Energy-Aware Encryption for Securing Video Transmission in Internet of Multimedia Things

Karthik Thiyagarajan, *Member, IEEE*, Rongxing Lu, *Senior Member, IEEE*, Kamal El-Sankary, *Member, IEEE*, and Hui Zhu, *Member, IEEE*

***Abstract*— High Efficiency Video Coding (HEVC) encryption, which has been proposed to encrypt intra prediction modes (structural information), transform coefficients (texture information), and motion related codewords (motion information), has received considerable attention recently. However, there is still the issue of efficiency when HEVC encryption is applied in the Internet of Multimedia Things (IoMT). Aiming at this challenge, in this paper, we propose a new low-overhead HEVC encryption scheme for energy-constrained IoMT. Concretely, the proposed scheme adjusts the selection of the aforementioned syntax elements to be encrypted according to the structure, texture, and motion energy present in each frame. It works as follows. The energy levels of quantized coefficients and motion vectors are calculated and compared with adaptive threshold values to classify the energy level in each video frame. When there is a high energy frame in the video, all the syntax elements are encrypted. When there is a low energy frame, alternate syntax elements are encrypted for achieving low encryption overhead. Moreover, in the case of transform coefficients, to withstand the interpolation attack, alternate coefficients are encrypted after correlating the frame with its neighboring coefficients. Extensive experiments were conducted, and the results demonstrate that the proposed scheme efficiently reduces the encryption overhead with low impact on the security level, making it suitable for IoMT.**

***Index Terms*— HEVC, Internet of Multimedia Things (IoMT), multimedia security, low encryption overhead.**

I. Introduction

HEVC (High Efficiency Video Coding) [1] is the latest video coding standard, which is used for compressing video and can provide efficient features adapted to different applications from large scale TV distributors to small scale multimedia networks. Compared with its predecessor H.264/AVC (Advanced Video Coding), HEVC achieves a 40-50% bit rate reduction by employing enhanced CABAC features and a hybrid spatial-temporal prediction model. Nevertheless, due to the larger prediction units and expensive motion estimation, HEVC is computationally expensive [1].

Video encryption provides security to video systems that can range from digital rights management to highly confidential military applications. However, encrypting the entire video stream is not advisable, as the decoding parameters get randomized, and it will result in an unexpected behavior of the decoder. It is feasible to encrypt certain syntax elements in the HEVC video stream. Then format compliance with no bit rate overhead and lower computational complexity can be achieved. However, integrating encryption with compression has a major effect on the overall computational cost of the system, which results in it not being suitable for some energy constrained systems that require low computational overhead, e.g., the Internet of Multimedia Things (IoMT) [2]. Actually, in order to adapt the IoMT environment, the following video encryption features need to be taken into consideration: i) security: visual scrambling of video after encryption; ii) overhead: time complexity caused by encryption; iii) format compliance: the encrypted video should be decodable by a standard HEVC decoder; and iv) statistics size-preservation: the bit rate should be preserved in the encrypted bit stream.

Selective encryption for HEVC stream was exploited from the work of Lian *et al.* [3] and Wang *et al.* [4]. They proposed encrypting intraprediction modes, syntaxes from transform coefficients and motion information. Wallendael *et al.* [5] provided an extensive analysis on cipher-able elements in the HEVC stream. Their work provides an analysis of efficient visual scrambling at the cost of bit rate increase. Shahid and Puech [6] presented a method to encrypt binstrings in a format compliant manner. The code words chosen for encryption are quantized transform coefficients and motion vector information, which can ensure a constant bit rate. In addition, they also proposed to convert dyadic to non dyadic encryption space suitable to create an input plaintext for AES CFB mode. Boyadjis *et al.* [7] extended the technique by encrypting code words in the CABAC: regular mode. This allows to encrypt the intraprediction luma modes with a trade-off in bitrate overhead. Their work proposed encrypting syntax elements such as intraprediction modes, quantized transform coefficients and motion vectors.

Several low computational encryption schemes for H.264/AVC as an extension of Lian *et al.* [3] have been

Manuscript received October 19, 2017; revised January 15, 2018 and February 5, 2018; accepted February 7, 2018. Date of publication February 20, 2018; date of current version March 7, 2019. This work was supported in part by the Natural Sciences and Engineering Research Discovery under Grant Rgpin 04009, in part by NBIF Start-Up under Grant Rif 2017-012, and in part by the National Natural Science Foundation of China under Grant 61672411.This paper was recommended by Associate Editor L. Zhou. *(Corresponding author: Karthik Thiyagarajan.)*

K. Thiyagarajan is with the Canadian Nuclear Laboratories, Deep River, ON K0J 1J0, Canada (e-mail: karthik.thiyagarajan@cnl.ca).

R. Lu is with the Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada (e-mail: rlu1@unb.ca).

K. El-Sankary is with the Department of Electrical Engineering, Dalhousie University, Halifax, NS B3H 4R2, Canada (e-mail: kamal.el-sankary@dal.ca).

H. Zhu is with the School of Cyber Engineering, Xidian University, Xi'an 710071, China (e-mail: zhuhui@xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2018.2808174

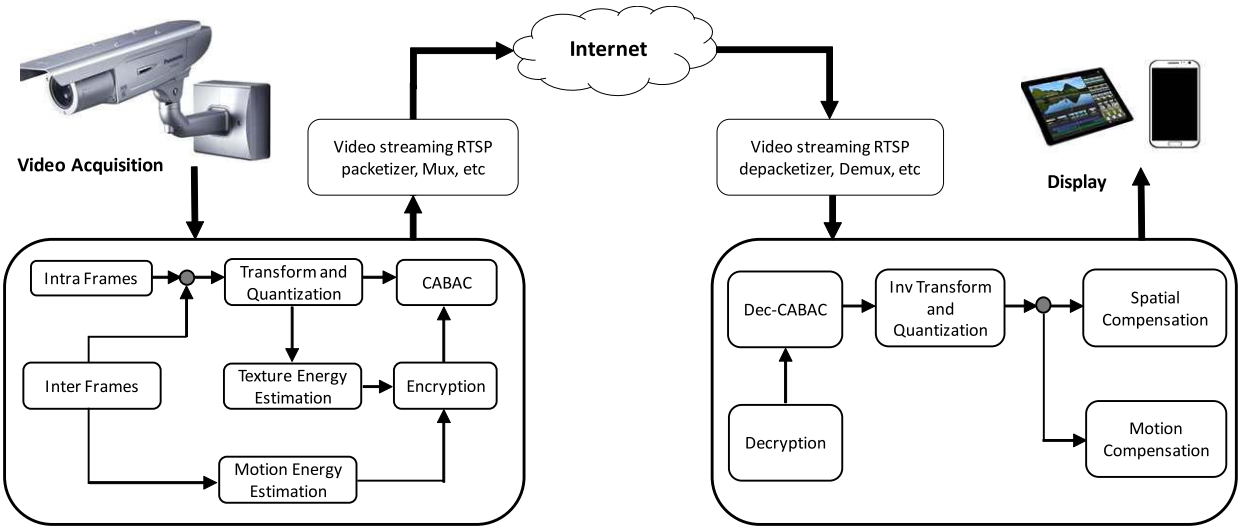


Fig. 1. System Model - Integrated Crypto Compression.

proposed in the literature. Wang *et al.* [8] proposed encrypting frames that were highly dependent on descendant frames. Wang's algorithm reduces the encryption overhead as the lower dependent frames are unencrypted. Zhao and Zhuo [9] proposed to chose syntax elements for encryption according to image statistical content in intra and inter frames. Only a small percentage of syntaxes are ciphered in frames with low level statistics, which provided low encryption overhead. Zhao *et al.* [10] also proposed an unequal encryption by parsing the background and foreground in video frames. Though the computational cost is reduced, the algorithm is only applicable in regions of interest that require high protection. Shen *et al.* [8] proposed encryptng syntax elements based on the inter frame dependency between adjacent frames. Shen's algorithm focuses on reducing error propagation due to encryption. Tosun *et al.* [12] introduced base layer encryption that divided the data and XORed it. Tomun's work achieves reasonable overhead. However, the unencrypted higher layers are vulnerable to leak visual information. Al-Hayani *et al.* [13] suggested an algorithm that applies compression on high frequency level 1 and 2 sub-bands and encrypts low frequency sub-band 3 without compression. Nazar's algorithm claims to have low computational cost, however the algorithm is proposed for wavelet based video coding. Thiyagarajan *et al.* [14] proposed chosing syntax elements for encryption based on scene change detection in P and B frames. Recently, Saleh *et al.* [15] encrypted moving object related information via motion syntaxes in HEVC stream. Although, mohamed's work secures motion information, I frame information is still visually insecure. Furthermore, in all the aforementioned works, several selective encryption algorithms have been designed to lower the encryption time cost with a tolerable trade off over security. In addition, other key requirements such as format compliance and statistical size-preservation were satisfied.

Uniquely, in this paper, we address the requirements of low computational overhead for a resource constrained IoMT and propose an automatic, tunable encryption algorithm with a tolerable trade-off in the security level. Multimedia IoT models have been discussed in the context of energy and security in the network and application layers. Zhou and Chao [16] and Zhou *et al.* [17] proposed an IoT architecture for securing multimedia transmission in the application layer through authentication, watermarking and key management. However, Liang's work does not discuss a multimedia encryption scheme in the application or presentation layer. Al-Turjman [18] proposed an energy-aware data delivery framework by optimizing routing paths for multimedia content delivery in the IoT's. Our work is similar to Fadi's except that we focus on energy-aware integrated encryption-compression in the presentation layer. Concretely, a selective encryption algorithm that chooses a certain number of syntaxes for encryption based upon frame level energy is proposed, which satisfies all the three video encryption properties mentioned above.

The remainder of the paper is organized as follows: Section II discusses the necessity and motivation for our algorithm and a brief overview of the contribution. In section III we perform preliminary analysis. An algorithm that classifies video frames in to low and high energy levels is also presented in this section. In section IV, a new selective encryption is proposed to optimize the computational cost while maintaining security. Section V shows details of the experimental results and analysis. Finally, conclusions and future work are given in section VI.

II. Problem Formulation

In this section, we propose a low complexity selective encryption frame work for secure video streaming over IoMT. Fig. 1 shows the system framework for an integrated HEVC-crypto framework for IoMT. The framework can be formulated as a security problem with energy consumption as the main constraint.

