I. J. Computer Network and Information Security, **2016, 4, 1-13**
Published Online April 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2016.04.01

Low Complexity Multimedia Encryption

Karthik. Thiyagarajan
Dalhousie University/Electrical and Computer, B3J3L3, Halifax, Canada
E-mail: kr318490@dal.ca

Kamal El-Sankary
Dalhousie University/Electrical and Computer, B3J3L3, Halifax, Canada
E-mail: Kamal.El-Sankary@dal.ca

Yongsheng Wang
Queens University/ECE, Belfast BT7 1NN, United Kingdom
E-mail: ywang26@qub.ac.uk

Issam Hammad
Dalhousie University/Electrical and Computer, B3J3L3, Halifax, Canada
E-mail: issam@dal.ca

Abstract—Selective encryption algorithms have been proposed to encrypt syntax elements such as intra prediction modes, the sign bit of nonzero DCT coefficients, along with the sign bit of motion vectors. These syntax elements are sensitive enough to produce effective scrambling effect with a relative low computational cost. In this paper, a novel scheme is proposed to further optimize the computational overhead incurred by the encryption for energy critical multimedia applications. The proposed scheme adjusts the selection of syntax elements to be encrypted according to the scene transitions within adjacent video frames. The ratio of intra-coded macroblocks in inter (P and B) frames is calculated and compared with an adaptive threshold value to detect the scene transitions. Furthermore, based on statistical analysis for a few video sequences, a dynamic threshold model to detect the scene transition is proposed. When there is a scene transition between the previous video frame and the current video frame, intra prediction modes and the sign bit of DCT coefficients in the current frame are chosen as syntax elements to be encrypted, whereas in the absence of a scene transition, the sign bit of motion vectors is chosen as the only sensitive syntax elements to be encrypted. Experimental results show that compared with previous work in this field, the proposed scheme can efficiently lower the computational cost incurred by the encryption while maintaining a similar perceptual scrambling effect.

Index Terms—Selective Encryption, Energy Critical Multimedia Devices, Scene Transition Detection, Intra Prediction Modes, Nonzero DCT Coefficients, Motion Vectors, Computational Cost, H.264/AVC.

I. INTRODUCTION

In broadcasting technology, with the proliferation of wireless multimedia applications, video compression standards have gained popularity. The advanced video compression standard, H.264/AVC, has been widely adopted in industry because of its significant improvement on the compression performance compared with previous MPEG compression standards [1] [2]. Meanwhile, the secure transmission of video streams is difficult to achieve in energy critical multimedia devices such as video surveillance systems, smart phones and wireless network cameras. Selective encryption is one of the most promising techniques for multimedia content protection, since it only involves a relative low computational cost while achieving an effective perceptual scrambling effect [3].

Several selective encryption schemes for H.264/AVC have been proposed in literature. Table I shows the ciphered data, advantages and disadvantages with respect to the contribution of this paper. Lian *et al.* [4] proposed to encrypt intra prediction modes, the sign bit of nonzero DCT (Discrete Cosine Transform) coefficients, and the sign bit of motion vectors. This method can effectively provide the perceptual protection without an impact on the compression ratio. Wang *et al.* [5] modified the encryption scheme in [4] to a user tunable encryption by employing three control factors for the three syntax elements to be encrypted. This scheme provides the flexibility to adjust the perceptual scrambling effect according to different multimedia applications.

Although previous work on selective encryption has reduced the computational cost compared with naive encryption [3], most of these previous work didn't provide the feasibility to further optimize the computational cost according to different applications, which could be very attractive for energy constrained applications, like wireless multimedia devices. There are only very few works which consider the necessity to further lower the computational cost of encryption for energy critical multimedia encryption. Wang *et al.* [8]

extended the work in [4] to make it more suitable for wireless sensor networks, which proposed to only encrypt frames that were highly dependent on by descendent frames. Their method can reduce the computational overhead caused by the encryption and reduce error propagation caused by an encryption technique.

However, the lower dependent frames are not encrypted and thus can leak the video information which significantly affects the security of their method. Zhao *et al.* [9] improved the method in [4] to lower the computational cost of the encryption. In this work, according to the texture complexity and motion intensity in each frame, certain amount of code words in that frame are encrypted, which can provide variant levels of the scrambling effect.

Table 1. Literature Review: L-Low, H-High, M-Moderate

Work	Intra Modes	Sign bit Residual	Sign bit Motion Vector	overhead cost	security
[4,5]	✓ ✓ ✓ ✓ ✓			M	H
[6]	Encrypts all the three syntaxes unequally based on object segmentation.			Lower than [4][5]	M
[7]	Encrypts all the three syntaxes based on inter frame dependency.			Lower than [4][5]	M
[8]	Encrypts all the three syntaxes, the syntaxes are selected based on frame level dependency and priority.			Lower than [4][5]	M
[9]	Encrypts all the three syntaxes, syntaxes a selected based on texture complexity and motion intensity.			Lower than [4][5]	M
Objectives in this paper	Encrypt all three syntaxes, syntaxes are selected based on scene transitions in inter frames.			Lower than [4][5]	H

If a frame just contains low texture complexity and motion intensity, only a small percentage of code words is encrypted, which weakens the security of the encryption in such frames. Zhao *et al.* [6] also developed an object based encryption algorithm, where the objects are segmented and unequally encrypted along with the background. Although the computational overhead is further reduced compared with work in [4], it is still too high for real time applications. Shen *et al.* [7] proposed a scheme which selects syntax elements for encryption based on statistical analysis of inter frame dependency between adjacent frames. The drawback of this method is similar to work in [8], as the lower dependent frames can leak video information. Those video encryption techniques listed in Table I can alleviate the encryption effect over compression ratio and format compliance, as the sign bits of nonzero DCT coefficients are encrypted. Most of the previous work on video encryption has mainly focused on the security in the application layer, and has less concern over optimizing the encryption cost for energy constrained wireless devices. Furthermore, in all the aforementioned works, various video encryption algorithms have been proposed to lower the computational overhead incurred by encryption but with a tradeoff over security. Previous works fail to provide an encryption algorithm with low computational overhead and strong scrambling effect. Therefore, this paper proposes a selective encryption with low computational overhead while maintaining the same security.

