LOW COMPUTATIONAL OVERHEAD VIDEO ENCRYPTION FOR WIRELESS MULTIMEDIA DEVICES

By

Karthik Thiyagarajan

Submitted in partial fulfilment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
September 2014

To my Parents P.Thiyagrajan, Malar Thiyagarajan and my sister Pavithra, Friends and Teachers

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Selective encryption is a promising technique which provides privacy protection for multimedia technology. In this thesis we propose a scheme to optimize computational cost for energy critical wireless multimedia applications. Latest research in selective encryption algorithm proposed to encrypt syntax elements such as intra prediction modes, sign bit of residual coefficients, along with sign bit of motion vectors. Such syntax elements are sensitive enough to provide effective scrambling effect with tractable computational cost. The proposed algorithm selects code words for encryption based on scene transitions in compressed frames. The ratio of intracoded macroblocks in inter frames is calculated and compared with a threshold value to detect scene transitions. Further, Based on statistical properties of the video frame, a dynamic threshold model for scene change detection has been proposed. In case of video frames with scene transition, intra modes and sign bit of residuals are chosen as code words to encrypt, whereas in case of no scene transitions, sign bit of motion vectors are chosen as sensitive code words for encryption. Experimental results showed that the proposed algorithm maintains guaranteed security with low computational overhead.

LIST OF ABBREVIATIONS AND SYMBOLS USED

MPEG	Moving Pictures Expert Group
AVC	Advanced Video Coding
WMSN	Wireless Multimedia Sensor Network
PSNR	Peak Signal to Noise Ratio
SSIM	Structural Similarity Index
AES	Advanced Encryption Standard
DCT	Discrete Cosine Transform
CAVLC	Context Adaptive Variable Length Coding
CABAC	Context Adaptive Binary Arithmetic Coding
ECB	Electronic Code Book
CFB	Cipher Feedback
OFB	Output Feedback
CTR	Counter mode
HDTV	High Definition Television
VoD	Video on Demand

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Dr. Kamal El-Sankary and Co-supervisor Issam Hammad, for their guidance, encouragement and support during my graduate study. Their deep insight and extensive knowledge in Embedded Systems and Video Technology helped me throughout my master's project. As well, I am would like to thank Dr. Jason Gu and Dr. William J. Phillips for being a part of my supervisory committee. Many thanks to the group mates in VLSI group for sharing their knowledge and experience. I would also like to express my gratitude to the department staff Nicole Smith and Caroline Burgess. Most of all, I would like to express my special thanks and appreciation to my parents for their support and encouragement throughout my 2 years master's degree. Also, many thanks to my friends in Halifax for their support.

Chapter 1 Introduction

1.1 Motivation

Recent advancements in computer technology, especially in communications have allowed digital multimedia applications to boom in the marketplace. The ease in compression, delivery and presentation of technology has broadened the applications of multimedia which includes

- Digital video Broadcasting
- Internet and mobile video streaming
- Video Calling
- DVD storage of video

Video compression is required to transmit video streams over band limited communication channels [1]. If videos are transmitted over insecure networks, an attacker could retrieve the video packets. Hence encrypting the video stream prior to transmission is becoming a popular research area. Encrypting the entire video stream has a great effect on the computational cost. Therefore selective encryption algorithms [3] are proposed to encrypt sensitive code word candidates in the compressed video. The encryption algorithms used are standard algorithms such as AES (Advanced Encryption Standard), DES (Data Encryption Standard)[3] etc. Selective encryption techniques provided low computational cost and acceptable security.

For an efficient selective encryption algorithm, some of the major constraints are

- The Number of sensitive code words selected in the video stream for encryption should be as low as possible.
- Should meet real time requirements such as low computational cost and energy efficient.
- High security and format compliance.

Many selective encryption algorithms were proposed in literature. Some of these algorithms had low computational costs [11][12] but were vulnerable to attacks. Recently [13] & [14] proposed to encrypt the intraprediction modes, sign bit of residual coefficients and sign bit of motion vectors in all the frames. The method in [13] & [14] degraded the perceptual quality without any impact on the compression performance and had low computational overhead suitable for real time multimedia encryption. However, the work in [13][14] had higher encryption overhead compared to the methods in [11][12]. There always existed a trade-off between low computational cost and security. Moreover, video encryption for wireless multimedia sensor networks and wireless multimedia applications has critical constraint over energy consumption as they are battery driven. Due to the importance of the Video encryption algorithm and the numerous applications that it has, the main concern of this thesis will be presenting new energy efficient selective encryption algorithms for wireless multimedia applications.

The JM (Joint Model) 18.5 reference software is the latest video codec released by the JVT (Joint Video Team) team. The entire Encoder and decoder have been implemented in C language. Because of this fact and most of the selective encryption techniques in literature have used JM codec's, the presented algorithms in have been implemented using the same.

1.2 Objective

Based on the above discussions, the main aim of my thesis is to have an energy efficient algorithm for wireless multimedia applications. The energy consumption of the node (multimedia processor) directly depends on the overall computational cost of the algorithm. The following goals are

- Analysis of code word (Intraprediction Modes, sign bit of residual coefficients and sign bit of motion vectors) sensitivity for encryption in H.264 compressed video stream.
- Analysis to reduce the number of code word (Intraprediction Modes, sign bit of residual coefficients and sign bit of motion vectors) candidates for encryption.
- Selective encryption algorithm based on scene transition detection.

Our selective encryption algorithm selects critical code word candidates for encryption based on scene transitions in the video stream. In case of P and B frames, if there is a scene transition intracoded macroblocks are chosen to encrypt, whereas in case of no scene transitions motion vectors are chosen to encrypt. This way the number of code words selected is reduced and further the computational cost and energy consumption.

1.3 Organization:

The thesis is organized as follows

Chapter 2 explains the H.264 standard in detail, about the various steps in the encoding process, CAVLC and CABAC entropy coding methods, AES encryption algorithm in the CTR mode. Furthermore, this chapter will discuss the terminologies used in video encryption.

Chapter 3, the proposed selective encryption algorithm is presented in detail, analysis of sensitive code word candidates and the analysis to reduce the number of selected code word candidates are also done in this chapter.

Chapter 4 shows the simulation results, discusses security of the proposed selective encryption algorithm in terms of PSNR and SSIM, computational overhead in terms of Total decoding time, Encryption data rate.

In chapter 5 we present the conclusion and future work.

Chapter 2 Video Compression and Encryption

In this chapter a brief overview of H.264 compression technology, definitions of H.264 frames (I, P and B), stream packetizing and definitions of intracoded macroblocks, residual coefficients and motion vectors are discussed. The advanced encryption standard in the CTR mode is presented. A literature review which studies previous proposed encryption algorithms for the selective encryption algorithm is presented. Their pros and cons relative with respect to our proposed algorithm is also discussed. A brief overview of metrics and parameters (PSNR and SSIM) used for evaluating the proposed video encryption is also discussed.

2.1 Introduction to Selective Video Encryption

In broadcasting technology, with the proliferation of wireless sensor multimedia networks, video compression protocols have gained popularity. The H.264 compression standard [1] was reported to have a promising compression performance better than previous compression standards [2]. Secured transmissions are becoming increasingly difficult to achieve in wireless multimedia applications, due to higher level of malicious attacks. The encryption/decryption algorithms used for data transmissions are not suitable for video applications, because in video we have large amount of data to transmit [3]. Selective encryption is one of the most determinant techniques for multimedia content protection in wireless sensor networks, as it can achieve real time processing requirements such as effective scrambling effect and low computational overhead [4].

2.2 Video compression Technology

Video compression has a very broad application range that covers all forms of digital compressed video from internet streaming to HDTV broadcast and digital cinema applications [1]. H.264 is an industry standard for video compression, the process of converting digital video into a format that takes up less bit rate/space when it is stored or transmitted. Fig. 1 shows the H.264 Encoder/Decoder Block diagram in detail.

