ROI encryption for the HEVC coded video contents

Conference Paper · September 2015

DOI: 10.1109/ICIP.2015.7351373

4 authors, including:

Mousa Farajallah
Palestine Polytechnic University
24 PUBLICATIONS **98** CITATIONS
SEE PROFILE

Olivier Déforges
Institut National des Sciences Appliquées de R…
173 PUBLICATIONS **693** CITATIONS
SEE PROFILE

Safwan El Assad
University of Nantes
170 PUBLICATIONS **453** CITATIONS
SEE PROFILE

ROI ENCRYPTION FOR THE HEVC CODED VIDEO CONTENTS

Mousa Farajallah, Wassim Hamidouche, Olivier Déforges and Safwan El Assad

IETR Lab CNRS 6164, France

ABSTRACT

In this paper we investigate privacy protection for the HEVC standard based on the tile concept. Tiles in HEVC enable the video to be split into independent rectangular regions. Two solutions are proposed to encrypt the tiles containing the Region Of Interest (ROI). The first solution performs encryption at the bitstream level by encrypting all HEVC syntax elements within the ROI tiles. The second solution enables a selective encryption of the ROI tiles under constant bitrate and format compliant requirements. To avoid temporal propagation of the encryption outside the ROI boundaries caused by inter prediction, the motion vectors of non ROI regions are restricted inside the non encrypted tiles in the reference frames.
Simulation results show that the proposed solutions perform secure and adaptive encryption of ROI in the HEVC video. Moreover, the bitrate overhead caused by the MVs restriction window varies between 1%-2.5% depending on both the video content and the number of tiles within the frame.

1. INTRODUCTION

The rapid development of multimedia devices and network infrastructure enables new video services in ultra high resolution. The new generation video coding standard, High Efficiency Video Coding (HEVC) [1], enables a gain of up to 50% in terms of subjective video quality, compared to the AVC high profile [2]. Moreover, the HEVC standard defines new tools such as Tile and wavefront to leverage multi-core architectures and accelerate the encoding process. The first HEVC version [3] finalized in January 2013 is expected to be promptly deployed in industry especially in new video surveillance products, where the importance of privacy protection is substantial. In this paper we investigate privacy in the HEVC standard through an efficient encryption of Region Of Interest (ROI). A number of studies have investigated privacy in video content [4, 5, 6, 7, 8, 9, 10]. Authors in [4] proposed a selective encryption solution of ROI based on Flexible Macroblock Ordering (FMO) concept in the H.264/AVC [11] and chaos encryption system. The macroblocks of each frame are mapped into two different slices, one regroups macroblocks within the ROI and other slice for macroblocks outside the ROI. Therefore, only the ROI slice is encrypted with a selective encryption solution based on chaos system. The ROI represents human faces which are detected in the video based on the skin color model [12]. Dufaux et al. [5] proposed a solution to hide ROI in MPEG-4 video content for privacy protection in video surveillance. This encryption solution is carried out at the transform domain by pseudo-randomly flipping the selected Transform Coefficient (TC) signs. To avoid the propagation of the encryption outside of the ROI, the macroblocks using the encrypted ROI as reference for inter prediction are rather Intra coded. Work in [6] enables rectangular region privacy by de-identifying faces. This solution guarantees that face recognition software cannot reliably recognize de-identified faces even though part of the facial details are preserved. Authors in [7] investigated privacy protection in the H.264/SVC (Scalable Video Coding). This solution first detects face regions (ROI) and then encrypts these ROI in the transform domain by scrambling the sign of the non-zero TCs at all SVC layers.
In this paper we propose two ROI encryption solutions in the HEVC standard based on Tile concept and the Advanced Encryption Standard (AES) system. The Tile concept [13] in HEVC splits the frame into rectangular regions. The ROI is included in a set of Tiles (called ROI tiles) and the background is within the rest of tiles (non ROI tiles). The first solution encrypts at the bitstream level all syntax elements within the ROI tiles, while the second solution performs format compliant and constant bitrate selective encryption of the ROI tiles. The selective encryption solution, carried out at the Context-Adaptive Binary Arithmetic Coding (CABAC) binstring level, encrypts a set of HEVC parameters including Motion Vector (MV) difference, MV signs, TCs and TC signs. To avoid the propagation of the encryption outside the ROI region caused by inter prediction, the MVs of non RIO region are restricted inside the background region. This may decrease the rate-distortion performance, but it enables a correct decoding of non encrypted region even when the ROI is not correctly decoded. Several works [14, 15, 16] have investigated selective encryption in the HEVC standard. To the best of our knowledge, this is the first work that investigates protection of ROI privacy in the HEVC standard.
The rest of this paper is organized as follows: Section 2 presents the encryption system including the HEVC standard, its entropy coding engine and the encryption algorithm. The

