Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings
Zafar Shahid, William Puech

To cite this version:

HAL Id: lirmm-01237024
https://hal-lirmm.ccsd.cnrs.fr/lirmm-01237024
Submitted on 2 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings

Zafar Shahid and William Puech

Abstract—This paper presents one of the first methods allowing the protection of the newly emerging video codec HEVC (High Efficiency Video Coding). Visual protection is achieved through selective encryption (SE) of HEVC-CABAC binstrings in a format compliant manner. The SE approach developed for HEVC is different from that of H.264/AVC in several aspects. Truncated rice code is introduced for binarization of quantized transform coefficients (QTCs) instead of truncated unary code. The encryption space (ES) of binstrings of truncated rice codes is not always dyadic and cannot be represented by an integer number of bits. Hence they cannot be concatenated together to create plaintext for the CFB (Cipher Feedback) mode of AES, which is a self-synchronizing stream cipher for so-called AES-CFB. Another challenge for SE in HEVC concerns the introduction of context, which is adaptive to QTC. This work presents a thorough investigation of HEVC-CABAC from an encryption standpoint. An algorithm is devised for conversion of non-dyadic ES to dyadic, which can be concatenated to form plaintext for AES-CFB. For selectively encrypted binstrings, the context of truncated rice code for binarization of future syntax elements is guaranteed to remain unchanged. Hence the encrypted bitstream is format-compliant and has exactly the same bit-rate. The proposed technique requires very little processing power and is ideal for playback on handheld devices. The proposed scheme is acceptable for DRM of a wide range of applications and diverse contents were used for experimental evaluations, since it protects the contour and motion information, along with the protection of the newly emerging video codec HEVC (High Efficiency Video Coding). Visual protection is achieved through selective encryption (SE) of HEVC-CABAC binstrings, as explained in detail in previous work of Shahid et al. [21], [22].

In comparison to SE-H.264/AVC, SE-HEVC poses two more challenges. The first is the non-dyadic encryption space (ES). For SE, ES of truncated rice binstrings in HEVC is not always dyadic. Hence we do not have an integer number of bits to prepare the plaintext. This challenge has been successfully addressed by conversion of non-dyadic ES to dyadic ES. The second challenge concerns the context modeling of truncated rice code. For truncated rice code, which is introduced for the first time in HEVC-CABAC, the context must remain unchanged during the encryption step for the sake of format compliance. In this paper, we have successfully addressed both of these challenges.

The rest of the paper is organized as follows. In Section II, overview of HEVC and its CABAC entropy engine is presented. This is followed by a description of recent work on selective encryption in Section III. We explain the proposed algorithm in Section IV. Section V contains an experimental evaluation and security analysis. There is also a comparison of the proposed technique with recent techniques. In Section VI, we present the concluding remarks about the proposed scheme.

I. INTRODUCTION

HEVC [8] (High Efficiency Video Coding) is the emerging video coding standard of ITU-T and ISO/IEC. HEVC achieves visual quality similar to its previous H.264/AVC High Profile, with around 30% bit-rate reduction for the low delay mode (I frame followed by a number of P frames) [25]. This mode is suitable for real-time applications like video-conferencing and on-line gaming. On the other hand, random access mode contains I, P and B frames and any part of the frame can be accessed randomly. This mode offers better compression and is suitable for storage applications and on-line movies like Netflix and Hulu. For this mode, HEVC offers around 20% bit-rate reduction on average, but with lower complexity than the H.264/AVC Baseline Profile [25].

With the inundation of digital content which can be copied and modified easily, concerns about the protection of digital content has been raised and it is relevant to analyze HEVC regarding its protection and authentication. Selective encryption (SE) is used to restrict access of video content to only authenticated users, wherein a small part of the bit-stream is encrypted with minimal resource overhead and sufficient protection is provided for most applications. In this work, we present a selective encryption method for CABAC binstrings of HEVC.

Arithmetic coding is very sensitive to change in a single bit and affects the format compliance of the whole bitstream. Format compliant SE is performed on CABAC-binstrings (instead of CABAC bitstreams), as explained in detail in previous work of Shahid et al. [21], [22].

II. HEVC AND ITS CABAC ENTROPY ENGINE

HEVC has better compression efficiency as compared to H.264/AVC due to more sophisticated compression tools. One of the main ways to improve the compression rate of high-resolution videos HD is to introduce larger block structures with flexible sub-partitioning [28]. In HEVC, MB is replaced by a coding unit (CU). CU divides a video frame into a number of rectangular regions. A CU contains one or several prediction
units (PUs) and transform units (TUs). The basic partition geometry of all of these elements is encoded by using quad tree coding.

Unlike H.264/AVC, the motion vector in HEVC is predicted from four spatial neighbors (top, top-right, left, left-bottom) and one co-located MV. Consequently, an encoder has to keep MVs of a co-located reference picture in order to find a correct MV predictor. This makes the design more complex. While in H.264/AVC, only in temporal skip mode, an encoder keeps co-located MVs. In spatial skip mode, the decoder keeps only a flag indicating whether the co-located MV is bigger or not. In I frame, spatial prediction is performed from samples of already decoded adjacent PUs, where the different modes are DC (flat average), horizontal, vertical, plane (amplitude surface), bi-linear, or one of up to 28 angular directions (number depending on the block size). In [13], Lainema and Unger proposed a directional intra prediction method for HEVC. They demonstrated that the proposed method outperforms the H.264/AVC intra prediction approach on average by 4.8%. While the coding efficiency gains become more significant and exceed 10% for sequences with dominant directional structures. In the entropy coding stage, HEVC introduced the concept of entropy slices. Entropy slices do not depend on information outside of the entropy slice and can be decoded independently. This enables parallelization of the entire entropy decoding loop, including context adaptation and bin coding.

