] \cdot P^{r_{max}} \tag{3}$$

with $r_{max} = 2^N - 1$.

In order to solve (3), the initial state vector $s[0]$ and the transition probability matrix P need to be determined. In Section IV-C1, we present our approach of determining $s[0]$. Notice that both $s[0]$ and the state space description of the MC are independent of the GOP structure. The transition probability P_{ij}, however, is highly dependent on the inter-prediction structure being used. In Section IV-B, we discuss the calculation of the transition probabilities of the MC in different GOP structures.

B. The transition probabilities P_{ij}

In this section, we first explore the principles of frame ID assignment and inter-frame prediction. Afterwards, a so-called *dependency matrix* is developed to describe the dependency between frames. Finally, the transition probabilities P_{ij} are computed based on the dependency matrix. Two GOP structures are studied in the sequel.

1) The hierarchical B-frame structure: Following the conventions in [3], inter-prediction in the hierarchical B-frame structure is jointly initiated from I-frames in the current and preceding GOP, see Fig. 1. Hence, in our model, the number of frames in a GOP is defined as $N = 2^m + 1$, including the I-frame preceding the current GOP. Frame dependency is indicated by the inter-prediction arrows in Fig. 1. For example, frame 2 is predicted from 1 if there exists a outgoing arrow from 1 to 2.

Denote $v^d[t]$ a *frame ID vector* in each temporal layer t of GOP d. An entry in $v^d[t]$, $v_i^d[t]$, refers to the i-th frame ID in layer t. As shown in Fig. 1, frames in the base layer are always assigned by $v^d[0] = [d \cdot 2^m, (d+1) \cdot 2^m]$, with $0 \leq d \leq n_{gop} - 1$. The assignment of frame IDs in layer t ($t \geq 1$) of GOP d obeys a general rule, that is, a frame ID in layer t is iteratively computed by adding or subtracting 2^{m-t} from the frame it depends on in layer $t-1$. By using such a rule, frames are ranked according to the displaying order. For instance, each frame in layer $t-1$, $v_i^d[t-1]$, determines two frame IDs in layer t on its left and right side by

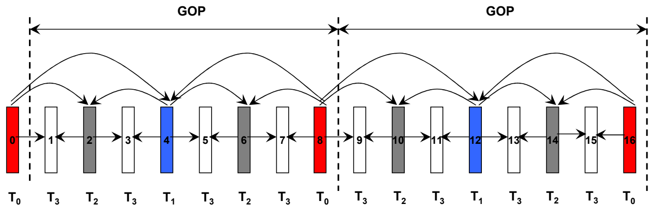


Figure 1. Hierarchical B-frame prediction structure of SVC temporal scalability. Frames are presented in display order, while the number below each frame indicates its corresponding temporal layer. Frames in the base layer T_0 are coded as I-frames. Those in the enhancement layers $T_1, T_2, \ldots$ are coded as B-frames.

$$\begin{cases} v_j^d[t] = v_i^d[t-1] - 2^{m-t} & \text{left side} \\ v_{j+1}^d[t] = v_i^d[t-1] + 2^{m-t} & \text{right side} \end{cases} \tag{4}$$

Notice that both I-frames determine a single frame in layer $t = 1$ of the current GOP by $d \cdot 2^m + 2^{m-1}$ or $d \cdot 2^m - 2^{m-1}$. The number of frames in layer t, $w[t]$, is given by

$$w[t] = \begin{cases} 2 & \text{if } t = 0 \\ 1 & \text{if } t = 1 \\ 2^{t-1} & \text{if } 2 \leq t \leq m \end{cases} \tag{5}$$