This work is supported by the European Celtic-Plus project 4KREPROSYS - 4K ultraHD TV wireless REmote PROduction SYStems -

proposed ROI encryption solutions in the HEVC standard are described in Section 3. The performance of these solutions is assessed in Section 4. Finally, Section 5 concludes the paper.

Fig. 1: Tile concept in the HEVC standard in red rectangles

2. SYSTEM DESCRIPTION

2.1. HEVC video standard

The significant coding gain [2] enabled by the HEVC standard is obtained thanks to new adopted tools, such as quadtree-based block partitioning, large transform and prediction blocks, accurate intra/inter predictions and the in-loop sample adaptive offset (SAO) filter [1]. Moreover, HEVC defines a highly adaptive entropy coding engine named the CABAC. The HEVC frame is partitioned into Coding Tree Units (CTUs). Each contains one luma Coding Tree Block (CTB) and two chroma CTBs. Recursive subdivision of a CTU results in Coding Unit (CU) leaves with the corresponding Coding Blocks (CBs). The CU can be split into Prediction Units (PUs), a basic entity for intra and inter predictions, and recursively split into Transform Units (TUs), a basic entity for residual coding [1]. The HEVC standard was designed with a particular attention to complexity, where several steps can be easily performed in parallel. Three high level parallel processing approaches, including independent slice, tile, and wavefront, can be used in HEVC to simultaneously process multiple regions of a single picture [17]. The tile concept splits the picture into rectangular groups of CTBs, called tiles. Tiles break the CABAC and the intra prediction dependencies that each tile can be independently processed. Figure 1 illustrates a HEVC picture composed of 15 tiles (red rectangles), where each tile consists of 9 CTBs. These tiles represent independent and contiguous regions of the HEVC frame. However, tile concept decreases the rate-distortion performance caused by the intra prediction limitation and the initialization of the CABAC probabilities.

2.2. CABAC binarization in HEVC

As illustrated in Figure 2, the CABAC engine in HEVC consists in three main functions: binarization, context modeling

