CONDITIONAL ACCESS TO H.264/AVC VIDEO BY MEANS OF REDUNDANT SLICES

Marco Grangetto, Enrico Magli, Gabriella Olmo

Dipartimento di Elettronica - Politecnico di Torino
Corso Duca degli Abruzzi 24 - 10129 Torino - Italy
Ph.: +39-011-5644195 - Fax: +39-011-5644099
firstname.lastname@polito.it

ABSTRACT

In this paper a novel conditional access scheme for the distribution of H.264/AVC video is presented. The algorithm permits to cypher the full quality video, while guaranteeing free access to a reduced quality video that can be used as preview. The scheme is based on the novel coding options, introduced in the H.264 standard, and therefore is fully compliant with the standard syntax. The secured video stream exhibits a very small rate overhead and requires limited supplementary computational cost at coding time.

Index Terms— Conditional access, security, encryption, H.264/AVC

1. INTRODUCTION

Security and digital rights management are becoming ever more important issues for the deployment of successful multimedia distribution systems across non trusted network, e.g. the Internet. Security can be obtained by means of recent encryption standards such as the advanced encryption standard (AES). AES can protect data communications, providing a high degree of security at a reasonable computational cost. However, direct encryption of compressed multimedia data exhibits a number of shortcomings. First of all, multimedia data are characterized by high bit-rates; as a consequence, the encrypting algorithm shall require excessive computational burden and/or power consumption. Moreover, multimedia compressed streams are characterized by particular data structures and syntax rules that cannot be preserved by direct encryption. The perturbation of the multimedia stream structure can prevent a number of possible operations, such as transcoding, rate shaping and quality scalability.

Selective encryption [1], which consists in ciphering only the subjectively most important compressed data, represents a possible solution in order to limit the computational cost. Encryption can be carried out at different stages of the compression process, e.g. the pixels, the transform coefficients, the quantization indexes, the bit-planes, the entropy coder, or the final codestream. It should be noted that encrypting data before the entropy coder, or during the entropy coding stage, may result in a loss in coding efficiency due to the modified data statistics [2]. When employing an international standard, the protected file should also be syntax-compliant with the standard, so that, if a decoder attempts to decode a protected file it will generate a meaningless content, but will not crash due to syntax errors; this problem is typically incurred if the compressed file undergoes encryption. In [3] an arithmetic coder is used to carry out joint compression and encryption. However, its application to H.264/AVC has turned out to be difficult, because decoding encrypted data such as the motion vectors or the coding modes can lead to out-of-range values. The same problem is encountered if one attempts to selectively applying an external encryption scheme such as AES to parts of the compressed file.

Another application, namely *conditional access*, has been proposed in [4]. In conditional access, a low-quality version of the image is left in the clear, and can be used to preview the multimedia content; the user can purchase a key, and then decode the content at full quality. In [5] conditional access to H.264 video is obtained using signal processing techniques. Two different techniques, respectively based on removal and separate encryption of DCT coefficients, and on random perturbation of the motion vectors, are proposed.

In this paper, we propose a novel conditional access scheme that is expressively designed for the H.264 standard. In particular, the low-quality version of the video sequence is obtained by injecting a controlled amount of drift in the decoder compensation loop. The proposed technique exploits the functionality of the *Redundant Slice* coding option and is fully compliant with the standard. In the following the algorithm performance is analyzed in terms of video quality, added complexity and rate overhead.

2. PROPOSED TECHNIQUES

The conditional access technique presented in the following is kept compliant with the H.264/AVC [6] video coding standard. In particular, the compressed and secured bitstreams are constrained to be playble by a standard H.264 decoder, which may be not aware of the securization process. This goal is

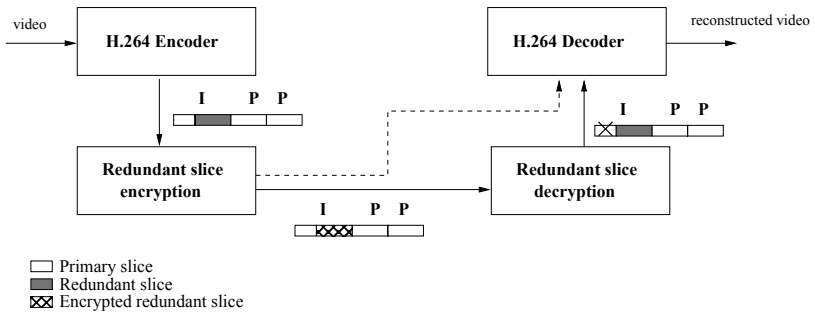

Fig. 1. Block diagram of the proposed conditional access scheme.

obtained using the concept of *primary* and *redundant* slices, defined in H.264/AVC standard.