Next, we investigate the frame prediction rule of the hierarchical B-frames as a function of frame IDs. Let k_{t_1} be an arbitrary frame in layer t_1 ($k_{t_1} \in v^d[t_1]$), and k_{t_2} an arbitrary frame in layer t_2, ($k_{t_2} \in v^d[t_2]$). Four observations are revealed from Fig. 1: 1) k_{t_2} is predicted from k_{t_1} if $t_2 > t_1$. 2) Bi-directional prediction applies to frames in $t \geq 1$. Frame k_{t_1} predicts one single frame in each succeeding layer along one direction. 3) Inter prediction of the two I-frames ($t = 0$) are bounded on their right- and left- side respectively. 4) Frames in layer $t_2 = m$ are not used to predict other frames. To summarize, a frame, k_{t_2}, is predicted from k_{t_1} if and only if the following relationships are satisfied:

$$k_{t_2|t_1} = k_{t_1} + \frac{2^m}{2^{t_2}} \quad \text{or} \quad k_{t_2|t_1} = k_{t_1} - \frac{2^m}{2^{t_2}} \tag{6}$$

where $0 \leq t_1 < t_2 \leq m$, and $t_2|t_1$ denotes the dependency of frame k_{t_2} on k_{t_1}. The number of frames, n_{dep}, that are predicted from an arbitrary frame in layer t is

$$n_{dep} = \begin{cases} m - t & \text{if } t = 0 \\ 2 \cdot (m - t) & \text{if } 0 < t < m \\ 0 & \text{if } t = m \end{cases} \tag{7}$$

To better illustrate the dependency between frames in (6), we develop a so-called $N \times N$ *dependency matrix* M. Each entry e_{xy} in M defines the dependency of frame y on frame x ($0 \leq x \leq 2^m + 1$ and $0 \leq y \leq 2^m + 1$). If the dependency of frame y on x is true, we have $e_{xy} = 1$. Otherwise, it is $e_{xy} = 0$. Combining (6) and (7), the entries in M are organized by

$$e_{xy} = \begin{cases} 1 & \text{if } x = y \\ 1 & \text{if } y = x - \frac{2^m}{2^{t_2}} \\ 1 & \text{if } y = x + \frac{2^m}{2^{t_2}} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

with $0 \leq t_1 < t_2 \leq m$, where t_1 is the temporal sub-layer that frame x belongs to, and t_2 is the temporal sub-layer that frame y belongs to. The first condition $x = y$ assures that a frame stays erroneous once it is contaminated. The second

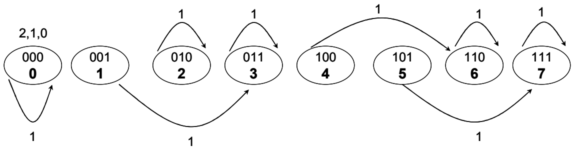


Figure 2. State diagram of the hierarchical B-frame structure with $m = 1$, $N = 3$ and $2^N = 8$ states. The number above state 0 indicates the frame ID with the right-to-left order. The arrows refer to the transition between states.

and third conditions includes all frames that are dependent on frame x.

The transition probability P_{ij} is derived together with the dependency matrix as

$$P_{ij} = \begin{cases} 1 & \text{if } \forall F_y \in Y_j\text{: } F_y = \bigvee_{x=0}^{N-1} F_x \cdot e_{xy} \text{ with } \forall F_x \in Y_i \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where Y_i and Y_j are the i-th and j-th states in Table I, and $F_y = \bigvee_{x=0}^{N-1} F_x \cdot e_{xy} = F_0 \cdot e_{0y} \vee F_1 \cdot e_{1y} \vee ... \vee F_{N-1} \cdot e_{N-1;y}$. The operation $A \vee B$ is 1 if A or B, or both are 1. If both are 0, $A \vee B$ is zero. Fig. 2 shows an example of the Markov state diagram with $N = 3$ frames and $2^N = 8$ states. As we can see from Fig 2, it is only possible to transit from state i to state j if $j \geq i$.

2) The IPPP prediction mode: The first frame in a GOP of the IPPP prediction mode is always an I-frame. All succeeding frames are encoded as P-frames, see Fig. 3. Note that there is no such concept of "temporal layer" in AVC, N is simply the number of frames in a GOP (including one I-frames and all succeeding P-frames).