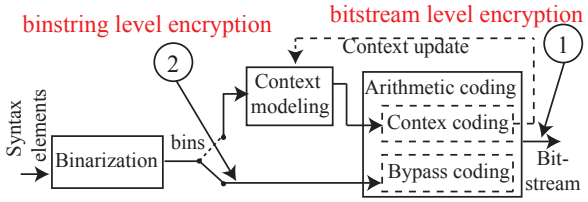

Fig. 2: Three main functions in the CABAC

and arithmetic coding [18]. First, the binarization step corresponds syntax elements to binary symbols (bin). Second, the context modeling updates the probabilities of bins, and finally the arithmetic coding compresses the bins into bits according to the estimated probabilities. Five binarization methods are used in HEVC namely Unary (U), Truncated Unary (TU), Fixed Length (FL), TRp, and EGk codes. The U code represents an unsigned integer Y with a bin string of length $Y+1$ composed of Y 1-bins followed by one 0-bin. The TU code is defined with the largest possible value of the syntax element $cMax$ ($0 \leq Y \leq cMax$). When the syntax element value $Y < cMax$, the TU is equivalent to U code, otherwise Y is represented by a bin string of $cMax$ 1-bins. The FL code represents a syntax element Y with its binary representation of length $\lceil \log_2(cMax+1) \rceil$. The TRp code is a concatenation of a quotient $q = \lfloor Y/2^p \rfloor$ and a remainder $r = Y - q2^p$. The quotient q is first represented by TU code as a prefix concatenated with a suffix r represented by the FL code of length p. The EGk code is also a concatenation of prefix and suffix. The prefix part of the EGk code is the U representation of $l(Y) = \lfloor \log_2(\frac{Y}{2^k}+1) \rfloor$. The suffix part is the FL code of $Y + 2^k(1-2^{l(Y)})$ with $cMax = k + l(Y)$. The arithmetic coding can be performed either by an estimated probability of a syntax element (context coded) or by considering equal probability of 0.5 (bypass coded).

3. ROI ENCRYPTION SOLUTIONS IN HEVC

In this section, we propose two solutions based on the tile concept to protect privacy in the HEVC standard. The first common step consists of the identification and tracking of the ROI in the video. This can be done in real time by any existing algorithm such as face identification and tracking for video surveillance applications [12, 10]. The second common step uses information of ROI localization in the frame provided by the first step to split the HEVC frame into tiles where all ROI are included in ROI tiles and the background in separated non ROI tiles. In Figure 1, tiles 1, 2, 6 and 7 including human face represents the ROI tiles and other tiles will represent background tiles. It should be noted that the tile repartition provided in Figure 1 with red edges is not optimal in terms of coding rate-distortion performance. In fact, more efficient tile repartition can be defined to minimize the number of tiles in the frame where all ROI and background regions are included

in a minimum number of tiles. For this example, the human face can be represented in one ROI tile which regroups only parts of tiles 1, 2, 6, 7 and the rest of tiles will represent the background. This repartition will provide more efficient coding performance since it reduces the number of tiles from 15 to only 6 tiles in the new repartition illustrated in Figure 1: ROI tile is highlighted in green dashed rectangle and non ROI tiles in blue dashed rectangles. The first solution, called Tile Naive Encryption - HEVC (TNE-HEVC), encrypts all syntax elements coded in CABAC within the ROI tiles. The encryption is performed at the bitstream level while the video headers, including VPS, SPS, PPS, and slice headers are not encrypted since they are not entropy coded with the CABAC. The second encryption solution, called Tile Selective Encryption - HEVC (TSE-HEVC), performs a selective encryption of ROI tiles in format compliant and at constant bitrate. The encryption is performed at the CABAC binstring level (see Figure 2) by encrypting only the most sensitive HEVC syntax elements to decrease the visual quality of the ROI. The encryption space is defined in the next section.

3.1. AES in block cipher mode

The AES encryption system [19] is used in cipher feedback (CFB) mode to encrypt the HEVC syntax elements. As shown in Figure 3, the AES in CFB mode introduces internal diffusion and external diffusion. The AES in CFB mode enables to produce the 128 bits (X_i) which can be than used to perform the selective encryption of HEVC syntax elements on the fly.

