

APPLICATION CONTROL FOR FAST ADAPTIVE ERROR RESILIENT H.264/AVC STREAMING OVER IP WIRELESS NETWORKS

C. Lamy-Bergot and B. Gadat

Thales Communications France, F-92704 Colombes, RCP/DT and HAT/SNE/SPM.

ABSTRACT

Following the joint source and channel coding paradigm, we propose in this article an application controlling strategy to allow fast adaptation of multimedia transmission to the impairments of an IP wireless network. To this purpose, a semi-analytical prediction algorithm of the video stream perceived quality of service (PQoS) expressed as the end-to-end distortion (EED) is introduced, that takes into account the distortions due to the diverse treatments of compression, transmission and decoding for a group of pictures (GOP) subject to an imperfect communication channel. This algorithm allows to define the parameters providing the best perceived quality at the receiver side in terms both of compression and optionally of error correcting redundancy insertion to minimise the error impact with a GOP adaptation rate, as demonstrated by the results presented for H.264/AVC streams transmitted in various channel configurations.

Index Terms— Optimised video transmission, cross-layer approach, wireless communication, joint source/channel coding, H.264/AVC.

1. INTRODUCTION

The delivery of digital content has been researched to provide the end-user better services and quality almost since its beginnings. Indeed, the IP packet-based transmission mode naturally leads to random losses in streams, which may affect the multimedia end-user perceived quality, while the possibility to rely on feedback channels leads to the problem of delay, despite the possible priority based solutions [1].

Offering the best quality of service to the end-user implies evaluating as accurately as possible the EED over an error prone channel. But while estimating the compression distortion is easy, as all information is available to the encoder, the estimation of EED is more difficult, as it depends on factors such as the transmission channel that are causal and not known at the encoder side. Nevertheless, many solutions have been proposed to estimate it, and define optimisation approaches to improve the perceived quality, from earlier

solutions aiming at adapting a stream at different pre-defined compression rates [2] to more elaborate rate-distortion (RD) prediction models [3], ultimately including error protection [4][5] [6][7][8].

In this article, we present an algorithm using a low complexity semi-analytical EED derivation based on [6] approach and show how it can be used to control from the application level the distortion and throughput of H.264/AVC [9] streams transmission over an imperfect channel.

This paper is organized as follows. Section 2 presents the considered transmission chain design principles. Section 3 details the application controller algorithm, and the derivation of the EED. Thanks to an abacus-driven calibration, including a “test & trials” phase to ensure that the semi-analytical formula is finely attuned to the stream being transmitted at the current moment of adaptation, the derivation complexity is low, and yet accurate. Section 4 presents examples of the experimental results that can be obtained with this controlling solution, and explains the interest and possible usage for real systems. Finally, conclusions are drawn in Section 5.

2. TRANSMISSION CHAIN DESIGN PRINCIPLES

Figure 1 presents the transmission chain considered in the paper, and outlines the key element called application controller which is in charge of driving the application layer operations, *i.e.* the source encoder (or data server) and the optional protection insertion module meant to combat effects of transmission over imperfect channels. Corresponding data and redundancy are then encapsulated for transmission over an IP-based network within RTP packets that are transmitted over an UDP-Lite/IPv6 socket directed towards the receiver side through the communication link model, and followed by the converse operations of IPv6/UDP-Lite/RTP decapsulation at the receiver side, with optionally the error correction decoding stage to combat errors due to the communication link, before the final operation of video decoding. Based on the quality of service provided by the transmission channel and its EED and throughput values prediction algorithm, the application controller sets the parameters to be used by the source encoder and error protection module to obtain the best possible PQoS for the end-user and/or meet throughput targets.

Our aim being to operate transparently over IP wireless

This work has been carried out thanks to French ANR through grant No 2009-VERS-019 (ARSSO project) and to the EC FP7 through grant agreement No INFSo-ICT-214625 (ICT OPTIMIX project).

networks, the controller does not modify nor interact with the lower access layers, except to collect their offered quality of service, represented by the couple (bitrate, error probability) offered above the link layer. While this quality of service may vary with time, it is considered as an input value, and is not negotiated specifically by the application. Nevertheless, the controller could be expanded to also control the lower layers, if desired.