Context adaptive binary arithmetic coding (CABAC) is the only entropy coding technique, which is supported in HEVC. HEVC-CABAC consists of three stages: 1) binarization, 2) context modeling, 3) binary arithmetic coding (BAC). In the binarization step, non-binary syntax elements are converted to binary form in so-called binstrings which are more amenable to compression by BAC. Binary representation for a non-binary syntax element is done in such a way that it is close to the minimum redundancy code. An individual bit in a binstring is called a bin. The bins are coded using either regular-BAC or bypass-BAC. In regular-BAC mode, bins are passed to the context modeling step followed by BAC. On the other hand, BAC is performed using a fixed context in bypass-BAC mode.

In HEVC-CABAC, there are five basic code trees for binarization, namely the unary code, truncated unary code, truncated rice code with context $p$ (TRp), $k$th order Exp-Golomb code (EGk) and fixed length code.

1) For an unsigned integer value $x \geq 0$, the unary code consists of $x$ 1’s plus a terminating 0 bit.
2) The truncated unary code (TU) is only defined for $x$ with $0 \leq x \leq \alpha$. For $x < \alpha$, the code is given by the unary code, whereas for $x = \alpha$ the terminating ‘0’ bit is neglected.
3) The truncated rice code with context $p$ (TRp) is introduced in HEVC for the first time. A TRp binarization is a concatenation of quotient ($q$) and remainder ($r$) for a context $p$. For an unsigned integer value $x \geq 0$, the quotient $q$ is given by $q = \lfloor x/p \rfloor$ and the remainder $r$ is given by $r = x - qp$. For $p = 0$, the TR0 binarization is exactly the TU binarization.
4) The EGk code is also a concatenation of prefix and suffix parts. For a given unsigned integer value $x > 0$, the prefix part of the EGk binstring consists of a unary code corresponding to the length $l(x) = \lfloor \log_2(x/2k+1) \rfloor$. The EGk suffix part is computed as the binary representation of $x + 2^k \{1 - 2^l(x)\}$ using $k + l(x)$ significant bits. Consequently, for EGk binarization, the code length is $2l(x) + k + 1$. When $k = 0$, the code length is $2l(x) + 1$.
5) The fixed length code is applied to syntax elements with a nearly uniform distribution or to syntax elements, for which each bit in the fixed length code binstring represents a specific coding decision e.g., coded block flag.

In HEVC-CABAC, the QTC syntax element is binarized by concatenation of the basic code trees. Binarization of QTC is done by REG0 (concatenation of TRp and EG0). It differs from H.264/AVC wherein QTC is binarized using UEG0 (concatenation of TU and EG0). For QTCs, binarization and BAC steps are applied to the syntax element $coe_f \_abs\_value\_minus\_3 = \_abs\_level - 3$, since QTCs of zero magnitude are encoded using a significant map. For motion vector differences (MVD), binstrings are constructed by EG1.

### III. Recent Work on Selective Encryption (SE) of Video Content

With the increase in digital video content, SE has attracted the attention of the research community for protection of copyright content. Since video content is huge in size and stored in compressed form, many researchers have proposed to perform encryption during the different compression steps, e.g., pixel domain, transform domain, quantized transform domain or bit-stream domain.

In [9], Jiang et al. propose to encrypt all intra prediction modes (IPMs) by chaotic pseudo-random sequence. It is followed by their scrambling by circulating sequences controlled by keys and gives a key distribution and synchronization scheme. The proposed scheme presents a good level of security but with a slight change in bit-rate. In [17], SE of H.264/AVC is carried out in some fields like intra-prediction mode, residual data, inter-prediction mode and motion vector difference (MVD). A scheme for commutative encryption and watermarking of H.264/AVC is presented in [16]. Here SE of some MB header fields is combined with watermarking DCT coefficient magnitude. This scheme presents a watermarking solution in an encrypted domain without exposing the video content. The drawback of the techniques proposed in [16], [17] is that they are not format compliant. Wang et al. [26] presented the partial encryption scheme on the codewords of $4 \times 4$ and $16 \times 16$ intra prediction mode (IPM), EGk code for MVD and level_suffix by using an RC4 stream cipher. Li et al. [14] devised a selective encryption technique for H.264/SVC on both entropy coders. The scheme encrypts IPM with signs of textures for base layers by using the stream cipher Leak Extraction (LEX) algorithm. Park and Shin [19] proposed a partial encryption of H.264/SVC scheme where the IPM, MVD and texture sign bits are encrypted. However, the IPM encryption proposed in [19], [26] affect the compression efficiency by negatively changing the video statistics.

Yeung et al. proposed perceptual video encryption at the transform stage by selecting one out of multiple unitary transforms [31]. The unitary transforms were significantly different from the discrete cosine transform (DCT) or discrete sine transform (DST), and the resulting coding efficiency is very close to
DCT. In [32], Yeung et al. extended their SE based on a unitary transform to the transforms of size 8 x 8 for high profiles of H.264 and MPEG4. The main drawback of a transform-coding based scheme is that it requires modification in the transform module of codec, which is very unlikely for hardware codec chips and even for DSP codecs. Moreover, it is a challenge to keep all transforms in the instruction cache, especially for embedded devices. Carrillo et al. proposed a partial encryption technique for H.264/AVC wherein they do permutations of pixels of MBs which are in the region of interest (ROI) [4]. The drawback of this scheme is that the bit-rate increases as the ROI size increases. This is due to a change in the ROI statistics as it is no longer a slow varying region, which is the basic assumption for video signals.