A. The Necessity of Low Complexity Encryption

Here we first point out the need for a low complexity encryption algorithm. Generally, multimedia networks

TABLE I
DESCRIPTION OF SYMBOLS USED IN THE ALGORITHM

Symbols	Defnitions
IPM, QTC, MV	Intraprediction modes, Quantized transform coefficients, Motion vectors
ξ_{TE}, ξ_{TET}	Texture energy, Threshold for texture energy
$\xi_M, \xi_{ME}, \xi_{MET}$	Motion block energy, Motion frame energy,Motion energy threshold
$Intra_{BLK}, Inter_{BLK}$	Intra block size, Inter block size
N_{QTC}	Number of QTC coefficients in a macroblock
QTC_I	QTC coefficients in a block
M_k, θ_k	Magnitude and phase of kth motion vector
B_{luma}, B_{chroma}	Sensitive bits in chroma and luma coefficients
B_{QTC}	Encrypt-able bits in QTC coefficients
$S_{IPM}, S_{QTC}, S_{mvd}$	Encrypt-able syntaxes in IPM , QTC and MV
N_{HTE}, N_{LTE}	Number of frames with high texture energy, low texture energy
N_{HME}, N_{LME}	Number of frames with high motion energy, low motion energy
MB_{Intra}, MB_{Inter}	Number of macroblocks in intra and inter frame
M, K, R, E	M frames in a video, Number of keys used in encryption , Number of rounds in encryption, Encryption cost

or IoMT have limited to severe energy constraints since most of the independent nodes are battery powered embedded devices [19]. Many recent efforts attempt to reduce power consumption and encoding compression cost to suit Internet-of-multimedia devices and networks [20]. However, encryption along with video compression increases the computational overhead and power consumption. Therefore, we should encrypt certain codewords that decreases the encryption overhead and the power consumption of the device.

B. Motivation for Selective Encryption Based on Energy Models

Human eyes are highly sensitive to texture patterns [21] and motion intensity in video frames [22]. Hence, we propose to provide variable selective encryption for frames with high texture and high motion activity. The texture energy in an intrablock is obtained as a product of the mean of transform coefficients and the size of the intra block. Whereas, the motion energy is obtained as a product of the motion vector magnitude, motion vector phase and the size of the inter block in an inter frame.

The main contributions of this paper are:

- First, a new energy evaluation model for the intra and inter frames is proposed. It is used to dynamically identify frames with high and low energy levels. In frames with high energy level, all the syntax elements in table I are encrypted. In frames with low energy, alternate syntax elements are chosen for encryption to reduce the computational cost.
- Second, to refrain interpolation attacks we propose encrypting alternate coefcient syntaxes after XORing the sign bit to its immediate neighboring sign bit.
- Third, as a compliment, we propose a combined encryption-permutation technique to refrain sketch attacks, i.e we randomly permute coding unit structures in each frame after the final entropy stage.

III. PRELIMINARIES

To investigate the texture and motion energy, two consolidated video sequences, Foreman(352 × 288) and Soccer(720 × 480), were chosen under the mainline profile with QP = 18 and a 4:2:0 sampling format.

A. Texture Energy Model

Frequency transforms are widely used in digital signal processing and especially for transform domain in video compression. The DCT transform can be represented as,

$$X(k,l) = \frac{2}{\sqrt{MN}} C_k C_l \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(m,n) \cdot \frac{\cos(2m+1)}{2M} \cdot \frac{\cos(2n+1)}{2N} \quad (1)$$

The DC coefficient corresponds to a zero horizontal and vertical frequency, which can be obtained by choosing $k = 0$ and $l = 0$. Most of the signal energy is concentrated in the DC coefficient. Coefficients with k and l being non-zero are AC coefficients which determine variations in gray values at certain rates and directions.

To investigate the textural characteristics of transform coefficients, consider the basis function in the discrete cosine transform

$$T(m,n) = \frac{\cos(2m+1)k\pi}{2M} \cdot \frac{\cos(2n+1)l\pi}{2N} \quad (2)$$

Equations 1 and 2 show that the transform coefficients are determined by the sum of the products of pixel value and basis functions. Each basis function coefficient describes textural and structural information in various directions [23]. Furthermore, the transformed coefficients contain texture distributions of various sub bands. This implies that smooth regions of dark and bright information reside in the low frequency coefficients and the sharp contour-edge information resides in the high frequency coefficients. Therefore, the texture energy of a block can be modelled using transform coefficients and size of an intrablock.

B. Motion Energy Model

A coding unit structure is composed of a number of sub-blocks. The displacement of a sub-block between two consecutive frames is represented as a motion vector. Motion compensation in HEVC can be represented as,

$$S'(x,y) = S''(x,y) + U'(x,y) \quad (3)$$

where, S' is the decoded unit, S'' is the motion compensated unit and U' is the predicted blocks. The motion compensated

(a) (b) (c) (d)

Fig. 2. Interpolation attack on frames after encrypting alternate sign bits of QTC's. (a) Foreman-Original Frame. (b) Foreman-Interpolation attack. (c) Football-Original Frame. (d) Football-Original Frame. (e) Football-Interpolation Attack.

coding unit can be represented as,

$$S''(x,y) = S''(x-dx, y-dy) \tag{4}$$

$$dx = \sum_{i=0}^{N-1} M_i \theta_i(x,y) \tag{5}$$

$$dy = \sum_{i=0}^{M-1} M_i \theta_i(x,y) \tag{6}$$

where, M_i is the magnitude of motion vectors and θ_i(x,y) is the phase of motion vectors. θ is defined in radians from $(-\pi, +\pi)$. Therefore, motion vectors contain descriptive information about the video as magnitude and phase angle. If the motion activity in the frame is high, temporal dependency between frames is exploited resulting in a larger motion vector magnitude and phase angle. If the motion activity is low, the magnitude and phase angle values are significantly lower. Therefore, the motion energy of an inter frame is defined as a factor of motion vector magnitude M(i,j), phase angle Q(i,j) and size of the interblock.

C. Encrypting Alternate Transform Coefficients (EATC)

For low complexity analysis, we chose to encrypt alternate intraprediction modes, transform coefficient-related codewords and motion vectors as proposed in [4]. Then, we implemented the transform coefficient interpolation attack [24] on encrypted transform units. The attack is as follows. First we recover the AC coefficients from known [Xmin,Xmax] values and use the linear programming method to recover encrypted coefficients. Finally, the DC coefficient can be recovered by taking the mean average of the interpolated AC coefficients. After interpolation, we set the encrypted intraprediction mode to its most probable mode. Fig. 2. shows the test results. The visually quality of the video frame can be improved when alternate transform coefficients are encrypted. This confirms that encrypting alternate transform coefficients for the sake of low complexity is insecure. Therefore, we propose a secure version of the same.

IV. Proposed Algorithm

Texture and motion energy models in section III are used to determine the energy level in video frames. Then, a multilevel encryption technique is variably applied to the video bit stream based on predicted energy levels. The following sections gives a detailed description of the algorithm.

A. Texture Energy in I Frames

I frames are completely self-referential and don't use information from neighbouring frames. For intracoding, pixels with in a block are predicted from adjacent reference pixels from neighbouring, previously decoded, blocks. HEVC utilizes 35 angular prediction modes to exploit spatial redundancy in still pictures in order to improve coding efficiency. Mode indices 0-17 use prediction units of 4 × 4 pixels and discrete sine transform(DST), whereas modes 0-34 use prediction units of size 8 × 8, 16 × 16 and 32 × 32 pixels and discrete cosine transform(DCT) [25].

As proposed in section III, the non-zero coefficients and prediction block size is used to identify the texture energy in the I frame.

The texture energy of a coding unit, is determined by

$$\xi_{TE} = \frac{Intra_{BLK}}{N_{QTC}} * \sum_{i=0}^{N_{QTC}-1} QTC_i \tag{7}$$

The dynamic threshold to classify a coding unit's energy is determined by the average of the texture energies of all block units in the frame, which is given as follows:

$$\xi_{TET} = \frac{1}{MB_{Intra}} * \sum_{i=0}^{MB_{Intra}-1} \xi_{TE}[i] \tag{8}$$

B. Motion Energy in P and B Frames

P and B frames are referential frames and make use of information from previously coded frames or futuristic frames. As mentioned in the previous section, the position of a block in a previously decoded frame with respect to the current block unit is given by motion vector(Ax,Ay). From section III, the energy of P and B frames is evaluated by motion activity.

The motion energy of a inter block unit is given by,

$$\xi_M = Inter_{BLK} * M_k(\delta_{xk}, \delta_{yk}) * \theta_k \tag{9}$$

and the motion energy of an inter frame with N compensated blocks can be

$$\xi_{ME} = \frac{1}{N} * \sum_{i=0}^{N-1} \xi_M[i] \tag{10}$$

The dynamic threshold to classify the motion energy of a M th frame is determined by the average motion energy

of a few previous inter frames:

$$\xi_{MET} = \frac{1}{M-t} * \sum_{i=(M-t)}^{M} \xi_{ME}[i] \tag{11}$$

C. Algorithm

The proposed encryption algorithm chooses syntax elements in frames for encryption based on texture and motion energy models. Table I shows the syntaxes and semantics. The texture energy (ξ_{TE}) of an I frame is calculated for all intra block units and compared with a threshold value (ξ_{TE}). The threshold value is obtained by averaging the texture energy of individual block units. Similarly, for inter frames, motion energy (ξ_M) is calculated for each motion compensated block and motion energy of an entire frame ξ_{ME} is calculated as the mean average of ξ_M. The threshold (ξ_{MET}) is the mean average of motion energy (ξ_{ME}), obtained from preceding inter frames with in a GOP. For coding units with a high texture energy, all the sensitive code words are chosen for encryption, whereas in coding units with low energy, alternate syntaxes are encrypted to minimize encryption overhead. If an inter frame contains less motion energy, all the motion related syntaxes are encrypted, while in the case of low energy frames, alternate syntaxes are encrypted.

Algorithm 1 Estimating Texture and Motion Energy HEVC Frames, Selecting Appropriate Syntaxes for Encryption

Input ← read Nth *Video Frame*
if Video frame is an intrapredicted frame **then**
  Begin-[Texture energy estimation]
  Calculate texture energy ξ_{TE} *(7)*
  Calculate threshold ξ_{TET} *(8)*
  if ξ_T greater than ξ_{TET} **then** Parse and encrypt all IPM and QTC related syntaxes
  else Parse and encrypt alternate IPM and QTC related syntaxes
if Video frame is an interpredicted frame **then**
  Begin-[Motion energy estimation]
  Calculate macroblock motion energy ξ_M *(9)*
  Calculate frame motion energy ξ_{ME} *(10)*
  Calculate motion energy threshold ξ_{MET} *(11)*
  if ξ_{ME} greater than ξ_{MET} **then** Parse and encrypt all IPM, QTC and MV related syntaxes
  else Parse and encrypt alternate IPM, QTC and,motion vector related syntaxes
End-[Energy estimation and encryption]
if End of Frame **then** END algorithm
else Read next *N+1 frame*.