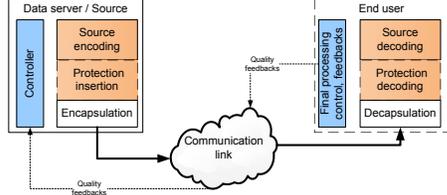


Fig. 1. Transmission chain principle.

3. APPLICATION CONTROLLER OPTIMISATION ALGORITHM

The considered concept of the application controller algorithm was initially introduced in [6] together with a semi-analytical expression of the EED \hat{D}_{gop} obtained for an H.264/AVC group of pictures IP_N of length $N + 1$ after it has been transmitted over a memoryless erroneous channel with bit error event probability $1 - P_c$. As shown in [6], \hat{D}_{gop} takes into account both distortions due to the compression and the channel imperfections. It is expressed depending on compression distortion D_0 , respective length n_i of the different frames of the GOP, and expressions of the different distortions D_{loss_i} observed when the i^{th} P-frame of the GOP is lost while the previous frames were correct. Taking finally into account numerical corrective factors β_i modelling the video sequence (limited) resilience ($\beta_0 \simeq 0.75$, $\beta_i \simeq 0.85$, $i > 0$), we have:

$$\hat{D}_{gop} = \prod_{i=0}^N P_c^{\beta_i n_i} \cdot D_0 + \sum_{i=0}^N \left[\prod_{j=0}^{i-1} P_c^{\beta_j n_j} \cdot (1 - P_c^{\beta_i n_i}) \cdot D_{loss_i} \right] \quad (1)$$

Corresponding throughput is consequently expressed as follows for a GOP of duration t seconds:

$$\hat{T}_{gop} = \sum_{i=0}^N n_i / t \quad (2)$$

In [6], the approach relied on the complete a priori knowledge of the multimedia data streams to be transmitted to derive the different coefficients of equation (1). Such a knowledge being in practice highly improbable, we show hereafter that only a minimal knowledge on the encoder is in fact necessary to derive the distortion and throughput, and consequently apply application controlling in realistic situations. This minimal knowledge consists of compression abacus curves (see Figure 2) drawn for the considered video

encoder to express its compression efficiency (throughput, *i.e.* length information on the GOP and its Intra frame) and resulting quality (expressed by objective metric such as the Peak Signal to Noise ratio (PSNR), *i.e.* distortion information) vs. considered quantization parameters. In order to limit the number of abacus, we enforced in our simulations a specific rule linking the quantization parameters (QP) for the different frames types (I, P and B (if used)) by a unique QP parameter, taken in the following equal to : $QP = QP_I = QP_P - 2 = QP_B - 2$. Derived off-line for different standard reference video streams, these abacus can then be used by the controller as models of the compression algorithm performance to derive the values necessary to equation (1), *i.e.* lengths n_i and distortions D_{loss_i} and D_0 .

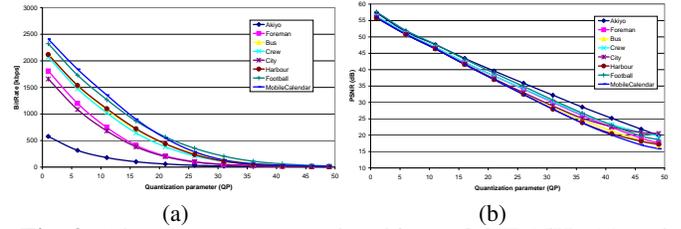


Fig. 2. Abacus curves examples: bitrate QCIF 15Hz (a) and PSNR CIF 30Hz (b) curves vs. QP for H.264/AVC videos.

For a given abacus curve, D_0 is naturally obtained from the quality abacus, while GOP total length $n_{tot} = \sum_{i=0}^N n_i$ and Intra length n_0 are obtained from the throughput abacus. Approximation based on experimental observations and knowledge of compression algorithm principles leads us to propose the complementary following expressions:

$$\begin{aligned} D_{loss_0} &= 10 \\ D_{loss_i} &= D_{loss_0} + i/N \times (D_0 - D_{loss_0}) \quad \forall i > 0 \\ n_i &= (n_{tot} - n_0) / (N - 1) \end{aligned} \quad (3)$$

Derivation of equation (1) coefficients being explained with the abacus curves, remains to determine the adapted curves at each control step. As shown by the algorithm flowchart in Figure 3, the controller operates first based on previous step/initial settings, then launches compression process and corresponding quality estimation. The obtained results (in terms of throughput and distortion) are compared with prediction and either adopted if they fit within acceptable margins (typically 10% of bitrate) or processed by the controller to determine the new reference curve it should use to relaunch the compression process. After a limited number of trials (typically 3 trials for a reasonable complexity), the algorithm converges and best settings are used for real transmission.