The $N \times N$ dependency matrix of the IPPP mode is

$$e_{xy} = \begin{cases} 1 & \text{if } y = x \\ 1 & \text{if } y = x + 1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

as a frame x only affect one single frame on its right side. Substitute (10) to (9), P_{ij} of the IPPP mode is computed. State diagram of the IPPP mode is not presented due to space limitations.

C. Performance evaluation

As shown in (3), the performance of error propagation is determined together by the initial state vector $s[0]$ and the transition probability matrix P. In this section, we discuss our approach to obtain $s[0]$ and the methodology to predict the number of erroneous frames affected by transmission errors both inside a GOP and within a video sequence consisting of multiple GOPs.

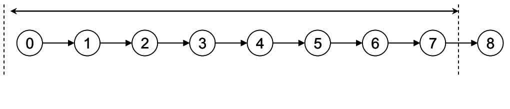


Figure 3. The structure of IPPP mode with $N = 8$ frames.

1) Estimating number of erroneous frames in a GOP: Given random packet loss with probability p over M packets, the number of frames affected by packet loss, denoted by Y_{pkt}, is a r.v. depending on n_k and p. Let Y_{dep} be the number of erroneous frames contaminated by frame dependency after error propagation. The sum of Y_{pkt} and Y_{dep} equals $Y[r_{max}]$ at step r_{max}.

The initial state vector $s[0]$ defines the probability that each state in Table IV-A may occur, where frame k is affected by packet loss with probability $1 - q^{n_k}$, and not affected with probability q^{n_k}. Given n_k packets in frame k and $q = 1 - p$, $s[0]$ is computed by

$$s_i[0] = \prod_{k=0}^{N-1} \left\{ 1_{f_{k;i}=1} \cdot (1 - q^{n_k}) + 1_{f_{k;i}=0} \cdot q^{n_k} \right\} \tag{11}$$

where $f_{k;i}$ denotes the status of frame k in state Y_i. The indicator function 1_y is defined as 1 if y is true, otherwise 1_y is zero [2, pp. 30].

Consequently, the pdf of Y_{pkt} is derived as

$$\Pr[Y_{pkt} = y] = \sum_{i=0}^{2^N-1} 1_{y_i=y} \cdot s_i[0] \tag{12}$$

where $y_i = \sum_{k=0}^{N-1} f_{k;i}$ is the number of frames with status 1 in state Y_i. The pdf of Y_{tot} is

$$\Pr[Y_{tot} = y] = \sum_{i=0}^{2^N-1} 1_{y_i=y} \cdot s_i[r_{max}] \tag{13}$$

with $y_i = \sum_{k=0}^{N-1} f_{k;i}$ and $s_i[r_{max}]$ is computed from (3). Let $i \to s$ be a *realization* (i.e., a sample path) of the error propagation process from an initial state i to the absorbing state s. The number of erroneous frames caused by inter-frame dependency for a single sample path from state i to state s is

$$y_{dep;i \to s} = y_s - y_i = \sum_{k=0}^{N-1} f_{k;s} - \sum_{k=0}^{N-1} f_{k;i} \tag{14}$$

Pdf of the number of frames influenced by frame dependency is therefore, determined by

$$\Pr[Y_{dep} = y] = \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} 1_{y_j - y_i = y} \cdot 1_{P_{ij}^{r_{max}}=1} \cdot s_i[0] \tag{15}$$

where $y_i = \sum_{k=0}^{N-1} f_{k;i}$, $y_j = \sum_{k=0}^{N-1} f_{k;j}$. Elements in the n-step transition probability matrix, $P_{ij}^{r_{max}} = \Pr[Y_{r_{max}} = j | Y_0 = i]$, defines the probabilities to move from initial state i to the steady-state j.