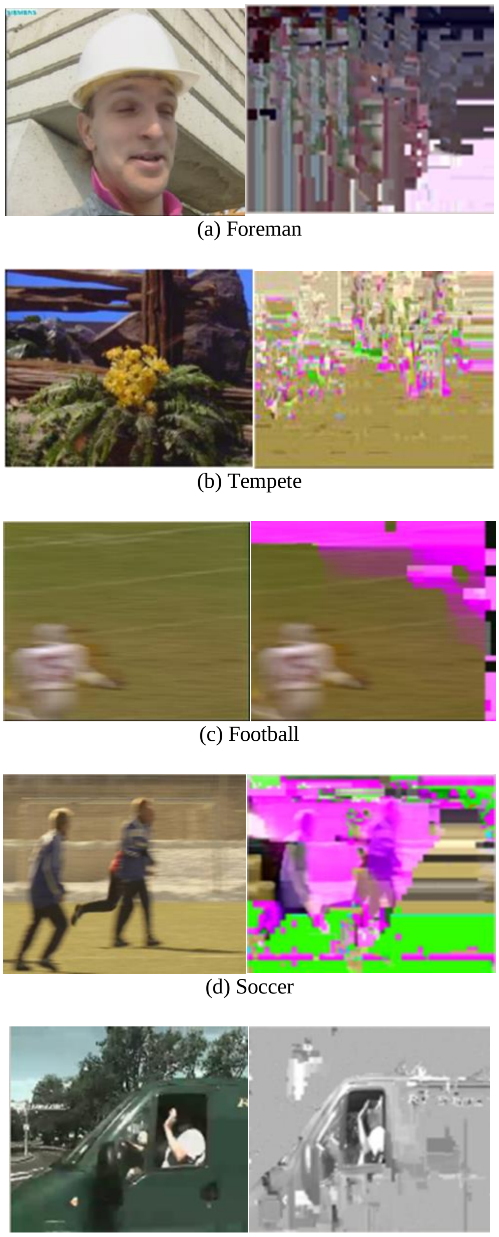
Due to the limitation of computing resources and available energy, selective encryption fits well in wireless multimedia applications [10]. Compared with the basic encryption which encrypts the whole video stream, selective encryption can efficiently reduce the computational overhead of the encryption by only ciphering the sensitive information which is only a small portion of a video stream.

Generally, wireless devices only have very limited energy resource, since most of such devices are battery powered [10]. Many recent efforts attempt to scale down the power consumption and encoding compression cost to suit wireless multimedia devices and networks [11]. However, the encryption scheme along with video compression increases the computational overhead and power consumption. Hence, the proposed work in this paper aims to reduce the encryption overhead and to make it more suitable for energy constrained multimedia devices. Distinguished from most of the aforementioned works, this paper proposes a new low complexity secure multimedia communication framework with a very similar scrambling effect as the one produced by the methods in [4][7]. A further lower encryption overhead can be effectively achieved to contribute the power efficiency for those wireless applications.

In summary, the main contribution of this paper includes:

1) A dynamic threshold model is proposed to adaptively calculate the threshold for detecting scene transitions in video frames. The proposed dynamic model overcomes the difficulty of choosing the suitable thresholds for different video streams.
2) The dynamic scene transition detection algorithm is used to identify frames with a scene transition and in these frames, intra-coded macroblocks are encrypted. In the frames without a scene transition, the sign bit of motion vectors are chosen to be encrypted. Intra-coded macroblocks that do not contribute to the scene transition can be left unencrypted to reduce the computational cost incurred by the encryption.

The rest of the paper is outlined as follows – In section II, we perform preliminary analysis. A dynamic threshold model for scene transition detection is presented in section III. In section IV, a new selective encryption scheme is proposed to optimize the encryption overhead while maintaining the resulting scrambling effect. Section V, shows details of the experimental results and analysis. Finally, the conclusions and future work are given in section VI.



(a) Foreman

(b) Tempete

(c) Football

(d) Soccer

(e) Horse Cab

Fig.1. Information Leaked During Gradual Scene Transition

II. Preliminary Analysis

To investigate the sensitivity of syntax elements, five consolidated video sequences, "Foreman", "Tempete" "Soccer", "Football" and "Horse Cab", were chosen at QCIF(Quarter Common Intermediate Format 176 x 144)) resolution under the mainline profile with QP=18 and 4:2:0 sampling format.

Encoder behaviour for preliminary analysis: The scene transition between frames may occur anywhere within a GOP (Group of Pictures). When there is an abrupt scene transition between frames, the first frame of the new scene is encoded as an I-frame since there is no temporal dependency to be exploited between the current frame in the current scene and previous frames in the previous scene. If certain gradual (non-abrupt) scene transition happens between frames, quite a lot of these frames are P or B frames. Especially, those objects and scenes that never appear in previous frames are very likely to be intra-coded, since under this scenario intra coding leads to a more efficient compression performance compared to the inter coding[12]. If there is no scene transition, only few of macroblocks are intra-coded and most are inter-coded since the latter can provide a better coding performance.

For preliminary analysis, the syntax elements chosen for encryption are intra prediction modes and sign bit of residuals in all the I frames, and motion vectors in P and B frames. AES (Advanced Encryption Standard) in counter mode was used to encrypt the chosen syntaxes. The results of preliminary analysis are shown in Fig. 1. Fig.1. (a) and (b) demonstrate that, in inter frames without scene transitions, ciphering the sign bit of motion vectors is enough to provide effective scrambling. However in case of inter frames with scene transitions, there are lot of intracoded macroblocks encoded. Therefore, as shown in Fig.1(c, d and e) even after encrypting the sign bit of motion vectors, some visual information can still be perceived in P and B frames.

A. Sample videos with no scene transition

$$\% = \frac{Number\ of\ intra\text{-}coded\ macroblocks\ in\ a\ frame}{Total\ number\ of\ macroblocks\ in\ a\ frame}$$

Even though there is no scene transition in "Foreman" and "Tempete", the percentage of IMB (intra-coded macroblocks) in P and B frames are 4.53% and 4.04%, respectively, as shown in Table II. Leaving the intra-coded macroblocks in P and B frames without scene transitions unencrypted does not affect the secrecy of the video and reduces the number of bits to be encrypted by 4.86% and 5.89% respectively.

B. Sample videos with gradual scene transition

In the other three video sequences, "Football", "Soccer" and "Horsecab", the percentage of intra-coded macroblocks (P and B frames) is 16.68%, 13.29% and 6.63% respectively as shown in Table II. However, the percentage of intra-coded macroblocks that contribute to scene transitions are 3.26% for "Football", 1.40% for "Soccer", and 0.11% for "Horse Cab". Encrypting intracoded macroblocks only in frames with scene transitions and reduces the number of bits to be encrypted by 11.2% for "Football", 13.01% for "Soccer", and 12.22% for "Horse cab".

Table 2. Percentage of Intrablocks and Intrablock (in bits)

Videos	Average Percentage		Frames with scene transition percentage of		Frames without scene transition percentage	
	IMB	IMB-Bits	IMB Bits	IMB-	IMB	IMB-Bits
Foreman	4.53	4.86	-	-	4.53	4.86
Tempete	4.04	5.89	-	-	4.04	5.89
Football	16.68	14.67	3.26	3.44	13.32	11.20
Soccer	13.29	18.54	1.40	5.53	11.89	13.01
Horse cab	6.63	15.32	0.11	3.10	6.52	12.22

III. Proposed Dynamic Scene Transition Detection

Several scene transition detection algorithms have been proposed in literature. Most previous schemes define a scene transition parameter and compare it with a fixed threshold [13] [14]. If the parameter is above the threshold, a scene transition is detected. However, a fixed threshold value cannot perform very well for different video sequences due to the diversity of the video content. The key problem for the scene transition detection is how to determine an optimal value for the threshold while different video contents contain substantial different perceptual information.