D. Low Complexity Analysis

The encrypt-able syntax elements in HEVC video stream are discussed and we analyze the encryption overhead in detail. In HEVC, to improve the compression efciency, CABAC uses regular mode to encode intraprediction modes. Their encoding depends on a flag namely the prev_Intra_Luma_pred_flag. Sensitive bits in the luma and chroma coefcients are,

$$B_{luma} = \begin{cases} mpm_idx, if\ \ prev\_Intra\_pred\_flag \\ rem\_Intra\_Luma\_pred\_Mode, otherwise \end{cases} \tag{12}$$

$$B_{chroma} = Intra\_Chroma\_Pred\_Mode \tag{13}$$

Bits for encryption,

$$S_{IPM} = \begin{cases} B_{Luma}, & if\ Luma\ components \\ B_{Chroma}, & Chroma\ components \end{cases} \tag{14}$$

Another new feature in CABAC is the dependency between the unit prediction mode and the scanning mode used for residual coding. This induces a special need for encryption as mentioned in [7]. Therefore, we follow the same method of IPM encryption as suggested in [7]. For binarization of quantized transform coefficients (QTC's), CABAC concatenates of truncated rice codes and exgolomb codes. Therefore, in the case of QTC's, the suffix of Trp code and the of EG0 codes are encrypted.

Binarization of QTC's is performed using a Trp code threshold as given by [26].

$$B_{QTC} = \begin{cases} EG0, & if\ QTC > TRPThreshold \\ TRP\ code, & otherwise \end{cases} \tag{15}$$

$$Bsign = < sign\ bits > \tag{16}$$

Encryptable syntaxes in quantized transform coefficients

$$S_{QTC} = \begin{cases} B_{QTC}, & if\ Trp\ components \\ B_{Sign}, & Sign\ components \end{cases} \tag{17}$$

As mentioned earlier, it is necessary to encrypt the sign bit of motion vectors for securing motion information. The syntax of motion information is given by,

$$< abs\_mvd\_greater0\_flag,\ abs\_mvd\_greater1\_flag,\ abs\_mvd\_minus2(EG1bins),\ mvd\_sign\_flag > \tag{18}$$

$$S_{MVD} = < abs\_mvd\_minus2(EG1bins), mvd\_sign\_flag > \tag{19}$$

Here, abs_mvd_greater0 and abs_mvd_greater1 specifies whether the absolute motion vector component is 0 or 1. Whereas, abs_mvd_minus2 represents the EG1 bins of a motion vector component and the mvd_sign_flag represents the sign bins. The EG1 and sign bins can be encrypted to provide temporal secrecy.

The encryption cost [27] of AES in counter mode can be represented as,

$$E = C * R + D * K \tag{20}$$

where C is the complexity of encrypting one syntax element, D is the cost of key schedule and the C and D constants depend on the hardware and software. The encryption cost of the proposed (E_P) and state-of-art (E_{SOF}) algorithms can be derived from the number of rounds R_P and state-of-art R_{SOF}

given in equation 21 and 22.

$$R_P = \sum_{i=0}^{M-1} S_{IPM}[N_{HTE}[i] + \frac{N_{LTE}[i]}{2}] + \sum_{i=0}^{M-1} S_{QTC}[N_{HTE}[i] + \frac{N_{LTE}[i]}{2}] + \sum_{i=0}^{M-1} S_{MV}[N_{HME}[i] + \frac{N_{LME}[i]}{2}] \quad (21)$$

$$R_{SOF} = \sum_{i=0}^{M-1} S_{IPM}[N[i]] + \sum_{i=0}^{M-1} S_{QTC}[N[i]] + \sum_{i=0}^{M-1} S_{MV}[N[i]] \quad (22)$$

$$Overhead = \begin{cases} R_P = R_{SOF}, if\ Lower\ energy\ frames = 0 \\ R_P = \frac{R_{SOF}}{2}, if\ Higher\ energy\ frames = 0 \end{cases} \quad (23)$$

The encryption algorithm automatically adjusts according to the energy content in the video frame. Equation 23 proves that the encryption cost can be lowered down to 50 % depending on the frame content and in the worst scenario, the encryption cost of the proposed algorithm is same as the state of art.

E. Security Analysis

The security of the proposed method of encrypting alternate intraprediction blocks (intraprediction modes and QTC) can be analyzed by using the spatial correlation between intrapredicted blocks and the temporal correlation between interpredicted blocks.

1) Transform Coefficient Interpolation Attack: A general method for recovering missing DCT coefficients and improving the encrypted video is demonstrated in section III. In this section a theoretical security evaluation on encrypting alternate syntax elements is provided. We now give a theoretical security evaluation of the proposed algorithm against an attack on the transform coefficients.

A residual block in the uncompressed domain can be represented as x(N,M),

$$x(N, M) = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1N} \\ \vdots & \ddots & \vdots & \\ P_{M1} & P_{M2} & \cdots & P_{MN} \end{bmatrix} \quad (24)$$

Similarly x(N,M) in the Fourier domain can be represented as,

$$Y(K, L) = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1K} \\ \vdots & \ddots & \vdots & \\ F_{L1} & F_{L2} & \cdots & F_{KL} \end{bmatrix} \quad (25)$$

and,

$$Y(K, L) = DCT_{MN} * x(M, N) \quad (26)$$

There are three cases:

a) Case 1: Encrypting all sign bits. Encrypting, Y(K,L) with an encryption E. we get ciphered block,

$$C(I, J) = Y(K, L) \oplus Key \quad (27)$$

where $C(I, J)$ is the encrypted version of $Y(K, L)$. In this case, it is impossible to obtain $Y(K, L)$ from $C(I, J)$ with out the encryption key Key.

b) Case 2: Encrypting alternate sign bits and applying an encryption function E to alternate quantized transform coefficients, the ciphered block can be rewritten as,

$$C(I, J) = E(K, L) + \sum_{k=\frac{K}{2}}^{K} \sum_{l=\frac{L}{2}}^{L} Y(k, l) \quad (28)$$

Where,

$$E(K, L) = \sum_{k=0}^{\frac{K}{2}} \sum_{l=0}^{\frac{L}{2}} Y(k, l) \oplus Key \quad (29)$$

In this case the encrypted coefficients E_{MN} can be obtained by interpolation attack as shown in section III.

c) Case 3: The proposed EATC ciphered block can be written as,

$$C(I, J) = E(K, L) + \sum_{k=\frac{K}{2}}^{K} \sum_{l=\frac{L}{2}}^{L} Y(k, l) \quad (30)$$

where,

$$E(K, L) = \sum_{k=0}^{\frac{K}{2}} \sum_{l=0}^{\frac{L}{2}} Y(k, l) \oplus \sum_{k=\frac{K}{2}}^{K} \sum_{l=\frac{L}{2}}^{L} Y(k, l) \oplus Key \quad (31)$$

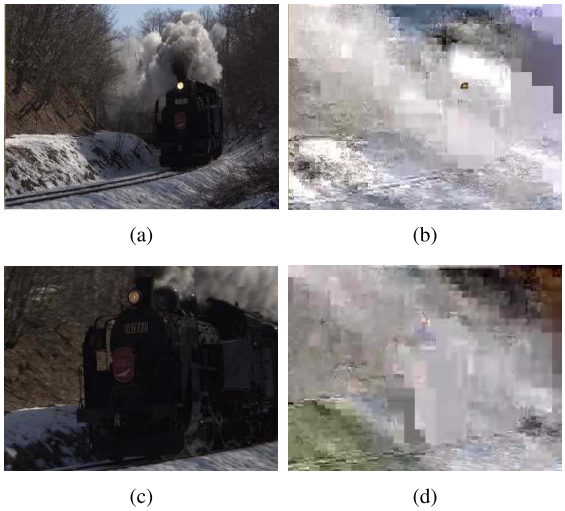
In case 3 it is impossible to obtain E(K,L) as neighboring coefficients are XOR'ed before encryption.

V. Experimental Results

In this section, the performance of the proposed scheme is analyzed. Reference implementation of HEVC HM 8.0 was used for simulation purpose. Low delay(Type I: IBBBPBBBP..) and random access(Type II: IPPPP..) structures are adopted as a group of pictures (GOP) with intra period equal to 10 in the encryption test bed. The set of standard video sequences and test bench of our simulation is shown in table II. To perform encryption on sensitive HEVC elements, we used the AES(Advanced Encryption Standard)in counter mode with an 128-bit initialization vector. We chose AES CTR as it allows to encrypt syntax elements of variable length and at any point in the stream. In this paper, the security is analyzed as visual perception and quality metrics as in [7]. In addition, several security attacks and analysis are implemented to conrm that the proposed encryption is secure.

TABLE II
EXPERIMENTAL SETUP

Parameters	Settings
Processor and RAM	2.4GHz CPU, 4Gb RAM
H.264/AVC software	HM reference 8.0
Number of frames encoded	100
GOP Type I	Random access mode (IBBBPBBBP...)
GOP Type II	Low delay mode (IPPPPPP...)
Intra Period	10
Videos	312x288 : Foreman,Football,Flowers 416x240 : BlowingBubbles, BQsquare, RaceHorses 720x480 : Driving, OpeningCeremony, Soccer 832x480 : Basketballdrill, BQmall 1920x1080 : Cactus, Kimono, Parkscene 1280x720 : KristenandSara , Johnny, Sideediting 2560x1600 : Steamlocomotivetrain , Traffic, PeopleonStreet

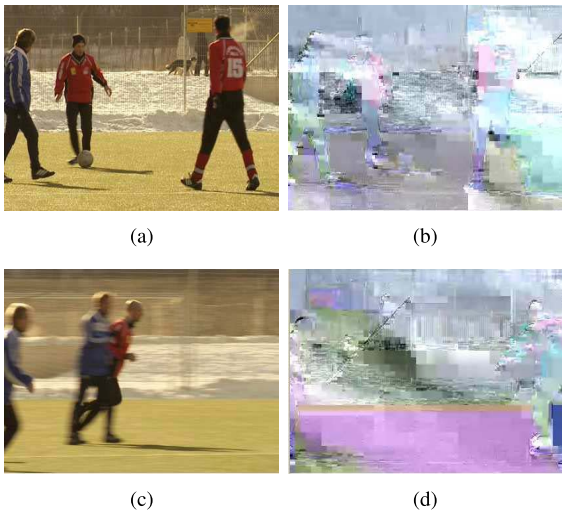


(a) (b) (c) (d)

Fig. 3. Proposed encryption applied to steam locomotive train, demonstrating scrambling effect on high and low enegy frames. (a) Steam locomotive train-High texture energy frame. (b) Steam locomotive train-Proposed encryption on frame in Fig. 3 (a). (c) Steam locomotive train-Low texture energy frame. (d) Steam locomotive train-Proposed encryption on frame in Fig. 3 (c).