The mean and variance of Y_{tot} are $E[Y_{tot}] = \sum_{i=0}^{N} i \cdot \Pr[Y_{tot} = i]$ and $\mathrm{Var}[Y_{tot}] =$

$\sum_{i=0}^{N} (i - E[Y_{tot}])^2 \cdot \Pr[Y_{tot} = i]$ respectively. The mean and variance of Y_{pkt} and Y_{dep} can be obtained in an analogous way.

2) Estimating number of erroneous frames in a video sequence: Performing an exact analysis about the erroneous frame estimation problem in a video sequence with multiple GOPs is very difficult. At step r, an arbitrary GOP d may enter N state: $G_d[r] = i$ with $0 \leq i \leq N$ and N is the total number of frames in a GOP.

The major challenge is to describe the entire system exactly. That is, to find all possible combinations the states that the n_{gop} GOPs can be at step r. As discussed in Section IV-A, we use 2^N states to describe N frames with two possible states. Given N states instead of 2 for each GOP, the exact analysis becomes more complex and requires a huge state space. Hence, we resort to a simple approximation to compute the number of erroneous frames in a video sequence.

Let Y_{pkt}^*, Y_{dep}^* and Y_{tot}^* be the number of frame affected by packet loss, frame dependency and the sum of the above two r.v. in the entire video sequence. Instead of seeking for the pdf of Y_{pkt}^*, Y_{dep}^* and Y_{tot}^*, we derive their expectations as

$$\mathrm{E}[Y_{tot}^*] = \sum_{d=0}^{n_{gop-1}} \mathrm{E}[Y_{tot;d}] \tag{16}$$

where $\mathrm{E}[Y_{tot;d}]$ is calculated in Section IV-C1 for a single GOP d. $\mathrm{E}[Y_{pkt}^*]$ and $\mathrm{E}[Y_{dep}^*]$ are computed analogously.

V. Results and Discussions

In this section, a high definition (HD) video sequence *old_town_cross* with the resolution of 1980x1080 and frame rate of 50fps is encoded with the SVC reference encoding software JSVM (Joint Scalable Video Model) [4] to generate the hierarchical B-frame and IPPP mode encoded frames (compatible with AVC). The version of JSVM under use was 9.18. We assign m=3 in the B-frame structure, which results in 9 frames in each GOP. To have fair comparison, the GOP size of the P-frame mode is set to 9 as well. Encoding parameters such as the quantization parameter are the same for both schemes.

We also developed a simulation program by using the C language to simulate IP fragmentation of the encoded bit-stream. The program also simulates packet loss over the fragmented packets. Since JSVM cannot decode frames containing errors, an extra script was developed to filter out damaged (hence undecodable) frames before decoding. The JSVM software, therefore, only decodes those frames that are error-free. Results obtained from the analysis are compared with the results derived from the simulated program. For each simulated result, 10^4 iterations are carried out.

A. The hierarchical B-frame structure

First of all, we compare the analytic $\Pr[Y_{pkt} > y]$ and $\Pr[Y_{dep} > y]$ with simulated results in Fig. 4(a) and (b). As we can see from Fig. 4, both curves match the simulated results very well. The tail probability $\Pr[Y_{pkt} > y]$ (or $\Pr[Y_{dep} > y]$) defines the probability that more than y frames are affected by packet loss (or frame dependency). Given the packet loss rate of $p = 0.1$, the probability that more than 3 frames are impacted by packet loss is approximately 0.32, shown in Fig. 4(a). The probability that more than 6 frames being affected by packet loss decreases to the order of 10^{-3} and 10^{-6} for $y = 8$ frames. Frame dependency seems to play an influential role, as illustrated in 4(b). In 90% of the cases, more than 4 frames are affected by frame dependency. In 28% of the cases, more than 6 frames are contaminated. With the exact analysis, we are able to evaluate the performance of $\Pr[Y_{pkt} > y]$ (or $\Pr[Y_{dep} > y]$) with high accuracy, which is normally very difficult to achieve with simulations. For instance, in order to have an accuracy of 10^{-6} for $y = 8$, as in Fig. 4(a), the simulation needs to be performed 10^{12} times, which is time consuming. In the following, only analytic results will be discussed except for Fig. 5(b) where the approximation in (16) is verified with simulated results for a video sequence with multiple GOPs.

