Sparse Selective Encryption for HEVC 4K Video Using Spatial Error Spread

Mengdie Huang[1], Cheng Yang[1], Hao Li[2], Jian Shen[3]

[1] School of Information and Communication Engineering, Communication University of China, China
[2] Da Hengqin Science and Technology Development Co., Ltd Postdoctoral Research Workstation, China
[3] College of Computer and Software, Nanjing University of Information Science and Technology, China
{mdhuang, chy, cuclihao}@cuc.edu.cn, s_shejian@126.com

Abstract

This paper proposes a universal HEVC SE method named Sparse Selective Encryption (SSE), which encrypts a few of bits on the bitstream without coupling with the encoder. High efficiency of the SSE results from using spatial correlation of prediction units to propagate the error from encrypted units. The sparse selection of key units is decided by two spatial influence models, which are designed according to the angular intra prediction mode and the motion vector prediction direction respectively. SSE allows user to customize the encryption density of the bitstream by adjusting the proportion of selected coding units. Additionally, the format compliance of the bitstream is kept by performing bit encryption on the bits of cipherable syntax elements of selected units using the modified AES CTR cipher. The paper presents the comprehensive security analysis and efficiency performance of the SSE, including visual perception, cryptography attack and computation cost. Achieved experiment results confirm that SSE is secure and effective to protect HEVC 4K videos and backwards compatible with lower resolution videos.

Keywords: 4K, HEVC, Selective encryption, Video security, Spatial correlation

1 Introduction

Video protection is a deep concern of multimedia content producers faced with emerging attack [1]. Demand for copyright protection increases among multimedia content creators, who devote themselves to produce innovative media programs [2]. Video encryption technique enables the access control of resources by scrambling the video stream. Given the expensive overhead of encrypting the entire stream, naive encryption is seldom adopted in video systems with the limit of delay and power consumption. Selective Encryption (SE) is suggested as a feasible encryption method on video compression domain, which can achieve the decrease on computation cost and keep the format compliance at the same time.

Under the huge investments on ultra-high definition (UHD) technology made by global content producers, both 4K TV sets and content have drawn more attention. Recently, 4K resources have been launched in Asia Pacific, North America and Eastern Europe. Compared with H.264/AVC [3], High Efficiency Video Coding (HEVC) improves the video compression ratio by 50% with maintaining the same video quality [4-5]. So, HEVC wins favor over many 4K broadcasting systems, websites and mobile applications (e.g. Netflix, Apple Facetime) [6]. SE for HEVC streams becomes an essential way to secure UHD videos.

SE in the compression domain can be classified into SE of bin and SE of bitstream. Innovations in previous SE algorithms are distributed as follows:

(1) Selection of syntax elements; different syntax elements have different impacts on the video rendering. Van Wallendael et al. [7] defined the cipherable syntax elements without influencing the compatibility of HEVC stream, including reference picture set, QP information, motion estimation and compensation information, parameters of deblock filter and sample adaptive offset. Boyadjis et al. [8] added the encryption of luminance prediction mode and remove the encryption of RPS, initial QP, and deblock filter parameters. This selection of syntax elements is adopted by Thiyagarajan et al. [9]. Tew et al. [10-11] presented a scrambling method of transform skip flag and AC coefficient sign in luminance channel. Shahid [12] proposed a dyadic encryption space for HEVC constituted by coefficient sign, MVD sign, suffix of coefficient bins. Sallam et al. [13] only encrypted bypass mode syntax elements to avoid influencing bit rate size. Mustafa et al. [14] suggested to encrypt the syntax elements of wave parallel process header.

(2) Decision on encryption density; decision rules to distinguish frames that require high protection from other frames are presented by Wang et al. [15].

*Corresponding Author: Cheng Yang; E-mail: chy@cuc.edu.cn

DOI: 10.3966/160792642019092005023

Thiyagarajan [9] improved the decision of frame by constructing the texture model and the motion model. Hole et al. [16] separated video frames by using seen change detection technique and then encrypted important frames. Encryption density reflects the ratio of words that are changed to those that are left in the data stream. Both the threshold and distinguish granularity have effect on video encryption density.

(3) Cryptographic algorithm; Boyadjis et al. [17] designed a transcoder set to parse and modify compressed syntax elements in the bitstream. Not long after, a stream cipher to improve the robustness of SE video in real time scene was proposed by them [18]. Sidaty et al. [19] presented an assessment methodology to score the perceptual security of HEVC SE algorithms. Sallam et al. [13] addressed the requirement of real time by replacing AES in SE of HEVC by RC6. An image encryption system using chaotic sequences was designed by Ye et al. [20]. Li et al. [21] performed chaotic encryption on the bypass mode bins. To scramble H.264 video content, Su et al. SE in [22] modified selected data in bitstream via information hiding technique. Mustafa et al. [23] proposed to perform bit-flipping to selected syntax elements. Sallam et al. [24] employed the chaotic logistic map in the SE to save the time of encryption compared with AES. Long et al. [25] encrypted the signs of residual coefficient with RC4. Hofbauer et al. [26] performed transparent encryption on a part of the coefficient signs of each block.

In this paper, we propose a Sparse Selective Encryption (SSE) technique to address the content security issue of HEVC 4K videos, and the algorithm backward compatible with low resolution videos. Contributions of the SSE mainly involves in four aspects: encryption efficient, format compliance, universality in HEVC coders, and adjustable security level. Refer to the conception of error propagation proposed in H.264 [27] and HEVC [28], spatial influence model for the intra frame is designed to measure the impact of intra CU on spreading distortion in this paper. Additionally, we put forward the spatial influence model for the inter frame to quantize the correlation of motion vector Fin space. Blocks with higher spatial impact are selected for encryption based on models. This paper presents a detailed description of principle, framework and practical performance of the proposed SE method.

2 Proposed Spatial Influence Models

Different from our research on H.264 video encryption [29] which detailed how did the distortion spread among frames, spatial correlation is a measure that looks at the relationship between close spatial units [30]. We worked out the spatial correlation among pixels in the intra frame and figured out the spatial correlation among motion vectors (MV) in the inter frame in this section. By means of encrypting the coding units (CU) that have high spatial influence, the distortion could be maximized to the entire frame with minimal encryption cost.

2.1 Constitution of Computational Matrix

A slice consists of several coding unit trees (CTU) in HEVC standard. CTU is divided into CUs in the size ranging from 64x64 down to 8x8. CU is further partitioned into prediction unit (PU) and transform unit (TU). In [28], analysis of error propagation is limited to the condition that adjacent PU size is the same as the predicted PU size. Here, computation unit of spatial correlation was mapped to the minimum PU size with 4x4 pixels, so that any split depth of neighbor CU is acceptable for the computation.

2.2 Spatial Influence Model for Intra Prediction

In the intra prediction, correlation of PUs in the spatial domain is used to predict the current pixel by referencing adjacent coded pixels within the same frame image. When a PU is being predicted, the spatial influence of adjacent PUs will be different if there is a difference of reference proportion between adjacent PUs. So, the spatial influence of the PU was proposed to measure the importance of PU, and related variables are listed in Table 1.