Huffman table based selective encryption has been studied in the literature in [30]. It encrypts by using different Huffman tables for different input symbols. The tables, as well as the order in which they are used, are kept secret. This technique is vulnerable to known plaintext attacks as explained in [10]. Key-based interval splitting of arithmetic coding (KSAC) has used an approach [11] wherein intervals are partitioned in each arithmetic coding iteration. A secret key is used to decide how the interval will be partitioned. The number of subintervals in which an interval is divided should be kept small as it increases the bitstream bit-rate. Randomized arithmetic coding [7] is aimed at arithmetic coding but instead of partitioning of intervals like in KSAC, a secret key is used to scramble the order of intervals. The drawback of these techniques is that the encrypted bitstream is not format compliant. Moreover, this resulted in a bit-rate overhead because of the change in the context model. In our previous work [21], [22], the novel idea of performing SE on CABAC binstrings instead of bitstreams for the sake of format compliance was presented. In this work, it is explained that format compliance can only be achieved if: 1) binstrings are selectively encrypted instead of bitstreams, 2) binstrings which we want to encrypt must be using bypass-BAC wherein a fixed context is used. Recently, Asghar et al. also presented an approach for the protection of scalable H.264/AVC while encrypting CABAC binstrings [2].

In this paper, we present a technique for HEVC content protection by selective encryption of HEVC-CABAC binstrings, while fulfilling real-time constraints, by transforming non-dyadic ES to dyadic and by extending SE to the header information. In Section IV, we describe our proposed approach to simultaneously apply format compliant SE and HEVC compression on video sequences in real time.

IV. THE PROPOSED SELECTIVE ENCRYPTION SCHEMES

Selectively encrypted HEVC bitstream will be format compliant and will fulfill real-time constraints provided the following conditions are fulfilled:

- **Same bit-rate:** the encrypted binstring must have the same length as the original binstring.
- **Format compliance:** the encrypted binstring must be valid and decodable by an entropy decoder.
- **Dyadic encryption space:** ES is defined as the number of valid values for a binstring. The candidate binstrings should have a dyadic ES, which can be represented by an integer number of bits. This is mandatory for real-time selective encryption using AES-CFB, wherein a number of binstrings are concatenated to prepare a plaintext for AES-CFB.

Format compliant SE is performed on a subset of binstrings, which fulfills the real-time constraints as stated above. Binstrings are then coded by BAC as shown in Fig. 1.

Among all five binarization techniques, the truncated rice code with context p (TRp) and kth order Exp-Golomb (EGk) code meet the conditions of SE, as shown in Fig. 1. The unary and truncated unary codes have different code lengths for each input value. They do not fulfill the first condition and their encryption will change the bitstream bit-rate. In fixed length code,
different bits indicate different information regarding the header and is not viable for format compliant SE. Suffixes of EGk and of TRp can be encrypted while keeping the bit-rate unchanged. TRp and EG0 are used for binarization of QTC. For MVD, EG1 binarization is used. Many of EG1 binstrings have the same length and hence the first and second conditions are fulfilled. Since MVDs are part of the CU (coding unit) header and are used for prediction of future motion vectors, it should be guaranteed that the encrypted motion vector will lie in the valid range.

A. Encryption Space in an HEVC Bitstream

The encryption space for HEVC is constituted by the sign bit of QTCs, suffix of TRp code, suffix of EG0 code, sign of MVDs and suffix of EG1 code. Context p must remain unchanged for both the TRp code and BAC step. This is in contrast to SE-H.264/AVC wherein BAC was the only component using context modeling. From the CABAC entropy engine standpoint, we can encrypt only binstrings which use bypass-BAC mode and fixed contexts, as highlighted in Fig. 1. Contexts for binstrings coded by regular-BAC are adaptive and their encryption makes the bitstream non-format compliant because of context mismatches on the encoder and decoder side. All binstrings which are coded by bypass-BAC are not encryptable. Rather, only binstrings which guarantee same bit-rate and same TRp context can be encrypted.

For binarization of QTCs, UEG0 code is used in H.264/AVC. It is replaced by REG0 (concatenation of TRp and EG0) in HEVC [18]. The main reason for using TRp code is to increase the number of bypass bits by coding QTCs up to a specific maximum value with TRp. Moreover, EGk code is optimally fit for distribution of H.264/AVC residuals, which is geometrical. While the distribution of HEVC residuals is such that they are better compressed by TRp codes.

The QTCs are denoted by levels in this section. Binarization of levels is performed using only TRp code up to a threshold, as shown in Fig. 2. The threshold depends on context p ([8, Table 9.35]) and is given by [12]:

\[ TR_{\text{limit}}[p] = \{7, 14, 26, 46, 78\}. \] (1)

1) Scenario I: \( \text{level} \leq TR_{\text{limit}}[p] \): If \( \text{level} \) is smaller than \( TR_{\text{limit}}[p] \) for context \( p \), it will be binarized using only TRp code, whose suffix will be encrypted as shown in Fig. 2. Context \( p \) for TRp code can increment by 1 at a time and is adaptive to \( \text{level} \) in the following manner:

**Algorithm 1** Selection of context \( p \) is adaptive to \( \text{level} \).

1: if \( \text{level} > 3 \times (1 \leq p_{\text{curr}}) \) then
2: \( p_{\text{next}} = \min(p_{\text{curr}} + 1, 4) \)
3: end if

For example, if \( \text{level} = 2 \) and the current context \( p_{\text{curr}} = 0 \), then \( p_{\text{next}} \) will remain unchanged. If the level is modified from 2 to 3 during the encryption step, then \( p_{\text{next}} \) will be 1. Context \( p_{\text{next}} \) must remain unchanged during encryption of QTCS otherwise the encrypted bitstream will not be format compliant. In a TRp code with context \( p \), we have \( p \) encryptable bits except for two limitations:

- \( TR0 \) (truncated rice with context 0) code is the same as truncated unary code and each binstring of \( TR0 \) has a different length and is not encryptable because of violation of the bit-rate constraint.
- The length of the last equal-length group of TRp codes is the same whether the EG0 code is appended or not. So if we encrypt the binstrings of this group, it may make the bitstream non-compliant. To fulfill this constraint, the last binstring is excluded from the set of equal-length binstrings, thus making the encryption space non-dyadic.