A. Visual Security

If the decrypted video is imperceptible or too scrambled to be understood, the video is considered to be perceptually secure. In Fig. 3 and Fig. 4, the visual results of the proposed encryption are presented. Fig. 3 (a) shows a steam locomotive train frame with high textural energy and Fig. 3 (c) shows another with low textural energy. Fig. 3 (b) and (d) show the respective frames encrypted by the proposed encryption algorithm. The high texture frame is visually secure as all the syntax elements are encrypted. In the case of frames with low texture energy, alternate syntax elements are encrypted after XORing neighbouring transform coefficients. Fig. 3 (c) and (d) justifies the security of EATC encryption. Fig. 4 (a) and (c) shows the original and encrypted version of a high motion inter frame in soccer video. Here, all the motion vectors are encrypted; therefore the visual security is high. In case of inter



(a) (b) (c) (d)

Fig. 4. Proposed encryption applied to soccer, demonstrating scrambling effect on high and low enegy frames. (a) Soccer- High motion energy frame. (b) Soccer-Proposed encryption on frame in Fig. 4 (a). (c) Soccer- Low motion energy frame. (d) Soccer-Proposed encryption on frame in Fig. 4 (c).

frames with low motion energy, alternate motion information syntaxes are encrypted which are shown in Fig. 4 (b) and (d). The correlation between inter frames and alternated motion syntax encryption results in an effective visual degradation of the video. It is clear that the proposed encryption conceals information by degrading the visual quality of video frames of all energy levels.

B. Metric Based Security Measures

To quantify the visual degradation shown previously, this section provides a numerical quality analysis of the proposed video encryption. PSNR and SSIM are the common metrics used to evaluate video encryption in [28] and [29]. PSNR describes the loss in visual quality of the video and SSIM indicates the structural coherence of a frame. The encryption distortion thresholds for PSNR and SSIM are 15db [30] and 0.5 [31] respectively. However, the performance of a PSNR/SSIM evaluation on video encryption is limited as their values are not suffice when comparing two highly scrambled videos. Therefore, our evaluation relies upon other two metrics mentioned in the literature namely, the LSS (luminance similarity score) and ESS (edge similarity score). LSS is a similarity measure of luminance blocks in a frame and ESS measures the degree of similarity of shapes and edges in images. The encryption distortion threshold for LSS and ESS is set to 0 and 0.5 [7]. For evaluation purposes we consider two QP values, 18 and 32. Table III compares the state-of-art encryption and the proposed encryption. Compared to the state-of-art encryption, there is little quality improvement, which is obvious as some syntaxes are left unencrypted to reduce encryption overhead. However, the quality improvement is still very low, which implies that our proposed encryption will protect the video well. Furthermore, the quality metrics of the proposed encryption are below the aforementioned threshold values confirming that the proposed algorithm can provide effective visual security.

TABLE III
METRIC ANALYIS FOR GOP TYPE II

Videos	QP	Original				State-of-Art-[7]				Proposed approach			
		PSNR	SSIM	LSS	ESS	PSNR	SSIM	LSS	ESS	PSNR	SSIM	LSS	ESS
Basket ball drill	18	36.59	0.943	1.0	0.943	11.42	0.323	-2.14	0.382	12.08	0.384	-2.54	0.401
	32	45.09	0.912	1.0	0.964	11.17	0.237	-1.50	0.346	12.88	0.306	-1.89	0.387
Basket ball	18	36.59	0.943	1.0	0.943	10.18	0.303	-1.60	0.415	12.38	0.358	-1.91	0.481
	32	43.88	0.948	1.0	0.992	10.08	0.222	-1.22	0.327	11.37	0.294	-1.63	0.371
BQmall	18	35.76	0.951	1.0	0.945	9.60	0.202	-2.91	0.312	11.67	0.289	-3.36	0.407
	32	43.42	0.943	1.0	0.943	8.42	0.317	-1.45	0.320	10.42	0.386	-1.98	0.391
Cactus	18	37.38	0.985	1.0	0.911	11.62	0.321	-1.78	0.364	11.97	0.403	-2.91	0.379
	32	44.66	0.911	1.0	0.960	10.75	0.420	-1.90	0.425	12.23	0.476	-2.18	0.431
Kimono	18	39.77	0.976	1.0	0.962	10.80	0.232	-1.11	0.391	13.42	0.284	-1.92	0.472
	32	41.70	0.983	1.0	0.925	12.07	0.328	-1.93	0.434	12.93	0.379	-2.38	0.455
Park scene	18	35.48	0.947	1.0	0.992	10.83	0.235	-1.83	0.377	11.61	0.296	-1.96	0.466
	32	40.00	0.922	1.0	0.923	9.55	0.245	-1.70	0.386	10.21	0.341	-1.85	0.434
Slide editing	18	38.86	0.940	1.0	0.932	12.29	0.272	-2.11	0.323	12.84	0.315	-2.88	0.416
	32	43.42	0.995	1.0	0.941	8.40	0.324	-1.22	0.468	9.17	0.391	-1.42	0.489
Kristen and sara	18	41.51	0.971	1.0	0.979	8.34	0.316	-1.67	0.313	10.66	0.388	-1.77	0.397
	32	46.12	0.956	1.0	0.901	9.21	0.219	-1.43	0.414	9.89	0.278	-1.81	0.481
Johnny	18	32.05	0.942	1.0	0.910	10.62	0.401	-1.27	0.327	11.15	0.476	-1.83	0.388
	32	42.14	0.977	1.0	0.917	8.99	0.320	-1.24	0.343	10.72	0.374	-1.60	0.395
Steam locomotive train	18	37.83	0.986	1.0	0.914	9.79	0.211	-1.32	0.437	10.82	0.285	-1.48	0.472
	32	45.35	0.957	1.0	0.968	8.26	0.417	-1.45	0.302	9.36	0.472	-1.70	0.386
Traffic	18	37.00	0.963	1.0	0.946	10.33	0.221	-2.10	0.321	12.80	0.314	-2.82	0.329
	32	44.45	0.931	1.0	0.960	7.99	0.234	-2.05	0.413	9.84	0.346	-2.73	0.446
People on street	18	37.95	0.958	1.0	0.932	10.74	0.310	-1.79	0.407	11.73	0.317	-1.99	0.448
	32	41.91	0.922	1.0	0.905	9.67	0.205	-1.23	0.377	10.23	0.309	-1.74	0.453
Blowing bubbles	18	33.86	0.973	1.0	0.947	8.97	0.419	-1.44	0.390	10.48	0.452	-1.80	0.401
	32	43.88	0.973	1.0	0.947	11.34	0.312	-1.52	0.407	12.39	0.388	-1.73	0.492
BQsquare	18	33.64	0.968	1.0	0.937	12.81	0.416	-1.37	0.445	13.19	0.455	-1.59	0.421
	32	40.78	0.975	1.0	0.901	11.78	0.224	-2.09	0.380	12.22	0.299	-2.50	0.424
Race horses	18	34.49	0.942	1.0	0.922	10.14	0.338	-1.49	0.329	11.25	0.381	-1.86	0.383
	32	44.86	0.927	1.0	0.923	10.62	0.202	-1.37	0.372	11.95	0.308	-1.75	0.417
Driving	18	32.97	0.932	1.0	0.958	8.20	0.327	-3.01	0.383	9.20	0.386	-3.46	0.422
	32	42.76	0.996	1.0	0.953	9.40	0.320	-1.78	0.321	11.98	0.465	-2.08	0.387
Opening ceremony	18	32.18	0.969	1.0	0.936	9.12	0.216	-1.60	0.428	10.16	0.265	-1.87	0.469
	32	32.41	0.932	1.0	0.972	11.30	0.210	-2.45	0.359	12.57	0.292	-2.77	0.451
Soccer	18	34.88	0.959	1.0	0.934	12.44	0.316	-1.58	0.370	12.90	0.410	-1.74	0.474
	32	43.40	0.933	1.0	0.981	12.79	0.213	-2.52	0.398	12.96	0.289	-2.92	0.463
Foreman	18	35.97	0.964	1.0	0.973	11.71	0.325	-1.91	0.362	13.87	0.379	-2.07	0.421
	32	46.08	0.996	1.0	0.954	11.00	0.406	-3.12	0.328	13.01	0.456	-3.48	0.418
Footaball	18	33.88	0.957	1.0	0.971	10.32	0.214	-1.77	0.296	10.44	0.313	-1.97	0.341
	32	45.34	0.976	1.0	0.937	9.22	0.314	-1.69	0.277	10.12	0.398	-1.79	0.352
Flowers	18	31.52	0.988	1.0	0.968	9.31	0.212	-2.19	0.328	9.46	0.295	-2.54	0.372
	32	42.22	0.981	1.0	0.957	8.47	0.319	-1.42	0.311	9.98	0.396	-1.84	0.406

C. Computational Complexity

The computational overhead of selective encryption is directly proportional to the number of syntaxs selected, which has an impact on the energy consumption, especially for internet of multimedia devices. The encryption/decryption over-head can be defined as the difference between the encoders or decoders coding time with and with out encryption/decryption.Encoding and decoding times were obtained from the HM reference console window. Tables IV and V show the computational overhead incurred by the state-of-art encryption versus the proposed algorithm. Two types of GOP are used in Table IV and V. Experimental results clearly show that the proposed method lowers the encryption over head by an average of 35-40% compared to the state-of-art approach while providing an effective scrambling.