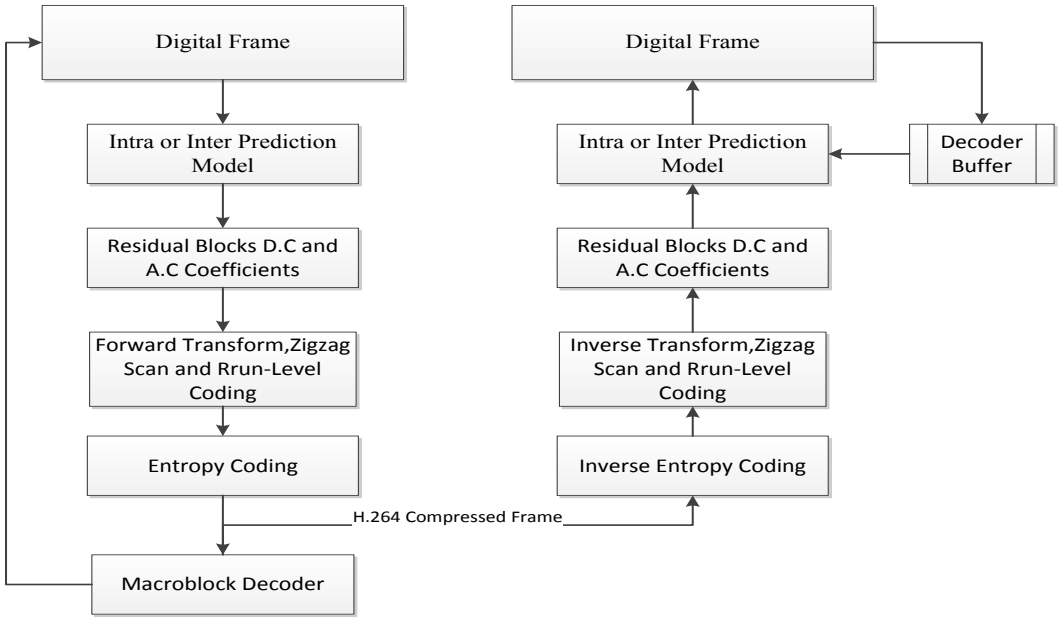

Figure.1 H.264 Block Diagram

The real time analog video is sampled and digitized into digital frames. Each digital frame consists of YUV samples, where Y represents the luminance component and UV represents the chrominance components in the frame. The next step is to compress the digital frames into IPB frames (H.264) as shown in fig. 2. The key parameters in the H.264 stream are the frame types (I, P and B frames), GOP (Group of Pictures), profiles and levels.

Figure 2. YUV to H.264 format

GOP: The GOP specifies the order of pictures in H.264 Video stream. The H.264 GOP consists of I, P and B frames just like any MPEG compressed video. An I frame indicates the beginning of a new GOP and end of the previous GOP as shown in figure 3. For example, in a sequence with pattern IPBPBI, the GOP size is equal to 5 (length between two I frames).

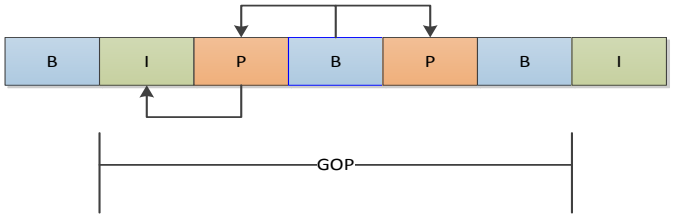

Figure 3. GOP size and frame level Dependency

In the compressed domain, there are three types of frames I, P and B frames.

I Frame: I frames are fully intracoded pictures i.e. they are predicted without motion compensation and are independent of other frames (I, P and B frames while decoding). I frames contain only intrapredicted macroblocks. As shown in figure 4 , each 16x 16 YUV samples (current macroblock) in the Nth frame are taken and subtracted with an approximately similar 16x 16 YUV samples(previously encoded macroblock with in the same Nth frame). The subtracted residuals are encoded as intracoded macroblocks in the compressed frame.

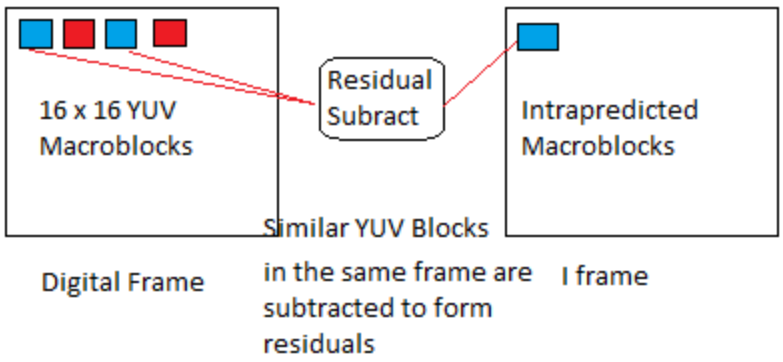

Figure 4. Intra Frames (I Frame)

P Frame: P frame holds the difference in motion change between the N-1th and Nth pframe. Figure 5 shows motion compensation technique used on macroblocks to obtain motion vectors.

The square block in the current frame is the target macroblock (current), an approximately similar block is searched in the reference frame. The similar macroblock found in the reference frame is the predicted macroblock. The search for this predicted macroblock is called motion compensation. The displacement of the predicted macroblock with respect to the current macroblock is represented as a motion vector. The predicted macroblock and current macroblock are subtracted to form residual coefficients. These residual coefficients along with motion vectors are encoded as interpredicted macroblocks.

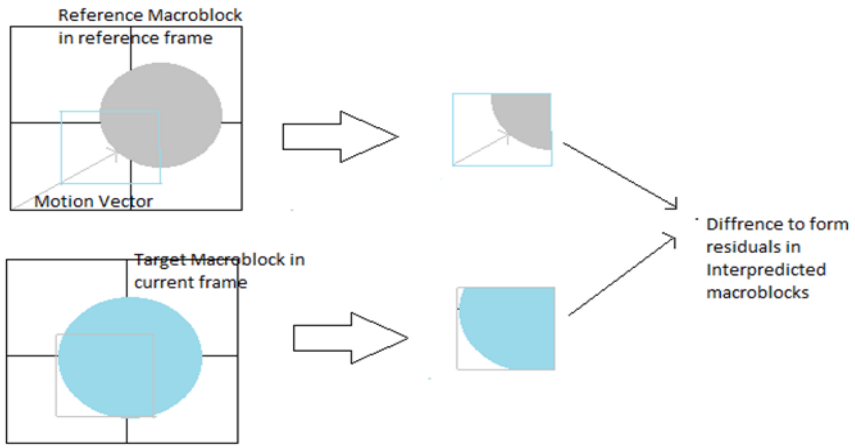

Figure 5. Inter Frames (P frames)

B Frame: B frames are predicted similar to P frames. The difference is , in case of B frames the motion differences between the N-1th (previous frame) and N+1 (next frame) is considered for encoding. Therefore, the B frames are dependent on the previous and next frames while decoding. Fig. 6 shows the Bi-Prediction in B frames.

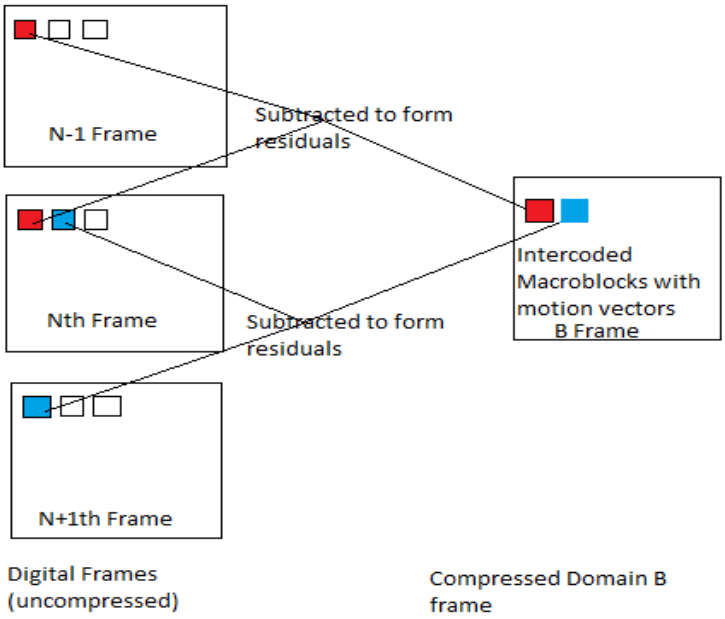

Figure 6. Inter Frame (B frames)

Stream Packetizing in H.264 Compression: Figure 7 shows a typical H.264 compressed video stream. The shaded blocks are the region of interests in the video stream. The sequence (seq 2) in figure 7 is the compressed video. It consists of GOP related parameters and a number of GOP's. Each GOP consists of one I frame and a number of dependent P and B frames. The frames are divided into number of slices. Sometimes a frame can have only one slice structure. The slices are further divided in to macroblocks which can be intracoded as well as intercoded. Incase of intracoded macroblock, the motion vectors are absent in the stream packet and are present in intercoded macroblocks.