In Fig. 5(a), the analytic results of $\mathrm{E}[Y_{pkt}]$, $\mathrm{E}[Y_{dep}]$ and $\mathrm{E}[Y_{tot}]$ are presented respectively. When packet loss rate is smaller than 0.1, frame dependency is a dominant factor in propagating errors. However, with $p > 0.2$, $\mathrm{E}[Y_{dep}]$ started to decrease and $\mathrm{E}[Y_{pkt}]$ begins to grow more drastically. This is because, given a fixed amount of frames in a GOP, more frames being affected by packet lost naturally leads to less frames being influenced by dependency. The inset of Fig. 5(a) plots $\mathrm{E}[Y_{tot}]$ together with its associated upper and lower bounds. The upper and lower bounds are computed by $\mathrm{E}[Y_{tot}] \pm \sigma$ respectively, where $\sigma = \sqrt{\mathrm{Var}[Y_{tot}]}$ is the standard deviation. The two bounds for $\mathrm{E}[Y_{pkt}]$ and $\mathrm{E}[Y_{dep}]$ are not presented due to space limit. Notice that the upper and lower bounds indicate the worst and optimal performance. For instance, we see that with $p = 10^{-2}$, the maximal number of erroneous frames reaches 9 and the minimal number of contaminated frames is around 6. From

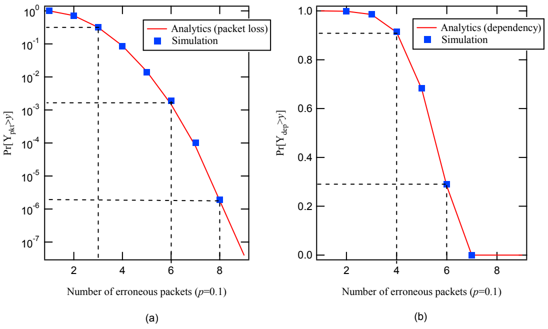
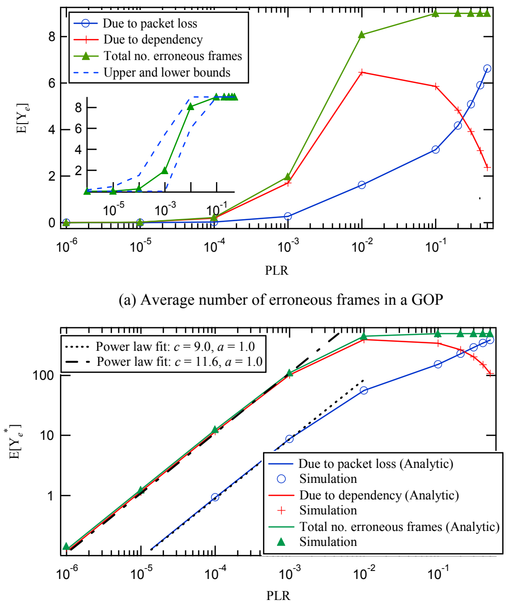


Figure 4. The tail behavior of $\Pr[Y_{pkt} > y]$ (a) and $\Pr[Y_{dep} > y]$ (b) versus number of erroneous frames y with $p = 0.1$.