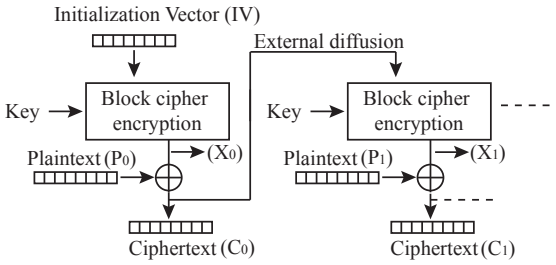

Fig. 3: AES encryption system in CFB mode

3.2. Encryption space

The TSE-HEVC solution encrypts only syntax element binarized in FL code and then bypassed. This restriction enables the TSE-HEVC solution to perform at the CABAC binstring level constant bitrate and format compliant encryption. The absolute value of MV difference minus 2 is binarized in EG1 code and then bypass coded in the CABAC. Thus, only the suffix of the MV difference can be safely encrypted. The signs of MV difference and the TCs are also encrypted since they are binarized in FL code with $cMax = 1$ and bypassed. Concerning the TCs, the remainig parameters ($TC - baselevel$) are bypass coded and binarized with a combination of TRp code with $p \in \{0, 1, 2, 3, 4\}$ and EGk code where $k = p+1$. The suffix EGk code can be safely encrypted while the encryption of the TRp suffix is not format compliant since the encryption of the whole suffix can affect the update of the p parameter. We use the algorithm proposed in [20] that enables the accurate determination of bins in the TRp suffix which can be encrypted without affecting the update of the p parameter. This allows the encryption space of TCs to be maximized while preserving format compliant and constant bitrate requirements.

3.3. MV restriction in the background tile

In the two ROI encryption solutions, the decoding of the background tiles can use, within the reference frames, decoded samples belonging to the encrypted tiles for inter layer prediction. Moreover, the encrypted MVs can also be used by the background tiles for inter layer prediction in the HEVC merge mode. Merge mode in HEVC derives the MVs information from a list of spatial neighbor and temporal candidates [1]. Therefore, these two decoding operations can propagate the encryption from the encrypted tiles to the background tiles when the ROI is not correctly decrypted. In the case of merge mode, we restrict the temporal candidates of the background tiles to be inside the background zone in the reference frame. In the case of regular inter prediction, the MVs differences are signaled. Therefore, we restrict the research window of the temporal candidates to be in the background tiles of the reference frames. It should be noted that the boundary of the research window is equal to non ROI boundaries narrowed by 3 pixels and 1 pixel in luma and chroma components, respectively. This enables to perform a safe interpolation process at the tiles boundaries. Finally, the in-loop filters (deblocking and SAO) are disabled at the tiles boundaries.

3.4. Encryption process

In the TNE-HEVC solution, the bitstream within the ROI tiles is encrypted block by block (P_i of 128 bits) as follows:

$$C_i = E_k(C_{i-1}) \oplus P_i \tag{1}$$

where $C_{-1} = IV$, C_i is the encrypted bits in the current ciphered block, C_{i-1} is the previous ciphered-block and E_k is the AES encryption function in CFB mode with the secret key k. The decryption at the decoder side is achieved with XOR operation in reverse using the same encryption algorithm:

$$P_i = E_k(C_{i-1}) \oplus C_i \tag{2}$$

In the TSE-HEVC solution, the encryption process is carried-out syntax element by syntax element after generating one AES block (X_i):

$$c_j = x_j \oplus p_j \tag{3}$$

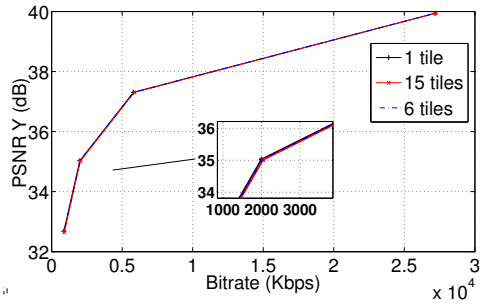

Fig. 4: Rate distortion performance

(a) TNE-HEVC PSNR1=10.95,(b) TSE-HEVC PSNR1=7.78, PSNR2=42.07, PSNR3= 21.15(dB) PSNR2=42.07, PSNR3=18.01 (dB)

Fig. 5: Visual quality of frame #1 in the $Kimono$ video sequence $QP = 27$ (PSNR1: PSNR Y ROI, PSNR2: PSNR Y background, PSNR3: PSNR Y frame).

where p_j is the encryptable bins of one syntax element and x_j is a group of bits, of the same size than p_j, from the vector (X_i) generated with the AES algorithm. The c_j bits are collected to construct the ciphered-block of 128 bits (C_i) used as an input in the next $(i+1)$ AES block cipher encryption.