Threshold settings for a particular set of sequences is likely to be different for another set of sequences. Selection of the optimal threshold may even require a repeated process of trials, which will significantly affect the real-time performance.

Table 3. Symbols and Definitions

Symbols	Definitions
N_I	Number of intra-coded macroblocks
N_{skip}	Number of skipped macroblocks
T_o	Total number of macroblocks in P frame
N_b	Number of backward inter-coded macroblocks
N_f	Number of forward inter-coded macroblocks
C, D	Constants depending on CPU cycles
K	Number of keys used
N	Number of blocks encrypted
E	Encryption cost
B_{ipm}	Bit number encrypted in intra-coded macroblocks
B_{snc}	Bit number of encrypted sign bits of DCT coefficients
B_{mv}	Bit number of encrypted sign bits of motion vectors
N_{intra}	Number of intra-coded frames
N_{inter}	Number of inter-coded frames
N_{sc}	Number of inter-coded frames detected with scene transition
MR_p and MR_b	Macroblock Ratios of P and B frame
TH	Threshold for the scene transition detection

Work in [15] proposed to detect scene transition based on the intensity levels of motion vectors. This approach can detect scene transition due to complex motion but fails in case of relatively gradual scene transitions with a low motion intensity. Work in [16] adopted the SAD (Sum of Absolute Differences) to detect scene transitions. However, calculating the SAD values is computationally intensive.

Furthermore, fames with high intensive motion can be easily false-detected. An approach in [17] is proposed to decide a dynamic threshold technique based on the bit rate fluctuation.

This method can efficiently detect gradual scene transitions and transitions due to high intensive motion but only suitable for the variable bit rate coding. To solve the threshold selection issue, an adaptive threshold calculation is proposed in this section, which can dynamically determine the threshold to detect scene transitions in different videos with varying characteristics. Symbols and definitions used in section III and IV are shown in Table III.

A. Adaptive threshold scene change detection algorithm

Inter prediction occurs only when neighboring frames have a strong temporal correlation. It is obvious that this correlation is reduced when a scene transition occurs. Hence, it can be expected that when there is a scene transition, more intra-coded macroblocks appear in P and B frames. Gradual scene transitions can be represented as,

$$\frac{t}{T}Y + (1-\frac{t}{T})X, 0 \le t \le T \tag{1}$$

Where t is the individual frame, X and Y are frames that denotes the start and end of a scene transition. T is the duration of the scene transition from frame X to Y. The ratio of intra-coded macroblocks and motion vectors in P and B frames are calculated, and compared to a threshold value to detect the scene transition. To detect any scene transition in P frames, the scene transition parameter MR_p is calculated and compared with the threshold TH. Similarly, to detect a scene transition in B frames, the scene transition parameter MR_b is calculated and compared with the threshold TH. TH for b frame is slightly lesser than P frame. B frames have two reference frames while P frames only have one reference frame, and B frames are nearer to their reference frames than P frames, B frames are easier to be intercoded, and hence require a smaller gradual scene change threshold. If the scene transition parameter is higher than the threshold, a scene transition is detected. Otherwise, it indicates the absence of a scene transition. Gradual scene transitions in P and B frames can be detected by the following macroblock ratios:

$$P\ Frame : N_{sc} = \begin{cases} Scene\ Change\ , MR_p > TH \\ No\ scene\ change, MR_p \le TH \end{cases}$$

$$Here, MR_P = \frac{N_I - N_{Skip}}{T_o} \tag{2}$$

$$B\ Frame: N_{sc} = \begin{cases} Scene\ Change, MR_b > TH \\ No\ scene\ change, MR_b \le TH \end{cases}$$

$$Here, MR_b = \frac{N_f - N_{Skip}}{N_b - N_{Skip}} \quad (3)$$

To adaptively adjust the threshold value, local statistical properties in a video sequence are used. The dynamic threshold selection for the ith inter frame can be calculated by the empirical mean value of macroblock ratios (MR_p /MR_b), of all the previous frames. The mean M_i is given as,

$$M_i = \frac{1}{i}\sum_{0}^{i} MR(i) \quad (4)$$

The scaling factor S in the equation (5) is used to determine the dynamic characteristic of the threshold value. A smaller S avoids missed detections. If S takes higher values, the threshold value becomes more rigid, which results in less frames to be detected with a false scene transition.

$$S = \frac{1}{MR(i)^{\frac{1}{x}}} \quad (5)$$

Parameter S is required to lower the threshold immediately after scene transition (shown in Fig. 2). x=1; Threshold is lowered abruptly, a lot of missed detection. x=3; Threshold is lowered smoothly, a lot of false detection. x=2; is the optimal value.

The threshold value for the ith inter frame is defined as follows:

$$TH_i = M_i \bullet S \quad (6)$$

B. Accuracy of the proposed scene transition detection

To validate the accuracy of the adaptive scene transition detection algorithm, three video sequences, “Soccer”, “Football”, and “Horse Cab”, were chosen. The coding sequence IPBPB… and the main profile are used in the test. Recall and precision have been used to evaluate the proposed scene transition detection. Recall and Precision are defined as in the equation (7) and (8), and denoted as R and P in Table IV and V,

$$Recall = \frac{N_c}{N_c + N_m} * 100 \quad (7)$$

$$Precision = \frac{N_c}{N_c + N_f} * 100 \quad (8)$$

Where N_c, N_m, and N_f are the number of correct, miss and false detection, respectively. N_T is the actual number of scene transitions present in the video. The above equation shows that N_m and N_f are inversely related to the accuracy of recall and precision, respectively. The fixed threshold used for comparison was chosen empirically after analyzing intra-coded macroblock ratios in all inter frames. Table IV and V show that the performance of the adaptive threshold scene transition detection algorithm is significantly improved compared to the algorithm with fixed threshold. Fig. 3. shows that the empirical analysis for fixing threshold can be avoided and the proposed adaptive threshold technique can be used to more accurately detect scene transitions.

As shown in Fig. 3, the adaptive threshold lowers its value for frames with scene transitions and vice versa for frames without scene transitions. Even though there are very few missed scene transitions, the critical frames due to scene transitions are always encrypted, which helps to maintain the security level of the proposed method while effectively reducing the number of syntax elements to be encrypted.