(a) Average number of erroneous frames in a GOP

(b) Average number of erroneous frames in a video sequence

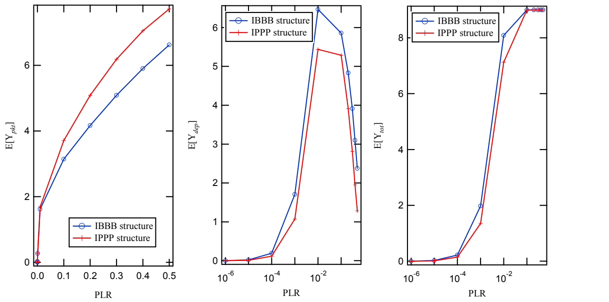
Figure 5. (a) Average number of erroneous frames in a single GOP as a function of p (lin-log scale). The inset plots the upper and lower bounds of the total number of erroneous frames. (b) Average number of erroneous frames in a video sequence with 62 GOPs as a function of p (log-log scale). Both figures present the analytic results. The approximation of (16) is verified with simulations in (b).

a statistic point of view, all results of individual packet loss events and the resulting erroneous frames are bounded by the two dotted curves.

In Fig. 5(b), we plot $\mathrm{E}\left[Y_{pkt}^{*}\right]$, $\mathrm{E}\left[Y_{dep}^{*}\right]$ and $\mathrm{E}\left[Y_{tot}^{*}\right]$ for a video sequence with 62 GOPs (approximately 9.3 seconds). As revealed from Fig. 5(b), equation (16) approximates simulated results very well. On a log-log scale, $\mathrm{E}\left[Y_{pkt}^{*}\right]$, $\mathrm{E}\left[Y_{dep}^{*}\right]$ and $\mathrm{E}\left[Y_{tot}^{*}\right]$ exhibit straight lines until the point of $p = 10^{-3}$, conforming to the power law distribution (defined as $y = c \cdot x^{a}$, where c is a normalization constant). The fitting parameter of a defines the slope of a curve. The fitting curves in Fig. 5(b) allow us to predict the average number of erroneous frames based on a and c without employing (16). However, the power law distribution fails to approximate the curves if $p > 10^{-3}$. We see clearly that, less than 20% of the frames ($\mathrm{E}\left[Y_{tot}^{*}\right]$) are affected by transmission errors up to $p = 10^{-3}$. With $p = 10^{-2}$, around 80% of the frames are affected. If $p \geq 10^{-1}$, almost all frames are contaminated.

B. Comparing the B-frame structure with the IPPP mode

In this section, the performance of the hierarchical B-frame structure is compared with the IPPP mode. As shown in Fig. 6, the IPPP mode is more sensitive to packet loss than the B-frame structure. This is because the IPPP mode has lower compression efficiency compared to the B-frame structure. A P-frame generally consists of more packets than a B-frame. According to (11), larger n_k incurs higher frame error probability. Therefore, the IPPP mode is more vulnerable to packet loss, especially with larger p.



(a) Due to packet loss

(b) Due to frame dependency

(b) Total no. erroneous frame

Figure 6. Average number of erroneous frames caused by packet loss (a), frame dependency (b) and both (c) of the IBBB and IPPP structure. Figure (b) and (c) are plotted on lin-log scale for easier reading.

The frame dependency of the B-structure is, by nature, higher than the IPPP mode, as plotted in Fig. 6(b). The maximal absolute difference between the two curves occurs at $p = 10^{-2}$, where the inter-frame prediction in the B-frame mode leads to one more erroneous frame than in the P-frame mode. The total number of erroneous frames in Fig. 6(c) is comparable between the IBBB and the IPPP structure. The average absolute difference between the B-frame and the P-frame mode is around 0.5, 0.7 and 0.2 frames in Fig. 6(a), (b) and (c) respectively. Considering the higher coding efficiency of the IBBB structure, and the marginal difference between the B- and P- frame mode against transmission errors, the SVC IBBB hierarchical coding structure appears to be a good candidate for video transmission in an error-prone environment with random packet loss.