The selection of encryptable bits of the TRp suffix is illustrated in Algorithm 2. Note that the encryption space for a binstring is selected in such a way that it does not affect context \( p \) and its length.

**Algorithm 2** Encryption space in a TRp suffix which fulfills format compliant constraints for selective encryption of HEVC.

Require: \( p > 0 \)

Require: \( \text{level} \leq TR_{\text{limit}}[p_{\text{curr}}] \)

1: \( TR_{\text{limit}}[p_{\text{curr}}] = \{7, 14, 26, 46, 78\} \)
2: \( \text{limit}_r = \text{para}[p_{\text{curr}}] = \{3, 5, 12, 24, \infty\} \)
3: $default_ES \leftarrow 2^p$

4: if $|level| \leq limit_{rice \_para}[p_{curr}]$

5: $upper_{bound} = limit_{rice \_para}[p_{curr}]$

6: $lower_{bound} = (prefix \ll p)$

7: else

8: $upper_{bound} = TR_{limit}[p_{curr}]$

9: $lower_{bound} = limit_{rice \_para}[p_{curr}] + 1$

10: end if

11: if $(prefix \ll p) + default_{ES} \leq upper_{bound}$ then

12: $upper_{bound} = (prefix \ll p) + default_{ES} - 1$

13: end if

14: if $(prefix \ll p) > lower_{bound}$ then

15: $lower_{bound} = (prefix \ll p)$

16: end if

17: $ES \leftarrow ((upper_{bound} - lower_{bound}) + 1)$

2) Scenario II: $|level| > TR_{limit}[p]$: If $|level|$ is higher than a $TR_{limit}[p_{curr}]$ limit, EG0 code is used for the binarization of $|level| - TR_{limit}[p_{curr}]$. In this case, the TRp code is fixed (all bits are ‘1’) and is not encryptable. The suffix of EG0 will be encrypted in this case, while guaranteeing the same binstring length and same future context for the TRp code, as shown in Fig. 2.

The sign and suffix of the MVD syntax element are also candidates for SE. HEVC uses EG1 code for binarization of MVDs. Encryption of MVD suffixes is crucial for protection of contours and motion information but in this case we cannot guarantee the format compliance.

B. Conversion of Non-Dyadic ES to Dyadic ES

To prepare plaintext for AES-CFB, we require suffixies having dyadic (i.e., power of 2, which can be constituted by integer number of bits) ES. In contrast to H.264/AVC, wherein ES was composed of EG0 suffixies and was always dyadic, it is a common scenario with TRp suffixies in HEVC to have non-dyadic ES. This is due to the fact that the threshold for the context $p$ of TRp binary codes lies at non-dyadic boundaries, as given in Algorithm 1. Real-time SE-HEVC using AES-CFB is only possible if non-dyadic ES is transformed to dyadic ES. This is achieved by decomposition of non-dyadic ES into small dyadic ES as explained in the following steps:

1) Let $L$ be the non-dyadic ES which is to be decomposed into $n$ smaller dyadic ES $ES_{dyadic_i}[n]$.

2) The first dyadic ES $ES_{dyadic_i}[0]$ consists of the first $2^k$ suffixies ($k \leq L$).

3) This process is repeated on the remaining ES $(L - 2^k)$ recursively to decompose $L$ into dyadic ES.

4) Dyadic ES $ES_{dyadic_i}[j]$ that contains the suffix to be encrypted, is the available dyadic ES for that suffix.

For example, a non-dyadic ES with $ES = 14$ will be decomposed into three dyadic ES with $ES = 8, 4$ and 2 respectively. If the suffix is on index 5, its ES will be the first dyadic ES $\{0, 1, 2, \ldots, 7\}$ with $ES = 8$. Similarly if the suffix to be encrypted has an index value of 9 or 13, their ESs will be $\{8, 9, 10, 11\}$ and $\{12, 13\}$, with $ES = 4$ and $ES = 2$ respectively. It is important to note that the last dyadic ES $ES_{dyadic_i}[n - 1]$ may have $ES = 1$. In that case, the suffix on the final index in the ES will not be encrypted. For example, for $ES = 15$, the dyadic ES will be of sizes 8, 4, 2 and 1. In this case, the last suffix cannot be encrypted because it has dyadic ES with $ES = 1$.

Fig. 3 illustrates the preparation of plaintext for the proposed method for SE of HEVC. For QTCs, the sign and suffix of either the TRp or EG0 binstring is a candidate for SE. While for motion information, the sign and suffix of EG1 binstrings are encrypted.

C. Selective Encryption of HEVC-CABAC Binstrings

HEVC has introduced the concept of entropy slices. Context models are reset at the start of every entropy slice. Moreover, entropy slices restrict the neighborhood definition. SE of HEVC is performed on each entropy slice independently. Let us consider $Y_i = X_i \oplus E_{q}[Y_i - 1]$ as the notation for the encryption of a $n$ bit block $X_i$, using the secret key $k$ with the AES-CFB. We have chosen to use this mode in order to keep the original compression rate. Indeed, with the CFB mode for each block, the size of the encrypted data $Y_i$ can be exactly the same as the size of the plaintext $X_i$. In this mode, the code from the previously encrypted block is used to encrypt the current one.