Table 4. Recall and Precision with Adaptive Threshold

Videos	N_T	N_c	N_m	N_f	R(%)	P(%)
Football	15	15	0	2	100	88.2
Soccer	38	36	2	3	94.7	92.3
Horse Cab	37	36	1	0	97.2	100

Table 5. Recall and Precision with Adaptive Threshold

Videos	TH	N_T	N_c	N_m	N_f	R(%)	P(%)
Football	0.35	15	14	1	6	93.3	70
Soccer	0.40	38	39	9	7	76.3	80
Horse Cab	0.38	37	33	4	2	89.1	94

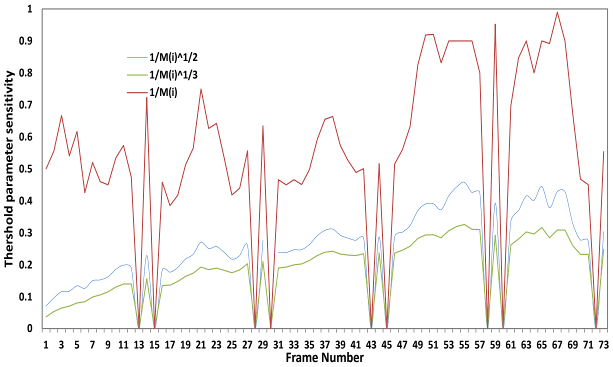


Fig.2. Threshold Sensitivity

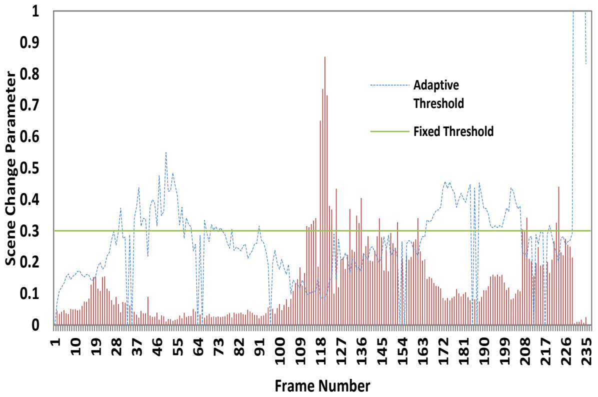


Fig.3. Fixed Threshold vs Adaptive Threshold

IV. Proposed Selective Encryption Scheme

According to the analysis in section I, the following conclusions can be made:

1) Unencrypted intra-coded macroblocks in P and B frames perceptually leak information, only during scene transitions. Thus, intra-coded macroblocks in frames with scene transitions are encrypted.
2) In order to keep low encryption overhead, the fewer the code-words of each syntax element are encrypted, the lower the overhead of the encryption is. Hence, intra-coded macroblocks in frames without scene transitions are left unencrypted.

The proposed encryption algorithm chooses syntax elements in inter frames for encryption based on scene transition detection. As shown in Fig. 4, In case of video frames with a scene transition, intra prediction modes and sign bits of non-zero DCT coefficients are chosen to be encrypted, whereas in case of no scene transition, only sign bits of motion vectors are chosen for encryption

Algorithm: Scene transition detection for P and B frames, selecting appropriate syntaxes for encryption.

```
1)  INPUT: Video Frame.
2)  IF current frame is an I frame
       Encrypt intra coded macroblocks,
       sign bit of DCT coefficients, goto
       1.
3)  [Detection of scene changes inter
    frames]

    BEGIN
          Calculate MR_p (P frame)/MR_b (B
          frame)
          Calculate mean $M_i$, scaling factor
    S
          IF Current Frame is P frame
          Update threshold $TH_i$;
          ELSE
          $TH_i$=$TH_i$-0.20
          Store MR(i) (for calculating next
          threshold).
    END
4)  IF scene change parameter MR_b /MR_p less
    than TH, encrypt sign bit of motion
    vectors

    ELSE Encrypt intra coded macroblocks,
    sign bit of DCT coefficient.

    END

5)  IF End of frame

       END algorithm.
    ELSE Go to step 2.
```

In H.264/AVC, when using CAVLC entropy coding, the intra prediction modes are encoded with exgolomb codes [18]. The intra prediction mode IPM consists of X bits of zero, one ‘1’ bit and X bits of information I which can be represented as

$$[X\ Zero's][1][M\ bits\ I] \tag{9}$$

In equation (9), it is required to encrypt the M bits of information “I” for securing the low resolution spatial information. Hence,

$$B_{ipm} = I \tag{10}$$

CABAC uses TU (truncated unary code) for encoding intraprediciton modes (intra_chroma_pred_mode) [19]. The suffix bits of TU codes are encrypted. H.264 offers two types of entropy coding to encode the quantized DCT coefficients CAVLC and CABAC. In case of nonzero quantized coefficients, as mentioned before, sign bits are chosen for encryption, Under CAVLC, the quantized coefficients are grouped as a series of syntax elements and runs of zero’s as shown in Fig. 5. The third syntax element is the level coefficient, which is the absolute value of the quantized coefficients. The level coefficients can be represented as <Prefix><Suffix>. If |a| is the magnitude of the DCT coefficient then,

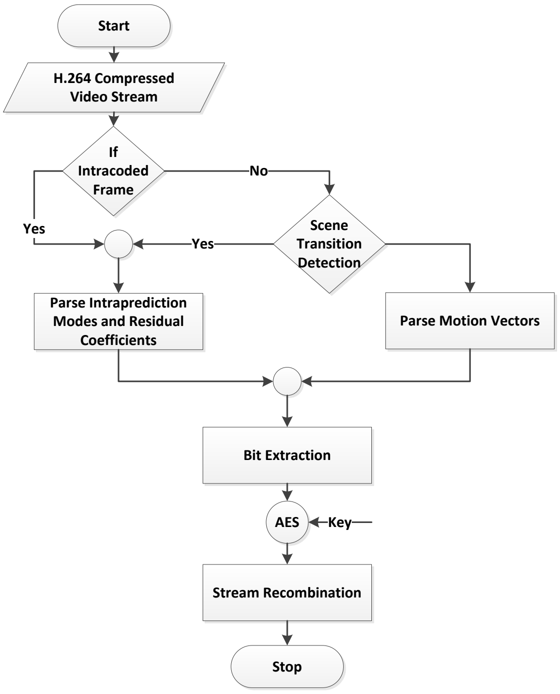
$$Prefix = <Zeros,1>, Suffix = <(|a| - 1), LSB> \tag{11}$$

In the above equation, LSB is the sign bit, the only information that is to be encrypted. Therefore,

$$CAVLC:\ B_{snc} = LSB \tag{12}$$

Whereas, if the entropy coding opted by the encoder is CABAC, then the binary arithmetic coding is adopted; unlike CAVLC, this adopts run length coding. The syntax for CABAC encoded quantized coefficients is shown in Fig. 6. Here, the level consists of two syntax elements

<Coef_abs_level_minus1, Coef_sign_flag> (13)



Start
H.264 Compressed Video Stream
If Intracoded Frame
No
Yes
Scene Transition Detection
Yes
Parse Intraprediction Modes and Residual Coefficients
Parse Motion Vectors
Bit Extraction
AES
Key
Stream Recombination
Stop

Fig.4. Functional Flowchart of Proposed Selective Encryption

Coeff_token
Sign_T1
Levels
Total Zeros
Run before Zeros

Fig.5. CAVLC-syntax

Significantbit_Coeff_Flag
Last_Significant_Coeff_Flag
Coeff_abs_level_minus1
Coeff_Sign_Flag