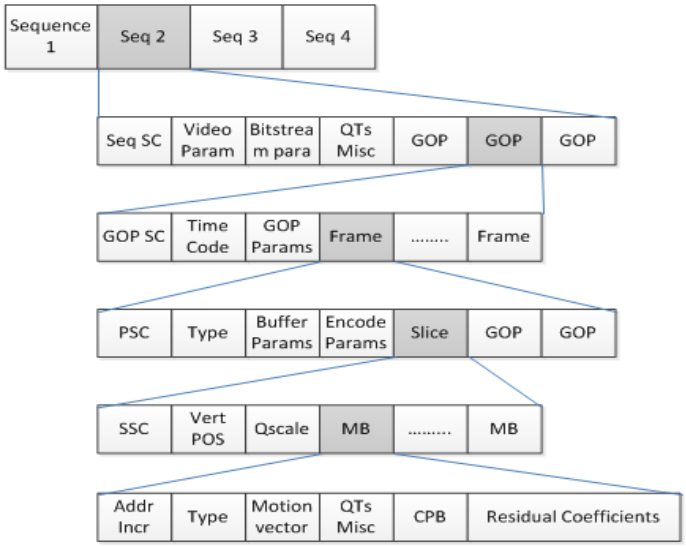

Figure 7. Frame level and Macro block Level Structure

Profile Types: Profiles define 21 sets of features/capabilities the encoder can choose for various levels of compression and optimization. H.264 encoder uses three major profiles based on the applications.

- **Baseline Profile (BP)**: Base line profile is used widely in video conferencing and mobile communications. They are extensively used for low computational resources.
- **Main Profile (MP)**: Main profile is used in mainstream consumer electronics for video broadcasting and storage.
- **Extended Profile (XP)**: Intended as the streaming video profile, this profile has relatively high compression capability and some extra tricks for robustness to data losses and server stream switching.

Levels: The level is a constraint on both frame rate and dimensions of the final video. More detailed descriptions of H.264/AVC (Encoder & Decoder) can be found in [2]. Figure.1 shows a brief overview of H.264 block diagram.

2.3 Cryptography

Cryptography is a part of engineering concerned in hiding data from malicious attackers, also called encryption, so that it can only be decrypted, by specific authorized individuals who have access to the secret key. A system for encrypting and decrypting information is a cryptosystem. Encryption usually involves an algorithm for combining the original data ("plaintext") with one or more "keys" — numbers or strings of characters known only to the transmitter and/or receiver. The resulting output of encryption is known as cipher text. As shown in figure 8 there are two main classes of cryptosystems, with different practical application. Public key methods use two different keys for encryption and decryption. On the other hand, secret key encryption methods use the same key for encryption and decryption. Cryptographic algorithms play an important role in security and resource conservation of real time applications.

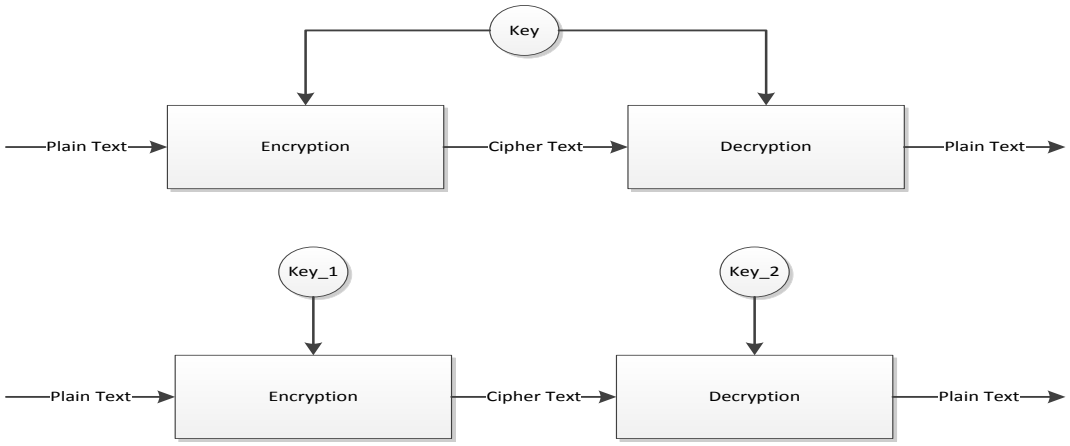

Figure.8 Symmetric and Asymmetric Encryption

2.3.1 Necessities for Video encryption [5]:

(i) High computational efficiency.

(ii) High compression efficiency due to encryption.

(iii) Low encryption ratio.

(iv) Speed.

(v) High security.

(vi) Low power consumption and memory.

Asymmetric encryption vs. symmetric encryption [5]:

(i) Asymmetric is more time and power consuming.

(ii) Symmetric is 200 times faster than asymmetric algorithms.

(iii) Asymmetric is about factoring of large prime numbers and derivation of two keys which is not suitable for real time applications.

Hence, symmetric algorithms are preferred rather than asymmetric key algorithms. The work in [5] demonstrates that the best algorithm for wireless sensor multimedia networks is RC5 (symmetric key cryptography having simpler structure than AES) as it is less energy consuming and time consuming in terms of both software and hardware, but however in terms of security AES is better than RC5. Both RC5 and AES algorithms have not been hacked up to date. Work in [6] shows that AES is the best encryption for real time video transmission, in terms of security, and time. We focus more on low computational cost; hence advanced encryption standard is the cryptography algorithm chosen for our purpose. AES is one of the strongest cryptographic algorithms known till date; it is based on substitution-permutation network, and is fast in both hardware and software. There are five modes of AES algorithm operation: ECB (Electronic Code Book), CBC (Cipher Block Chaining), OFB (Output Feed Back), CFB (Cipher Feed Back) and CTR (Counter Mode). ECB and CBC operate as block ciphers. The former is the naive AES whereas, in the later, the current plaintext is XORed with previous cipher text before encryption. In the OFB mode, the output is fed back to the input. CFB is similar to CBC but is operated in stream mode. CFB, CTR and OFB operate as stream ciphers.

AES CTR mode has been chosen for the following reasons: Low Latency, Low Error Propagation and High speed compared to other modes. Encryptor supports both encryption and decryption, No zero padding, counter is used rather than a feedback. OFB and CFB suffer from cycle length changes. Hence, we have chosen CTR mode in our implementation. The cryptographic algorithm chosed for practical encryption is completely application dependent (i.e based on required security and complexity)

2.3.2 Advanced Encryption Standard

AES is the standard encryption standard adopted by the NIST (National Institute of Standards and Technology) for securing data while communication. AES works on substitution permutation network. AES has a fixed block size of 128 bits and a key size of 128, 192 or 256 bits. The entire algorithm is divided in to two sections, the Key expansion unit and the state processing unit (Figure 9). The number of rounds is 10 in case of 128 bits key (12 when key length is 192 bit and 14 when the key length is 256). For encryption, each round consists of the following four steps: 1) Substitute bytes, 2) Shift rows, 3) Mix columns, and 4) Add round key. The last step consists of XORing the output of the previous three steps with four words from the key schedule.

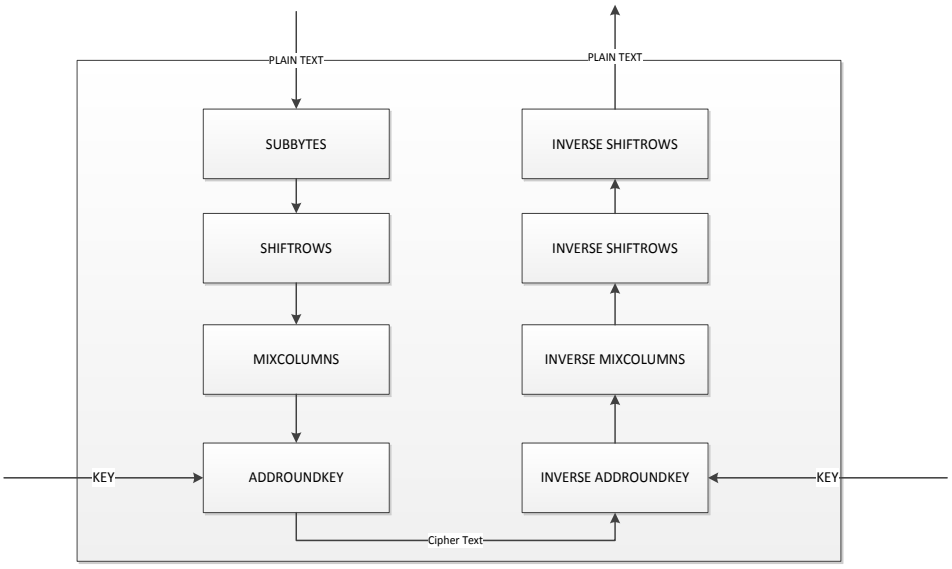

Figure 9 AES- State processing unit

2.4 Video Encryption Standard

2.4.1 Applications:

The potential application scenarios for multimedia encryption are vast and are often application specific with diverse functionality of the encrypted video stream.

