MARKOV MODEL OF H.264 VIDEO SOURCES PERFORMING BIT-RATE SWITCHING

Stefania Colonnese, Gianpiero Panci, Stefano Rinauro and Gaetano Scarano

Dip. INFOCOM, “Sapienza” Universita´ di Roma
via Eudossiana 18 00184 Roma, Italy
(colonnese, gpanci, rinauro, scarano)@infocom.uniroma1.it

ABSTRACT

Fast and bit-saving video bit-rate switching is an important issue in video streaming systems on a time varying channel as the one offered by a wireless mesh network, or the one sensed during a vertical handover. The recent H.264 video coding standard supports the seamless switching among bitstreams coded at different bit-rates by means of suitably coded frames, named Switching Pictures. This work addresses the modelling of the traffic generated by a H.264 source performing bit-rate switching using SP frames. The H.264 source is modelled by a Markov chain where each state models the generation of an entire Group Of Pictures (GOP), and is characterized by the kind of SP frame encoded in the GOP. Interframe correlation, typical of video sources, is suitably taken into account by the interstate dependence. The accuracy of the model is assessed by comparison of the cell loss rate of a fixed size buffer filled with a synthetic source according to the model herein proposed, with a state of the art AR model and with a real H.264 video codec.

Index Terms— Source modelling, bit-rate switching, H.264.

1. INTRODUCTION

Multimedia traffic is typically expensive from a bandwidth allocation point of view and it suffers vulnerability to errors, delay and jitter. Hence, a good modelling of a video source is important in network design to predict the impact of the video traffic on the network performances and to optimize parameters such as allocated bandwidth and buffer sizes in order to provide a certain degree of QoS (Quality of Service) in terms of packet loss, delay and jitter. The most recent video coding standard, namely ITU-T Rec. H.264 or ISO/IEC MPEG-4/Part 10-AVC is a hybrid video codec making use on a number of novel coding functionalities and syntactic structures [1]. Thanks to the innovative compression tools, H.264 is expected to be the best solution for video transmission in a number of services, especially over wireless networks, including the emerging technology of the wireless mesh networks. A relevant and innovative feature of H.264 is the availability of two new coding schemes, and of the corresponding syntactic structures named Switching Pictures (SP) that can be perfectly reconstructed even when different reference frames are used for their prediction [7]. Each switching frame has a primary and secondary representation. The primary representation is sent along each bitstream and it provides the virtual access point to the bitstream for users incoming by other bitstream. The secondary representation is sent only at the switching phase, and it allows decoding exactly the same frame as the primary representation, while using different reference frames. In other words the primary and secondary coded frames are two alternative representations of the same frame, differing only in the prediction step. In a realistic streaming session, primary SP frames are periodically inserted in order to provide random access points to video sequence. The employment of SP frames is particularly useful in streaming applications, since it provides virtual access points to each bitstream at a lower cost than I frames. In literature a lot of work has been done at the aim of modelling a video source. In particular, in [2] an overview of the most used models employed in video source modelling is presented, investigating Markovian, TES (Transform Expand Sample), and self-similar models. In [3] an MPEG1 video source in synthesized by correlating 3 different stochastic processes in discrete time (AR - Auto Regressive - models), one for each frame (I, P and B) employed in the standard, whereas in [4] a representation of an MPEG4 and H.264 VBR (Variable Bit Rate) source is developed via a wavelet and time domine combined method both for inter and intra-GOP (Group of Picture) correlation. In [5] and [6] the authors investigate a Markovian representation of an H.264 stream based on a Gamma like marginal frame size distribution to fit the I, P, and B distribution. In the recent literature the dynamical behavior of the H.264 source during a bitstream switching hasn’t still been analyzed deeply. In this work, we address the modeling of a VBR H.264 source transmitting allowing bitstream switching using SP frames. We resort to a Markovian Model (MM) to represent the dynamical behavior of a video source switching between different bit-rates. In literature Markov chains are often employed to model the interframe correlation. Here we resort to a finite state machine to exploit the Group of Picture (GOP) structure of the sequence. More specifically each state models the generation

978-1-4244-1764-3/08/$25.00 ©2008 IEEE

of an entire Group Of Pictures (GOP), and is characterized by the kind of SP frame encoded in the GOP. Interframe correlation, typical of video sources, is suitably taken into account by the interstate dependence. The model is validated by comparing the buffer loss rate for the MM traffic, the real source and a state of the art AR model [3]. The remainder of this paper is organized as follows: in Section 2, we will describe the analyzed Markovian model; in Section 3 we will introduce the model validation based on a network point of view while in Section 4 we will show the main simulation results; Section 5 concludes the paper.