It must be pointed out that other RD curves could be used for our on-the-fly calibration and consequent controlling algorithm. Typically, more complex models such as those of [3][5] or simplified models such as [10] which uses residual complexity of intra-pictures to estimate the type of content, from which the RD curve to use could be deduced.

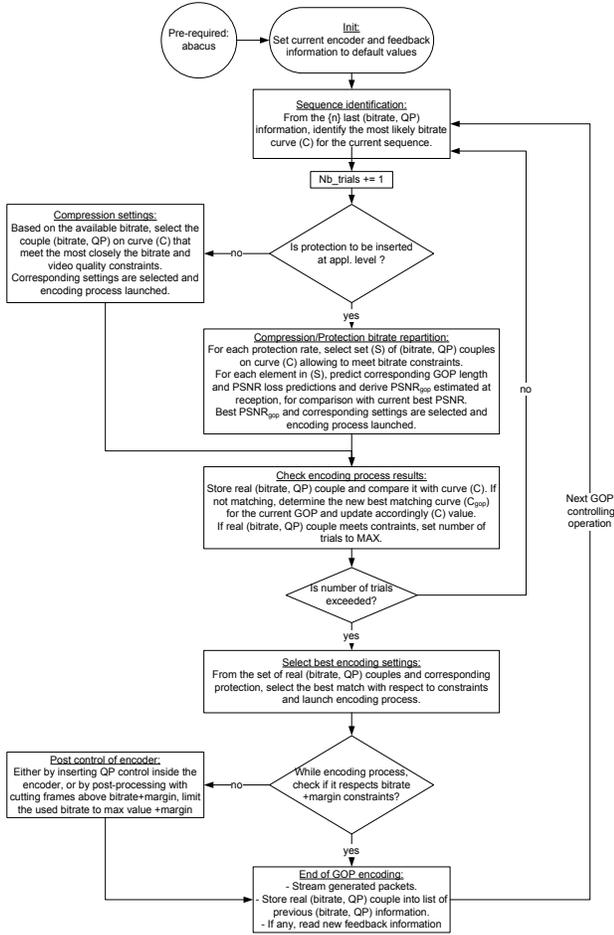


Fig. 3. Flowchart for application adaptation through "partial and trial" abacus driven algorithm.

4. SIMULATION EXAMPLES

In our experimentations, the communication link is a bandwidth unlimited wireless channel, corresponding to an optional additive white Gaussian noise (AWGN) insertion, the video encoder is the x264[11] real-time H.264/AVC encoder and systematic convolutional error correcting codes are considered as in [6] (similar results were also obtained with Reed-Solomon codes). Beside a first test with the well-known 'Foreman' sequence, we consider in the following a concatenation of different ITU-T sequences, created to alternate complex and more simple scenes. The so-called 'MixsequenceCIF' video is a compilation of 9 ITU-T CIF at 30 Hz reference sequences, for a total of 2730 frames ('Akiyo', 'Children', 'Container', 'Foreman', 'Hall Monitor', 'Mobile Calendar', 'Mother and daughter', 'Stefan', 'Table tennis'). For QCIF resolution, the sequences 'Carphone' (third place) and 'Trevor' (last place), representative of low rate mobile applications were also added, leading to the 'Mixsequence-QCIF' compilation of 11 ITU-T QCIF at 15 Hz reference sequences for a total of 3360 frames.