SE is performed on binstrings before compression by BAC. Non-binary syntax elements are transformed into binstrings through binarization and at the same time we fill $X_i$ with encryptable bits until either the vector $X_i$ is completely filled or the slice boundary is reached. Let $I(X_i)$ be the length up to which vector $X_i$ is filled. In case of a slice boundary, if $I(X_i) < C$, we apply a padding function $p(j) = 0$, where $j \in \{I(X_i) + 1, \ldots, C\}$, to fill in vector $X_i$ with zeros up to $C$ bits. Note that the main difference with respect to H.264/HVC is that we have to prepare the plaintext while getting encryptable bits from one entropy slice and we have to pad it with zeros if the end of the entropy slice is reached.

In AES-CFB, the previously encrypted block $Y_{i-1}$ is used as the input of the AES algorithm in order to create $Z_i$. Then the current plaintext $X_i$ XORed with $Z_i$ in order to generate
Moreover, generation of the encrypted stream $Z_i$ depends on the previous encrypted block $Y_{i-1}$. Consequently, if two plaintexts are identical $X_i = X_j$ in CFB mode, then the two corresponding encrypted blocks are always different, $Y_i \neq Y_j$.

The decryption process for AES-CFB works in the same fashion except that the input is encrypted and the output will be the plaintext. The decoded plaintext vector is split into suffixes in order to substitute the encrypted binstrings with the original ones. The bitstream is then further decoded to get the decrypted video content using standard HEVC decode steps.

V. EXPERIMENTAL RESULTS

In this section, different aspects of the proposed schemes for SE-HEVC are analyzed. For simulation purposes, reference implementation of HEVC HM 8.0\textsuperscript{1} was used. The set of benchmark video sequences along with their resolutions is given in Table I. This is the same set of sequences that is being used in the HEVC standardization process \cite{HEVC}.

We simultaneously applied our proposed SE scheme and HEVC compression as described in Section IV, on all the benchmark video sequences along with their resolutions is given in Table I. This is the same set of sequences that is being used in the HEVC standardization process \cite{HEVC}.

We simultaneously applied our proposed SE scheme and HEVC compression as described in Section IV, on all the benchmark video sequences along with their resolutions is given in Table I. This is the same set of sequences that is being used in the HEVC standardization process \cite{HEVC}.

Table II provides an analysis of the effect of QP on the ES for the kimono video sequence. Note that the ES slightly decreases with an increase in the QP value. In a video bitstream, the data part decreases with an increase in QP, while the proportion of video header (NAL header, slice header, CU/PU/TU headers etc.) increases with an increase in the QP value. For variations in the QP range from 18 to 42, the change in ES for SE-HEVC in low delay mode is from 19.93\% to 16.56\%, as shown in Table III. This is in contrast to the SE-CABAC of H.264/AVC wherein ES decreases from 19.97\% to 9.46\% (\cite{Shahid2015, Shahid2016}), for the same QP range. This is because of the replacement of truncated unary (TU) coding with truncated rice (TR) coding for QTC binarization.

Tables II and III also provide a comparison of low delay (I, P frames) and random access (I, P, B frames) modes for PSNR and ES. Generally, the PSNR of B frames is slightly less as compared to that of P frames, which results in an overall decrease in PSNR for random access mode. ES in low delay mode is also slightly higher than in random access mode. Since random access mode is more generic and contains I, P and B frames, the results will be presented only for random access mode from hereon for the sake of brevity.

A. Encryption Space for SE-HEVC

Encryption space (ES) is defined as the percentage of encryptable bits in a video bitstream. It varies from one sequence to another based on the video content. In Table II, ES for different benchmark video sequences is presented for the QP value 18. Note that ES varies from 16.96\% to 20.08\% based on the video content. Video sequences with either static background or translational movement (e.g., Vidyo1) have less ES. BQMall also has less ES despite its complex background. This is due to the fact that the background is static and coded by skip blocks and translational motion in the foreground is very efficiently predicted. On the other hand, video sequences which have complex motion and moving background have high ES. For example, RaceHorses contains walking horses, camera movement and high-texture grass in the background and has higher ES. The ES for EG1 suffixes of MVD syntax elements depends on the complexity of the motion in the video sequence. For a simple translational motion, which can be easily estimated, ES for EG1 suffixes of an MVD syntax element will be less, while it will be higher for video sequences with complex motion.

B. Visual Protection of SE-HEVC for I, P and B Frames

To demonstrate the visual protection offered by our proposed scheme, we compressed 50 frames in random access mode with 2 B frames between consecutive P frames. Video frame areas that contain many details and texture will have lot of non-zero QTCs and consequently will be strongly encrypted. On the other hand, homogeneous areas in a video frame, i.e., areas containing series of identical pixels, are less ciphered. The SE of MVD syntax elements is helpful for the protection of motion information in a video sequence. Table IV compares the average PSNR of all benchmark video sequences without encryption and with

\begin{table}[h]
\centering
\caption{The set of benchmark video sequences used to evaluate the performance of the proposed SE-HEVC technique \cite{HEVC}.
\label{tab:benchmark_sequences}
\begin{tabular}{|c|c|c|c|}
\hline
Class & Resolution & Frame rate & Videos \\
\hline
A & 2560×1600 & 30 & Traffic (S01), PeopleOnStreet (S02) \\
B1 & 1920×1080 & 24 & Kimono (S03), ParkScene (S04) \\
B2 & 1920×1080 & 50-60 & Cactus (S05), BasketballDrive (S06), BQTerrace (S07) \\
C & 832×480 & 30-60 & BasketballDrill (S08), BQMall (S09), PartyScene (S10), RaceHorses (S11) \\
D & 416×240 & 30-60 & BasketballPass (S12), BQSquare (S13), BlowingBubbles (S14), RaceHorses (S15) \\
E & 1280×720 & 60 & Vidyo1 (S16), Vidyo2 (S17), Vidyo3 (S18) \\
\hline
\end{tabular}
\end{table}
TABLE II
ES ANALYSIS OF SE-HEVC FOR BENCHMARK VIDEO SEQUENCES AT A QP VALUE OF 18 FOR LOW DELAY AND RANDOM ACCESS MODES. ES IS LESS FOR SEQUENCES WITH EITHER STATIC BACKGROUND OR TRANSLATIONAL MOTION, WHILE IT IS HIGHER FOR COMPLEX VIDEO CONTENT.