VI. Conclusion

This paper presents an exact analysis to examine the impact of packet loss on the H.264 scalable video coding. With a simple approximation, the model is extended to predict the performance of a video sequence consisting of multiple GOPs. Major conclusions from the performance analysis are: 1) Frame dependency in SVC B-frame structure is a dominant factor in propagating transmission errors with packet loss rate $p < 0.1$. 2) The upper and lower bounds obtained from the analysis suggests the worst and optimal performance of individual loss events. In order to satisfy users with the worst performance, it is important to look at the upper bounds and thereafter enhance the robustness of the bit-stream to be delivered under packet loss. 3) When $p \leq 10^{-3}$, the average number of frames affected by transmission errors can be approximated by the power law distribution. To avoid drastic increment in the number of erroneous frame, the packet loss rate should be controlled as $p \leq 10^{-3}$ (depending on system requirements). 4) Despite the hierarchical frame dependency, the overall performance of the B-frame structure is, in fact, comparable with the IPPP mode under random packet loss.

Performance evaluation presented in this paper is based on the number of frames affected by transmission errors. Examining other metrics, such as the PSNR, that can properly reflect the pixel-level quality degradation is the focus of our future work. The error propagation process investigated in this paper can be directly employed to describe the inter-frame motion compensation in the future model. Besides, adapting the initial state vector $s[0]$, the burst packet loss process can be incorporated.

VII. Acknowledgment

The authors would like to thank W. Acke for proofreading this paper.

References

[1] ITU-T and ISO/IEC JTC1, JVT-W201, “JointDraft 10 of SVC Amendment”, Apr. 2007.

[2] P. Van Mieghem, “Performance Analysis of Communications Networks and Systems”, Cambridge University Press, 2006.

[3] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H. 264/AVC standard”, IEEE TCSVT, 2007. 17(9): pp. 1103–1120.

[4] Team, J.V., H. 264/SVC reference software (JSVM 9.18) and manual. CVS sever at garcon. ient. rwth-aachen. de, 2009.

[5] S. Tang, E. Jaho, I. Stavrakakis, I. Koukoutsidis, and P. Van Mieghem, “Modeling gossip-based content dissemination and search in distributed networking”, Computer Communications, 2011. 34(6): pp. 765–779.

[6] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels”. IEEE Journal on Sel. Areas in Communications, 2000. 18(6): pp. 1012–1032.

[7] Y.J. Liang, J.G. Apostolopoulos, and B. Girod, “Analysis of packet loss for compressed video: Does burst-length matter?”, Proc. IEEE ICASSP, Apr. 2003. vol. 5: pp. 684-687.

[8] S. Tao, J. Apostolopoulos, and R. Guérin, “Real-time monitoring of video quality in IP networks”, IEEE/ACM Trans. on Networking (TON), 2008. 16(5): pp. 1052–1065.

[9] Z. Chen and D. Wu, “Prediction of Transmission Distortion for Wireless Video Communication: Analysis”, IEEE Trans. on Image Processing, 2012. 21(3): pp. 1123–1137.

[10] J.M. Monteiro, C.T. Calafate, and M.S. Nunes, “Evaluation of the H. 264 scalable video coding in error prone IP networks”, IEEE Trans. on Broadcasting, 2008. 54(3): pp. 652–659.

[11] M. Ghareeb, A. Ksentini, and C. Viho, “Scalable Video Coding (SVC) for multipath video streaming over Video Distribution Networks (VDN)”, Int. Conf. on Information Networking (ICOIN), 2011: pp. 206–211.

[12] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, “Modeling of loss-distortion in hierarchical prediction codecs”, IEEE Int. Symp. on Signal Processing and Information Technology, 2006: pp. 536–540.

[13] Y. Wang, Z. Ma, and Y.F. Ou, “Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation”, IEEE Int. Packet Video Workshop, 2009: pp. 1–9.

[14] M.R. Ardestani, A. Shirazi, and M.R. Hashemi, “Rate-distortion modeling for scalable video coding”, IEEE 17th Int. Conf. on Telecommunications (ICT), 2010: pp. 923–928.