4. RESULTS AND ANALYSIS

4.1. Experimental configuration

The HEVC reference software model version 15 (HM 15.0) is used to encode and decode the video sequences. Encryption and decryption algorithms are integrated into the codec. We consider three HD video sequences from the common HEVC text conditions [21]. These video sequences are coded in the low delay P configuration at different quantization parameters $QP \in \{22, 27, 32, 37\}$ with enabling the tile tools.

4.2. Results

Figures 4 shows the average rate distortion performance of the three video sequences in three coding configurations: 1 tiles, 15 tiles and 6 tiles with MVs limitations and disabling the in-loop filters across the tile edges. The rate distortion loss of the two tiles configurations is provided in Table 1 in terms of Bjontegaard's difference. In the 6 tiles repartition, the RD loss, caused by these three limitations, remains low and does not exceed 2.5% in the three video sequences. Figures 5a and 5b illustrate the visual quality of frame #1 of *Kimono* video sequence where the ROI is encrypted with the TNE-HEVC and TSE-HEVC solutions, respectively. In both solutions, the visual quality is drastically decreased with Peak Signal to Noise Ratio (PSNR) of the ROI less than 11 dB, while the background region remains clear. However, the TNE-HEVC solution is not HEVC format compliant, thus the encryption desynchronizes the decoder inside the ROI: desynchronization of the CABAC from the start green part in the ROI. The decoder is re-synchronized at the next clear tile thanks to the access points at the non encrypted slice header signaling tiles position in the bitstream. Table 2 gives the average PSNR Y and the Encryption Space (ES) of the three video sequences. The average PSNR inside the ROI remains low for all sequences and does not exceed 11.5 dB. The ES of the TNE-HEVC solution is on average equal to 14.28% since all syntax elements in the ROI are encrypted while it represents on average only 2.5% in the TSE-HEVC solution. Finally, face recognition in the *Kimono* video sequence is assessed with using Principal Components Analysis (PCA) algorithm [22] based on Mahalanobis Cosine (MAHCOS) distance. The first rank is used which corresponds the best match of the test image compared to the training one. The recognition rate decreases from 68% in the original video to 0.5% in the video encrypted with TSE-HEVC solution.

5. CONCLUSION

In this paper we have proposed two encryption solutions based on the HEVC tile concept and the AES algorithm in CFB mode to protect privacy in the HEVC video content. The first solution performs encryption at the bitstream level while the second solution carries out format compliant encryption at the CABAC binstring level. Restrictions are introduced in the HEVC coding process to prevent the propagation of the encryption outside the ROI region but at the expense of rate-distortion loss. Experimental results showed that both solutions perform a secure protection of privacy in the HEVC video content; and the TSE-HEVC solution has a low ES and prevents unexpected behavior of the decoder.

Schemes	15 tiles per frame			Optimal (6 tiles)		
	Y	U	V	Y	U	V
Kimono	-3.27%	-3.26%	-3.03%	-1.49%	-1.63%	-1.56%
ParkScene	-1.41%	-2.34%	-2.07%	-0.51%	-1.21%	-1.09%
BQTerrace	-1.37%	-3.89%	-5.09%	- 0.5%	-1.64%	-2.54%

Table 1: Bjontegaard's difference of HEVC tile repartitions

Schemes	TNE-HEVC		TSE-HEVC		PSNR2
QP=22	PSNR 1 (dB)	ES (%)	PSNR1 (dB)	ES(%)	(dB)
Kimono	9.04	9.82	10.18	2.05	41.67
ParkScene	10.91	12.7	11.27	2.02	39.81
BQTerrace	10.79	20.32	10.89	3.42	38.72

Table 2: ES and PSNR of the ROI encryption solutions

6. REFERENCES

[1] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1648–1667, December 2012.