<table>
<thead>
<tr>
<th>Class</th>
<th>Sequence</th>
<th>low delay</th>
<th>random access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR(Y)</td>
<td>ES</td>
<td>ES-Class (%)</td>
</tr>
<tr>
<td>A</td>
<td>Traffic</td>
<td>44.08</td>
<td>18.03</td>
</tr>
<tr>
<td></td>
<td>PeopleOnStreet</td>
<td>43.89</td>
<td>18.91</td>
</tr>
<tr>
<td>B1</td>
<td>ParkScene</td>
<td>42.38</td>
<td>18.36</td>
</tr>
<tr>
<td></td>
<td>Kimono</td>
<td>43.37</td>
<td>18.5</td>
</tr>
<tr>
<td>B2</td>
<td>BasketballDrive</td>
<td>42.63</td>
<td>18.22</td>
</tr>
<tr>
<td></td>
<td>BQTerrace</td>
<td>43.4</td>
<td>19.95</td>
</tr>
<tr>
<td></td>
<td>Cactus</td>
<td>41.92</td>
<td>19.6</td>
</tr>
<tr>
<td>C</td>
<td>BasketballDrill</td>
<td>43.6</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>BQMall</td>
<td>42.22</td>
<td>17.38</td>
</tr>
<tr>
<td></td>
<td>PartyScene</td>
<td>43.05</td>
<td>19.32</td>
</tr>
<tr>
<td></td>
<td>RaceHorses</td>
<td>43.15</td>
<td>20.08</td>
</tr>
<tr>
<td>D</td>
<td>BasketballPass</td>
<td>44.65</td>
<td>17.88</td>
</tr>
<tr>
<td></td>
<td>BQSquare</td>
<td>42.41</td>
<td>18.94</td>
</tr>
<tr>
<td></td>
<td>BlowingBubbles</td>
<td>41.91</td>
<td>18.44</td>
</tr>
<tr>
<td></td>
<td>RaceHorses</td>
<td>42.94</td>
<td>18.66</td>
</tr>
<tr>
<td>E</td>
<td>Vidyo1</td>
<td>45.34</td>
<td>16.96</td>
</tr>
<tr>
<td></td>
<td>Vidyo3</td>
<td>44.86</td>
<td>18.1</td>
</tr>
<tr>
<td></td>
<td>Vidyo4</td>
<td>45.35</td>
<td>17.57</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>43.39</td>
<td>18.45</td>
</tr>
</tbody>
</table>

TABLE III
ES ANALYSIS OF THE PROPOSED SCHEME OVER THE WHOLE RANGE OF QP VALUES FOR THE KIMONO VIDEO SEQUENCE FOR LOW DELAY AND RANDOM ACCESS MODE.

<table>
<thead>
<tr>
<th>QP</th>
<th>low delay ES (%)</th>
<th>random access ES (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>19.93</td>
<td>19.81</td>
</tr>
<tr>
<td>18</td>
<td>18.50</td>
<td>18.81</td>
</tr>
<tr>
<td>24</td>
<td>18.34</td>
<td>18.75</td>
</tr>
<tr>
<td>30</td>
<td>18.08</td>
<td>18.43</td>
</tr>
<tr>
<td>36</td>
<td>17.77</td>
<td>17.89</td>
</tr>
<tr>
<td>42</td>
<td>16.56</td>
<td>17.05</td>
</tr>
</tbody>
</table>

For an analysis of visual protection by SE-HEVC over the whole range of QP values, Table V compares the average PSNR of kimonos over the whole range of QP values without encryption and with SE-HEVC. Non-zero QTCs decrease with an increase in QP and the contribution of SE-MVD in the overall protection increases with an increase in the QP value. Note that PSNR values of luma and chroma components remain in the lower range for all QP values. Fig. 5 shows original as well as SE video frame #8 of kimonos for QP values of 18, 30 and 42. It is evident that the encrypted video frame is quite secure over all QP values in terms of texture, motion, color and luminance.

C. Computational Overhead

Computational overhead is an important factor especially for smartphones and hand-held devices which have limited processing power. It can be calculated through an analysis of the additional time required for encoding and decoding encrypted bitstreams. The computers used for this simulation had Intel Core 2 Duo T8100 processor with 3072 MB random access memory. Table VI shows the encoding and decoding time for the kimonos sequence with an intra period of 10. The additional processing time for encoding (column 4) and decoding (column 7) is negligible as compared to the overall processing time, thus confirming the processing efficiency of the proposed scheme.

D. Security Analysis

This section presents a security analysis of the proposed SE-HEVC scheme in various aspects. The analysis includes a
TABLE IV