D. Bit Rate Analysis

Video bit rate can be defined as the bits per second in the entropy stage. Encrypting syntax elements for improved visual security can affect the bit rate of an encrypted video. This section analyzes the bit rate increase between the original and encrypted videos. Table VI and VII shows the bit rate increase (in percent). Motion vector signs and residual signs are CABAC encoded using bypass bins, therefore encrypting MVD's and QTC's cause no increase in the bit rate [6]. However, encrypting IPM related syntax's, which are coded

TABLE IV
COMPUTATIONAL COMPLEXITY FOR GOP TYPE I, QP = 32

Videos	Original(S)		State-of-Art-[7](S)		Proposed approach(S)		State-of-Art-[7] overhead(S)		Proposed approach overhead(S)	
	Encoder	Decoder	Encrypt-Encode	Derypt-Decode	Encrypt-Encode	Derypt-Decode	Encryption overhead	Deryption overhead	Encryption overhead	Deryption overhead
Basket ball drill	680.83	1.72	706.10	1.81	696.18	1.78	25.27	0.060	15.35 0	0.040
Basket ball	669.46	1.57	695.36	1.63	684.56	1.60	25.90	0.060	15.10	0.036
BQmall	689.11	1.64	712.93	1.70	705.00	1.67	23.82	0.062	15.90	0.038
Cactus	3139	6.81	3262.63	7.07	3211.86	6.96	123.63	0.261	72.86	0.150
Kimono	3854.67	7.82	3999.85	8.30	3942.71	8.12	145.18	0.481	88.04	0.307
Park scene	3162.50	7.80	3281.36	8.09	3234.13	7.98	118.86	0.292	71.63	0.187
Slide editing	979.26	2.35	1018.54	2.43	1001.49	2.40	39.28	0.088	22.23	0.054
Kristen and sara	1067	2.37	1109.10	2.46	1091.75	2.42	42.10	0.091	24.75	0.054
Johnny	1018.05	2.27	1057.37	2.35	1041.32	2.32	39.32	0.087	23.27	0.052
Steam locomotive train	5824.67	13.11	6056.12	13.56	5957.31	13.42	231.45	0.459	132.64	0.311
Traffic	5400.85	12.43	5612.96	12.91	5522.73	12.72	212.11	0.486	121.88	0.297
People on street	10971.76	16.12	11413.13	16.82	11224.16	16.54	441.37	0.704	252.4	0.425
Blowing bubbles	169.102	0.56	175.74	0.58	172.87	0.574	6.031	0.021	3.76	0.013
BQsquare	147.76	0.55	153.65	0.58	151.21	0.56	5.89	0.033	3.45	0.0126
Race horses	222.37	0.65	231.43	0.67	227.61	0.66	9.06	0.023	5.24	0.015
Driving	743.44	1.80	772.23	1.87	760.82	1.84	28.79	0.070	17.38	0.047
Opening ceremony	489.10	1.44	508.12	1.50	500.22	1.47	19.02	0.0509	11.62	0.0234
Soccer	612.54	1.49	638.43	1.54	626.76	1.52	25.89	0.055	14.22	0.033
Foreman	173.97	0.56	179.94	0.58	177.98	0.57	5.97	0.021	4.01	0.013
Footaball	130.62	0.44	135.66	0.45	133.64	0.45	5.04	0.183	3.02	0.010
Flowers	186.64	0.65	194.09	0.67	190.87	0.66	7.45	0.026	4.22	0.015

TABLE V
COMPUTATIONAL COMPLEXITY FOR GOP TYPE II, QP = 32

Videos	Original(S)		State-of-Art-[7](S)		Proposed approach(S)		State-of-Art-[7] overhead(S)		Proposed approach overhead(S)	
	Encoder	Decoder	Encrypt-Encode	Derypt-Decode	Encrypt-Encode	Derypt-Decode	Encryption overhead	Deryption overhead	Encryption overhead	Deryption overhead
Basket ball drill	687.22	2.03	714.36	2.11	702.09	2.08	27.14	0.078	14.87	0.048
Basket ball	678.95	2.05	705.79	2.13	696.61	2.10	26.84	0.076	17.66	0.051
BQmall	688.81	1.71	715.89	1.77	704.89	1.75	27.08	0.065	16.08	0.040
Cactus	3156.80	7.21	3274.91	7.48	3223.74	7.38	118.11	0.269	66.94	0.167
Kimono	3862.25	8.45	4008.12	8.78	3943.97	8.65	145.87	0.331	81.72	0.211
Park scene	3254.54	8.29	3383.07	8.62	3331.04	8.48	128.53	0.332	76.5	0.190
Slide editing	1135.72	2.64	1181.09	2.74	1161.65	2.70	45.37	0.101	25.93	0.063
Kristen and sara	1185.36	2.90	1231.21	3.01	1212.97	2.96	45.85	0.112	27.61	0.067
Johnny	1171.29	2.71	1215.12	2.82	1198.08	2.77	43.83	0.111	26.79	0.062
Steam locomotive train	6105.28	14.34	6334.17	14.98	6238.70	14.72	228.89	0.640	133.42	0.380
Traffic	6155.13	14.13	6394.63	14.69	6289.63	14.46	239.51	0.560	134.56	0.330
People on street	13267.83	16.07	13796.12	16.71	13581.93	16.45	528.29	0.640	314.10	0.381
Blowing bubbles	170.16	0.70	176.92	0.73	174.13	0.720	6.76	0.027	3.97	0.016
BQsquare	148.27	0.54	154.03	0.56	151.73	0.55	5.76	0.021	3.46	0.012
Race horses	218.83	0.67	227.63	0.70	224.00	0.69	8.80	0.027	5.17	0.016
Driving	727.27	1.75	775.81	1.82	754.12	1.79	48.54	0.071	26.85	0.041
Opening ceremony	500.44	1.68	520.55	1.75	512.06	1.72	20.11	0.067	11.16	0.038
Soccer	627.65	1.64	652.91	1.70	642.08	1.678	25.26	0.065	14.43	0.037
Foreman	187.66	0.59	195.01	0.80	192.03	0.72	7.35	0.208	4.37	0.130
Footaball	134.31	0.37	139.80	0.39	137.42	0.38	5.49	0.015	3.11	0.010
Flowers	180.03	0.59	187.22	0.61	184.22	0.60	7.19	0.021	4.19	0.014

using regular mode can affect the bit rate of the encrypted video. Table VI and VII show that the proposed encryption has an average bit-rate fluctuation of (0.1%) and is with-in the tolerable range as demonstrated in the state-of-art analysis [7]. Further, the bitrate of the proposed encryption is reduced by 20% compared with the state-of-art encryption as there is a reduction in the number of IPM syntaxes chosen for encryption.

E. Brute Force Attack

Protection against a brute force attack is guaranteed if the decryption is highly sensitive to the key change, i.e., the ciphered elements should not be decrypted correctly for a one bit variation between the encryption and decryption keys. For this purpose a key change test is implemented as demonstrated in [6]. A 128 bit key is chosen for encryption and the most significant bit is changed in the decryption key and the decrypted/decoded bit stream. Fig. 5 shows the brute force attack implemented on the original steam locomotive train and soccer frames shown in Fig. 3 and 4. Fig. 5 (a), (b), (c), (d) shows the brute force attack on frames with different energy levels. This indicates that the proposed encryption algorithm can resist brute force attack. Fig. 6 (a) and (b) shows the PSNR values of individual steam locomotive train and soccer frames decrypted with the wrong key and the key with a 1-bit change. Note that the key with one bit change produces a similar video

TABLE VI
BITRATE ANALYSIS, GOP TYPE I, QP = 32

Videos	State-of-art[7](%)	Proposed Algo-rithm(%)	Bit Rate Reduc-tion(%)
Basket ball drill	0.21%	0.17%	-19.04%
Basket ball	0.30%	0.23%	-23.33%
BQmall	0.28%	0.24%	-14%
Cactus	0.21%	0.18%	-14.28%
Kimono	0.036%	0.027%	-24.99%
Park scene	0.19%	0.16%	-15.78%
Slide editing	0.11%	0.10%	-9.09%
Kristen and sara	0.36%	0.29%	-19.44%
Johnny	0.08%	0.06%	-25.0%
Steam locomotive train	0.42%	0.33%	-21.42%
Traffic	0.12%	0.10%	-16.66%
People on street	0.060%	0.048%	-19.5%
Blowing bubbles	0.075%	0.061%	-22.95%
BQsquare	0.244%	0.20%	-16.65%
Race horses	0.12%	0.10%	-16.66%
Driving	0.072%	0.056%	-22.85%
Opening ceremony	0.091%	0.073%	-19.78%
Soccer	0.183%	0.15%	-18.03%
Foreman	0.134%	0.10%	-25.37%
Football	0.152%	0.13%	-15.63%
Flowers	0.15%	0.14%	-7.28%

TABLE VII
BITRATE ANALYSIS,GOP TYPE II,QP = 32

Videos	State-of-art[7](%)	Proposed Algo-rithm(%)	Bit Rate Reduc-tion(%)
Basket ball drill	0.19%	0.15%	-21%
Basket ball	0.16%	0.13%	-18.75%
BQmall	0.037%	0.029%	-21.62%
Cactus	0.093%	0.076%	-18.27%
Kimono	0.115%	0.10%	-9%
Park scene	0.06%	0.052%	-13.3%
Slide editing	0.14%	0.11%	-21%
Kristen and sara	0.11%	0.10%	-9%
Johnny	0.28%	0.24%	-14%
Steam locomotive train	0.042%	0.038%	-5%
Traffic	0.043%	0.031%	-27.90%
People on street	0.16%	0.15%	-6.25%
Blowing bubbles	0.30%	0.23%	-23%
BQsquare	0.18%	0.15%	-16.5%
Race horses	0.031%	0.027%	-12.90%
Driving	0.11%	0.09%	-18%
Opening ceremony	0.13%	0.11%	-15%
Soccer	0.12%	0.09%	-25%
Foreman	0.17%	0.12%	-29%
Football	0.096%	0.076%	-18.75%
Flowers	0.162%	0.154%	-4.93%

with an average PSNR value of 9db, which is similar to the video decrypted with the wrong key.

F. Transform Coefficient Interpolation Attack

The problem of recovering missing transform coefficients recovery problem is reported in [24]. Although the problem focuses on video compression and error concealment, it also

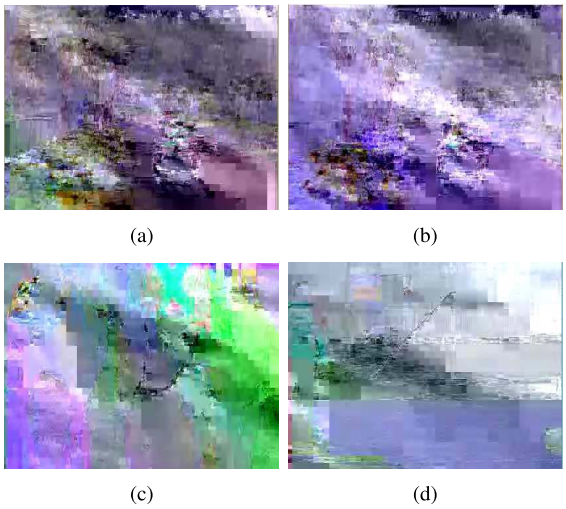


Fig. 5. Brute force attack implemented on steam locomotive train and soccer, original frames are shown in Fig. 3 and 4. (a) Steam locomotive train - Brute force attack applied on frame in Fig. 3 (b). (b) Steam locomotive train - Brute force attack applied on frame in Fig. 3 (d). (c) Soccer - Brute force attack applied on frame in Fig. 4 (b). (d) Soccer - Brute force attack applied on frame in Fig. 4 (d).