Table 1. Related variables used in spatial influence models

Variables	Definitions
Y_i	Spatial influence of the intra PU_i on a frame
N_{buttef}	The times of the intra PU_i be referenced by other PUs
N_{block_k}	The number of the 4x4 pixels block located in the bottom edge of PU_i
$W_{prefect}$	The size of the predicted PU_i which references the PU_i
$W_{s,u}$	The size of the PU_i
P_k	The proportion of the pixels in the predicted PU_i which references the pixels in $block_k$
β_i	Spatial influence of the inter PU_i on a frame
P	$P=\begin{cases}1, & \text{MV of } PU_i \text{ is selected}\\ 0, & \text{else}\end{cases}$
ε	Spatial influence of the CU on a frame
$N_{s,u}$	The number of PU in the CU

On one hand, the range of reference pixels is determined by the intra prediction mode, which indicates the direction of reference pixels. HEVC standard supports 35 kinds of intra prediction mode with the index ranging from 0 to 34 [31]. Figure 1 presents the directions for angular intra prediction modes. Under a certain intra prediction mode, the reference proportion of the adjacent PU is calculated according to the projected area of the adjacent PU in predicted PU, seen in Figure 2. Since the value of proportion was only related to the mode, reference proportion of adjacent partition under 35 intra prediction modes was worked out in advance and stored into a three-dimensional array named percentage. On the other hand, in terms of the SE decision, when the reference proportion is the same, the adjacent PU with smaller size is more valuable for encryption because of the higher propagation times based on the size of the adjacent PU itself. Thus, spatial influence model of the PU in intra prediction was defined as (1).

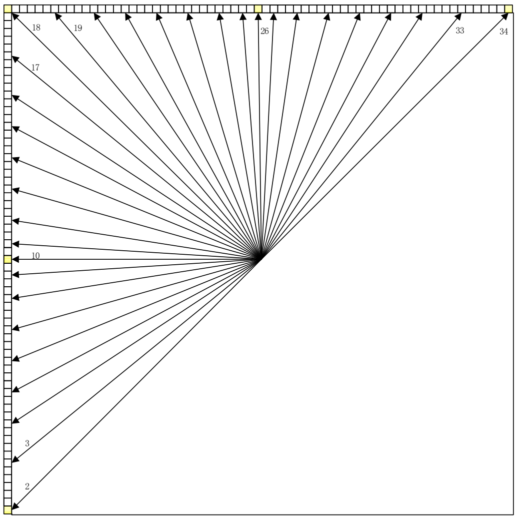


Figure 1. Reference directions for 33 kinds of angular intra prediction mode in HEVC. Mode number is 2 to 34

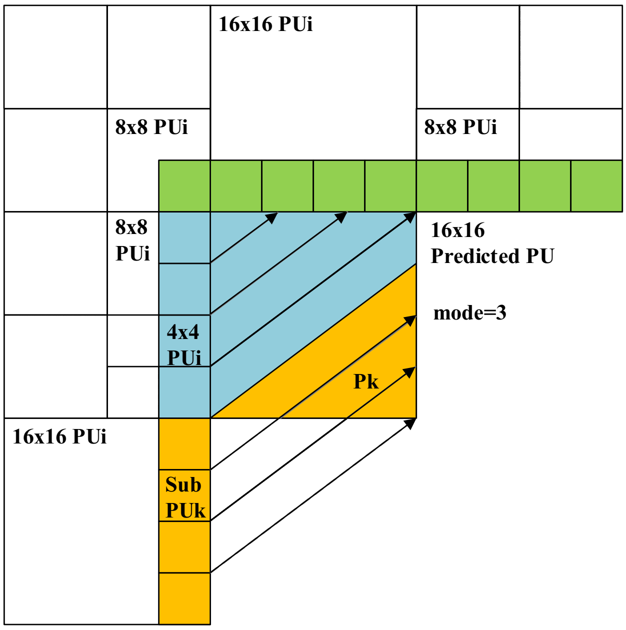


Figure 2. Reference proportion of adjacent PU pixels in the predicted PU when the angular intra prediction mode is 3

$$\gamma_i = \sum_{j=0}^{N} bekef^{-1} \frac{\sum_{k=0}^{N} block^{-1} * W_{pre} p \cup j}{W_{PU}} \tag{1}$$

2.3 Spatial Influence Model for Inter Prediction

When an inter PU is predicted, only one of MVs of adjacent PUs will be selected as the predicted MV of the predicted PU. Thus, spatial influence of adjacent PUs is different as well for inter prediction.

Three inter prediction modes are defined in HEVC to predict the MV of a PU, namely advanced motion vector prediction (AMVP) mode, merge mode and skip mode. Skip mode can be looked as an exception of merge mode. In MV prediction, there is a MV candidate list constructed by the MV of coded PUs in spatial domain (i.e. top, top right, top left, bottom left and top left), seen in Figure 3. MVP index indicates the index of the selected reference MV in the list and further demonstrates the direction of selected PU because the order of the MVs in the candidate list is arranged according to the direction. Here, the spatial influence model of the PU in inter prediction was defined as (2).

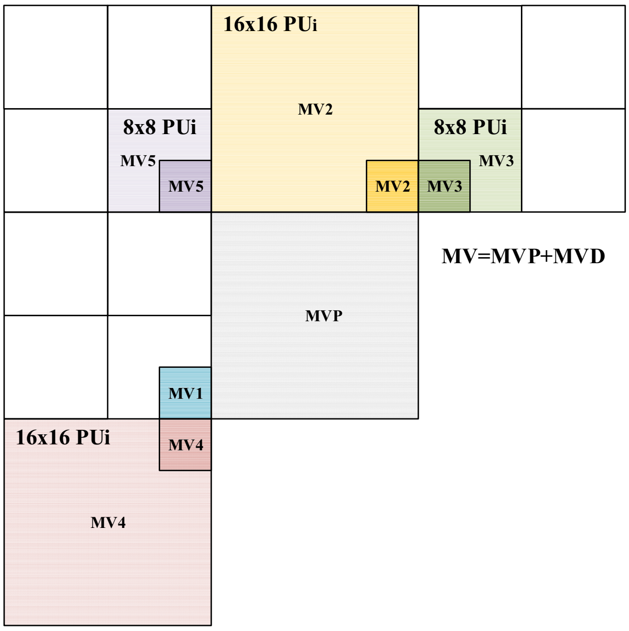


Figure 3. Five spatial candidate MVs for MV prediction

$$\beta_i = \begin{cases} \sum_{j=0}^{N} bekef\ P * \beta_i & \beta_i \neq 0 \\ & \beta_i = 0 \\ P, & \end{cases} \tag{2}$$

The iteration of β_i enabled the maximum spatial influence to map to the starting point of each MV reference chain, which represented the important PU in the inter prediction.

For intra prediction and inter prediction, spatial influence of the CU was defined as the average of spatial influence of PUs in the CU.

3 Proposed Selective Encryption Algorithm

In this section, a novel HEVC SE algorithm named SSE was proposed based on designed models. SSE is independent of the encoder without weakening the encoding efficiency. Details of parse, selection and encryption in SSE are described in following.

3.1 Parse HEVC Bit Stream

Parsing HEVC bit stream is aimed at deciphering the meaning of syntax element bits on the bit stream. On one hand, incoming parameters of spatial influence model need to be deciphered from the entropy coded bit stream. On the other hand, both the offset of selected syntax element bin string in bit stream and the bit length of it should be worked out accurately for extracting the plaintext of encryption. Parameters that need to be decoded are listed in Table 2.