Fig.6. CABAC-syntax

The Coef_sign_flag is the syntax element that represents the sign bit of the quantized coefficients. Hence,

$$CABAC: \quad B_{snc} = Coef\_sign\_flag \tag{14}$$

As mentioned earlier, it is necessary to encrypt the sign bit of motion vectors for securing the temporal information. The motion vectors can be represented as,

$$< M\ Zeros >< 1 >< Suffix\_info > \tag{15}$$

The LSB of the Suffix_Info represents the sign bit of the motion vector. Hence,

$$CAVLC: \quad B_{mv} = LSB\,of\ (Suffix\_info) \tag{16}$$

In case of CABAC, the MVD can be represented as prefix and suffix in UEG3 (Unary Exgolomb-3) binstring. The sign bit is present in suffix only if the following conditions hold

$$< Pre - Suffix, Sign\,bit > = \begin{cases} \dfrac{0 < |MVD| < 9}{|MVD| \geq 9} \end{cases} \tag{17}$$

The entire suffix or the LSB sign bit can be encrypted to provide temporal secrecy.

$$CAVLC: \quad B_{mv} = sign\,bit\ of\ (UEG3\,binstring) \tag{18}$$

The encryption cost [20] of the AES algorithm in CTR mode can be represented as

$$E = C * N + D * K \tag{19}$$

Where, C is the cost of encrypting a single block. D is the cost of converting one key to a key schedule. The exact values of C and D depend on the hardware and software. Let E_T be the encryption data rate or cost of the selective encryption algorithm when, the intra prediction modes, sign bit of coefficients and sign bit of motion vectors are encrypted in all the frames. Let E_P be the encryption data rate or cost of the proposed algorithm. Then the number of rounds N_T and N_P in E_T and E_P respectively are,

$$N_P = \sum_{0}^{M-1} B_{ipm}[N_{Intra} + N_{Sc}] + \sum_{0}^{M-1} B_{snc}[N_{Intra} + N_{Sc}] + \sum_{0}^{M-1} B_{mv}[N_{Inter} - N_{Inter}]$$

$$N_T = \sum_{0}^{M-1} B_{ipm}[N_{Intra} + N_{Inter}] + \sum_{0}^{M-1} B_{snc}[N_{Intra} + N_{Inter}] + \sum_{0}^{M-1} B_{mv}[N_{Inter}] \tag{20}$$

Comparing N_T and N_P we observe that,

$$E_T > E_P \tag{21}$$

	N_T		N_P
IPM	$N_{Intra} + N_{Inter}$	>	$N_{Intra} + N_{Sc}$
SNC	$N_{Intra} + N_{Inter}$	>	$N_{Intra} + N_{Sc}$
MVD	N_{Inter}	>	$N_{Inter} - N_{Sc}$

It is to be noted that in spite of any scene transition in an I frame, all the sensitive syntax elements are encrypted. The objective is to reduce the number of code words selected in inter coded frames. Hence, scene transition detection is applied only for the P and B frame. Abrupt scene transitions are encoded as an I frame; therefore, abrupt scene cut detection is not implemented in our algorithm. The sensitive syntax elements are often considered to be variable length, to achieve the length-kept encryption, AES in CTR mode is used.

V. EXPERIMENTAL RESULTS

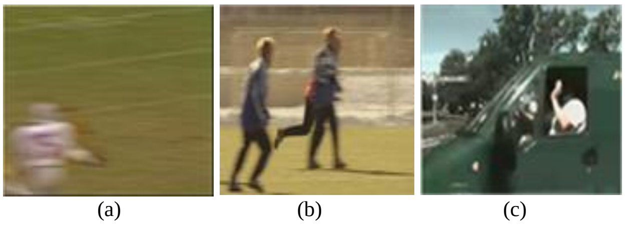
Five video sequences, “Foreman”, “Tempete” “Football”, “Soccer” and “Horse cab”, were chosen to evaluate the proposed algorithm. Here, the proposed method is compared with the approaches in [4] and [5] with respect to perceptual security and computational cost. The perceptual security is evaluated in terms of PSNR and SSIM, and the total decoding time is used to evaluate the computational cost. The test bench to implement the proposed methods and the previous approaches in [4] and [5] is specified in Table VI. Two types of GOP structures have been used. Both of them are extensively adopted in wireless multimedia applications [21].

Table 6. Video Test Bench

Experimental Set Up	
Processor &RAM	2.4GHz CPU, 4Gb RAM
H.264/AVC Reference Software Number of Frames Encoded 260	JM (Joint Model) 18.5
Format CIF	
GOP Type I	IPBPBPBPBP………The first frame is I frame followed by P and B frames. Inserts I frame during an abrupt scene transition.
	Entropy Coding :CABAC
GOP Type II	IPPPPPPP…….with Intra refresh Mode
	Entropy Coding: CAVLC
Frame Rate	30 fps

A. *Perceptual Security*

If a video encrypted by an encryption algorithm is decoded without the correct secret key and the resulting scrambling effect of the decoded video is too chaotic to be understood, it is considered that this encryption algorithm is of high perception security. Fig. 7 shows the scrambling effects of the proposed methods for different sample videos. In Fig. 7, sub figures a, b and c shows the original inter frames and the sub figures d, e and f shows the same frame, encrypted with the proposed encryption algorithm. Fig. 7 (d, e, and f) show that the inter frames with gradual scene transition, they are perceptual secure, because the scene transitions are detected and intracoded macroblocks are encrypted. In these three figures, Fig. 7 (d), (e), (f), the motion vectors are not encrypted to reduce the bit number of syntax elements to be encrypted.



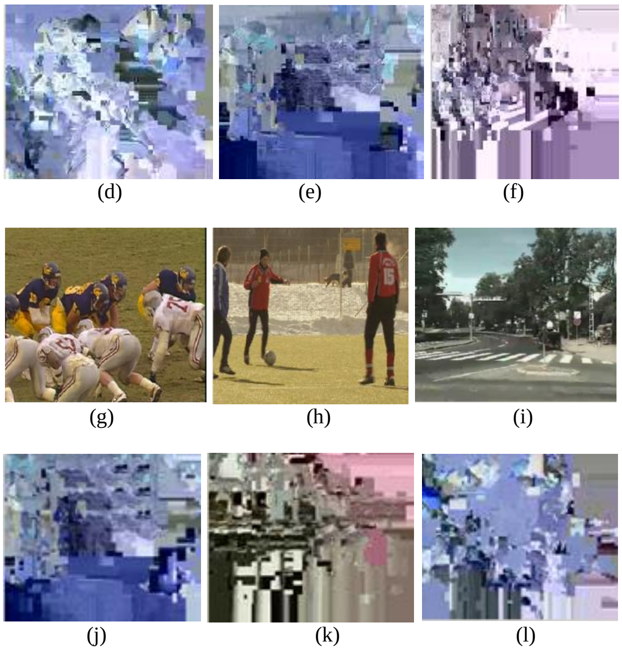


Fig.7. Encrypted Video with Proposed Algorithm

Table 7. PSNR Comparison for GOP Type I

Videos	QP	Approach in [4][5]			Proposed Approach		
		Y	U	V	Y	U	V
Foreman	18	8.00	8.32	9.13	7.87	8.90	9.44
	24	5.73	12.18	12.9	6.23	8.10	9.64
	30	6.12	9.75	11.05	7.00	8.46	9.12
Tempete	18	8.47	6.15	6.62	9.11	6.17	6.59
	24	6.54	6.60	14.08	6.66	6.65	6.39
	30	7.05	6.05	6.35	5.67	6.08	6.24
Football	18	8.90	11.32	12.10	9.15	11.14	12.28
	24	5.47	5.56	6.30	7.23	8.33	8.00
	30	4.60	5.26	7.22	5.60	6.65	7.79
Soccer	18	9.12	11.23	12.70	10.199	10.06	11.41
	24	8.20	7.54	6.15	9.38	10.33	9.30
	30	6.69	9.41	8.60	7.00	10.21	8.60
Horse Cab	18	6.78	7.73	7.47	6.89	7.83	12.08
	24	11.34	11.89	12.80	12.10	11.98	12.867
	30	5.38	8.23	9.44	4.87	8.43	9.11