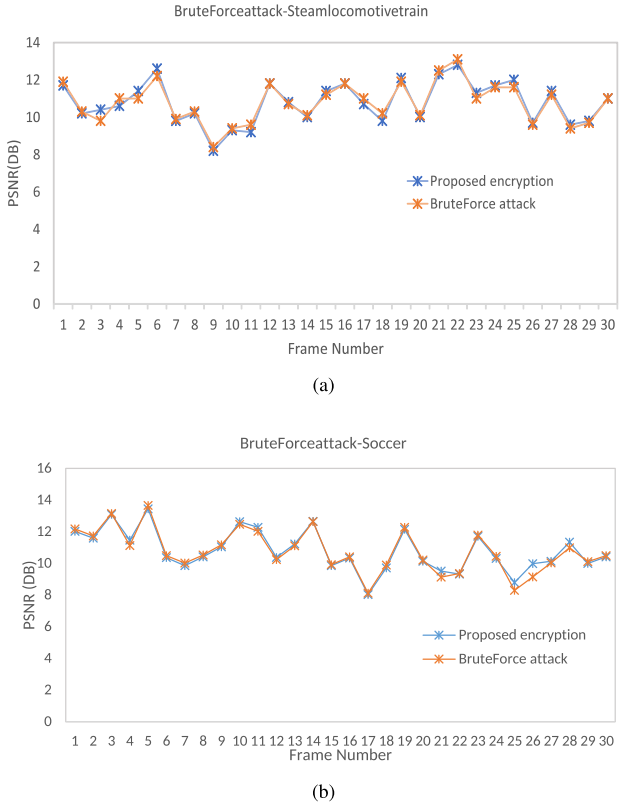
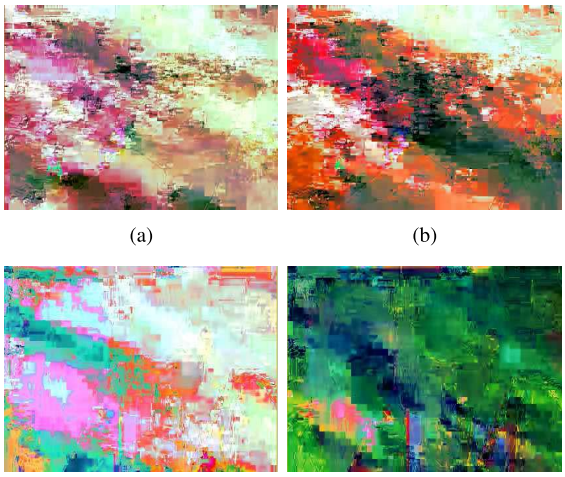


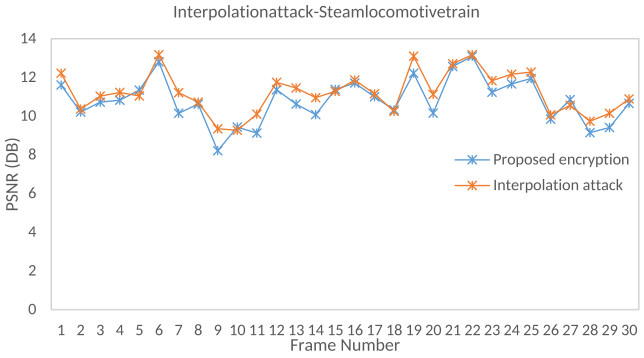
Fig. 6. PSNR of encrypted video and brute force attack on the encrypted video, first 30 frames of steam locomotive train and soccer. (a) Steam locomotive train - Brute force attack (Frame Number vs PSNR DB). (b) Soccer-Brute force attack (Frame Number vs PSNR DB).

can be studied in the context of video encryption. As mentioned earlier, an interpolation attack can breach the encryption when alternate transform coefficients are encrypted. Therefore, we propose encryptng alternate QTC's after XORing

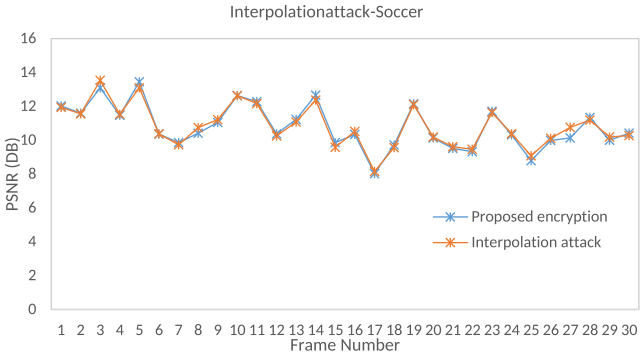


(a) (b) (c) (d)

Fig. 7. Interpolation attack on steam locomotive train, foreman, football videos after applying proposed encryption. (a) Steam locomotive train - Interpolation attack applied on frame in Fig. 3 (b). (b) Steam locomotive train - Interpolation attack applied on frame in Fig. 3 (d). (c) Foreman - Interpolation attack applied on frame in Fig. 2 (a). (d) Football - Interpolation attack applied on frame in Fig. 2 (c).



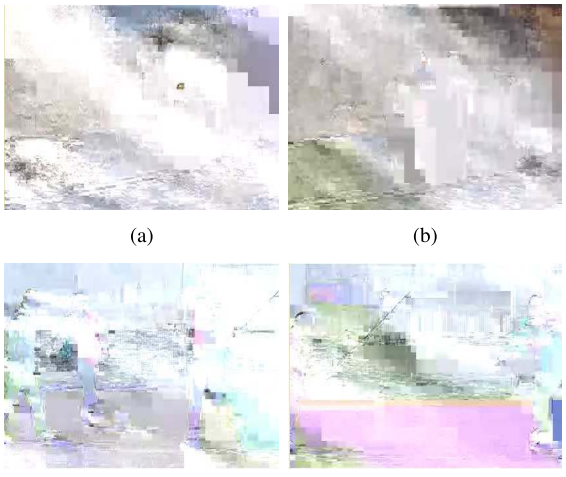
(a)



(b)

Fig. 8. PSNR of encrypted video and interpolation attack on the encrypted video, first 30 frames of steam locomotive train and soccer. (a) Steam locomotive train -Interpolation attack (Frame Number vs PSNR DB). (b) Soccer-Interpolation attack (Frame Number vs PSNR DB).

neighboring coefficients. We have implemented the interpolation attack mentioned in [24]. First we recover the AC coefficients from known [Xmin,Xmax] values and the linear programming to recover encrypted coefficients. Fig. 7 shows



(a) (b) (c) (d)

Fig. 9. Replacement attack on steam locomotive and soccer videos after applying proposed encryption. (a) Steam locomotive train - Replacement attack applied on frame in Fig. 3 (b). (b) Steam locomotive train - Replacement attack applied on frame in Fig. 3 (d). (c) Soccer - Replacement attack applied on frame in Fig. 4 (b). (d) Soccer - Replacement attack applied on frame inFig. 4 (d).

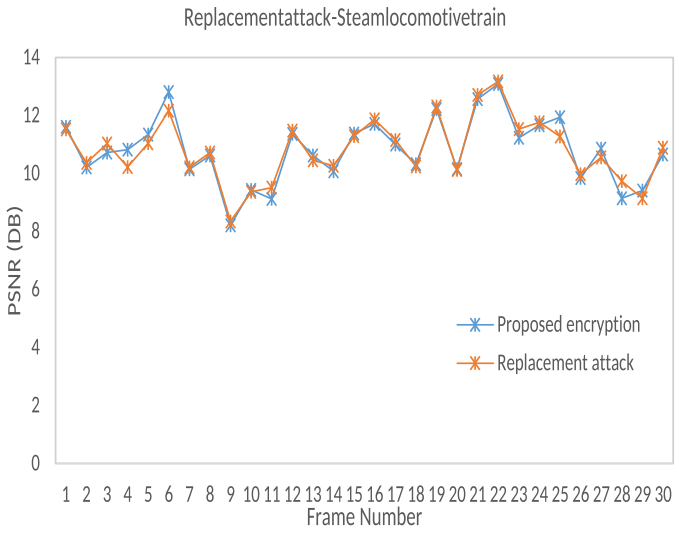
the interpolation attack implemented on video frames with different energy levels. Fig. 7 and (a) are frames with high textural energy and Fig. 7 (b) is the frame with low textural energy. Further, figure Fig. 7 (c) and Fig. 7 (d) shows the interpolation attack on foreman and football videos after applying the proposed encryption. The original frames are shown in Fig. 2. PSNR values of the encrypted videos with replacement attacks are shown in Fig. 8. The error propagation induced by XORing neigboring coefficients ensures that the proposed transform coefficient encryption and can withstand an interpolation attack.

G. Replacement Attack

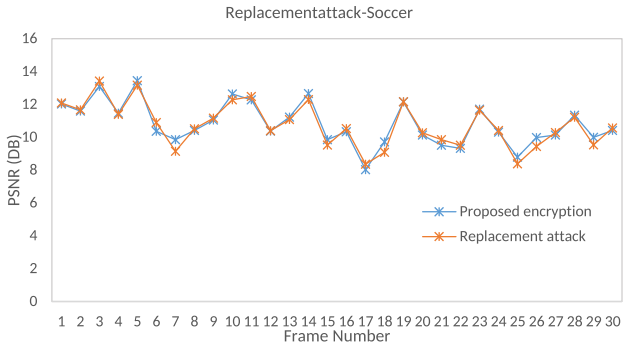
In the replacement or known plain text attack, the encrypted bits are replaced with known non encrypted bits in the HEVC stream. Fig. 10 shows the PSNR values of the encrypted video with and with out the replacement attack. The average PSNR value of the encrypted video is 10 db and that of the attacked video is 11 db. Furthermore, the replacement attack is implemented on frames with high and low textural and motion energies as shown in Fig. 9. In frames of all energy level, the encrypted video can with stand replacement attacks. This confirms that the robustness of the proposed encryption against replacement attacks.