Table 2. Parameters obtained by parsing bit stream

Needed parameters of spatial influence models	Value domain
CU pixels location in picture	x⊆ [0, PicWid], y⊆ [0, PicHei]
CU partition shape	2Nx2N/2NxN/Nx2N/NxN/2NxnU/2NxnD/nLx2N/nRx2N
CU prediction mode	Intra/Inter
Intra PU prediction mode	0~35
MV prediction reference type	P/B
Inter PU MV prediction direction	Left/Above/Above Right/ Below Left/Above Left
m_fifo_idx	0~length of slice byte stream)
m_bitsNeeded	-8~0

In the HM, an opened HEVC framework, m_fifo is a buffer for storage of byte stream. In syntax element parsing, m_fifo_idx indicates the index of read bytes and m_bitsNeeded points to the bit offset. Depending on two variables, bit length of the syntax element bin string was figured out according to the (3). The length of the current syntax element bin string i is signed as l_i. mbN_i is the value of m_bitsNeeded after decoding and mfi is the value of m_fifo_idx after decoding. Offset of the decoded syntax element in bit stream file is the accumulation of the length of former decoded syntax elements.

$$l_i = (mbN_i - mbN_{i-1}) + 8*(mfi_i - mfi_{i-1}) \qquad \textbf{(3)}$$

3.2 Select Specific Syntax Elements of Partial CUs

SSE chose CUs for encryption based on spatial influence models. The spatial influence of CU represents the effect of CU distortion in frame on condition that CU decoded error. First selection was at CU level. After figuring out the spatial influence of each CU in frame, CUs were sorted form large to small according to the value of spatial influence. Sorted CU was represented as CU_n with n referring to the ranking of CU spatial influence. CUs with high spatial influence were picked out for encryption because of more serious error propagation resulted from encrypting them. The encryption density of CUs in the frame was determined by the parameter α, seen in (4). α presents the proportion of selected CUs in total CUs of a frame. The number of CUs in a slice and selected CUs are denoted as Num_{CU} and Sec_{CU} respectively. According to the experiment result, α was set as 0.04 in the proposed SE, which is sparse for encryption density.

$$Sec_{CU} = \sum_{n=0}^{\alpha * Num_{CU} - 1} CU_n \qquad \textbf{(4)}$$

To avoid modifying syntax elements with influence on the format compliance of bitstream, syntax elements of selected CUs were further selected by referencing the cipherable syntax elements summarized in [7-9], encrypted syntax elements in the SSE are listed in Table 3. Notably, considering human eyes are more sensitive to luminance than to chrominance, which was proved in [8], only prediction mode of the intra CU in the luminance channel was encrypted in SSE.

Table 3. Syntax elements in selected CU for encryption

Syntax Element Name
IPCM information
Intra Luma Prediction mode
Merge Index
Skip Flag
Inter Dir PU
Ref Frame Idx
MVD PU
AMVP Idx
QT Coefficient (QTC)

3.3 Encrypt Selected Bits

Proposed SSE algorithm used the AES-CTR cipher for encrypting the HEVC bitstream. Block cipher mode that operated on fixed length plaintext block is not applicable to proposed SE because selected bits could not always be represented by an integer number of bytes. Among five modes of AES {i.e. CBC, ECB, CTR, CFB, OFB}, practicable modes for encrypting plaintext that is smaller than group length are CTR, CFB and OFB. Under the same provable security, the encryption efficiency of these three mods from high to low was CFB, CTR and OFB respectively, gave by Sallam et al. [13]. But, generation of ciphertext block is relying on serial computation when using CFB mode. AES-CTR was chose for encrypting HEVC bitstream because it is amenable to parallelization and enable the SSE to further apply to large-scale video data. In modified AES-CTR mode, n bits within high position of pseudorandom sequence generated by AES cipher were extracted. Then a bitwise XOR was performed on the extracted bits and n-bit plaintext block to form a n-bit ciphertext block.

The flowchart of the proposed encryption algorithm is presented in Figure 4. In HEVC standard, slice is packaged into NAL unit for transmission in complicated network. Key distribution was based on the slice rather than the video, so that the encrypted NAL could be decrypted independently as soon as received by receiver, without waiting for the next NAL.

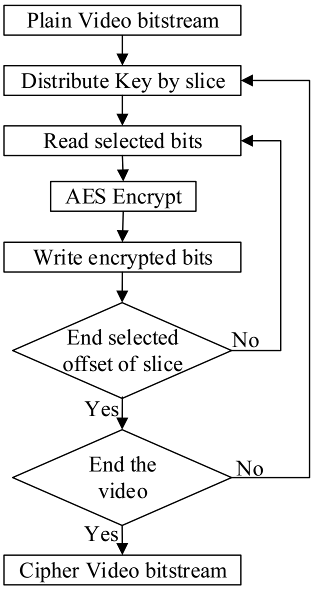


Figure 4. Encryption flowchart of the proposed SE method

4 Experiment Results

4.1 Experiment Preparation

Rich test sequences were adopted to evaluate security and performance of the SSE on various quality videos, as shown in Table 4. Both 4K UHD videos and official standard test videos in different resolutions were used. Different from videos in group B, C, D and F which could be used to test both Random Access (RA) coding structure and Low Delay (LD) coding structure, sequences in group A and group E were used to specially test RA and LD respectively. To assess the performance of SSE on different bitstream structures, sequences were encoded into two groups for encryption with parameter setting shown in Table 5.

Table 4. Test videos used in proposed SE experiment

Type	Resolution	Video	fps	Bit Depth
4K	3840x2160	*Bosphorous*	120	8
		Honeybee	120	8
		Jockey	120	8
		ShakeNDry	120	8
		YachtRide	120	8
A	2560x1600	*SteamLocomotiveTrain*	30	10
		PeopleOnstreet	30	8
B	1920x1080	*Kimono*	24	8
		ParkScene	24	8
		Cactus	50	8
E	1280x720	*Johnny*	60	8
		KristenAndSara	60	8
		FourPeople	60	8
F	1280x720	*SlideEditing*	30	8
C	832x480	*BasketBallDrill*	50	8
		BQMall	60	8
D	416x240	*BlowingBubbles*	50	8
		BQSquare	60	8
C	416x240	*RaceHorses*	30	8
Other	1920x1080	*Jockey*	120	8

Table 5. Experiment preparation

Attribute	Value
Processor and RAM	Intel® Core™2 CPU 2.83GHz, 4G RAM
IDE	Microsoft Visual Studio 2012
Software	HM reference 18.6
Configuration File	encoder_lowdelay_main.cfg
Frame Number	100
Intra Frame Period	16
GOP size	4
GOP Ⅰ (RA)	IPPPP
GOP II (LD)	IBBBP

4.2 Decision on The Threshold of Encryption Density

The domain of α was [0.04, 1] and the default α was decided as 0.04 in the SSE. Why is the value of α not smaller? In terms of used test sequences, resolution varied from 416x240 pixels to 3840x2160 pixels. The number of CUs per frame was limited, as presented in (5). To ensure the selected CU was valid in the frame, α should meet the (6). Since the minimum number of CU for a 416x240 frame was limited to 25, the value of α should be greater than 0.04.

$$25=\frac{416\times 240}{64\times 64}\le N_{CU}\le\frac{3840\times 21600}{8\times 8}=129600 \quad \textbf{(5)}$$

$$1\le\alpha\cdot\frac{minimum\ resolution}{maximum\ CU\ size}\le\frac{minimum\ resolution}{maximum\ CU\ size} \quad \textbf{(6)}$$

The encryption distortion thresholds for SSIM and PSNR are 0.5 and 15db [9]. As shown in Figure 5 and Figure 6, the degree of encryption distortion under different settings of α were reflected in the decrease in

SSIM and PSNR. Since the real total number of CU in the frame are usually larger than minimum, several smaller α {0.001, 0.003, 0.005, 0.007, 0.009, 0.01, 0.02, 0.03} were computed to delineate the trend of change on SSIM and PSNR. As the encryption density increased, the whole perceptibility of the picture tended to decrease (seen in Figure 7.), occasionally fluctuated. Obviously, encrypted video has been strong scrambled when 4 percentage of CUs were encrypted.