1) DRM

The most common application of video encryption is in digital rights management (DRM), more clearly copyright protection, in which the product providers aim to secure their content value, i.e., they want to prevent redistribution of their content, very frequently videos.

2) Pay TV Providers

Video encryption is also commonly used by cable TV providers [7]. In the application scenario of perceptual encryption (only visual encryption is considered) the availability of a public low quality version is a requirement and the threat is that an attacker is able to compute a reconstruction of the original content with higher quality then the available public version.

3) Video Surveillance

Privacy protection in video surveillance system is also a common application, where encryption plays a key role in privacy preservance; here the privacy of the people and objects in the video should be preserved. The security threat in the video privacy surveillance is the identification of a human person or object, e.g., a license plate, in the video, and thus has to be prevented.

4) Wireless Sensor Networks

Video conferences [8] are another major application scenario in which videos are encrypted for privacy and content protection. Secured Scalable Streaming (SSS) of video in wireless streaming is one of the frequently discussed applications in video encryption. Mobile computing is not a distinct multimedia encryption scenario but can be considered as uncommonly used applications.

5) Mobile Computing

Mobile computing is not a distinct scenario from WMSN (Wireless Sensor Multimedia Networks). Video encryption for mobile devices mainly focuses on energy consumption as it imposes strict constraints on low computational complexity, which interests low computational encryption approaches.

2.4.2 Parameters of video encryption

Certain properties and associated functionalities of the encrypted bit stream should be preserved and solely depends on the type of application. Thus the key properties to be considered in video encryption scheme are as follows [3]:

1) Computational Complexity :

   It may be defined in terms of memory requirement and encryption overhead of encryption algorithms. Memory requirement of encryption is determined by memory occupied by code and data while time overhead measures the time requirement for encryption/decryption. Software and hardware implementation of cryptographic algorithms exits in plenty and both have different and sometimes contrary characteristics. Here we will restrict our discussion to software implementation of video encryption techniques i.e. encryption overhead with respect to time.

2) Security Offered:

   The security requirement of a video encryption algorithm solely depends on the type of application. Some applications (Pay TV) requires loose security needs with perceptual degradation only while video conferencing may require totally closed communication for all others out of communication group. The security of the video encryption is measured in terms of scrambling effect of the video after encryption. Recent works in video encryption have used PSNR and SSIM to evaluate their results.

3) Compression Efficiency:

   Real time videos are very large and hence needed be compressed before storage for bandwidth effective transmission. Encryption algorithm can be before compression, after compression or embedded in compression. During the design of encryption algorithm it is major concern that the size of compressed video should not be increased by the encryption. The syntax and semantics of the bit stream should remain the same after encryption.

4) PSNR (Peak Signal to Noise Ratio)[9]

PSNR is a metric used to evaluate the reconstructed image or video after compression. It determines the ration between the peak original data and the noise induced PSNR is an approximation of human perception in terms of visual quality. PSNR is measured in logarithmic decibel scale. If M is the monochrome image and N is its noisy approximation then the MSE (Mean Square Error) is given by,

$$MSE = \frac{1}{m}\frac{1}{n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[M(i,j) - N(i,j)]^2 \qquad (2.3)$$

$$PSNR = 10log_{10}\frac{MAX^2{}_M}{MSE} \qquad (2.4)$$

Where MAX_M is the maximum pixel value of the image M

5) SSIM (Structural Similarity Index)[10]

Structural similarity index is a metric widely used to measure the perceptual video quality of the video. It is a method used to measure the similarity between two frames. SSIM is designed to improve the traditional methods in PSNR and MSE (Mean Square Error). SSIM Index is based on measuring of three components (luminance similarity, contrast similarity and structural similarity) and combining them into result value. The difference between PSNR and SSIM is that the former measures the perceivable errors between the two frames used for comparison whereas the later measures the structural changes (differences) between the two frames. If x and y are two windows of size NxN then SSIM is calculated as,

$$SSIM(x,y) = \frac{(2\,\mu(x)\,\mu(y)+C1)\ .\ (2\,\sigma(x,y)+C1)}{(\mu^2(x)+\mu^2(y)+C1)\,.(\sigma^2(x)+\sigma^2(y)+C2)} \qquad (2.5)$$

Where $\mu(x)$ is the average of x, $\mu(y)$ is the average of y, $\sigma^2(x)$ the variance of x , $\sigma^2(y)$ the variance of y, $\sigma^2(xy)$ the covariance of x and y.

6) Replacement Attack:

Replacement attack is an attack done on encrypted video by replacing all the encrypted bits by 1.This enhances the perceptual quality of the encrypted video and further visually leaks some information.

2.5 Literature review

Several selective encryption algorithms have been proposed in literature. The selective encryption algorithms proposed to encrypt intraprediction modes, sign bits of residual coefficients and sign bit of motion vectors. Their work focuses over low computational cost and maximum security.

2.5.1 Encrypting DCT Coefficients:

Shahid et al [11] [12] proposed to encrypt all nonzero coefficients (Figure 10) and sub suffix in the suffix of nonzero along with the sign bits. The method demonstrated effective scrambling without degrading the compression performance. However, the computational complexity was increased. Encrypting the DCT coefficients alone does not provide sufficient security as encrypted video is vulnerable to replacement attack as shown in [13]. Our work has encrypted the DCT coefficients along with the intraprediction modes which with stands from replacement attack.