Let us first illustrate the accuracy of the semi-analytical quality estimation obtained with equation (1) when using abacus prediction and equations (3) expressions. Figure 4 presents the PSNR quality obtained for 'Foreman' QCIF 15Hz first GOP, and compares these results to the experimental values obtained with the real compression process. As can be seen from the figure, the prediction is quite accurate and justifies our abacus-based quality estimation. The results also highlights the fact that the theoretical formula is neither a lower or an upper bound, but an approximation, which explains why some results can be better than the theoretical ones. Interestingly, the abacus used by the algorithm for this prediction is not 'Foreman' but 'Akiyo', which reflects the fact that our prediction is really based on the current GOP and not on an average behaviour of the transmitted sequence.

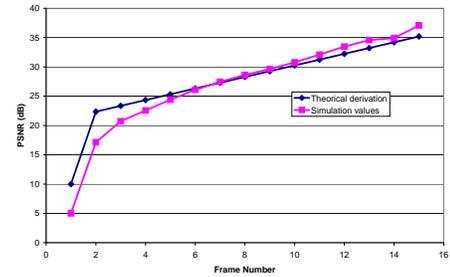


Fig. 4. Comparison of prediction vs. experimental results for 'Foreman' sequence, first GOP, QCIF 15 Hz.

Let now consider the efficiency of the throughput control algorithm, in the case where the channel condition are very good (corresponding to a minimal protection decision for our convolutional codes). Figures 5 and 6 present the obtained variation of throughput with respect to time for MixSequence CIF video sequence, and the corresponding PSNR variation for a network target fixed to 300kb/s (corresponding to 270kb/s for the data before encapsulation), with a tolerance margin of 10%, derived over a sliding windows to compensate for eventual peaks. Figure 5 shows the limited number of 3 trials is sufficient for almost all control step to remain within the tolerance margin, and shows a final average throughput of 263.78kb/s only at 2.5% of the target.

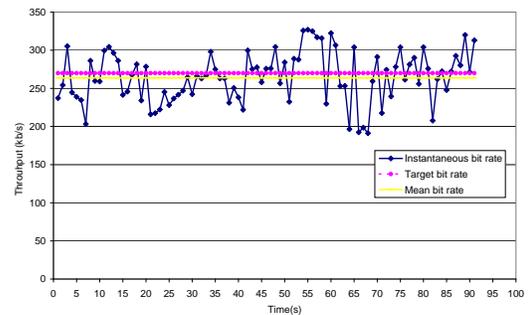


Fig. 5. Results for 'MixSequenceCIF' sequence, for a target of 270 kb/s and perfect channel: bitrate evolution.

To better illustrate the interest of the algorithm, let consider now a case where the throughput evolves with time. We

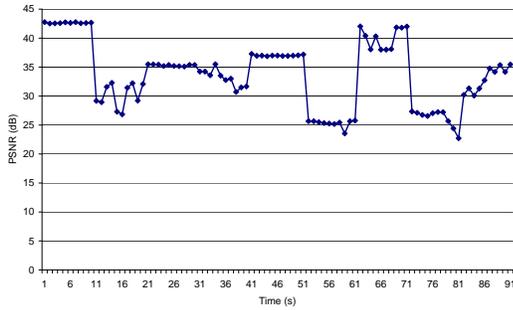


Fig. 6. Results for 'MixSequenceCIF' sequence, for a target of 270 kb/s and perfect channel: quality evolution.

use for Figures 7 and 8 a throughput pattern obtained from the ICT WiMAGIC project, that models the evolution of MAC throughput offered by a WiMAX radio access in mobility conditions. Beside this throughput evolution, we have also added in the simulation an emulated variation of channel, with equivalent signal-to-noise ratio E_s/N_0 going from bad (2dB) to good (8dB) and average (5dB), with changes time instants illustrated by the arrows above the x-axis.

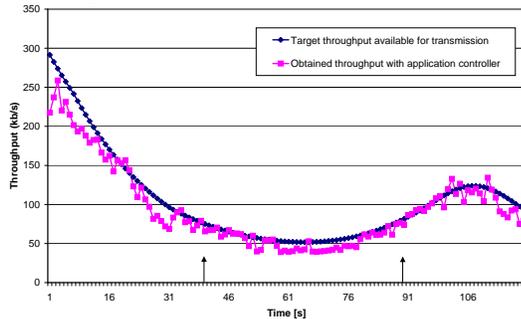


Fig. 7. Bitrate variations for a WiMAX user profile.