<table>
<thead>
<tr>
<th>Class</th>
<th>Sequence</th>
<th>QP 18</th>
<th>QP 32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR (Y) (db)</td>
<td>PSNR (U) (db)</td>
<td>PSNR (V) (db)</td>
</tr>
<tr>
<td></td>
<td>Orig.</td>
<td>SE</td>
<td>Orig.</td>
</tr>
<tr>
<td>A</td>
<td>Traffic</td>
<td>44.08</td>
<td>7.99</td>
</tr>
<tr>
<td></td>
<td>PeopleOnStreet</td>
<td>43.88</td>
<td>8.23</td>
</tr>
<tr>
<td>B1</td>
<td>ParkScene</td>
<td>42.38</td>
<td>10.25</td>
</tr>
<tr>
<td></td>
<td>Kimono</td>
<td>43.37</td>
<td>9.55</td>
</tr>
<tr>
<td></td>
<td>BasketballDrive</td>
<td>42.67</td>
<td>8.95</td>
</tr>
<tr>
<td></td>
<td>BQTerrace</td>
<td>43.42</td>
<td>7.89</td>
</tr>
<tr>
<td></td>
<td>Cactus</td>
<td>41.93</td>
<td>7.38</td>
</tr>
<tr>
<td>C</td>
<td>BasketballDrill</td>
<td>43.61</td>
<td>11.35</td>
</tr>
<tr>
<td></td>
<td>BMall</td>
<td>42.21</td>
<td>10.86</td>
</tr>
<tr>
<td></td>
<td>RaceHorses</td>
<td>43.17</td>
<td>10.64</td>
</tr>
<tr>
<td>D</td>
<td>BasketballPass</td>
<td>44.66</td>
<td>13.76</td>
</tr>
<tr>
<td></td>
<td>BSquare</td>
<td>42.39</td>
<td>6.72</td>
</tr>
<tr>
<td></td>
<td>BlowingBubbles</td>
<td>41.92</td>
<td>12.15</td>
</tr>
<tr>
<td></td>
<td>RaceHorses</td>
<td>42.94</td>
<td>8.75</td>
</tr>
<tr>
<td>E</td>
<td>Vido1</td>
<td>45.35</td>
<td>10.32</td>
</tr>
<tr>
<td></td>
<td>Vido3</td>
<td>44.86</td>
<td>9.42</td>
</tr>
<tr>
<td></td>
<td>Vido4</td>
<td>45.35</td>
<td>8.98</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>43.40</td>
<td>9.67</td>
</tr>
</tbody>
</table>

Fig. 4. Frames #0, #3, #8 (I, P, and B respectively) of BasketballDrive sequence for SE-HEVC for QP value 18. (a) Original #0. (b) Original #3. (c) Original #8. (d) SE-HEVC #0. (e) SE-HEVC #3. (f) SE-HEVC #8.

histogram comparison of original and encrypted video frames in Section V-D.1, encryption quality analysis in Section V-D.2, edge and structural protection in Section V-D.3, entropy and standard deviation analysis in Section V-D.4, key sensitivity analysis in Section V-D.5, and protection against known plaintext attack in Section V-D.6.
1) Histogram Analysis: A histogram of a video frame gives the frequency distribution of the intensity levels. For a good encryption system, histograms of original and encrypted video frames should differ from each other. A histogram of the original *kimono* #8 video frame is shown in Fig. 6(a), while histograms of the encrypted version of the same video frame with different bit-shifts are presented in Fig. 6(b)–(d). It is evident from Fig. 6(a)–(d) that the histograms of original and encrypted frames are entirely different.

2) Encryption Quality Analysis: The encryption quality (EQ) represents the average number of changes to each gray level \( L \) and the selective encrypted video frame (cipherframe) of a video sequence of resolution \( m \times n \) with \( L \) intensity levels. \( V(x, y) \), \( V'(x, y) \in \{0, \ldots, L - 1\} \) are the gray levels of video frames \( V \) and \( V' \) at position \((x, y)\) \((0 \leq x < m - 1, 0 \leq y < n - 1)\). Let \( H_L(V) \) denote the number of occurrences of each grey level \( L \) in the plainframe \( V \). Similarly, \( H_L(V') \) denotes the number of occurrences of each grey level \( L \) in cipherframe \( V' \). The EQ can be defined as:

\[
EQ = \frac{\sum_{L=0}^{255} H_L(V') - H_L(V)}{256}.  \tag{2}
\]
The encryption quality for frame #8 of *kimono* is shown for a QP value of 18 in Table VII for a different number of shifts. Note that the encryption quality remains in the higher range for all the shifts.

3) Protection of Edges and Structural Information: A good visual encryption system aims to make the visual content incomprehensible by protection of edges and contour information. The degradation in encrypted video frame can effectively be evaluated by measuring the distortion introduced at the edges. The edge differential ratio (EDR) [24] depicts a deviation in the location of edge formation contributing pixels in the original image and its encrypted video frame. The mathematical representations of EDR can be expressed as:

$$ EDR = \frac{\sum_{i,j=1}^{N} |B(i,j) - \hat{B}(i,j)|}{\sum_{i,j=1}^{N} |B(i,j) + \hat{B}(i,j)|}, $$

where $B(i,j)$ and $\hat{B}(i,j)$ denote the pixel values in the edge detected binary version for the original and encrypted images, respectively. Fig. 7 shows the binary images after Laplace edge detection for the original and encrypted video frames. It is evident that the structural information is heavily distorted in the encrypted video frames. Higher EDR indicates (close to ‘1’) better protection of the structural information of a video frame, whereas a lower value (close to ‘0’) indicates similarity between the original and encrypted video frames. EDR for frame #0 of *Kimono, BasketballDrive* and *RaceHorses* is 0.97, 0.92 and 0.94 respectively, which confirms that the original and encrypted video frames are entirely different. It is important to note that EDR is performed for I frames. The values would be even higher if conducted on P or B frames because of encryption of suffixes of MVD syntax elements.

4) Analysis of Entropy and Local Standard Deviation: The security of the encrypted image can be measured by consid-
Fig. 8. Key sensitivity test for *kimono* #0. The SE frame is decrypted with: (a) Original key: PSNR (YUV) = \{45.78, 47.04, 48.88\} dB, (b) 1-bit different key: PSNR (YUV) = \{5.45, 19.11, 13.63\} dB.

Fig. 9. Known plaintext attack for *kimono* #0: (a) Encrypted frame: \(I^{PSNR (YUV)} = \{10.12, 20.65, 14.23\}\) dB, (b) Video frame with encrypted bits set to 0: PSNR (YUV) = \{11.12, 20.19, 14.69\} dB.