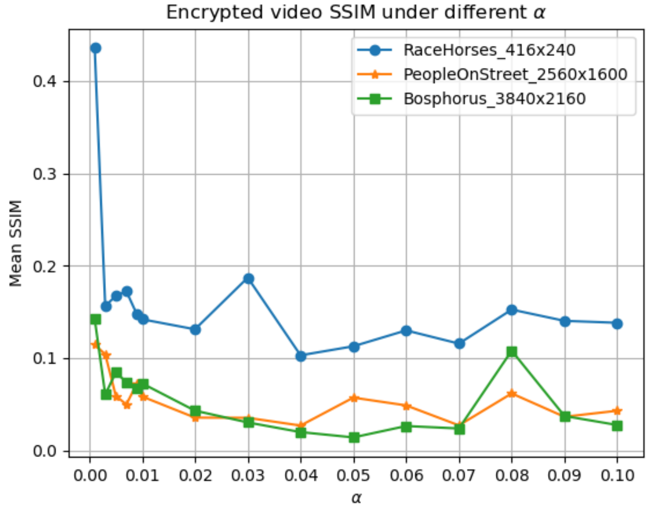


Figure 5. Mean SSIM of the test video under different values of α

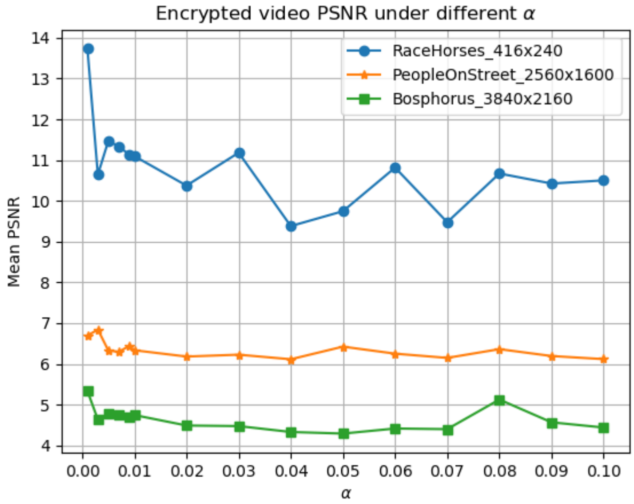
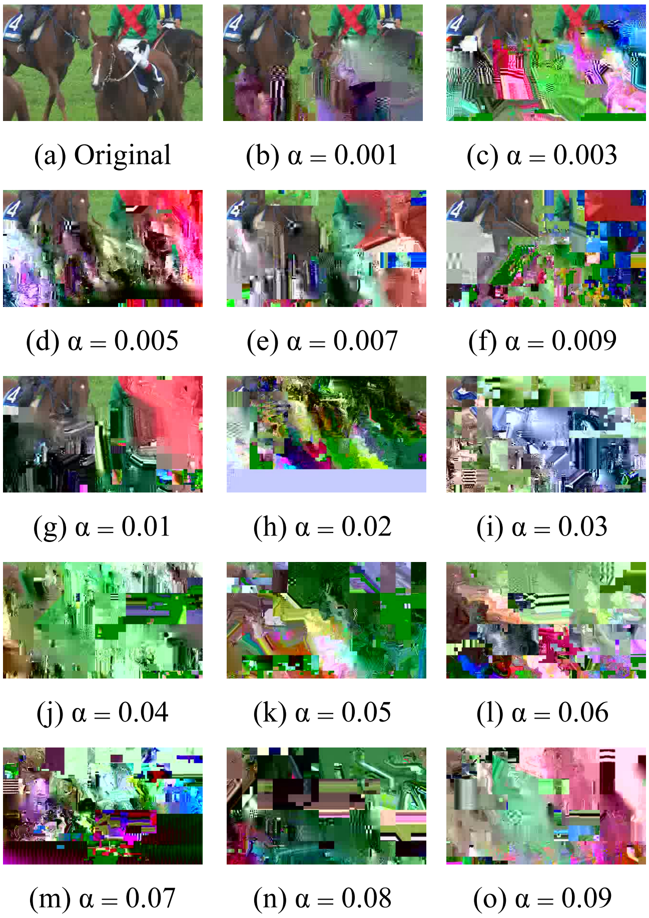


Figure 6. Mean PSNR of the test video under different values of α

4.3 Validity of Spatial Influence Model

Spatial influence model was designed to evaluate the spatial correlation of CU in a frame. Validity of the model for intra prediction is presented in Figure 8. The numerical SSIM of encrypted *Bosphorous_3840x2160* reflected that the perception of frame image was stronger destroyed while encrypting intra frame CUs with high spatial influence. Similar conclusion was drawn about spatial influence model for inter prediction, as shown in Figure 9. Spatial influence model is applicable to HEVC video of different resolutions, seen in Figure 10. On one hand, encryption of partial CUs (α=0.04) with low spatial influence aroused efficient scrambling, which reflected the necessity of selection on CU level with finer granularity. On the other hand, under the same encryption proportion, frames encrypted based on spatial influence model resulted in more decrease of SSIM.