Table 8. PSNR Comparison for GOP Type II

Videos	QP	Approach in [4][5]			Proposed Approach		
		Y	U	V	Y	U	V
Foreman	18	7.23	8.12	9.00	7.11	9.56	10.12
	24	6.10	8.45	10.13	6.56	9.23	10.43
	30	6.85	9.39	9.80	7.10	8.48	9.78
Tempete	18	8.32	7.12	7.48	6.23	6.92	7.12
	24	7.72	11.43	12.56	8.14	11.12	12.87
	30	6.63	10.21	9.27	7.08	9.33	10.19
Football	18	9.78	7.33	13.14	9.15	11.10	12.33
	24	6.78	7.89	8.13	6.45	7.56	8.12
	30	9.13	8.14	8.16	9.45	9.16	10.75
Soccer	18	8.12	7.43	10.25	9.34	8.12	11.65
	24	9.13	10.12	11.43	9.13	10.89	10.92
	30	10.13	11.24	12.45	10.53	11.12	12.23
Horse Cab	18	7.43	8.23	8.90	7.21	9.14	8.65
	24	9.34	9.87	10.37	8.12	8.86	9.73
	30	6.12	10.1	11.8	6.92	11.0	12.8

Table 9. SSIM Comparison for GOP Type I

Videos	QP	Approach in [4][5]	Proposed Approach
Foreman	18	0.12	0.14
	24	0.21	0.24
	30	0.25	0.26
Tempete	18	0.18	0.18
	24	0.12	0.23
	30	0.14	0.11
Football	18	0.19	0.23
	24	0.27	0.29
	30	0.11	0.15
Soccer	18	0.16	0.12
	24	0.15	0.20
	30	0.14	0.19
Horse Cab	18	0.17	0.19
	24	0.23	0.27
	30	0.18	0.13

Table 10. SSIM Comparison for GOP Type II

Videos	QP	Approach in [4][5]	Proposed Approach
Foreman	18	0.11	0.10
	24	0.10	0.10
	30	0.17	0.17
Tempete	18	0.12	0.15
	24	0.13	0.13
	30	0.11	0.14
Football	18	0.18	0.2
	24	0.14	0.11
	30	0.09	0.09
Soccer	18	0.19	0.17
	24	0.12	0.26
	30	0.13	0.2
Horse Cab	18	0.12	0.14
	24	0.21	0.24
	30	0.25	0.26

This can lead to a further lower computational cost as indicated in section V.B. From these three figures, it can be observed that no significant perceptual information of the video is leak while having a relatively lower computational cost. Fig. 7 (g), (h) and (i) are the original inter frames without scene transitions, the same frames encrypted with the proposed encryption algorithm are shown in in Fig. 7 (j), (k) and (l). In case of inter frames without a scene transition, the motion vectors are chosen as syntax elements for encryption while the intra coded macroblocks are not encrypted, since the error propagation through the inter prediction can already result in an effective scrambling effect to provide the perceptual security, as shown in Fig. 7 (j), (k) and (l). In P and B frames without scene transitions, actually there is a considerable percentage of intra coded macroblocks, leaving those intra coded macroblocks unencrypted can achieve a lower encryption overhead with almost similar perceptual security. The most common means to evaluate the scrambling effect of encrypted video is through PSNR [22]. SSIM [23] is another very useful metric which gives better analysis compared to PSNR [24].

In the experiments, PSNR and SSIM under different QP settings are used to evaluate the perceptual security of the proposed methods for several video sequences.

The PSNR for three different chrominance component (Y, U, and V) are provided. The experimental results for different GOP settings are shown in Table VII and VIII. It can be observed that in terms of PSNR, the proposed selective encryption method can achieve a very similar scrambling effect as the work in [4] and [5] under different QP sand GOP settings.

In terms of SSIM, the same conclusion can be obtained, are shown in Table IX & X.

B. Computational Cost

The computational cost of the video encryption algorithm highly depends on the volume of bits to be encrypted, which not only means a more real time performance but also has a vital impact for the energy consumption in those energy-critical applications, like wireless multimedia devices. Real time performance and lower energy consumption can be effectively achieved by the proposed video encryption methods. The number of bits encrypted is given by EDR (Encrypted Data Rate), which is the ratio of encrypted bits to the total number of bits in the video stream. The encryption overhead (the incurred computational cost) is defined as the difference between the total processing times with and without encryption.

Table 11. Computational time, unit: s.

Videos	GOP Type I		GOP Type II	
	Approach [4][5] (s)	Proposed Approach (s)	Approach [4][5] (s)	Proposed Approach (s)
Foreman	1.87	1.32	1.72	1.23
Tempete	2.00	1.50	1.93	1.56
Football	2.09	1.41	2.06	1.46
Soccer	1.75	1.12	1.71	1.15
Horse cab	2.61	1.92	2.77	1.92

Table 12. Encrypted Data Rate, unit: %

Videos	GOP Type I		GOP Type II	
	Approach [4][5] (%)	Proposed Approach (%)	Approach in [4][5] (%)	Proposed Approach (%)
Foreman	6.43	3.84	6.17	2.56
Tempete	3.22	1.88	3.12	1.73
Football	10.01	4.56	9.98	4.40
Soccer	11.14	6.54	10.89	6.12
Horse cab	16.86	10.77	15.32	9.34

The total processing time is obtained in terms of seconds by running the JM 18.5 codec [25]. he encryption overhead incurred by the proposed algorithm is compared with the normal decoding time, as the decoder is much faster than the encoder. The experimental results clearly show that the computational cost of the proposed methods are significantly lower than the previous approaches in [4] and [5] in terms of both the computational time and EDR which are listed in Table XI and XII, respectively. Different GOP types are used in the experiments and the same conclusion can be obtained. The proposed encryption method can save an average computational time by 0.59s compared to the approach in [4][5] while providing a similar scrambling effect.