C. Key Sensitivity Test

To guarantee the security of a cryptosystem against brute-force attacks, the cryptosystem should be highly sensitive towards the key. So the ciphertext cannot be decrypted correctly although there is only a slight difference between the encryption or decryption keys. For this purpose, a key sensitivity test is assumed where we picked one key and then applied the proposed technique for encryption and then made a one bit change in the key and decoded the bitstream. The numerical results show that the proposed technique is highly sensitive towards the key change, that is, a different version of the encrypted video sequence is produced when the keys are changed, as shown in Fig. 8. PSNR of *luma* of decrypted frames with 1-bit different key is 9.45 dB, which lies in the same lower range as the encrypted video frames.

E. Comparative Analysis

The proposed algorithm is pioneer work on SE-HEVC and has successfully addressed the challenges posed by its entropy engine. In this section, a comparison of our proposed SE scheme with recent work is performed to verify the effectiveness of the proposed scheme. Table VIII presents a detailed comparison between the proposed SE-HEVC and recent work. Previous methods on selective encryption of video codecs have disadvantages over compression ratio, format compliance and/or security of implemented schemes by choosing weak encryption parameters and/or weak cipher algorithms. Only residual data and sign bit encryption in [22], [23], [27], [29] are considered to provide low level perception security, and the video can be recovered by brute force attacks.

Although, the pixel domain encryption proposed by Carrillo et al. [4] is robust to transcoding, it has the disadvantage of being independent of the video compression system and requires complete re-encoding. Yeung et al. [31] proposed encryption in the transform domain. This technique has the limitation of requiring complete re-encoding of the video bitstream and thus a lot of processing power. Moreover, this technique is not compliant with respect to the respective video standard and requires alternating transforms. IPM encryption is proposed by several researchers [9], [14], [19], [26]. It changes the video statistics in a negative manner. This is because absolute values are used and thus the change in any bit of the absolute value causes bit-rate fluctuation.

Along with its processing efficiency and format compliance, the proposed scheme has several distinct features. For example, ES is only slightly affected over the whole range ofQP values. When the QP varies from 18 to 42, ES of SE-HEVC varies from from 19.93% to 16.56% (Table III). While for H.264/AVC, ES decreases from 19.97% to 9.46% (Table II) over the same QP range. Moreover, SE-HEVC has addressed one more challenge successfully, which is conversion of non-dyadic ES to dyadic ones for encryptable suffixes. Non-dyadic ES does not allow formation of a plaintext for AES-CFB from encryptable binstrings.

The motion and contour information of the video content is better protected in the proposed scheme as compared to previous methods based on residual data and sign bit encryption. Fig. 10
shows the edge information in the original and encrypted frame \#0 of foreman sequence taken from [22]. It is evident that edges and contours are not as well protected as in Fig. 7 for SE-HEVC. Moreover, the EDR (edge differential ratio) for SE-CABAC of H.264/AVC is 0.8, which is much lower in comparison to 0.94 of SE-HEVC and verifies that the proposed scheme offers better protection by distortion of the structural information.

VI. CONCLUSION

The paper presents a pioneer scheme for format compliant visual protection of HEVC using selective encryption. It starts with an in-depth analysis of HEVC-CABAC from an encryption standpoint. It is followed by the proposed algorithm for SE of HEVC, which fulfills all the real-time constraints, including conversion of non-dyadic ES to dyadic ES. The SE is performed on the entropy slices independently in HEVC. In this way, the proposed SE method does not affect the parallelism of HEVC. Moreover, SE is performed in CABAC binstrings such that they remain valid binstrings thereafter having exactly the same length. The proposed method has the advantage of being suitable for streaming over heterogeneous networks because of no change in bit-rate. The experiments have shown that we can achieve the desired level of protection, both for texture information using SE of QTC syntax elements and motion information using SE of MVD syntax elements, under a minimal set of computational requirements.

REFERENCES


TABLE VIII

<table>
<thead>
<tr>
<th>SE Scheme</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeung et al. [31]</td>
<td>No</td>
<td>No</td>
<td>Transform</td>
<td>Yes</td>
<td>No</td>
<td>Permutation</td>
<td>No</td>
</tr>
<tr>
<td>Jiang et al. [9]</td>
<td>No</td>
<td>No</td>
<td>IPM</td>
<td>Yes</td>
<td>No</td>
<td>Chaotic based PRNG</td>
<td>No</td>
</tr>
<tr>
<td>Dufaux and Ebrahimi [6]</td>
<td>Yes</td>
<td>No</td>
<td>Transform</td>
<td>Yes</td>
<td>Yes</td>
<td>Pseudo random sign inversion</td>
<td>No</td>
</tr>
<tr>
<td>Li et al. [15]</td>
<td>No</td>
<td>No</td>
<td>NAL</td>
<td>No</td>
<td>No</td>
<td>Stream cipher</td>
<td>No</td>
</tr>
<tr>
<td>Lian et al. [16]</td>
<td>No</td>
<td>No</td>
<td>Transform</td>
<td>No</td>
<td>No</td>
<td>Stream cipher</td>
<td>No</td>
</tr>
<tr>
<td>Carrillo et al. [4]</td>
<td>Yes</td>
<td>Yes</td>
<td>Pixel</td>
<td>Yes</td>
<td>Yes</td>
<td>Pseudo random pixel permutations</td>
<td>No</td>
</tr>
<tr>
<td>Shahid et al. (SE-CABAC) [22]</td>
<td>Yes</td>
<td>No</td>
<td>Binstrings</td>
<td>No</td>
<td>No</td>
<td>AES-CFB</td>
<td>No</td>
</tr>
</tbody>
</table>

Proposed scheme | Yes | No  | Binstrings | No  | No  | AES-CFB | Yes  |