[2] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparaison of the Coding Efficiency of Video Coding standards including High Efficiency Video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1969–1684, December 2012.

[3] "High Efficiency Video Coding," in *Rec. ITU-T H.265 and ISO/IEC 23008-2*. Sapporo, JP, January 2013.

[4] Fei Peng, Xiao wen Zhu, , and Min Long, "An ROI Privacy Protection Scheme for H.264 Video Based on FMO and Chaos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 10, pp. 1688–1699, October 2013.

[5] F. Dufaux and T. Ebrahimi, "Scrambling for Privacy Protection in Video Surveillance Systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1168–1174, 2008.

[6] E. M. Newton, L. Sweeney, and B. Malin, "Preserving Privacy by de-identifying Face Images," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 232 – 243, February 2005.

[7] H. Sohn, E.T. AnzaKu, W. De Neve, and Y. M. Ro, "Privacy Protection in Video Surveillance Systems Using Scalable Video Coding," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, September 2009, pp. 424 – 429.

[8] M. Saini, P. K. Atrey, S. Mehrotra, and M. Kankanhalli, "Adaptive Transformation for Robust Privacy Protection in Video Surveillance," *Hindawi Journal Advances in Multimedia*, February 2012.

[9] K. Martin and K. N. Plataniotis, "Privacy Protected Surveillance Using Secure Visual Object Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1152 – 1162, August 2008.

[10] M. Qi, X. Chen, J. Jiang, and S. Zhan, "Face Protection of H.264 Video Based on Detecting and Tracking," in *International Conference on Electronic Measurement and Instruments (ICEMI)*, July 2007, pp. 172 – 177.

[11] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[12] E. N. Lorenz, "Deterministic Nonperiodic Flow," *Journal of Atmospheric Sciences*, vol. 20, pp. 130 – 141, 1963.

[13] R. Sjoberg, Y. Chen, A. Fujibayashi, M. M. Hannuksela, J. Samuelsson, T. K. Tan, Y.-K. Wang, and S. Wenger, "Overview of HEVC High-Level Syntax and Reference Picture Management," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1969–1684, December 2012.

[14] G. V. Wallendael, A. Boho, J. D. Cock, A. Munteanu, and R. V. d. Walle, "Encryption for High Efficiency Video Coding with Video Adaptation Capabilities," *IEEE Transactions on Consumer Electronics*, vol. 59, pp. 634 – 642, August 2013.

[15] Z. Shahid and W. Puech, "Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings," *IEEE Transactions on Multimedia*, vol. 16, pp. 24 – 36, January 2014.

[16] W. Hamidouche, M. Raulet, and O. Deforges, "Real time SHVC Decoder: Implementation and Complexity Analysis," in *IEEE International Conference on Image Processing (ICIP)*, October 2014.

[17] C. C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schier, "Parallel Scalability and Efficiency of HEVC Parallelization Approaches," *IEEE TCSVT*, vol. 22, pp. 1827–1838, December 2012.

[18] V. Sze and M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," *IEEE TCSVT*, vol. 22, pp. 1778–1791, December 2012.

[19] N. Ferguson, B. Schneier, and T. Kohno, "Cryptography Engineering: Design Principles and Practical Applications," in *Wiley Publishing Inc.,*. ISBN 978-0-470-47424-2, 2010.

[20] W. Hamidouche, M. Farajallah, M. Raulet, O. Déforges, and S. El Assad, "Selective Video Encryption using Chaotic System in the SHVC extension," in *40^{th} IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane Australia, April 2015.

[21] B. Frank, "Common Conditions and Software Reference Configurations," Document JCTVC-H1100, JCT-VC of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, San Jose, CA, February 2012.

[22] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Advances in Signal Processing*, vol. 2010, pp. 26, 2010.