(a) Original (b) α = 0.001 (c) α = 0.003

(d) α = 0.005 (e) α = 0.007 (f) α = 0.009

(g) α = 0.01 (h) α = 0.02 (i) α = 0.03

(j) α = 0.04 (k) α = 0.05 (l) α = 0.06

(m) α = 0.07 (n) α = 0.08 (o) α = 0.09

Figure 7. Frame 50 of encrypted *RaceHorses_416x240* under different values of α

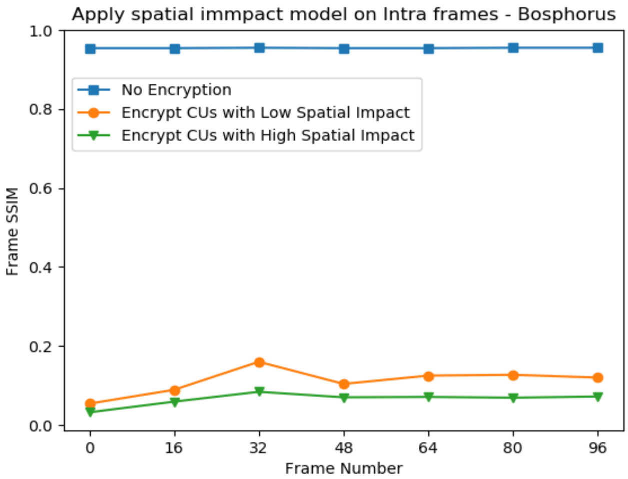


Figure 8. Apply spatial influence model to intra prediction

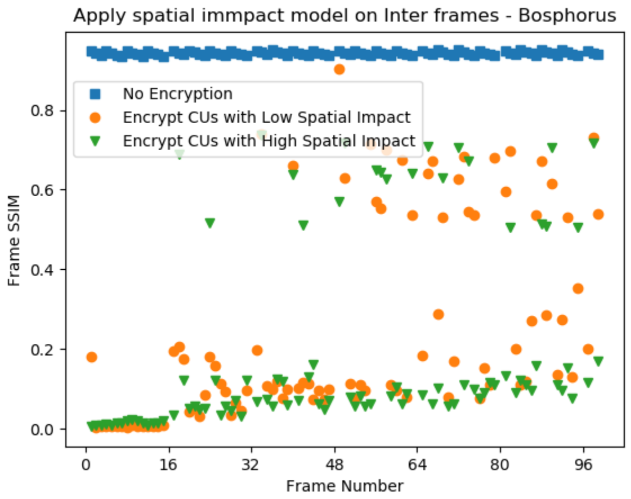


Figure 9. Apply spatial influence model to inter prediction

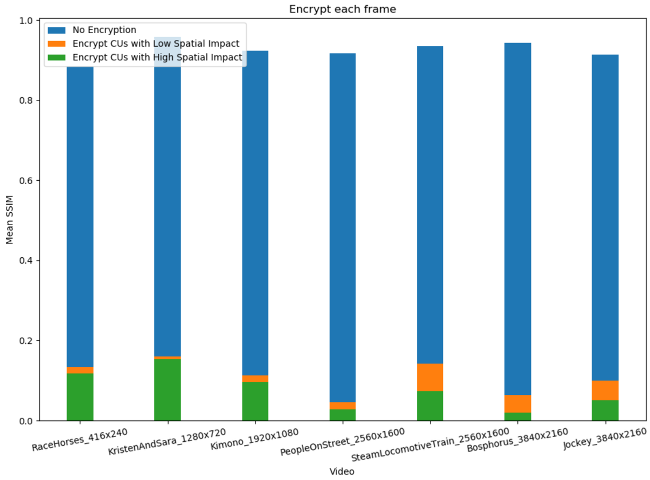
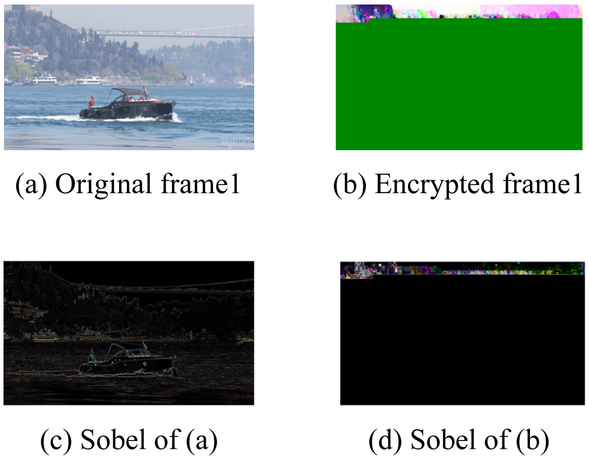


Figure 10. Compliance of the spatial influence model

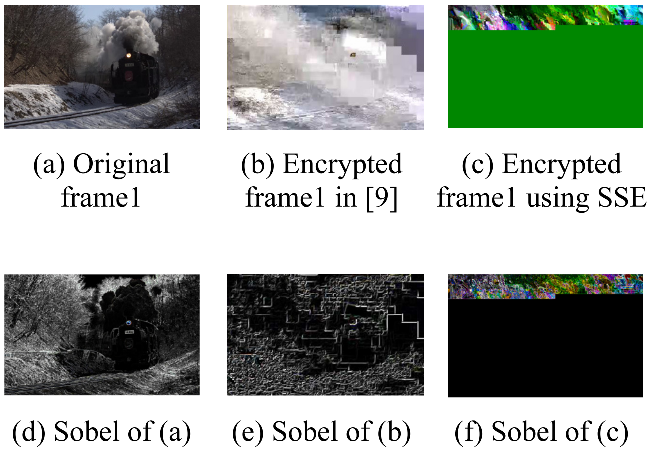
4.4 Perception Security

Imperceptibly of the encrypted video content is an essential requirement of the SE because the bits without encryption may lead to the leakage of visual feature of the encrypted video during decoding, like structural information, textual information and motion information. Figure 11 shows the perceptually secure encrypted 4K video by proposed SE. In order to emphasize the visual security of proposed SE, Sobel filter, an edge detection application, was performed on the protected frame. Obviously, structures and edges were almost completely concealed. Not only 4K videos, but also the video with 2560x1600 or smaller resolution can be effective protected by our method. In Figure 12, visual results of proposed SE were compared with that of the state-of-art scheme [9]. Since the distinguish on frames by the texture energy in [9], a bit of edge information in encrypted low energy frame was leaked in the Sobel filter result. In our method, encryption strength of the frame is adaptive, more total number of CU in the frame, more encrypted CU. Thus, more thorough protection for each frame was achieved.


(a) Original frame1 (b) Encrypted frame1

(c) Sobel of (a) (d) Sobel of (b)

Figure 11. Encrypted *Bosphorus3840x2160* using SSE


(a) Original frame1 (b) Encrypted frame1 in [9] (c) Encrypted frame1 using SSE

(d) Sobel of (a) (e) Sobel of (b) (f) Sobel of (c)

Figure 12. Encrypted *SteamLocomotiveTrain2560x1600* using SSE

4.5 Metric Analysis

4.5.1 Encrypted Video Quality Measures

PSNR (Peak Signal to Noise Ratio) and SSIM (structure similarity index) were used to measure the perceptual distortion of encrypted video. SSE results were demonstrated in Table 6, Table 7 and Table 8. Compared to [13], there is a little improvement in SSIM of encrypted videos as some CU are left unencrypted. But the SSIM still is further smaller than the threshold, and encryption time for each frame is less than 1/2800 of that in [13]. Encryption result of standard HEVC test videos were compared to the state-of-art scheme [9] and striking advantages were achieved owing to the finer selection on CU level. Numerical metrics further confirmed the visual security of encrypted video by SSE.