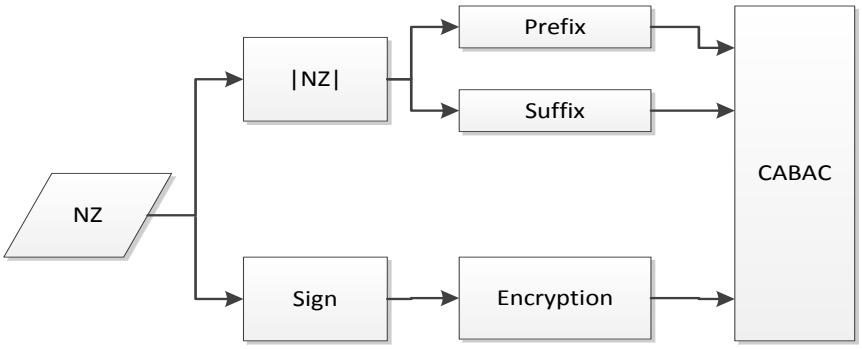

Figure 10: Sign Bit Encryption of CAVLC Coefficients

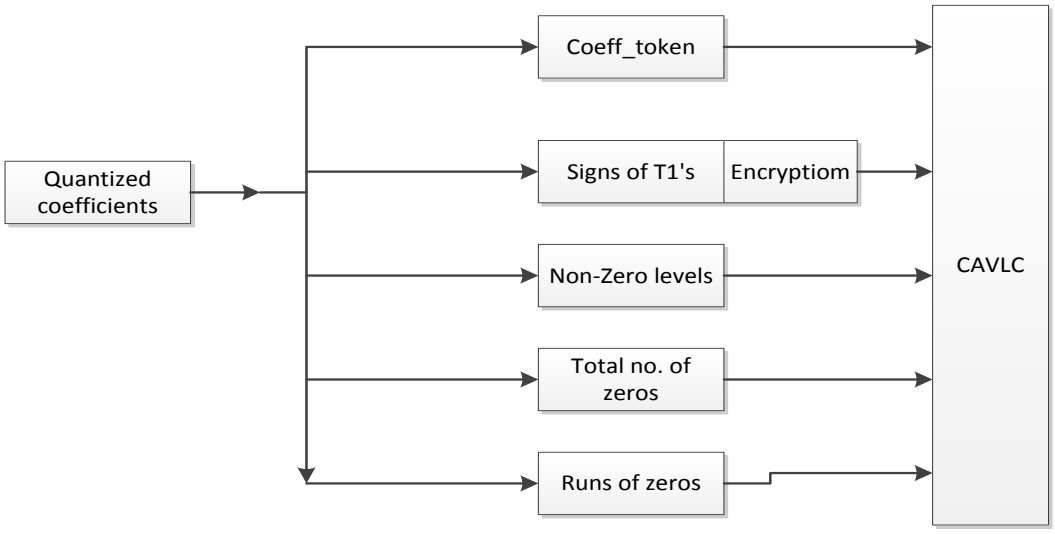

Figure 11: Sign Bit Encryption of CABAC Coefficients

2.5.2 Encrypting Intraprediction Modes, Sign bit of Transform Coefficients and Sign bit of motion vectors

Finally Lian et al [13] proposed encrypting Intraprediction Modes, DC coefficients and sign bit of residual information along with sign bit of motion vectors in P and B frames (shown in figure 12). The method degraded the perceptual quality without any impact on the compression performance and had low computational overhead suitable for real time multimedia encryption. Wang et al [14] demonstrated that the perceptual scrambling effect in these two methods [11][12] is mainly produced by encrypting the sign bits of non-zero coefficients. Further, Wang et al modified the encryption scheme in [13] to a user tunable encryption employing a factor N. The method claimed to have high security and adequate overhead suitable for real time applications. The code word (Intraprediction modes, sign bit of residuals and sign bit of motion vectors) selection for encryption sustains itself against any replacement attack. The encrypted code words (Intraprediction modes, sign bit of residuals and sign bit of motion vectors) sustains against replacement attacks because both the low and high spatial information is encrypted. In our work an adaptive scene change detection algorithm is proposed and further used for video encryption, where only certain code words among intraprediciton modes, sign bit of residuals and sign bit of motion vectors are selected for encryption based on scene transitions. In our case we are able to achieve 5 – 10% lower computational cost with acceptable security.

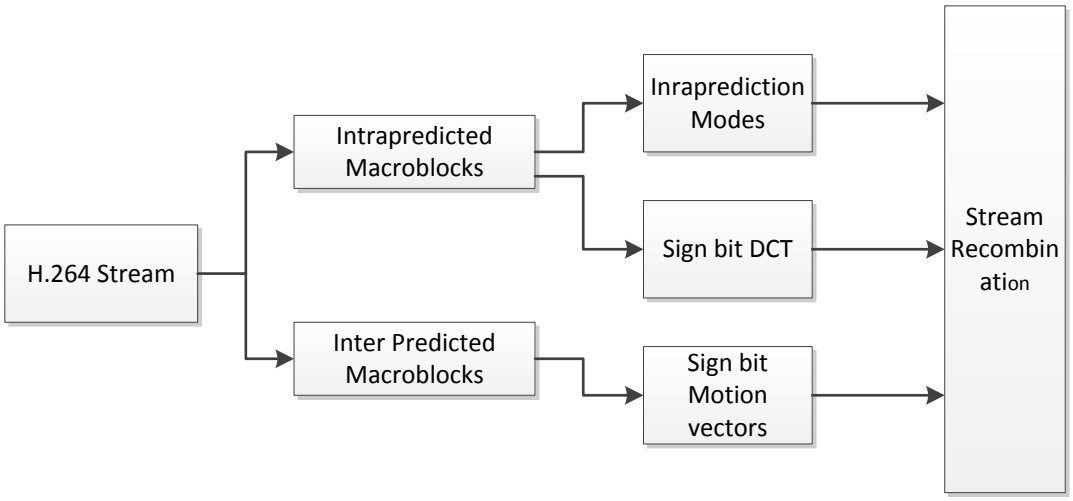

Figure 12: Sign bit encryption in [16][19]

2.5.3 Energy Efficient Encryption Algorithms:

Wei Wang et al [15] extended the encryption of the sign bits of nonzero coefficients, intra modes and sign bit of motion vectors, and the selective encryption included the effect of encryption in wireless sensor networks. The work [15] intended to optimize encryption overhead and scale down error propagation with an encryption technique based on interframe dependency and priority. Their method proposed to only encrypt frames that were highly dependent by descendent frames for decoding. However, the lower dependent frames become insecure. Zhao et al [16] improved the method in [13], for achieving lower complexity. Depending on the frames texture complexity and motion intensity, certain amount of code words are encrypted, which provided varying levels of scrambling effect. Lower percentage of code words is encrypted in case of frames with low texture complexity and motion intensity, which leaves the frame insecure. The motion intensity indexing algorithm [17] is limited to GOP size of 16 [33] and fail to detect transitions due to gradual frame transition (Camera motion) [34].

The carefully designed scrambling techniques mentioned in Table I alleviate the encryption effect over compression ratio and format compliance, as the sign bits are encrypted. Most of the previous works on multimedia encryption have focused on application layer cryptography, and they did not consider optimizing encryption cost for energy constrained wireless devices. Furthermore, in all the aforementioned works, various challenging algorithms for optimizing the encryption overhead were proposed, with a trade-off over security.

TABLE 1: LITERATURE REVIEW

Work	Intra Modes	Sign bit Residual	Sign bit of MVD	overhead cost	Security
[13]	✓	✓	✓	S	L
[11,12]		✓		S	L
[14]	✓	✓	✓	S	L
[15]	Encrypts all the three syntaxes, the syntaxes are selected based on frame level dependency and priority			Lower than [16][19]	S
[16]	Encrypts all the three syntaxes, syntaxes a selected based on texture complexity and motion intensity			Lower than [16][19]	S
Our Objective	Encrypt all three syntaxes, syntaxes are selected based on scene transitions in P and B frames			Lower than [16][19]	H

2.6 Literature review on scene Change Detection Algorithms

Several scene change detection algorithms have been proposed in literature. Computational schemes define a scene change parameter and compare it with a fixed threshold. If the parameter reveals a substantial change, a scene transition is detected. However, a fixed threshold value cannot perform well for different videos due to the diversity of pixel characteristics. The key problem is to obtain an optimal value for the threshold. Threshold settings for particular set of sequences may vary for another set of sequences. Selection of right threshold may require a process of trial and error. Work in [17] proposed to detect scene changes based on intensity levels of motion vectors. This approach can detect changes due to complex motion but fails in case of gradual transitions with low motion intensity.

Work in [18] demonstrated to detect scene transitions using sum of absolute differences, however, calculating the SAD (Sum of Absolute Differences) values is computationally intensive. Further, fames with high intensive motion can be easily missed. Approach in [19] proposed a dynamic threshold technique based on bit rate fluctuations. This method can efficiently detect gradual scene transitions and transitions due to high motion but suitable only for variable bit rate coding. To solve the threshold selection issue, adaptive threshold based scene change detection has been proposed in section III that can dynamically fix threshold for detecting scene changes in videos with varying characteristics.

2.8 Conclusion

There is less research on multimedia encryption focusing on wireless sensor networks and wireless multimedia devices. For any encryption algorithm in wireless multimedia sensor networks, security, computational cost, and energy consumption are important constraints [20]. Computational cost means that the encryption or decryption process must have low processing time. Energy consumed is also an important constraint especially in wireless devices and wireless sensor multimedia networks, as they are battery powered [20][21]. Most of the video sensors used in a WMSN are embedded processors such as Intel Strong Arm RISC, TI davinci processors or Samsung Processors [22][23]. In spite of being powerful, these processors are battery driven; hence the computational cost must be as minimum as possible while also ensuring high security. Many recent efforts attempt to scale down the power consumption and encoding compression cost to suit wireless multimedia sensor networks [24][25]. However, the encryption scheme along with video compression increases the computational overhead and power consumption. The works in [13] & [14] have demonstrated to have high security and claim to have tractable computational cost suitable for real time encryption applications. However the number of code words selected for encryption can be further reduced to optimize the complexity. The literature review in the previous section shows that there are issues of security when the computational cost is reduced and vice versa. The previous works fail to provide a selective encryption algorithm with high security and low computational cost hence, my thesis aim to reduce the computational overhead with high security suitable for energy constrained multimedia devices.