2. MARKOV MODEL OF H.264 VIDEO SOURCE

Let us now model a VBR H.264 coded video sequence characterized by periodic SP primary frame insertion at a nominal average bit rate. The GOP structure at the nominal bit-rate R_1 can be represented as:

$$\underbrace{\Big(SP_{prim}(R_1),\ P(R_1),\ P(R_1),\ \cdots,\ P(R_1)\Big)}_{N_{gop}} \quad (1)$$

where the primary SP frame at rate R_1 $SP_{prim}(R_1)$ is followed by $N_{gop}-1$ P frames at the same rate[1]. N_{gop} represents the length in frames of the GOP. When a bit-rate switching between the rate R_1 and a rate R_2 is considered, a secondary SP coded frame $SP_{sec}(R_1, R_2)$ is sent instead of the primary SP frame. Then the following frames are extracted from the coded bitstream at the nominal bit-rate R_2. Hence the structure of the GOP that realizes the switching can be expressed as:

$$\underbrace{\Big(SP_{sec}(R_1, R_2),\ P(R_2),\ P(R_2),\ \cdots,\ P(R_2)\Big)}_{N_{gop}} \quad (2)$$

Generally speaking, a H.264 video source performing bit-rate switching among the rates $\{R_i, \quad i = 1, \cdots, L\}$ will present the following $L + L \cdot (L-1)$ alternative GOP structures:

$$\begin{array}{l} \underbrace{\Big(SP_{prim}(R_i), P(R_i), P(R_i), \cdots, P(R_i)\Big)}_{N_{gop}} \ i = 1 \cdots L \\ \underbrace{\Big(SP_{sec}(R_i, R_j), P(R_j), P(R_j), \cdots, P(R_j)\Big)}_{N_{gop}} \ i, j = 1 \cdots L \end{array} \quad (3)$$

Here we consider a Markovian source model build up by $N_s = L + L \cdot (L-1)$ states; each state is associated to different GOP structures. The Markov chain modelling the coded bitstream structure with reference to the case $L = 2$ is shown in Fig. 1. After a random number of consecutive GOPs at the rate R_1 (R_2) a bit-rate switching is performed with probability $\pi_{1,2}$ ($\pi_{2,1}$).

Let us observe that in literature a Markov model (MM) is introduced [3] to model interframe correlation, while here we are mostly interested in representing the inter-GOP correlation. The state machine is expected to capture well the dynamical behaviour of the source performing bitstream switching. A Markov chain is completely described by a transition matrix Π, whose element π_{ij} equals the probability of transition from the i-th state to the j-th one.

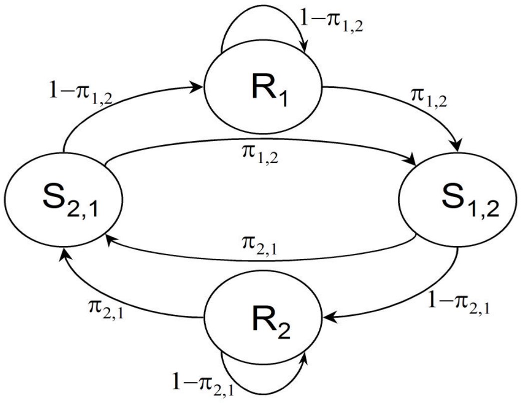

Fig. 1. *H.264 video source Markov Model for L=2.*

In the simple case of $L = 2$, represented in Fig. 1, the transition matrix Π will exhibit the following form:

$$\begin{array}{c|cccc} & R_1 & R_2 & S_{1,2} & S_{2,1} \\ \hline R_1 & 1-\pi_{1,2} & 0 & \pi_{1,2} & 0 \\ R_2 & 0 & 1-\pi_{2,1} & 0 & \pi_{2,1} \\ S_{1,2} & 0 & 1-\pi_{2,1} & 0 & \pi_{2,1} \\ S_{2,1} & 1-\pi_{1,2} & 0 & \pi_{1,2} & 0 \end{array}$$