Primary slices are used to code the primary picture, and are associated to a normative decoding procedure. On the other hand, redundant slices have been introduced as an error resilience tool and represent an alternative representation of a picture; as an example, redundant slices can be created by increasing the quantization parameter QP, so as to generate a coarser approximation of the image. The H.264 recommendation does not specify a normative decoder behavior in presence of redundant slices. Clearly, when some of the samples in the decoded primary picture cannot be correctly decoded, due to errors or transmission losses, the decoder shall replace the samples of the decoded primary picture with the corresponding ones of the decoded redundant slice. Moreover, it is possible to create more than one redundant representation of each slice. Each slice is marked by its *redundant_pic_cnt* counter, which is coded in the slice header. The primary slice has *redundant_pic_cnt*=0, whereas alternative and coarser representations are given increasing numbers of *redundant_pic_cnt*. In this work, redundant slices are used to provide conditional access to the H.264 video stream. In particular, redundant slices are used to convey encrypted data that allow to improve the decoded quality only when the cyphering key is given to the final user. In the following, only two levels of access quality are considered for simplicity, by employing a single redundant slice representation; nevertheless, the proposed scheme can be extended to an arbitrary number of quality levels by using more than one redundant representation.

The proposed securization system is graphically sketched in Fig. 1. The proposed encoder is based on a H.264/AVC encoder followed by a simple cyphering post-processing. The input video is coded according to the H.264/AVC standard. Every Intra (I) slice is coded twice; in the first pass a primary representation with reduced quality and rate is generated by using a larger $QP_p = QP + \Delta QP$, where QP is the quantization parameter specified by the user or determined by the rate control stage. Then, a redundant slice (*redundant_pic_cnt=1*) is formed by coding the same macroblocks with quantization parameter QP. The redundant representation of each picture is stored in the reference buffer for future motion estimation and compensation. This final step is not compliant with the standard H.264 decoding procedure; in fact, standard decoding assumes that the primary representation of each picture is the best one, and must be used for future motion compensation. As a consequence, the motion estimation, based on the low quality primary picture, will generate drift in the decoder compensation loop. The outcome is a reduction of the video quality obtained by a standard H.264 decoder, which generates a low-quality preview of the video sequence. The H.264 coding stage is then followed by the post-processing tool that encrypts the redundant slices, so as to prevent unauthorized use of the conveyed data. The input H.264 plaintext bitstream is a sequence of Network Access Layer Units (NALU), each one containing a coded slice. The cyphering stage parses the NALU and reads the *redundant_pic_cnt* field from each slice header; only the payload of redundant slices is encrypted by means of a Vernon stream cypher. The seed of the pseudorandom number generator must be communicated to the decoder, in order to allow it to decode the video sequence at full quality. Initializing the generator with the given seed, the decoder will reproduce the same sequence of numbers used by the encoder, and by ex-oring them with the received bits, it will be able to reconstruct the plaintext exactly. Clearly, other stream cyphers, such as RC4, could also be used; the AES standard in output feedback mode could also be employed. It is important to notice that only the slice payloads are encrypted so as to guarantee that the H.264 header syntax is not violated.

In presence of the primary slices the secured H.264 bitstream is compliant with a standard decoder (see dash line in Fig. 1). In fact, a standard H.264 decoder can correctly parse all slice headers and consequently discards all redundant slices. The decoding of the low quality reference pictures generates drift in the motion compensation loop, thus limiting the quality of the whole decoded video. An exception is represented by I coded macroblock in predicted slices that can be selected in high motion areas. It is evident that such I macroblocks can locally improve the image quality. This effect will be analyzed in Sect. 3. A possible solution

Table 1. Experimental results on Foreman sequence, $R = 115$ kbps, $PSNR(Y) = 35.73$ dB (full quality).