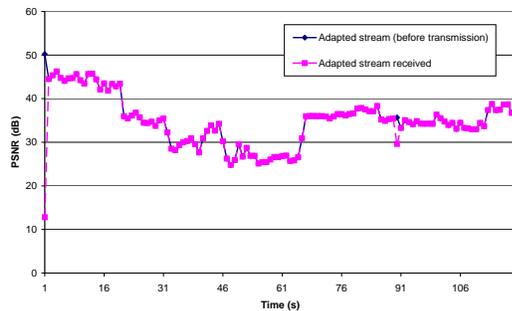


Fig. 8. Quality variations for a WiMAX user profile.

The available throughput for the application and the corresponding achieved throughput evolution obtained through application controlling is given in Figure 7, while the corresponding quality evolutions before transmission and at the receiver side are given in Figure 8. It can be observed first that the throughput target here again is met with remarkable accuracy, despite the quick throughput variation, and the additional variation of the wireless link quality. Secondly, it must be noted that except in two occasions, corresponding to channel quality degradations, that cannot be anticipated by our causal algorithm, the remaining error rate due to the

radio access is completely taken care of by the application controller. The quality variations naturally evolve in relation with the bitrate constraints, but without any dramatic quality fall drops that would be unacceptable for a system. This is very important to mention, as a streaming UDP or even TCP based system without adequate rate control would generally result in losing packets either due to congestion or excess of transmission delay. In such cases, the quality drops can be very important, and of course problematic for the end user.

5. CONCLUSIONS

An application control algorithm to drive efficiently H.264/AVC transmissions over IP wireless network is presented in this paper. This method relies on a semi-analytical expression for the end-to-end distortion due to video transmission over an error-prone channel, and on a calibration thanks to an iterative usage of pre-calculated abacus modelling the used video encoder performance. The algorithm was shown to be able to adapt extremely quickly (typical adaptation rate being one GOP) to both throughput and error probability evolutions due to the transmission channel. Future works include the adaptation of this algorithm to scalable video schemes, but also to limited bandwidth channels.

6. REFERENCES

- [1] F. De Angelis, I. Habib, F.A.M. Davide, and M. Naghshineh, "Intelligent content aware services in 3g wireless networks," *IEEE J. Sel. Areas Comm.*, vol. 23, no. 2, pp. 221–234, 2005.
- [2] L. Hanzo and J. Streit, "Adaptive low-rate wireless videophone systems," *IEEE TCSVT*, vol. 5, no. 4, pp. 305–318, Aug. 1995.
- [3] Z. He, J. Cai, and C.W. Chen, "Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding," *IEEE TCSVT*, vol. 12, no. 6, pp. 511–523, June 2002.
- [4] M. Gallant and F. Kossentini, "Rate-distortion optimized layered coding with unequal error protection for robust internet video," *IEEE TCSVT*, vol. 11, no. 3, pp. 357–372, March 2001.
- [5] M. Van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 w lans," *IEEE J. Sel. Areas Comm.*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.
- [6] C. Bergeron and C. Lamy-Bergot, "Modelling h.264/avc sensitivity for error protection in wireless transmissions," in *Proc. of IEEE MMSP'06*, Canada, Oct. 2006, pp. 302–305.
- [7] H. Yang and K. Rose, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in h.264/avc," *IEEE TCSVT*, vol. 17, no. 7, pp. 845–856, July 2007.
- [8] Y. Zhang, W. Gao, Y. Lu, Q. Huang, and D. Zhao, "Joint source-channel rate-distortion optimization for h.264 video coding over error-prone networks," *IEEE Tr. Mult.*, vol. 9, no. 3, pp. 445–454, Apr. 2007.
- [9] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE TCSVT*, vol. 13, no. 7, pp. 560–576, July 2003.
- [10] G.G. Lee, H.-Y. Lin, and M.-J. Wang, "Rate control algorithm based on intra-picture complexity for h.264/avc," *IET Image Processing*, vol. 3, no. 1, pp. 26–39, 2009.
- [11] VideoLan, *H.264/MPEG-4 AVC encoder: x264*, available at [git://git.videolan.org/x264.git](http://git.videolan.org/x264.git), Jan. 2010.