Chapter 3 Proposed Selective Encryption Algorithm

In this chapter a new selective encryption algorithm has been proposed for energy constrained multimedia devices. Parameter sensitivity in H.264 video stream is discussed. A preliminary analysis has been done to show that intracoded macroblocks encoded in P and B frames only at scene transitions. Based on the analysis a selective encryption algorithm has been proposed which adapts and encrypts code words based on scene transitions in video sequences. The algorithm was implemented and tested in JM 18.5 reference software. The results show that the proposed algorithms has low computational cost and suitable for energy constrained multimedia device.

3.1 Introduction:

To investigate the sensitivity of syntax elements, four consolidated videos "Foreman","Soccer","Football" and "Horse Cab" were chosen at QCIF resolution under the mainline profile, IPB, with QP=18 and 4:2:0 sampling format. The scene changes between frames can occur anywhere within a GOP. The H.264 encoder uses many intra-coded macroblocks for these scenes even if they are P- or B-frames. In case of abrupt scene transitions, the first frame of the new scene is encoded as an I-frame in order to improve coding efficiency. On the other hand, if a scene contain gradual scene transitions, it is more efficient to intercode for coding performance [26].

Analysis: The syntax elements chosen for encryption are intra modes and sign bit of residuals in all the I frames, and motion vectors in P and B frames. AES stream cipher was used to encrypt the chosen syntaxes. The results in Fig .13 demonstrate that, unencrypted intracoded macroblocks in P and B frames leak information, only if there is a gradual scene transit

(a)-Foreman (b)-Football (c)-Soccer (d)-Horsecab

Figure 13: Information leaked during gradual scene transition

Table II shows the average percentage of intracoded macroblocks in the four sample videos. Table III shows the bit rate savings in the four sample videos due to intracoded macroblocks.

A: Sample videos with No Scene Transitions

Even though there are no scene transitions in "Foreman" and "Tempete" videos, the total percentage of intracoded macroblocks (P and B frames) is 4.53% and 4.04% respectively as shown in Table II. Leaving the intracoded macroblocks in P and B frames unencrypted does not affect the secrecy of the video and can achieve a bit rate savings of 4.86% and 5.89% respectively.

B: Sample videos with Scene Transitions

In "Football", "Soccer" and "Horsecab" videos, the percentage of intracoded macroblocks (P and B frames) is 16.68%, 13.29% and 6.63% respectively as shown in Table II. However, the percentage of intracoded macroblocks that contribute to scene transitions are "Football=3.26%", "Soccer=1.40%" &"Horse Cab=0.11%". Encrypting intracoded macroblocks only in frames with scene transitions can achieve a bitrate savings of "Football=11.2%", "Soccer=13.01%" & "Horsecab=12.22%".

TABLE 2 PERCENTAGE OF INTRA BLOCKS IN P AND B FRAMES

Videos	Average Percentage of IMB	Percentage of IMB in frames with Scene Transition	Percentage of IMB in frames with no Scene Transition
Foreman	4.53	-	4.53
Tempete	4.04	-	4.04
Football	16.68	3.26	13.32
Soccer	13.29	1.40	11.89
Horse cab	6.63	0.11	6.52

TABLE 3 PERCENTAGE OF INTRA BLOCKS IN P AND B FRAMES

Videos	Average Bit Rate of IMB (%)	Bitrate of IMB in frames with Scene Transition (%)	Bitrate of IMB in frames with no Scene Transition
Foreman	4.86	-	4.86
Tempete	5.89		5.89
Football	14.67	3.44	11.20
Soccer	18.54	5.53	13.01
Horse cab	15.32	3.10	12.22

3.2 Proposed Work

The contribution of this research can be depicted in two steps

1) A dynamic threshold model that adaptively selects threshold for scene change detection is proposed. The algorithm overcomes the difficulty of empirical analysis to fix the threshold.

2) The dynamic scene change detection algorithm is used to identify frames with scene transitions and encrypt intra predicted macroblocks. In the absence of scene transitions, sign bit of motion vectors are chosen as sensitive code words for encryption. Intracoded macroblocks that do not contribute to the scene transition can be left unencrypted; to achieve low complexity. Table IV shows the symbols and definitions used in sections II & III

TABLE 4 SYMBOLS AND DEFINITIONS

Symbols	Definitions
N_I	Intracoded Macroblock
N_{skip}	Skip Macroblock
T_o	Total No: of Intracoded Macroblock in P frame
N_b	No: of Backward Intercoded Macroblock
N_f	No: of Forward Intercoded Macroblocks
M	Total number of Bits encrypted in each frame
C,D	Constants depending on CPU cycles
F	Offset Flag
K	Number of Keys used
N	Number of Blocks Encrypted
E	Encryption Cost
IPM	Intraprediction Modes
SNC	Sign Bit of quantized Coefficients
MVD	Sign Bit of Motion Vectors
N_{intra}	No: of Intracoded Frames in video stream
N_{inter}	No: of Intercoded Frames in Video Stream
N_{sc}	No: of Intercoded Frames with Scene Transition
B_{ipm}, B_{mv}, B_{snc}	Bits encrypted in IPM,SNC and MVD
TH	Threshold for scene change detection

3.2.1 Proposed Dynamic Scene Change Detection

Interprediction occurs only when neighboring frames have a correlation. It is obvious that this correlation is reduced when a scene transition occurs; and hence it is concluded that when there is a scene transition, many intracoded macroblocks are predicted in P and B frames. Gradual scene transitions can be represented as,

$$\frac{t}{T}Y + \left(1 - \frac{t}{T}\right)X, 0 \leq t \leq T \qquad (1)$$

Where, T is the duration of the scene change from frames X to Y. The ratio of intracoded macroblocks and motion vectors in P and B frames are calculated, and compared to a threshold value to detect a suspected scene transition. To detect any suspected scene change 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 frame, the scene transition parameter MR_b is calculated and compared with the threshold TH. Threshold for b frame is slightly lesser than the threshold for P frame because B frames have two reference frames (I and P), whereas P frames have only one reference frame, and further, B frames are easier to intercode. If the scene change parameter is higher than the threshold, a scene transition is detected. If the scene transition parameter is low, then it indicates the absence of scene transition. Gradual scene transitions in P ($N_{sc}(p)$) and B($N_{sc}(b)$) frames can be detected by,

$$N_{sc}(p) = \begin{cases} MR_p = \frac{N_I + N_{skip}}{To} > TH \\ else, no\ scene\ change \end{cases} \qquad (2)$$

$$N_{sc}(b) = \begin{cases} MR_b = \dfrac{N_f - N_{skip}}{N_b - N_{skip}} > TH \\ else, no\ scene\ change \end{cases} \qquad (3)$$

For the adaptive threshold function, local statistical properties of the sequence have been used. The dynamic threshold selection for ith inter frame can be calculated by the empirical mean value of previous macroblock ratios. The mean M_i is given as,

$$M_i = \frac{1}{i}\sum_0^i MR(i) \qquad (4)$$

The Scaling factor S is an important function as it determines the characteristics of the thresholding function. If S takes higher values, the threshold value becomes more rigid, keeping higher values of threshold, without approaching the current scene change parameter value. A lower value of S allows detecting scene transitions efficiently, and lowering of the threshold is done to avoid false detections immediately after scene changes.

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

The threshold value for the i^{th} inter frame is given by,

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

3.2.2 Accuracy of Scene Change detection

To validate the accuracy of the adaptive scene change detection algorithm three "Foreman","Soccer","Football" and "Horse Cab" were chosen in the mainline profile. Recall and precision have been used to evaluate the proposed scene change detection. Recall and Precision are defined as follows,

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

$$Precision = \frac{N_c}{N_c + N_f} * 100$$

Where N_c, N_m, and N_f are the number of correct , miss and false detection, respectively. The above equation shows that N_m and N_f is inversely related to the accuracy of recall and precision.