H. Edge Detection Structural Analysis

The degradation of edge and contour information in frames can be measured using the EDR(Edge Differential Ratio) and by applying laplacian edge detection. Fig. 11 shows the laplacian edge detection of original and encrypted video frames with high and low textural energy. It is evident that the proposed encryption distorts the edge and contour information. The EDR determines the deviation in pixels that contribute to edge information between the original and

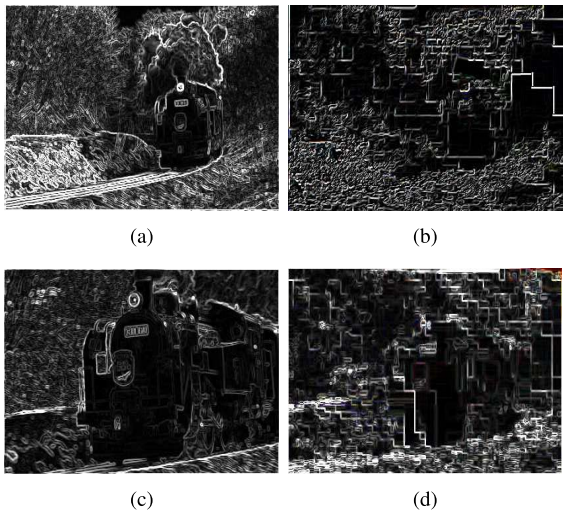


(a)



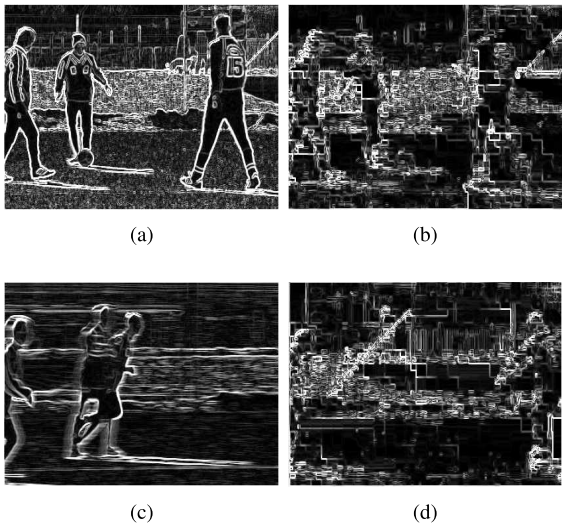
(b)

Fig. 10. PSNR of encrypted video and replacement attack on the encrypted video, first 30 frames of steam locomotive train and soccer. (a) Steam locomotive train - Replacement attack (Frame Number vs PSNR DB). (b) Soccer - Replacement attack (Frame Number vs PSNR DB).



(a) (b) (c) (d)

Fig. 11. Sobel's Edge detection applied on original and encrypted steam locomotive train frames to verify structural encrypted content. (a) Steam locomotive train- Edge detection on original high texture energy frame in Fig. 3 (a), EDR = 0.72. (b) Steam locomotive train- Edge detection on encrypted high texture energy frame in Fig. 3 (b), EDR = 0.77. (c) Steam locomotive train- Edge detection on original low texture energy frame in Fig. 3 (c), EDR = 0.64. (d) Steam locomotive train- Edge detection on encrypted low texture energy frame in Fig. 3 (d), EDR = 0.72.



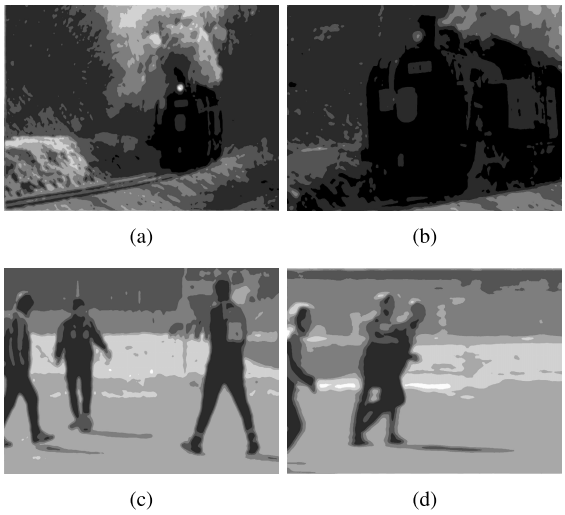
(a) (b) (c) (d)

Fig. 12. Sobel's edge detection applied on original and encrypted soccer frames to verify structural encrypted content. (a) Soccer- Edge detection on original high motion energy frame in Fig. 4 (a), EDR = 0.93. (b) Soccer- Edge detection on encrypted high motion energy frame in Fig. 4 (b), EDR = 0.90. (c) Soccer- Edge Detection on original low motion energy frame in Fig. 4 (c), EDR = 0.95. (d) Soccer- Edge Detection on encrypted low motion energy frame in Fig. 4 (d), EDR = 0.94.

encrypted video frame. The closer EDR is to 1, the more the edge information is distorted, whereas an EDR close to 0, indicates poor encryption. The EDR values of the edge detected frames are shown in Fig. 12. It is evident that the structural information is fully distorted when using the proposed encryption scheme.

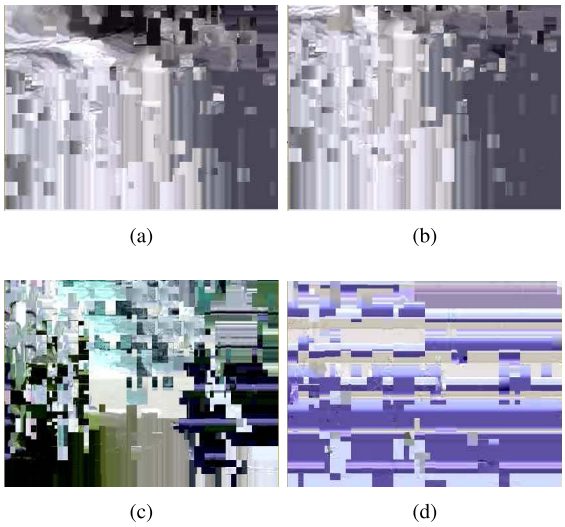
I. Sketch Attack With Permutation [32]

A sketch attack is a signal processing operation to sketch the out line of encrypted frames based upon the coding units bit stream size. Minemura *et al.* [32] demonstrated the sketch attack on format compliant encryption proposed by Wang *et al.* [4]. Miemuras attack was able to retrieve visual information from encrypted frames. He suggested that the sketch attack relies on information retrieved from bit stream size information and when the coding or transform units are permuted or diffused the sketch attack can be prevented. Permuting or diffusing coding units in the prediction domain can cause serious compression degradation as the correlation between frames are disturbed. For validation of our encryption, we implement the sketch attack on frames encrypted by the proposed algorithm which is shown in Fig. 13. It is evident that the syntax elements chosen for encryption are vulnerable to sketch attacks. Therefore, we propose to permute transform units in the entropy domain to retain the compression efficiency of the codec. We apply random permutations to each frame in the entropy domain and encrypt the permutation key. Fig. 14 shows the video frame with combined permutation and encryption. Permutation adds complexity to encryption. However, the overall computational



(a) (b) (c) (d)

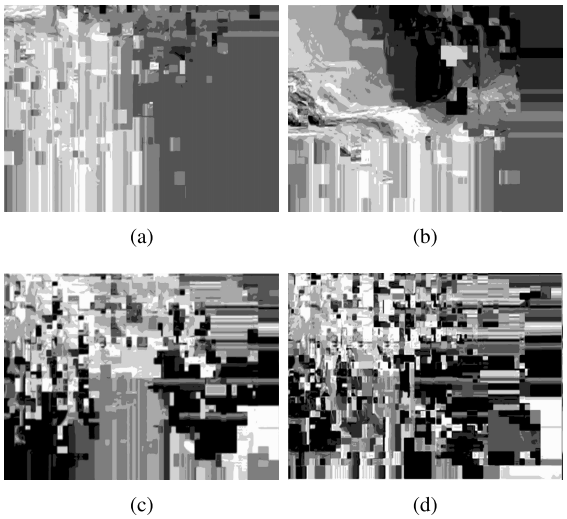
Fig. 13. Sketch attack implemented on frames encrypted by the proposed algorithm, steam locomotive train and soccer. (a) Steam locomotive train- Sketch attack after proposed encryption on high energy frame in Fig. 3 (b). (b) Steam locomotive train- Sketch attack after proposed encryption on low energy frame in Fig. 3 (d). (c) Soccer- Sketch attack after proposed encryption on high energy frame in Fig. 4 (b). (d) Soccer- Sketch attack after proposed encryption on low energy frame in Fig. 4 (d).



(a) (b) (c) (d)

Fig. 14. Proposed encryption algoithm with permutation in entropy domain on steam locomotive train and soccer videos. (a) Steam locomotive train- Proposed encryption with permutation on high energy frame in Fig. 3 (a). (b) Steam locomotive train- Proposed encryption with permutation on low energy frame in Fig. 3 (c). (c) Soccer- Proposed encryption with permutation on high energy frame in Fig. 4 (a). (d) Soccer- Proposed encryption with permutation on low energy frame in Fig. 4 (c).

complexity is still lower than [7] as shown in the simulation results. Table VIII shows the security evaluation with PSNR. The permutation degrades the quality of video much more compared to encryption alone. Table IX shows the encryption and decryption overhead when encryption and permutation together are used. As shown, the complexity increases 10% compared to overhead results in Table V. Fig. 13, 14 and 15 show the sketch attack implemented on frames with different texture and motion characteristics. The result confirm that permutation and encryption together refrain a sketch attack.



(a) (b) (c) (d)

Fig. 15. Sketch attack implemented on frames encrypted by the proposed algorithm and permuted on SteamLocomotiveTrain and Soccer videos. (a) Steam locomotive train- Sketch attack after applying proposed encryption with permutation on high energy frame in Fig. 14 (a). (b) Steam locomotive train- Sketch attack after applying proposed encryption with permutation on low energy frame in Fig. 14 (b). (c) Soccer- Sketch attack after applying proposed encryption with permutation on high energy frame in Fig. 14 (c). (d) Soccer- Sketch attack after applying proposed encryption with permutation on low energy frame in Fig. 14 (d).