where each row and each column has been labeled with the related state. More in general, the matrix is built up by N_s^2 elements; let us observe that the rows pertaining to the states $S_{j,i}, j = 1, \cdots, L$ are equal to the rows pertaining to the state R_i. Let us now consider a N_{gop}-dimensional random variable representing the sizes of the frame of the GOP associated to the λ-th state, $\lambda = 1, \cdots, N_s$:

$$\mathbf{X}[n] \overset{\text{def}}{=} [X_0[n], \cdots, X_M[n]]^{\mathrm{T}}$$

being $X_i[n]$ the size of the i-th frame of the n-th GOP of the coded video sequence. The more general model accounts for a correlation between the size of a frame and the sizes of the ν preceding frames. Such an interframe correlation has been exploited in literature via the employment of AR

[1] Without any loss of generality the P frames can be substituded by any other kind of non switching frame, that is, B-frames or I-frames.

models and was proved to be useful in modelling VBR video sources. Here we will consider interframe correlation also in the modelling a source performing bitstream switching. Under the hypothesis that the size of a particular frame depends on a number $\nu < N_{gop}$ of preceding frames, the state machine representing the video sequence introduced above satisfies the first order Markovian property:

$$P(\mathbf{X}[n]|\mathbf{X}[n-1], \mathbf{X}[n-2] \cdots) = P(\mathbf{X}[n]|\mathbf{X}[n-1]) \quad (4)$$

Let us assume the following model for the variate $\mathbf{X}$:

$$\mathbf{X}[n] = \mathbf{A}_\lambda \mathbf{X}[n-1] + \mathbf{B}_\lambda \mathbf{E}[n] + \mathbf{C}_\lambda \qquad (5)$$

where $\mathbf{E}[n]$ is a zero-mean suitably distributed random vector with statistically independent components and the matrixes $\mathbf{A}_\lambda, \mathbf{B}_\lambda$ and the vector $\mathbf{C}_\lambda$ are constants typical of the λ-th state. Only the last ν components of the matrix $\mathbf{A}_\lambda$ are non-zero, *i. e.*:

$$\mathbf{A}_\lambda = \begin{bmatrix} 0 & \cdots & a_{1,N_{gop}-\nu+1} & \cdots & a_{1,N_{gop}} \\ & \vdots & \vdots & \vdots & \vdots \\ & \cdots & \vdots & \vdots & \vdots \\ 0 & \cdots & a_{N_{gop},N_{gop}-\nu+1} & \cdots & a_{N_{gop},N_{gop}} \end{bmatrix}$$

The matrix $\mathbf{B}_\lambda$ takes into account the dependence of each variate $X_i[n]$ with the preceding variates $X_j[n], j < i$ belonging to the same GOP. The constant non-zero vector $\mathbf{C}_\lambda$ drives the expected value of $\mathbf{X}[n]$. The distribution of $\mathbf{E}[n]$ is chosen so to assure that, in each state λ, $p\left(X_1[n]|X_{N_{gop}}[n-1]\right)$ follows the distribution of SP frames and that $p\left(X_i[n]|X_{i-1}[n]\right) i = 2, \cdots, N_{gop}$ follows the distribution of P frames. The distribution of the number of bits for I and P and B frames is proved to be [6] well approximated by a Gamma distribution. In [8] it is proved that this approximation stands also for SP primary pictures [2]. Furthermore in [8] it is observed that the histogram of the size of SP secondary frames - instead - exhibits short tails and it can be approximated by a Gaussian distribution.

The expected value and the autocorrelation function for the variate $\mathbf{X}[n]$ in (5) are defined as follows:

$$\begin{aligned} \underline{m}_x &\overset{\text{def}}{=} E\left\{\mathbf{X}[n]\right\} \\ \mathbb{R}_x[m] &\overset{\text{def}}{=} E\{\mathbf{X}[n]\mathbf{X}^T[n-m]\} \end{aligned} \qquad (7)$$

and can be proved to be:

$$m_x = \left(I - \sum_{\lambda=1}^{N_s} p_\lambda \mathbf{A}_\lambda\right)^{-1} \left(\sum_{\lambda=1}^{N_s} p_\lambda \mathbf{C}_\lambda\right) \qquad (8)$$