Δ QP	Rate [kbps]	ρ	PSNR(Y) [dB]
10	120	0.048	30.78
15	118	0.031	27.60
20	117	0.017	24.51
25	116	0.009	22.51

Table 2. Experimental results on Mobile sequence, $R = 402$ kbps, $PSNR = 31.54$ dB (full quality).

Δ QP	Rate [kbps]	ρ	PSNR(Y) [dB]
10	440	0.094	26.58
15	433	0.077	23.66
20	430	0.067	20.48

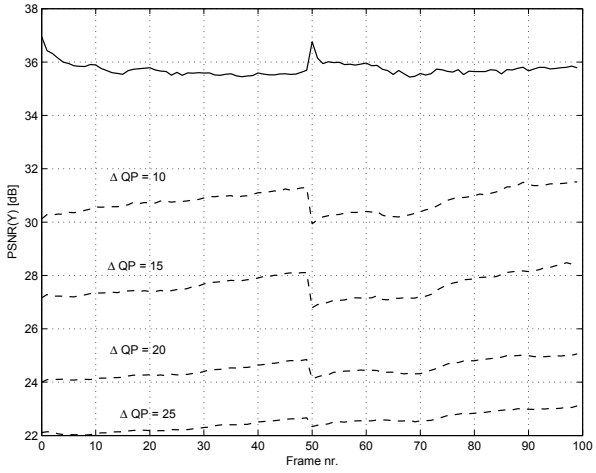

Fig. 2. Luminance PSNR of Foreman sequence versus frame number: full quality (solid), standard decoder (dash).

can be obtained by generating a redundant representation of predicted slices when the number of I coded macroblock exceeds a given threshold.

In order to access the high quality video a pre-processing step must be used before standard H.264 decoding (see solid line in Fig. 1). The pre-processing stage parses the secured NALU sequence, decrypts the redundant slices and removes the corresponding primary ones. As a consequence, the following H.264 decoder is forced to use the high quality redundant pictures, thus generating the full quality video. Alternatively, a nonstandard decoder can be employed, which checks for the presence of redundant slices on-the-fly.

The proposed conditional access scheme exhibits a certain rate penalty due to the use of two compressed representations of the same slice. It is clear that this penalty depends on the selected value of ΔQP. In the following the rate penalty will be measured in terms of the redundancy $\rho = \Delta R/R$, being ΔR the extra rate used to code the low quality primary slices and R the rate needed to code the full quality video.

3. EXPERIMENTAL RESULTS

The proposed technique has been tested on standard QCIF and CIF sequences. In particular, the results reported in the following are worked out employing 100 frames of the QCIF Foreman sequence at 30 Hz and 100 frames of the Mobile CIF sequence at 10 Hz. Both sequences have been encoded using an Intra refresh of 50 frames (GOP=50), no B slices and CABAC entropy coding.

The full quality video for the Foreman sequence has been obtained with a fixed $QP = 28$, corresponding to $R = 115$ kbps and yielding an average PSNR for the luminance component of 31.54 dB. Slices corresponding to the whole frame are employed. In Tab. 1 the coding rates and their corresponding redundancy ρ are shown when the proposed technique is used with $\Delta QP = 10, 15, 20, 25$. The obtained redundancy ranges from about 5% to less than 1% for the largest value of ΔQP. In this latter case QP_p is saturated to its maximum value of 51 and the redundancy reaches its lowest bound. These results show that the proposed scheme introduces a limited rate penalty. Moreover, the value of ρ also represents the percentage of the bitstream that has been cyphered, thus requiring additional computational cost with respect to standard video coding. The rightmost column in Tab. 1 shows the average luminance PSNR obtained when decoding the secured H.264 stream with a standard decoder and allow one to verify the amount of drift introduced in the compensation loop in absence of the access key. Clearly, the amount of drift, i.e. the quality of the freely available video, is controlled by means of ΔQP. In Fig. 3 the luminance PSNR is reported as function of the frame number for all simulated cases (dash curves) and compared with the full quality video. It is worth noticing that the quality of the predicted frames slightly improves along the GOP; this behavior is due to the leakages in the prediction loop, represented by I refreshed macroblocks and drift error reduction because of sub-pixels interpolation and loop filter [7]. As already mentioned, this gain can be limited by inserting cyphered redundant slices for selected P slices as well. This will inevitably impacts on the amount of extra rate; in the present work we considered negligible the slight quality improvement experimented along the GOP, while we preferred to keep the redundancy as limited as possible. Different choices may be driven by practical application constraints; as an example, in the case of wireless video streaming it may vital to save bandwidth. Finally, in Fig.3 the visual quality of the 40th frame of the Foreman sequence is shown; the full quality frame (top-left) is compared with the frame obtained by a standard decoder when $\Delta QP = 10$ (top-right),