TABLE VIII

PSNR ANALYSIS,GOP-TYPE II, QP = 32

Videos	PSNR-encryption(DB)	PSNR-encryption with permutation(DB)
BQmall	10.42	9.48
Park scene	10.21	8.31
SlideEditing	9.17	7.86
Steam locomotive train	9.36	7.97
People on street	12.57	9.81
Race horses	11.95	9.06
Soccer	12.96	10.52
Foreman	9.98	8.13

TABLE IX

ENCRYPTION WITH PERMUTATION OVERHEAD ANALYSIS, GOP TYPE II, QP = 32, SECONDS

Videos	Encoder (S)	Decoder (S)	Encoder overhead Increment(%)	Decoder overhead Increment(%)
BQmall	706.46	1.669	10.30%	9.96%
Park scene	3339.36	8.50	10.21%	10.48%
Slide editing	1165.01	2.71	9.64%	8.42%
Steam locomotive train	6286.20	15.32	9.21%	9.42%
People on street	13609.84	16.52	8.89%	8.42%
Race horses	224.55	0.705	10.59%	11.70%
Soccer	644.33	1.682	13.58%	10.66%
Foreman	192.59	0.609	12.03%	13.98%

Permutation and encryption together can be a new research focus and indeed an in-depth analysis is required.

VI. CONCLUSION

In this paper, we have analyzed the motion and texture energy models in video frames and proposed a new energy

aware encryption scheme for securing video transmission in IoMT. Extensive experiments were conducted, and the results show that the proposed scheme fulfils all the constraints required for a video encryption. Since we encrypt elements in the context model, there is a slight change in bit rate. However, the variation in bit-rate is tolerable [7]. In our future work, we plan to provide a comprehensive analysis on encrypting every 3rd, 4th, and 5th syntax elements after XORing the neighbouring coefficients, and provide an exhaustive comparative analysis between low complexity video encryption algorithms in IoMT.

REFERENCES

[1] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in *Proc. Picture Coding Symp. (PCS)*, San Jose, CA, USA, Dec. 2013, pp. 394–397, doi: https://doi.org/10.1109/PCS.2013.6737766

[2] R. Pereira and E. Pereira, "Video streaming: H.264 and the Internet of Things," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Mar. 2015, pp. 711–714.

[3] S. Lian, Z. Liu, Z. Ren, and H. Wang, "Secure advanced video coding based on selective encryption algorithms," *IEEE Trans. Consum. Electron.*, vol. 52, no. 2, pp. 621–629, May 2006.

[4] Y. Wang, M. O'Neill, and F. Kurugollu, "A tunable encryption scheme and analysis of fast selective encryption for CAVLC and CABAC in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1476–1490, Sep. 2013.

[5] G. Van Wallendael, A. Boho, J. De Cock, A. Munteanu, and R. Van de Walle, "Encryption for high efficiency video coding with video adaptation capabilities," *IEEE Trans. Consum. Electron.*, vol. 59, no. 3, pp. 634–642, Aug. 2013.

[6] Z. Shahid and W. Puech, "Visual protection of HEVC video by selective encryption of CABAC binstrings," *IEEE Trans. Multimedia*, vol. 16, no. 1, pp. 24–36, Jan. 2014.

[7] B. Boyadjis, C. Bergeron, B. Pesquet-Popescu, and F. Dufaux, "Extended selective encryption of H.264/AVC (CABAC)- and HEVC-encoded video streams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 892–906, Apr. 2017.

[8] H. Shen, L. Zhuo, and Y. Zhao, "An efficient motion reference structure based selective encryption algorithm for H.264 videos," *IET Inf. Secur.*, vol. 8, no. 3, pp. 199–206, May 2014.

[9] Y. Zhao and L. Zhuo, "A content-based encryption scheme for wireless H.264 compressed videos," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2012, pp. 1–6.

[10] Y. Zhao, L. Zhuo, M. Niansheng, J. Zhang, and X. Li, "An object-based unequal encryption method for H.264 compressed surveillance videos," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2012, pp. 419–424.

[11] W. Wang, M. Hempel, D. Peng, H. Wang, H. Sharif, and H.-H. Chen, "On energy efficient encryption for video streaming in wireless sensor networks," *IEEE Trans. Multimedia*, vol. 12, no. 5, pp. 417–426, Aug. 2010.

[12] A. S. Tosun and W.-C. Feng, "Lightweight security mechanisms for wireless video transmission," in *Proc. Int. Conf. Inf. Technol., Coding Comput.*, Apr. 2001, pp. 157–161.

[13] N. Al-Hayani, N. Al-Jawad, and S. Jassim, "Simultaneous video compression and encryption for real-time secure transmission," in *Proc. 8th Int. Symp. Image Signal Process. Anal. (ISPA)*, Sep. 2013, pp. 240–245.

[14] K. Thiyagarajan, K. El-Sankary, Y. Wang, and I. Hammad, "Low complexity multimedia encryption," *Int. J. Comput. Netw. Inf. Secur.*, vol. 8, no. 4, p. 1, 2016.

[15] M. A. Saleh, N. M. Tahir, and H. Hashim, "Moving objects encryption of High Efficiency Video Coding (HEVC) using AES algorithm," *J. Telecommun., Electron. Comput. Eng.*, vol. 8, no. 3, pp. 31–36, 2016.

[16] L. Zhou and H. C. Chao, "Multimedia traffic security architecture for the Internet of Things," *IEEE Netw.*, vol. 25, no. 3, pp. 35–40, May 2011.

[17] L. Zhou, D. Wu, B. Zheng, and M. Guizani, "Joint physical-application layer security for wireless multimedia delivery," *IEEE Commun. Mag.*, vol. 52, no. 3, pp. 66–72, Mar. 2014.

[18] F. Al-Turjman, "Energy-aware data delivery framework for safety-oriented mobile IoT," *IEEE Sensors J.*, vol. 18, no. 1, pp. 470–478, Jan. 2018.

[19] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 4, pp. 18–39, 4th Quart., 2008.

[20] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 5, pp. 645–658, May 2005.

[21] J. Bergen, "D. in regan (ed.), vision and visual dysfunction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10B, no. 5, pp. 114–134, 1991.

[22] Q. Qu, Y. Pei, and J. W. Modestino, "An adaptive motion-based unequal error protection approach for real-time video transport over wireless IP networks," *IEEE Trans. Multimedia*, vol. 8, no. 5, pp. 1033–1044, Oct. 2006.

[23] E. Y. Lam, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[24] S. Li, A. Karrenbauer, D. Saupe, and C.-C. J. Kuo, "Recovering missing coefficients in DCT-transformed images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1537–1540.

[25] R. Sjöberg *et al.*, "Overview of HEVC high-level syntax and reference picture management," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1858–1870, Dec. 2012.

[26] J. Sole *et al.*, "Transform coefficient coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1765–1777, Dec. 2012.

[27] D. F. García, "Performance evaluation of advanced encryption standard algorithm," in *Proc. Int. Conf. Math. Comput. Sci. Ind. (MCSI)*, Aug. 2015, pp. 247–252.

[28] A. Massoudi, F. Lefebvre, C. De Vleeschouwer, B. Macq, and J.-J. Quisquater, "Overview on selective encryption of image and video: Challenges and perspectives," *EURASIP J. Inf. Secur.*, vol. 2008, p. 179290, Dec. 2008. [Online]. Available: https://doi.org/10.1155/2008/179290

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] L. Dubois, W. Puech, and J. Blanc-Talon, "Smart selective encryption of CAVLC for H.264/AVC video," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Nov./Dec. 2011, pp. 1–6.

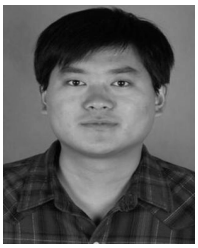
[31] S.-K. A. Yeung, S. Zhu, and B. Zeng, "Quality assessment for a perceptual video encryption system," in *Proc. IEEE Int. Conf. Wireless Commun., Netw. Inf. Secur. (WCNIS)*, Jun. 2010, pp. 102–106.

[32] K. Minemura, K. Wong, R. C.-W. Phan, and K. Tanaka, "A novel sketch attack for H.264/AVC format-compliant encrypted video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2309–2321, Nov. 2017.

Karthik Thiyagarajan was born in Kuwait. He received the B.E. degree in electronics and communication engineering from Anna University, Chennai, India, in 2010, the Post Graduate Diploma degree in embedded system design from NIIT, Calicut, India, and the master's degree from Dalhousie University, Halifax, NS, Canada. He is currently a Senior Control Systems Engineer (cryptography and security) with the Canadian Nuclear Laboratories–Atomic Energy of Canada Limited, Canada. His research interests are critical infrastructure security, securing cyberphysical embedded systems, and video processing/compression.

Rongxing Lu received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Canada, in 2012. He was a Post-Doctoral Fellow with the University of Waterloo from 2012 to 2013. He was an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2013 to 2016. He has been an Assistant Professor with the Faculty of Computer Science (FCS), University of New Brunswick (UNB), Canada, since 2016. His research interests include applied cryptography, privacy enhancing technologies, and Internet of Things big data security and privacy. He is currently a Senior Member of the IEEE Communications Society (ComSoc). He was a recipient of the most prestigious Governor Generals Gold Medal in 2012 and the 8th IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2013. He received eight best student paper awards from some reputable journals and conferences. He was the Winner of the 2016–2017 Excellence in Teaching Award from FCS, UNB. He currently serves as the Vice-Chair (Publication) for the IEEE ComSoc Communications and Information Security Technical Committee. He has published extensively in his areas of expertise (with citation 11,100+ and H-index 51 from Google Scholar as of 2018).

Kamal El-Sankary (M'07) received the B.Eng. degree from Lebanese University, Tripoli, Lebanon, in 1997, the M.A.Sc. degree in electrical engineering from the University of Quebec, Montreal, QC, Canada, in 2002, and the Ph.D. degree in electrical engineering from Ecole Polytechnique, University of Montreal, Montreal, in 2006. He joined the Department of Electrical and Computer Engineering, Dalhousie University, Halifax, NS, Canada, in 2006, where he is currently an Associate Professor. His current research interests include mixed-signal, analog, digital, and RF integrated circuit designs and embedded systems. He is the Chair of the Atlantic Canada IEEE Circuits and Systems and the Solid State Circuits Joint Chapter.



Hui Zhu (M'13) received the B.Sc. degree from Xidian University, China, in 2003, the M.Sc. degree from Wuhan University in 2005, and the Ph.D. degree from Xidian University in 2009. In 2013, he joined the School of Electrical and Electronics Engineering, Nanyang Technological University, as a Research Fellow. Since 2016, he has been a Professor with the School of Cyber Engineering, Xidian University. His research interests include the areas of applied cryptography, data security, and privacy.