The autocorrelation function is proven to be, for $m >> 0$:

$$\mathbb{R}_x[m] = \left(\sum_{\lambda=1}^{N_s} p_\lambda \mathbf{A}_\lambda\right) \mathbb{R}_x[m-1] + \left(\sum_{\lambda=1}^{N_s} p_\lambda \mathbf{C}_\lambda\right) \underline{m}_x^T \qquad (9)$$

where the $p_\lambda, \lambda = 1, \cdots, N^s$ represent the limit state probability of the Markov chain and can be found as the eigenvector related to the unitary eigenvalue of the transition matrix Π under the hypothesis that the parameters assure the system's stability which in turn is assured if all of the eigenvalues of the matrix $\left(\sum_\lambda p_\lambda \mathbf{A}_\lambda\right)$ are less that one.

3. MODEL VALIDATION

The accuracy of the proposed model will be evaluated from a network analysis point of view; we compare the buffer load generated by a real H.264 video source and the Markovian source model. Such a validation approach is presented and validated in [3]. The H.264 video source is formed by the Video Coding Layer (VCL) and Network Adaptation Layer (NAL). The VCL is assumed to generate data organized in Slices, with one Slice for every coded frame. The Slices are then encapsulated by the NAL into NAL Units (NALU) according to the NALU format for packet-oriented networks. We assume that the NALUs are transmitted using a buffer. A NALU is stored in the buffer only if there is available space for it. Otherwise, it is lost. The buffer is managed with a first-in first-out (FIFO) policy, $i.e.$ the NALUs are transmitted in the same order as they enter the buffer. In order to analyze the model of a H.264 source without bitstream switching, the buffer output rate is assumed to be constant and equal to the average VBR source rate. On the other hand, the model of a H.264 source performing bitstream switching is analyzed by varying of the buffer output rate, according to the initial and final average rates of the VBR source.

4. SIMULATIONS

In this Section we present the simulation results referring to the source model described in the preceding Sections under the mentioned buffer settings. We referred to the simple case of $L = 2$ and $\nu = 1$, thus considering the switchng between only two possible bit-rates and restricting the inter-frame dependence to only one preceeding frame. The real coded sequences are generated using the H.264 encoder JM version 11.0 [9], extended profile. The MPEG-4 class A test sequence Akiyo (low spatial detail and low amount of movement) and the MPEG-4 class B test sequences Foreman and News (medium spatial detail and low amount of movement

[2] The probability density function of the gamma distribution can be expressed in terms of the gamma function $\Gamma(z)$:

$$\begin{aligned} g(x, \alpha, \beta) &= x^{(\alpha-1)} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} u_{-1}(x) \\ \Gamma(z) &= \int_0^\infty t^{z-1} e^{-t} dt \end{aligned} \qquad (6)$$

where α and β are parameters that can be calculated from the mean and variance.

or viceversa) in QCIF format were encoded at different bitrates with ten frame per second and one SP picture every five P frames; to simplify the simulations we didn't account for the presence of B frames. Such parameters reflect typical coding settings for a wireless video communication. Those sequences (100 frames each) have been concatenated so to obtain a longer sequence - that we call the "aggregated" sequence - exhibiting a few scene changes.

We performed two sets of simulations, one to test the behaviour of the model when the source doesn't perform bitstream switching, and one where the switching is taken in to account. The experimental results are obtained comparing the cell loss rate observed filling the buffer model presented in Section 3 with the real encoded data and with the data generated according to the Markovian model herein presented. We considered four target bit-rates (20, 48, 100 and 320 kbps); the size of the buffer's cell is 384 bits (48 bytes being the size of an ATM cell without the overhead). The statistical parameters to drive the Markov model described in Section 2 are obtained averaging the values of the frame's size over the aggregated sequence. For sake of comparison we also present the results obtained by a well known literature model, that is the AR(1) model presented in [3]. Fig. 2 compares the cell loss rate obtained by the real H.264 source, the MM here proposed and the AR(1) model in source when bitstream switching is not performed [3]. It is interesting to note a good correlation between the cell loss rate estimated on the base of the synthetic source and the one evaluated on the H.264 codec. Fig. 3 instead depicts the cell loss rate observed in presence of bitstream switching between different bit-rates. The switching is performed with probability $\rho = \mu = 0.5$. The results in Fig. 3 show that the dynamical structure of the Markov chain in Fig.1 allows to capture favourably the behaviour of a real source performing bit-rate switching outperforming the model in [3].