Table 6. Performance of SSE on 4K videos

4K video	GOP	PSNR		SSIM		Encryption time per frame (sec)
		Original	Proposed	Original	Proposed	
Bosphorus	IPPPP	39.81	4.56	0.944	0.038	0.75095
	IBBBP	39.96	5.96	0.945	0.056	0.84467
Honeybee	IPPPP	38.43	9.32	0.899	0.047	0.60285
	IBBBP	38.59	9.47	0.900	0.076	0.62633
Jockey	IPPPP	38.9	6.45	0.913	0.045	0.68284
	IBBBP	39.02	6.79	0.915	0.065	0.72684
ShakeNDry	IPPPP	36.69	8.34	0.893	0.027	0.93629
	IBBBP	36.79	8.48	0.895	0.035	0.83074
YachtRide	IPPPP	38.18	5.05	0.941	0.042	1.59774
	IBBBP	38.3	4.99	0.942	0.039	1.54537

Table 7. Comparison between [13] and proposed SSE

Resolution	Video (IPPP)	SSIM		PSNR		Encryption time per frame (sec)	
		[13]	Proposed	[13]	Proposed	[13]	Proposed
3840x2160	*Bosphorus*	0.186	**0.04**	10.83	**4.56**	2320	**0.75095**
1920x1080	*Jockey*	0.020	0.12	8.63	**7.7**	505	**0.17903**
1280x720	*FourPeople*	0.064	0.13	10.6	**8.77**	248	**0.08741**

Table 8. Distortion comparison between [9] and proposed SSE

Video (IPPP)	SSIM			PSNR		
	Original	[9]	Proposed	Original	[9]	Proposed
SteamLocomotiveTrain	0.94	0.472	**0.07**	37.38	9.36	**7.69**
PeopleOnstreet	0.92	0.309	**0.06**	34.52	10.23	**6.24**
Kimono	0.92	0.379	**0.08**	37.48	12.93	**11.04**
ParkScene	0.9	0.341	**0.06**	34.5	10.21	**9.76**
Cactus	0.89	0.476	**0.07**	34.66	12.23	**6.44**
Johnny	0.95	0.374	**0.19**	39.29	10.72	**8.5**
KristenAndSara	0.96	0.278	**0.21**	39.16	9.89	**8.98**
SlideEditing	0.99	0.391	**0.03**	38.88	9.17	**4.07**
BasketBallDrill	0.89	0.306	**0.13**	34.65	12.88	**8.83**
BQMall	0.92	0.386	**0.26**	33.81	10.42	11.28
BlowingBubbles	0.89	0.388	**0.11**	31.93	12.39	**8.65**
BQSquare	0.91	0.299	**0.07**	31.65	12.22	**6.99**
RaceHorses	0.89	0.308	**0.12**	32.01	11.95	**9.93**

4.5.2 Bitrate Analysis

Bitrate describes the rate of bits per second in the entropy stage of video coding. One of the common requirements of the video SE is minimizing the impact on the original bit rate. There are two entropy modes provided for syntax element bin strings in the CABAC, regular mode and bypass mode. Since the context model is only used in the regular mode, the encryption of bypass will cause no increase in the bitrate [4]. However, when the encryption is performed on the regular bins such as intra prediction mode, bitrate will be easily affected because of the modification of relative statics in the context model. To protect the structure information of the frame, intra prediction mode related syntax elements were encrypted in [9] and the proposed SE scheme. Compared with the average bitrate raise in [9], the maximum and minimum of which are 0.33% and 0.027% respectively, no increase of bitrate was caused by the proposed method because SSE uncoupled from the entropy coding totally. Thus, not only the static bitrate but also the compatibility of different HEVC codes were realized.

4.6 Computation Complexity

4.6.1 Encryption Cost

A metric for measuring the video encryption cost in the AES CTR mode was described as (7), proposed in [32] and adopted by [9]. R and K represent the number of encryption rounds and the number of keys respectively. Since C and D, which refer to the complexity of encrypting one syntax element and the cost of key schedule respectively, depended on the performance of hardware and software, only parameters R and K were studied in this part to evaluate the AES encryption cost. Under the circumstance of the same value of K shown in (8), the difference in encryption cost between [9] and proposed SE resulted from the discrepancy in the number of rounds, as given in (10) and (11). In the state-of-art scheme, 4 kinds of syntax elements, intra luminance prediction mode, intra chrominance prediction mode, MVD and QTC were encrypted, were encrypted. In the proposed SE, the plaintext consists of the bits of 9 kinds of syntax elements, including IPCM, intra luminance prediction mode, merge index, skip flag, inter prediction direction, reference frame index, MVD, AMVP index and QTC. However, by reason of the novel selection algorithm in CU level, encryption cost still decreased more than 75% in our method compared with that in the best scenario of the state-of-art SE. Here, the lower computation complexity of proposed SE was demonstrated in the theory.

$$E = C * R + D * K \tag{7}$$

$$K_{SSE} = K_{SOA} = N_{frame} \tag{8}$$

$$\beta[i] = \begin{cases} 0, \mathit{higi\ energy\ frame}\,[i] \\ 1, \mathit{low\ energy\ frame}\,[i] \end{cases} \tag{9}$$

$$\begin{aligned} R_{SOA} &= \sum_{i=0}^{N} \mathit{frame}^{-1}(1-\beta[i]*0.5)*N_{CU}[i]*N_{secSyntax} \\ &= \begin{cases} \sum_{i=0}^{N} \mathit{fram}^{-1}\ 4*N_{CU}[i], \mathit{worst\ scenario} \\ \sum_{i=0}^{N} \mathit{fram}^{-1}\ 2*N_{CU}[i], \mathit{best\ scenario} \end{cases} \end{aligned} \tag{10}$$

$$\begin{aligned} R_{SSE} &= \sum_{i=0}^{N} \mathit{fram}^{-1}\ \alpha * N_{CU}[i] * N_{sceSyntax} \\ &= \sum_{i=0}^{N} \mathit{fram}^{-1}\ 0.04 * N_{CU}[i] * 9 \\ &= \sum_{i=0}^{N} \mathit{fram}^{-1}\ 0.36 * N_{CU}[i] \end{aligned} \tag{11}$$

4.6.2 Encryption Time

For video systems, especially the system providing real-time video service, lower delay in encrypting will benefit both the server-side efficiency and user-side experience [33]. Experiments on 4K videos were seldom covered by previous SE schemes, except [13]. The comparison of encryption time in [13] and proposed SE is shown in Table 7. Even if AES CTR cipher was proved slower than RC6 in [13], our SE approach achieved more efficient performance as a result of improvement on SE framework and selection algorithm while ensuring perceived security. Table 6 lists the computational results of SSE on 4K videos. Clearly, mean delay of each frame (0.91 seconds) from SE is far smaller than the average encoding time (260 seconds) per frame of original HEVC videos.

In the analysis of other sequences within lower resolution, proposed SE showed an overall advantage as well on the overhead in HEVC RA mode and LD mode, seen in Table 9 and Table 10. Compared to the state-of-art scheme [9], the actual time spent on encryption in proposed SE was reduced by an average of 57%. Due to the weak spatial correlation of the objects on the structure and motion characteristics in *PeopleOnstreet*, spatial influence model lost the superiority under the setting of low encryption density, resulting in poor performance of the algorithm on the sequence. However, strength of the SE still was obvious in most sequences, saving an average of 50% total time taken on the encryption and decryption.