C. *Replacement Attack*

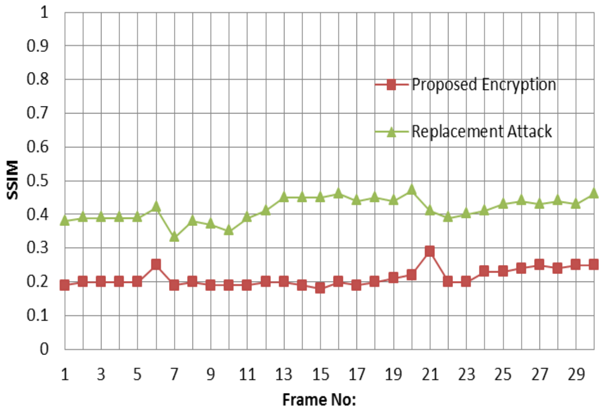
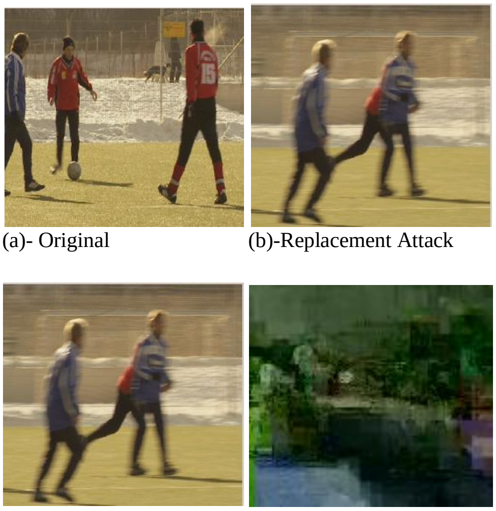


Fig.8. Replacement Attack Soccer frame 100-130 (SSIM Value)



(a)- Original (b)-Replacement Attack

(c)- Original (d)-Replacement Attack

Fig.9. Security against Replacement Attacks

The security of the proposed scheme depends on the adopted AES cipher. To date, the AES cipher is not vulnerable to any kind of known attacks and provides a strong security [26]. Due to the characteristics of video contents, video encryption is also vulnerable to specific attacks, such as the replacement attack [27]. In this attack, the encrypted syntax elements are replaced by fixed values to intend to obtain a video with an acceptable perceptual quality.

The replacement attack is demonstrated on the soccer video which is encrypted by the proposed algorithm. All the sign bits are set to positive values and intra prediction set to the most probable predicted IPM. Fig. 8 shows the SSIM values of the encrypted video with and without the replacement attack. It can be seen that the perceptual quality of the encrypted video can be improved to certain extent but still has a relatively strong perceptual security after the replacement attack. Fig. 9 (a) shows the original P frame without a scene transition and Fig.9 (b) shows the same P frame with replacement attack. In case of no scene transitions, even though the intracoded macroblocks are left unencrypted, no information can be retrieved after the replacement attack as shown in Fig.9 (b). Fig. 9 (c) shows the original frame with scene transition and Fig.9 (d) shows the same P frame with replacement attack. The frame with replacement attack is well scrambled, in which both the intra prediction modes and sign bits of the non-zero DCT coefficients are encrypted. Hence, the proposed algorithm can withstand replacement attacks.

D. *Deactivated Motion Compensation attack*

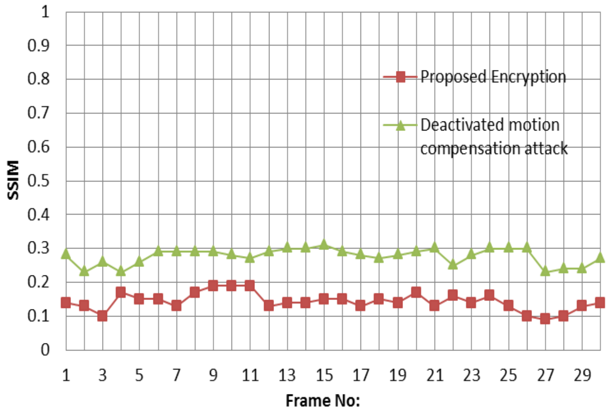
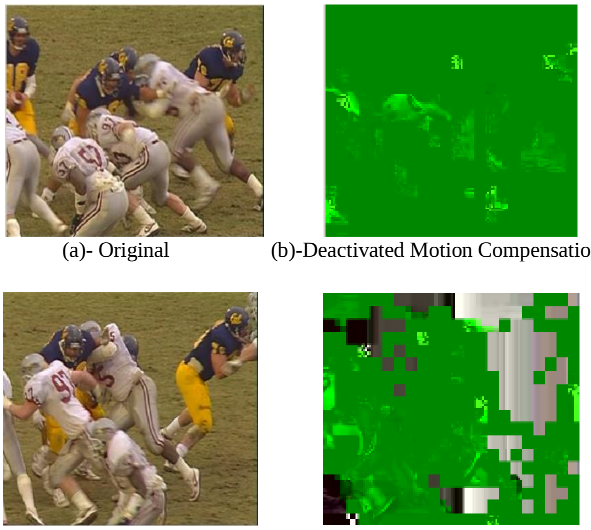


Fig.10. Deactivated Motion Compensation Attack Soccer frame (SSIM Value)

To analyze the fraction of intra coded macroblocks that are not encrypted, deactivated motion compensation attack has been implemented on football video.



(a)- Original (b)-Deactivated Motion Compensation

(c)- Original (d)-Deactivated Motion Compensation

Fig.11. Security against Deactivated Motion Compensation Attack

In this attack, the encrypted video is decoded with deactivated motion compensation i.e. only non-encrypted intra coded macroblocks are used to construct the video frame. Hence, in this attack, each block is constructed based on the latest's encoded and non-encrypted intra coded macroblocks. Fig.10 shows the SSIM values of the encrypted video and the encrypted video with deactivated motion compensation attack.

The SSIM evaluation indicates that the deactivated motion compensation attack does not improve the visual quality of the football video. Fig.11 (a) & (c) shows the original inter frames in which the intracoded macroblocks were left unencrypted. Fig.11 (b) & (d) shows the same interframe with deactivated motion compensation attack. From these figures, it is clear that the unencrypted intracoded macroblocks does not affect the security of the video.

VI. Conclusion

In this paper, a new selective encryption algorithm was proposed, with low computational cost to optimize energy consumption in energy critical wireless sensor multimedia networks and wireless multimedia devices. The algorithm aims to reduce the computational cost by selecting sensitive code word candidates based on scene transitions. The Encryption cost (E) is directly dependent on the number of scene transitions (Nsc) in the video stream. An adaptive threshold function for scene change detection was also proposed. Experimental results clearly indicate that the proposed algorithm can provide scrambling levels equivalent to the previous approaches with low computational overhead.

A security analysis of the proposed scheme was also given, which indicates that the scheme is secure against replacement attack and deactivated motion compensation attack.

References

[1] ITU-T. Rec. (ISO/IEC 14496-10): 2010, *Advanced Video Coding for Generic Audio Visual Services*.

[2] Y. Shi and H. Sun, Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards. Boca Raton. FL: CRC Press, 2000.

[3] T. Stutz and A.Uhl, “A survey of H.264 AVC/SVC encryption”, *IEEE Trans. Circuits syst. Video techno.*, vol, 22, no. 3, pp. 325-339, ju. 2011.