TABLE 5 RECALL AND PRECISION WITH FIXED THRESHOLD

Videos	N_{sc}	N_c	N_m	N_f	**R(%)**	**P(%)**
Football	15	14	1	10	93.3	58.3
Soccer	38	26	15	11	63.4	73.3
Horse Cab	37	33	4	2	89.1	94.2

TABLE 6 RECALL AND PRECISION WITH ADAPTIVE THRESHOLD

Videos	N_{sc}	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

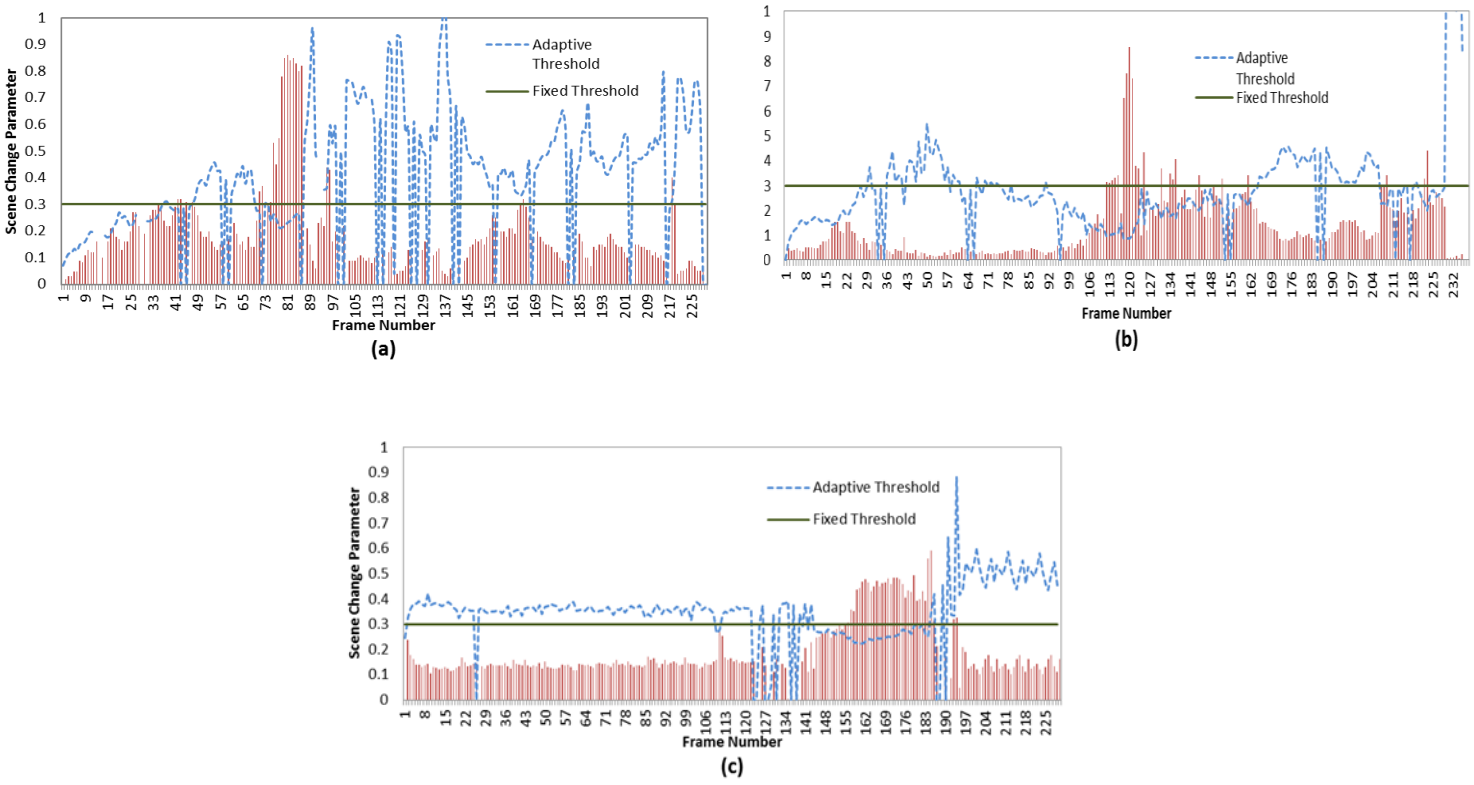

Figure 14: Proposed adaptive threshold vs fixed threshold (a)-Football, (b)-Soccer, (c)- Horsecab

The fixed threshold used for comparison was chosen empirically after analyzing macroblock ratios in all inter frames. Table V & VI shows that, the performance of the adaptive threshold scene change detection algorithm is improved compared to the algorithm with fixed threshold. It is clear that empirical analysis for fixing threshold can be avoided and adaptive threshold technique can be used to detect scene transitions. Even though, there is missed scene transition

detection, the critical frames that are perceptually insecure due to scene transition have higher percentages of intracoded macroblocks. The higher values of parameter S assures these frames are encrypted. Fig. 14 shows the versatility of the proposed adaptive threshold with different videos. Further, proposed dynamic threshold method has the advantage of low complexity, as it uses previous macroblock ratios for fixing threshold, which makes it suitable for real time applications.

3.2.3 Proposed Selective Encryption Scheme

According to the analysis in section I, the following conclusions can be made,Unencrypted intracoded macroblocks in P and B frames leak information, only during scene transitions.In order to keep low encryption overhead, the fewer the syntaxes are encrypted, the lower the time complexity is. Hence, intracoded macroblocks that do not contribute to scene transitions are left unencrypted. Thus, our encryption algorithm choses syntax elements in inter frames for encryption based on scene transition detection. As shown in fig. 15, In case of video frames with scene transition, intra modes and sign bit of residuals are chosen to encrypt, whereas in case of no scene transitions, sign bit of motion vectors are chosen for encryption.

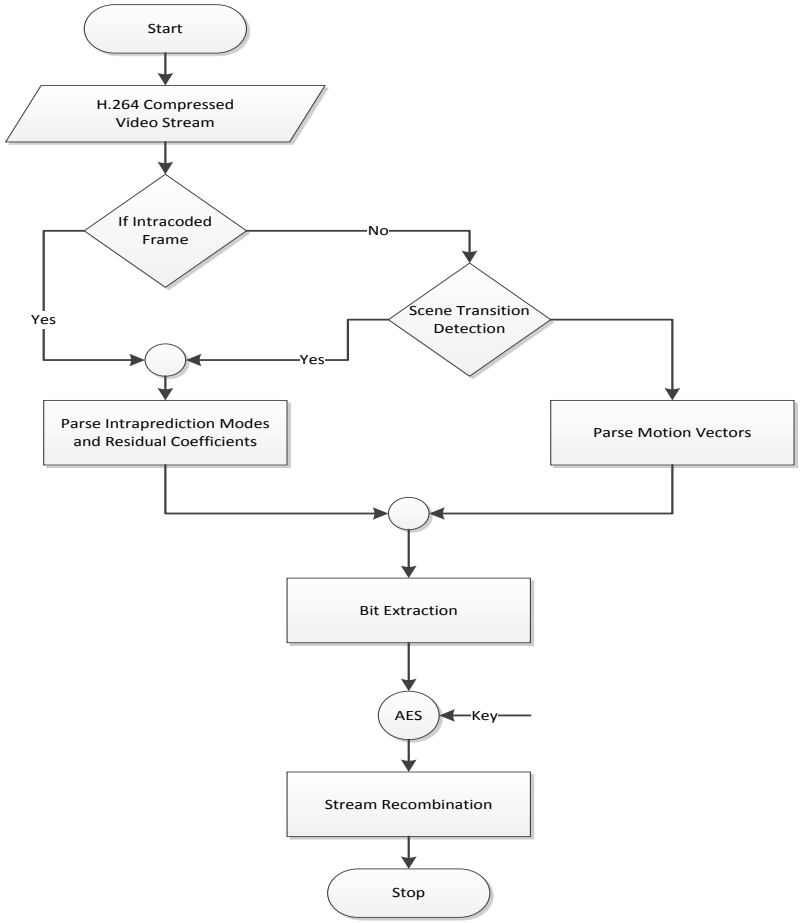

Figure 15. Functional Flowchart of Proposed Selective Encryption

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

1) **INPUT:** Video Stream, Key.
2) **IF** Current frame is an I frame
   Go to step 5 for encryption.
3) [Detection of Scene changes inter frames]
   INPUT: IF P Frame
   BEGIN
   Calculate (MRp), Go to step 2 in Adaptive threshold
   IF Scene Change Parameter Rp less than TH, Encrypt MVD
   ELSE
   Encrypt IPM, SNC
   END
   INPUT: IF B Frame
   BEGIN

```
            Calculate (MRb), Go to step 2 in Adaptive threshold
           IF Scene Change Parameter Rb less than TH, Encrypt MVD
            ELSE
          Encrypt IPM, SNC.
4) IF End of frame
     END
    ELSE Go to step 2.
```

Algorithm: Adaptive Threshold

1) **Input** MR (P or B) for i^{th} frame
2) [Calculate Threshold parameters]
   Calculate Mean M_i, Scaling factor S
3) [Update Threshold]
   IF Current Frame is P frame Update threshold TH_i
   ELSE
   TH_i=TH_i-0.20
   Store MR(i) (for calculating next threshold)
   Continue