Table 9. Efficiency comparison between [9] and SSE in random access mode

Video (IPPPP)	Encrypt time (sec)		Decrypt time (sec)		Total time		
	[9]	Proposed	[9]	Proposed	[9]	Proposed	Reduction
SteamLocomotiveTrain	133.42	**49.362**	0.38	2.193	133.8	**49.74**	62.82%
PeopleOnstreet	314.1	432.947	0.381	8.027	314.481	433.33	-37.79%
Kimono	81.72	**22.89**	0.211	1.832	81.931	**23.10**	71.80%
ParkScene	76.5	**52.888**	0.19	0.823	76.69	**53.08**	30.79%
Cactus	66.94	**44.747**	0.167	2.353	67.107	**44.91**	33.07%
Johnny	26.79	**5.519**	0.062	0.067	26.852	**5.58**	79.22%
KristenAndSara	27.61	**6.379**	0.067	0.064	27.677	**6.45**	76.71%
SlideEditing	25.93	**13.251**	0.063	0.361	25.993	**13.31**	48.78%
BasketBallDrill	14.87	**5.777**	0.048	0.176	14.918	**5.83**	60.95%
BQMall	16.08	**6.206**	0.04	0.163	16.12	**6.25**	61.25%
BlowingBubbles	3.97	**1.87**	0.016	0.115	3.986	**1.89**	52.68%
BQSquare	3.46	**1.75**	0.012	0.045	3.472	**1.76**	49.25%
RaceHorses	5.17	**2.284**	0.016	0.199	5.186	**2.30**	55.65%

Table 10. Efficiency comparison between [9] and SSE in low delay mode

Video (IBBBP)	Encrypt time (sec)		Decrypt time (sec)		Total time		
	[9]	Proposed	[9]	Proposed	[9]	Proposed	Reduction
SteamLocomotiveTrain	132.64	**48.08**	0.311	2.027	132.951	**50.11**	62.31%
PeopleOnstreet	252.4	429.731	0.425	6.385	252.825	436.12	-72.50%
Kimono	88.04	**22.999**	0.307	0.687	88.347	**23.69**	73.19%
ParkScene	71.63	**51.251**	0.187	1.285	71.817	**52.54**	26.85%
Cactus	72.86	**42.421**	0.15	0.59	73.01	**43.01**	41.09%
Johnny	23.27	**5.491**	0.052	0.096	23.322	**5.59**	76.04%
KristenAndSara	24.75	**6.423**	0.054	0.173	24.804	**6.60**	73.41%
SlideEditing	22.23	**12.935**	0.054	0.098	22.284	**13.03**	41.51%
BasketBallDrill	15.35	**5.69**	0.04	0.157	15.39	**5.85**	62.01%
BQMall	15.9	**6.046**	0.038	0.161	15.938	**6.21**	61.06%
BlowingBubbles	3.76	**1.876**	0.013	0.129	3.773	**2.01**	46.86%
BQSquare	3.45	**1.804**	0.012	0.169	3.4626	**1.97**	43.02%
RaceHorses	5.24	**2.236**	0.015	0.123	5.255	**2.36**	55.11%

4.7 Cryptography Security

The symmetrical key of 128 bits was used in SSE and was distributed by the frame, resulting a key space containing $(2^{128})^{N_{frame}}$ possible keys for a video sequence, as expressed in [12]. It is large enough to prevent an adversary from using a brute-force attack to find the encryption key. The symmetrical key of 128bits was selected randomly for generating the persuade random sequence which plays the role of sub key for encrypting message one time. Plaintext space of the sequence was equivalent to the space of sub key, as demonstrated in [13]. Parameter M[i][j] refers to the ceil result of syntax element[i][j] bit length divided by 128. Under the circumstance of one-time pad, overall ciphertext has been theoretically unbreakable for known plaintext attack because of the uniqueness of subkey. [14] presents the ciphertext space of the SSE, where parameter L[i][j] refers to the bit length of syntax element[i][j]. Both the plaintext space and ciphertext space are closely related to the number of selected CUs, reflecting flexibility and controllability of the algorithm security.

$$KS = (2^{128})^{N_{frame}} \quad \textbf{(12)}$$

$$PS = \prod_{i=0}^{N} frame^{-1} \prod_{j=0}^{0.36} N_{CU[i]-1}\ 2^{128*M[i][j]} \quad \textbf{(13)}$$

$$CS = \prod_{i=0}^{N} frame^{-1} \prod_{j=0}^{0.36} N_{CU[i]-1}\ 2^{L[i][j]} \quad \textbf{(14)}$$

5 Conclusion

This paper details a novel HEVC SE algorithm named SSE, which fills the gap of research in the content protection problem of 4K-UHD video. SSE achieves the high efficiency on videos with high resolution and high frame rate by processing blocks with higher spatial correlation. Additionally, a user parameter, α, is derived from the algorithm to support flexible adjustment of encryption strength.

Proposed SE is not only for 4K videos, but also for backward resolutions. We implemented proposed SE and made rich comparison with state-of-art analysis given in [9] and [13]. The validity of designed models and the basis for the determination of parameter α are demonstrated in analysis part. Experiment result implies that SSE saves 50% of the execution time on average.

In future work, effort would be taken into the implement of SE module in video transmitting system. Additionally, proposed SE scheme would be adjusted for scalable video coding (SVC) standard [34] to accommodate more video scenarios.

Acknowledgements

Thanks are due to C. Yang, H. Li and J. Shen for valuable discussion. The Paper is supported by the Fundamental Research Funds for the Central Universities (CUC2019A009).

References

[1] K. Minemura, K. Wong, R. C.-W. Phan, K. Tanaka, A Novel Sketch Attack for H.264/AVC Format-Compliant Encrypted Video, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 27, No. 11, pp. 2309-2321, November, 2017.

[2] A. Noore, Secure Distribution of Heterogeneous Multimedia Content on the Internet, *International Journal of Internet Protocol Technology*, Vol. 1, No. 3, pp. 198-203, May, 2006.

[3] M. Jiang, X. Yi, N. Ling, Quantizer Control for H.264/AVC Streaming Over Networks, *Journal of Internet Technology*, Vol. 5, No. 3, pp. 301-304, July, 2004.

[4] M. Asghar Naveed, M. Ghanbari, An Efficient Security System for CABAC Bin-Strings of H.264/SVC, *IEEE Transactions on Circuits and Systems for Video Technology*,

Vol. 23, No. 3, pp. 425-437, March, 2013.

[5] K. Tai, M. Chen, X. Li, Content Adaptive Intra Prediction Algorithm for HEVC Encoder, *Journal of Internet Technology*, Vol. 17, No. 3, pp. 609-618, May, 2016.

[6] S. H. Bae, J. Kim, M. Kim, S. Cho, J. Soo Choi, Assessments of Subjective Video Quality on HEVC-Encoded 4K-UHD Video for Beyond-HDTV Broadcasting Services, *IEEE Transactions on Broadcasting*, Vol. 59, No. 2, pp. 209-222, June, 2013.

[7] G. Van Wallendael, A. Boho, J. De Cock, A. Munteanu, R. Van de Walle, Encryption for High Efficiency Video Coding with Video Adaptation Capabilities, *IEEE Transactions on Consumer Electronics*, Vol. 59, No. 3, pp. 634-642, August, 2013.

[8] B. Boyadjis, C. Bergeron, B. Pesquet-Popesc, F. Dufaux, Extended Selective Encryption of H.264/AVC (CABAC)- and HEVC-Encoded Video Streams, *IEEE Transactions on Circuits & Systems for Video Technology*, Vol. 27, No. 4, pp. 892-906, April, 2017.

[9] K. Thiyagarajan, R. Lu, K. El-Sankary, H. Zhu, Energy-Aware Encryption for Securing Video Transmission in Internet of Multimedia Things, *IEEE Transactions on Circuits & Systems for Video Technology*, Vol. 29, No. 3, pp. 610-624, May, 2018.

[10] Y. Tew, K. Minemura, K. Wong, HEVC Selective Encryption Using Transform Skip Signal and Sign Bin, *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Hong Kong, China, 2015, pp. 963-970.