[4] Shiguo Lian, Zhongxuan Liu, Zhen Ren and Haila Wang, “Secure Advanced Video Coding Based on Selective Encryption Algorithms,” *IEEE Transaction on Consumer Electronics*, Vol. 52, No. 2 ,2006, pp.621-629, 2006.

[5] Yongsheng Wang, Maire O'Neill, Fatih Kurugollu, “A tunable encryption scheme and analysis of fast selective encryption for CAVLC and CABAC in H.264/AVC”, *IEEE Trans. Circuits syst. Video techno.*, Vol 23, No. 9, pp. 1476-1490, 2013.

[6] Y. Zhao, L. Zhuo , N. Mao, J. Zhang and X. Li "An object-based unequal encryption method for H.264 compressed surveillance videos", Proc. 2012 IEEE Int. Conf. Signal Processing, Communication and Computing (ICSPCC), pp.419 -424 2012.

[7] Haojie Shen; Li Zhuo; Yingdi Zhao, "An efficient motion reference structure based selective encryption algorithm for H.264 videos," *Information Security, IET*, vol.8, no.3, pp.199, 206, May 2014. doi: 10.1049/iet-ifs.2012.0349.

[8] 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.

[9] Yingdi Zhao; Li Zhuo, "A content-based encryption scheme for wireless H.264 compressed videos," *Wireless Communications & Signal Processing (WCSP)*, 2012 International Conference on , vol., no., pp.1,6, 25-27 Oct. 2012 doi: 10.1109/WCSP.2012.6543022.

[10] Misra *et al.*: A survey of Multimedia Streaming in wireless sensor networks, *IEEE Communications Survey & tutorial*, Vol. 10 No.4 Fourth Quarter 2008.

[11] 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*., 15(5):645–658, May 2005.

[12] J. Lee, I. Shin, and H. Park, “Adaptive intra-frame assignment and bitrate estimation for variable GOP length in H.264,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 10, pp. 1271–1279, Oct. 2006.

[13] Zhenyu Yu; Zhiping Lin, "Scene change detection using motion vectors and dc components of prediction residual in H.264 compressed videos," *Industrial Electronics and Applications (ICIEA)*, 2012 7th IEEE Conference on, vol., no., pp.990, 995, 18-20 July 2012.

[14] Bohyun Hong; Minyoung Eom; Yoonsik Choe, "Scene Change Detection using Edge Direction based on Intra Prediction Mode in H.264/AVC Compression Domain," TENCON 2006. *2006 IEEE Region 10 Conference* vol., no., pp.1, 4, 14-17 Nov. 2006.

[15] J.-R. Ding and J.-F. Yang, “Adaptive group-of-picture and scene change detection methods based on existing H.264 advanced video coding information”, *IET Image Processing*, Vol. 2, No. 2, pp. 85-94, 2008.

[16] S. Youm and W. Kim, “Dynamic threshold method for scene change detection”, In Proc. *ICME2003*, Vol. 2, pp. 337-340, July. 2003.

[17] H. Li, G. Liu, Z. Zhang and Y. Li, “Adaptive scene-detection algorithm or VBR video stream”, *In IEEE Trans. Multimedia*, Vol. 6, No. 4, pp.624-633, Aug. 2004.

[18] Wiegand, T.; Sullivan, G.J.; Bjontegaard, G.; Luthra, A., "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.13, no.7, pp.560, 576, July 2003.

[19] Marpe, D.; Schwarz, H.; Wiegand, T., "Context-based adaptive binary Arithmetic coding in the H.264/AVC video compression standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.13, no.7, pp.620, 636, July 2003.

[20] FIPS 197 (Advanced Encryption Standard), NIST Publications, 2001.

[21] Multimedia over IP and Wireless Networks: Compression, Networking, and Systems by Mihaela van der Schaar, Philip A Chou. Academic press, 2011.

[22] 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 Journal on information security* 2008:179290.

[23] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity*,” IEEE Trans. Image Process*, vol. 13, no. 4, pp. 600– 612, 2004.

[24] Z. Wang, A.Bovik, “Mean Squared error: Love it or leave it? A new look at signal fidelity measures,” *IEEE Signal Processing*. Mag., vol., 26, no.1, pp. 98-117, Jan. 2009.

[25] JM Reference ver 18.5 (2012) http://iphome.hhi.de/suehring/tml.

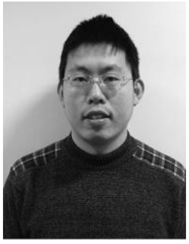
[26] NIST Special Publication 800-57, Recommendation for Key Management, 2012.

[27] M. Podesser, H.Schmidt and Uhl, “Selective bitplane encryption for secure transmission of image data in mobile environments”, in Proc. *5th IEEE Nordiac signal Process*. Symp. oct. 2002, pp. 4-6.

Authors' Profiles

Karthik Thiyagarajan was born in Kuwait. He received his bachelors Electronics and Communication Engineering in Anna University, Chennai, India in 2010 and Post Graduate Diploma in embedded system design from NIIT, Calicut, India. He finished a master's degree in Dalhousie University Halifax, NS, Canada. Currently he works as a Algorithm engineer (Cryptography and Security) in NRC, Canada. His research interests are embedded systems, cryptography and Compression for video.

Kamal El-Sankary Received the B.Eng degree from the Lebanese University in 1997 and the M.A.Sc degree in electrical engineering from University of Quebec in 2002 and the Ph.D. degree in electrical engineering from Ecole Polytechnique, University of Montreal in 2006. Dr. El-Sankary is an Associate Professor in the Electrical and Computer Engineering Department at Dalhousie University. His research interests include Embedded Systems, RF, and analog and mixed-signal circuits design.



Yongsheng Wang (S'12) received the B.E. degree with the highest honor in automation engineering from Xidian University, Xi'an, China, in 2006, and the M.Eng. degree in control science and engineering from Tsinghua University, Beijing, China, in 2009. He obtained his Ph.D. degree in electronic engineering from the Center for Secure Information Technologies, Queen's University, Belfast, UK in 2013. Before joining the Queen's University, he was an Embedded Engineer on circuits and system design based on DSP in the Second Academy of China Aerospace and Industry Corporation, Beijing, China. He currently works as a system design engineer for Shrader electronics, Antrim, Northern Island His current research interests include video encryption, privacy protection, multimedia.

Issam Hammad was born in Amman, Jordan in 1986. He received his B.Sc Degree in Computer Engineering from Princess Sumaya University for Technology, Amman, Jordan, in 2008 and his M.A.Sc Degree in Electrical and Computer Engineering from Dalhousie University, Halifax, NS, Canada, in 2010. . He is currently working as a technical consultant for SWI System ware Innovation Corporation, Toronto, Ontario, Canada. His research interests include video processing, cryptography and computer architecture.

How to cite this paper: Karthik. Thiyagarajan, Kamal El-Sankary, Yongsheng Wang, Issam Hammad,"Low Complexity Multimedia Encryption", International Journal of Computer Network and Information Security(IJCNIS), Vol.8, No.4, pp.1-13, 2016.DOI: 10.5815/ijcnis.2016.04.01