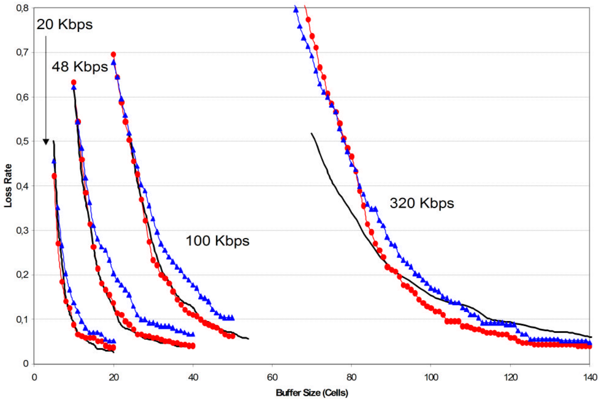

Fig. 2. *Cell loss rate for the MM (blue - triangle), the AR(1) model in [3] (solid line) and real (red - circle) H.264 source without bitstream switching*

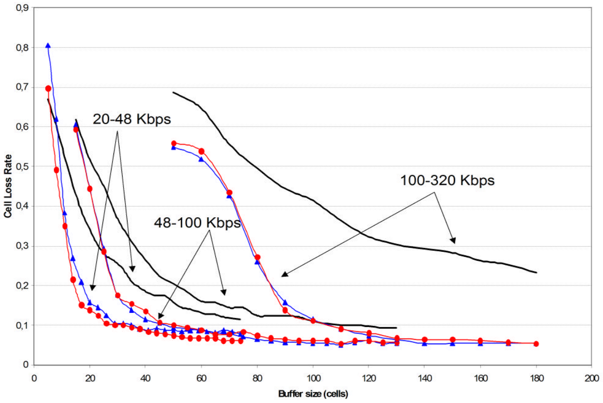

Fig. 3. *Cell loss rate for the MM (blue - triangle), the AR(1) model in [3] (solid line) and real (red - circle) H.264 source when bitstream switching is performed*

5. CONCLUSION

The H.264 video coding standard introduces a compression tool for fast bit-rate switching, based on the syntactic element Switching Pictures (SP). In this work a Markov model of a H.264 source performing bit-rate adaptation using Switching Pictures is analyzed. In the model each state represent the generation of an entire GOP. We assessed the model performance by evaluating the cell loss rate of a fixed size buffer filled with the synthetic source implementing the model. Numerical experiments show that the Markov GOP model provides a good approximation of the source behaviour, at least at the aim of the network resource allocation.

6. REFERENCES

[1] T. Wiegand *et all.*, "Overview of the H.264 Video Coding Standard", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 13, No. 7, pp. 560-576, July 2003.

[2] M.R. Izquierdo, D.S. Reeves, "A survey of statistical source models for variable-bit-rate compressed video", *Multimedia Systems* , Vol.7, No 3, May 1999, pp. 199-213, May 1999.

[3] N.D. Doulamis *et all.*, "Efficient modeling of VBR MPEG-1 coded video sources", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 10, No. 1, pp. 93-112, Febraury 2000.

[4] M. Dai, D. Louguinov, "Analysis and Modeling of MPEG-4 and H.264 Multi-Layer Video Traffic", INFOCOM-2005, Miami, Florida, U.S., March 13-17, 2005.

[5] H. Koumaras *et all.*, "A Markov Modified Model of H.264 VBR Video Traffic", IST Mobile Summit 2006, Mykonos, Greece.

[6] H. Koumaras *et all.*, "Analysis of H.264 Video Encoded Traffic", International Network Conference 2005, Samos, Greece.

[7] M. Karczewicz, R. Kureren, , "The SP and SI frames design for H.264/AVC", *IEEE Trans. on Circ. and Sys. for Video Tech.*, Vol. 13, No. 7, pp. 637-644 July 2003.

[8] S. Colonnese, G. Panci, S. Rinauro, G. Scarano, "Modeling of H.264 Video Sources Performing Bitstream Switching", PCS-2007, Lisbon, Portugal, November 7-9, 2007.

[9] H.264 codec JM11.0 available at http://iphome.hhi.de/suehring/tml/