[11] Y. Tew, K.-S. Wong, R. C.-W. Phan, Joint Selective Encryption and Data Embedding Technique in HEVC Video, *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Jeju, Korea, 2016, pp. 1-5.

[12] Z. Shahid, W. Puech, Visual Protection of HEVC Video by Selective Encryption of CABAC Binstrings, *IEEE Transactions on Multimedia*, Vol. 16, No. 1, pp. 24-36, January, 2014.

[13] A. I. Sallam, O. S. Faragallah, E.-S. M. El-Rabie, HEVC Selective Encryption Using RC6 Block Cipher Technique, *IEEE Transactions on Multimedia*, Vol. 20, No. 7, pp. 1636-1644, July, 2018.

[14] A. Mustafa, Hendrawan, Secure HEVC Video by Encrypting Header of Wavefront Parallel Processing, *2017 International Conference on Telecommunication Systems Services and Applications*, Lombok, Indonesia, 2017, pp. 1-4.

[15] W. Wang, M. Hempel, D. Peng, H. Wang, H. Sharif, H.-H. Chen, On Energy Efficient Encryption for Video Streaming in Wireless Sensor Networks, *IEEE Transactions on Multimedia*, Vol. 12, No. 5, pp. 417-426, August, 2010.

[16] R. N. Hole, M. Kolhekar, Robust Video Encryption and Decryption Using Selective Encryption, *2017 International Conference on Nascent Technologies in Engineering Field*, Navi Mumbai, India, 2017, pp. 1-4.

[17] B. Boyadjis, M.-E. Perrin, C. Bergeron, S. Lecomte, A Real-Time Ciphering Transcoder for H.264 and HEVC Streams, *IEEE International Conference on Image Processing*, Paris, France, 2014, pp. 3432-3434.

[18] B. Boyadjis, C. Bergeron, S. Lecomte, Auto-synchronized Selective Encryption of Video Contents for an Improved Transmission Robustness over Error-prone Channels, *IEEE International Conference on Image Processing*, Quebec, Canada, 2015, pp. 2969-2973.

[19] N. Sidaty, W. Hamidouche, O. Deforges, A New Perceptual Assessment Methodology for Selective HEVC Video Encryption, *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, Los Angeles, USA, 2017, pp. 1542-1546.

[20] R. Ye, H. Lan, Q. Wu, A Fractal Interpolation Based Image Encryption Scheme, *2018 IEEE International Conference on Computer and Communication Engineering Technology*, Beijing, China, 2018, pp. 291-295.

[21] J. Li, C. Wang, X. Chen, Z. Tang, G. Hui, C.-C. Chang, A Selective Encryption Scheme of CABAC Based on Video Context in High Efficiency Video Coding, *Multimedia Tools & Applications*, Vol. 77, No. 10, pp. 12837-12851, May, 2018.

[22] P.-C. Su, T.-F. Tsai, Y.-C. Chien, Partial Frame Content Scrambling in H.264/AVC by Information Hiding, *Multimedia Tools and Applications*, Vol. 76, No. 5, pp. 7473-7496, March, 2017.

[23] A. Mustafa, Hendrawan, Secure HEVC Video by Utilizing Selective Manipulation Method and Grading Level Model, *2017 International Conference on Wireless and Telematics*, Palembang, Indonesia, 2017, pp. 1-6.

[24] A. I. Sallam, E.-S. M. El-Rabaie, O. S. Faragallah, Efficient HEVC Selective Stream Encryption Using Chaotic Logistic Map, *Multimedia Systems*, Vol. 24, No. 4, pp. 419-437, July, 2018.

[25] M. Long, F. Peng, H.-Y. Li, Separable Reversible Data Hiding and Encryption for HEVC Video, *Journal of Real-Time Image Processing*, Vol. 14, No. 1, pp. 171-182, Janaury, 2018.

[26] H. Hofbauer, A. Uhl, A. Unterweger, Transparent Encryption for HEVC Using Bit-Stream-Based Selective Coefficient Sign Encryption, *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014, pp. 1986-1990.

[27] Q. Wang, G. Liu, Z. Liu, H. Liu, W. Zuo, N. Wang, Network Acknowledgement-based and Error-propagation-aware Importance Modelling for H.264/AVC Video Transmission Over Wireless Networks, *IET Communications*, Vol. 8, No. 15, pp. 2737-2750, April, 2014.

[28] P.-C. Chang, K.-L. Chung, J.-J. Chen, C.-H. Lin, T.-J, Lin, An Error Propagation Free Data Hiding Algorithm in HEVC Intra-coded Frame, *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Chiang Mai, Thailand, 2013, pp. 1-9.

[29] M. Huang, C. Yang, Y. Zhang, Selective Encryption of H.264/AVC Based on Block Weight Model, *2018 IEEE 18th International Conference on Communication Technology*, Chongqing, China, 2018, pp. 1368-1373.

[30] L. J. Hubert, R. G. Golledge, C. M. Costanzo, Generalized Procedures for Evaluating Spatial Autocorrelation, *Geographical Analysis*, Vol. 13, No. 3, pp. 224-233, September, 2010.

[31] J. Guo, G. Chen, L. Huang, H. Chao, Fast Intra Coding Algorithm Based on Image Texture Analysis for HEVC, *Journal of Internet Technology*, Vol. 16, No. 6, pp. 1033-1047, November, 2015.

[32] D. F. Garca, Performance Evaluation of Advanced Encryption Standard Algorithm, *2015 Second International Conference on Mathematics and Computers in Sciences and in Industry*, Sliema, Malta, 2015, pp. 247-252.

[33] K. R. Vijayanagar, J. Kim, Real-time Low-bitrate Multimedia Communication for Smart Spaces and Wireless Sensor Networks, *IET Communications*, Vol. 5, No. 17, pp. 2482-2490, November, 2011.

[34] M. H. Hajiesmaili, M. S. Talebi, A. Khonsari, Joint Multipath Rate Control and Scheduling for SVC Streams in Wireless Mesh Networks, *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 15, No. 4, pp. 239-251, May, 2014.

Biographies

Mengdie Huang was born in Nanjing, China. She received her bachelor Communication Engineering in HYIT, Huai'an, China in 2017. Currently, she is pursuing a graduate degree in electronic information engineering at the Communication University of China (CUC), Beijing. Her research interests are multimedia security, H.264 and HEVC, and secure video communication.

Cheng Yang received the Ph.D. degree in Cryptography from Beijing University of Posts and Telecommunications. He is currently a Professor and Head of the Multimedia Security and Smart Interaction Lab in Communication University of China, Beijing.

Hao Li received the M.S. and Ph.D. degree in electronics and communication engineering from Communication University of China. Now, he is a postdoctor in Da Hengqin Science and Technology Development Co.,Ltd. His current research interests include digital right management, cloud security, attribute-based encryption.

Jian Shen received the B.E. degree from Nanjing University of Information Science and Technology, Nanjing, China, in 2007 and the M.E. degree in Computer Science from Chosun University, Gwangju, Korea, in 2009. Since 2009, he is working toward the Ph.D degree in Computer Science from Chosun University, Gwangju, Korea. Currently, he is a professor at Nanjing University of Information Science and Technology. His research interests include network security, security systems, mobile computing and networking, ad hoc networks and systems, and ubiquitous sensor networks.