Fig. 3. Visual quality of the 40th frame: full quality (top-left), $\Delta QP = 10$ (top-right), $\Delta QP = 15$ (bottom-left), $\Delta QP = 20$ (bottom-right).

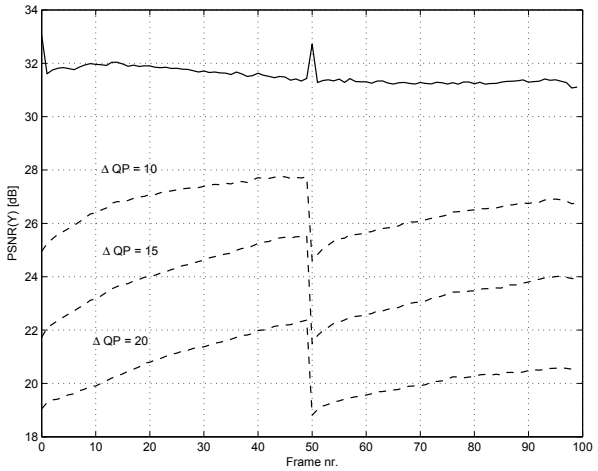

Fig. 4. Luminance PSNR of Mobile sequence versus frame number: full quality (solid), standard decoder (dash).

$\Delta QP = 15$ (bottom-left) and $\Delta QP = 20$ (bottom-right).

In Tab. 2 the results obtained on the Mobile CIF sequence are reported. In this case we set $QP = 31$, corresponding to a coding rate $R = 402$ kbps, and yielding an average luminance PSNR of 31.54 dB. For the CIF sequence 5 slices per frame have been formed. In Fig. 4 the instantaneous PSNR for the luminance component is reported. If compared with the Foreman case, the quality improvement along the GOP is quite significant due to a larger percentage of I refreshed macroblocks.

In conclusion, the presented algorithm yields conditional access to the H.264 compressed video without requiring any modification of the standard syntax. Moreover both the rate overhead and the added computational cost are very limited. Finally, the encrypting tool is implemented by an external post-processing step, making the scheme very flexible with respect to the adopted cyphering technique. Future research will also investigate the possibility to obtain conditional access by means of H.264 switching pictures. Moreover, the performance in terms of image quality, complexity and rate overhead will be compared with conditional access schemes adopting scalable video coding.

4. REFERENCES

[1] H. Cheng, X. Li, "Partial encryption of compressed images and videos," *IEEE Transactions on Signal Processing*, vol. 48, n. 8, pp. 2439-2451, Aug. 2000.

[2] M. Wu, Y. Mao, "Communication-friendly encryption of multimedia," *Proc. of IEEE MMSP 2002*.

[3] M. Grangetto, E. Magli, G. Olmo, "Multimedia selective encryption by means of randomized arithmetic coding," *IEEE Trans on Multimedia*, vol. 8, n. 5, pp. 905-917, Oct. 2006.

[4] R. Grosbois, P. Gerbelot, T. Ebrahimi, "Authentication and access control in the JPEG 2000 compressed domain," *Proc. of the SPIE 46th Annual Meeting*, USA, 2001.

[5] E. Magli, M. Grangetto, G. Olmo, "Conditional access to H.264/AVC video with drift control,", *Proc of. IEEE ICME 2006*.

[6] Joint Video Team JVT of ISO/IEC MPEG and ITU-T VCEG, *International Standard of Joint Video Specification (ITU-T Rec. H.264 , ISO/IEC 14496-10 AVC)*, Mar 2003.

[7] Y. Wang, Z. Wu, J.M. Boyce, "Modeling of transmission-loss-induced distortion in decoded video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, n. 6, pp. 716-732, Jun. 2006.