In H.264/AVC, the intraprediction modes are encoded with exgolomb codes, CABAC or CAVLC. The intraprediction mode IPM consists of X zero, One '1' bit and X bits of Information I which can be represented as

$$IPM=2^X+I-1 \quad (8)$$

In equation (1), it is required to encrypt the information "I" for securing the low resolution spatial information. Hence,

$$B_{ipm}= I \quad (9)$$

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. 16 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,

$$\text{Prefix} = < \text{Zeros}, 1> \quad (10)$$

$$\text{Suffix} = < (|a|-1), \text{LSB}> \quad (11)$$

$$\text{LSB} = \begin{cases} |a| > 0, 1 \\ |a| < 0, 0 \end{cases} \quad (12)$$

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

$$\text{CAVLC:} \quad B_{snc} = \text{LSB} \quad (13)$$

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. 17. Here, the level consists of two syntax elements <Coef_abs_level_minus1, Coef_sign_flag>. Furthermore, <coef_abs_level_minus1> can be represented as <Prefix><Suffix> where,

$$\text{Prefix} = <|a| \text{ one's}> \quad (14)$$

$$\text{Suffix} = <|a|-14-2^{suffixlength}> \quad (15)$$

Significantbit_Coeff_Flag
Last_Significant_Coeff_Flag
Coeff_abs_level_minus1
Coeff_Sign_Flag

Figure 16.CAVLC-syntax

Coeff_token
Sign_T1
Levels
Total Zeros
Run before Zeros

Figure 17.CABAC-syntax

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

$$\text{CABAC:} \qquad B_{snc} = \text{Coef\_sign\_flag} \qquad (16)$$

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,

$$\text{<M Zeros><1><Suffix\_Info>} \qquad (17)$$

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

$$\text{CAVLC:} \qquad B_{mv} = \text{LSB of (Suffix\_Info)} \qquad (18)$$

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

$$< \text{Pre} - \text{Suffix, Sign bit} > = \begin{cases} 0 < |MVD < 9 \\ |MVD| \geq 9 \end{cases} \qquad (19)$$

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

$$\text{CAVLC:} \quad B_{mv} = \text{sign bit of (UEG3 binstring)} \qquad (20)$$

The encryption cost [30] of the AES algorithm in CTR mode can be represented as

$$E = C * N + D * K \quad (21)$$

Let E_T be the encryption data rate or cost of the selective encryption algorithm when, the intraprediction 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,

$$N_T = \sum_0^{M-1} B_{ipm} [N_{Intra} + N_{Inter}] + \sum_0^{M-1} B_{snc} [N_{Intra} + N_{Inter}] + \sum_0^{M-1} B_{mv} [N_{Inter}] \quad (22)$$

$$N_P = \sum_0^{M-1} B_{ipm} [N_{Intra} + N_{Sc}] + \sum_0^{M-1} B_{snc} [N_{Intra} + N_{Sc}] + \sum_0^{M-1} B_{mv} [N_{Inter} - N_{Sc}] \quad (23)$$

Comparing N_T and N_P we observe that,

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

Likewise we can conclude that,

$$\boldsymbol{E_T > E_P} \quad (24)$$

It is to be noted that in spite of any scene transition in an I frame, all the sensitive syntax elements are encrypted. Our aim 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. AES assures guaranteed security rather than any mathematically developed algorithm.

Chapter 4 Simulation Results

Five consolidated videos “Foreman”, “Tempete” “Football”, “Soccer” and “Horse cab” have been taken for evaluating the proposed algorithm. Here, the proposed method is compared with the approaches in [13] & [14] with respect to security and computational cost. Total decoding time is used to evaluate computational complexity, and security is evaluated by perception security and perception quality (PSNR and SSIM). The test bench for simulating our work and the approaches in [13][14] is shown in table VII. For evaluation purposes, two types of GOP structures have been used. Both these GOP structures are extensively used in wireless sensor networks and wireless multimedia applications [27].

TABLE 7 VIDEO TEST BENCH

Experimental Set Up	
Processor &RAM	2.4 GHz I-52430M processor, 4Gb RAM
Software	JM18.5
Number of Frames Encoded	260
Format	CIF
GOP Type I	IPBPB……One I frame at the start and P and B frames in the format PBPBPBPB. Inserts I frame during abrupt scene cut.
	Entropy Coding :CABAC
GOP Type II	PPPPPPP……..No I frame , with Intra refresh Mode
	Entropy Coding :CAVLC
Frame Rate	30 fps

4.1 Perceptual Security

If an encryption algorithm is too chaotic to be understood then it is considered to be an algorithm of high perception security. Fig. 18 shows the proposed encryption results for different sample videos. Fig. 4 (a)(b)(c) shows the inter frames with gradual scene transition, they are perceptual security because, the particular P and B frames with scene changes are detected and encrypted. Leaving the motion vectors unencrypted does not disturb the secrecy of the video. In case of inter frame without scene transitions, the motion vectors are chosen as syntax elements for encryption, leaving the intracoded macroblocks. Even though, there is considerable percentage of intracoded macroblocks (above 30%) in P and B frames without scene transitions, leaving the intracoded macroblocks unencrypted achieves low computational cost with acceptable secrecy as shown in figure 18-(d), (e) and (f).

(a) **(b)** (c)

(d) **(e)** **(f)**

Figure 18: Encrypted Video with proposed Algorithm

4.2 Perceptual Quality

One of the reasonable means to evaluate scrambling effect of encrypted video is through PSNR [9]. SSIM [10] is another useful metric which gives better analysis compared to PSNR [28]. However, we have considered both PSNR and SSIM metrics to evaluate the propose algorithm. Taking several videos, we evaluate the PSNRs (YUV) and SSIM for different QP parameters of the encrypted videos. The experiment's results are shown in Table VIII & IX. They show that the PSNR's of the encrypted video (proposed selective encryption algorithm) are almost similar to the PSNRS in the work of [13][14]. In case of SSIM it can be clearly seen from Table X and Table XI that even though the number of critical code word candidates, the scrambling affect is almost maintained.

TABLE 8 PSNR COMPARISON GOP TYPE I

Videos	**QP**	**Approach in [13][14]**			**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	12.90	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	10.19	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	11.3	11.89	12.80	12.10	11.98	12.8
	30	5.38	8.23	9.44	4.87	8.34	9.11

TABLE 9 PSNR COMPARISON GOP TYPE II

Videos	QP	Approach in [13][14]			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	10.1	11.24	12.45	10.5	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	10.12	11.87	6.92	11.00	12.82

TABLE 10 SSIM COMPARISON

Videos	QP	Approach in [13][14]	Proposed Approach
Foreman	18	0.11	0.1
	24	0.1	0.1
	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	0.098	0.091
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

TABLE 11 SSIM COMPARISON

Videos	QP	Approach in [13][14]	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

4.3 Computational Cost

The computational time and the volume of bits encrypted in selective encryption algorithm determines the energy consumption and speed of the encryption algorithm. High speed and low energy consumption are critical requiremements of multimedia encrption. This can be achieved by lower encryption overhead.

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 . Encryption overhead can be defined as the difference between the total processing time, with and without encryption.The total processing time is obtained from the Jm 18.5 codec[29]. Table XII and XIII shows the Computational overhead and EDR of the proposed algorithm. The simulation results clearly show that the computational complexity and EDR are lower for the proposed selective encryption scheme compared to the approaches in [4]&[7]. We observe that whichever GOP type is used, our algorithm has a low computational complexity compared to the approach in [4]&[7]. Foreman and tempete videos have low computational cost due to the absence of gradual scene transitions. However, in Football, Soccer and Horsecab we encrypt only those intracoded macro blocks that contribute to the scene transition. Like wise, the computational cost is reduced compared to the works in [4]&[7], where all the intracoded macroblocks in P and B frames are encrypted.

TABLE 12 COMPUTATIONAL COMPLEXITY

	GOP Type I		**GOP Type II**	
	Work[13,14]	**Proposed**	**Work[13,14]**	**Proposed**
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	2.61	1.92	2.77	1.92

TABLE 13 ENCRYPTED DATA RATE

	GOP Type I		GOP Type II	
	Work[13,14]	**Proposed**	**Work[13,14]**	**Proposed**
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	16.86	10.77	15.32	9.34

4.4 Replacement Attack

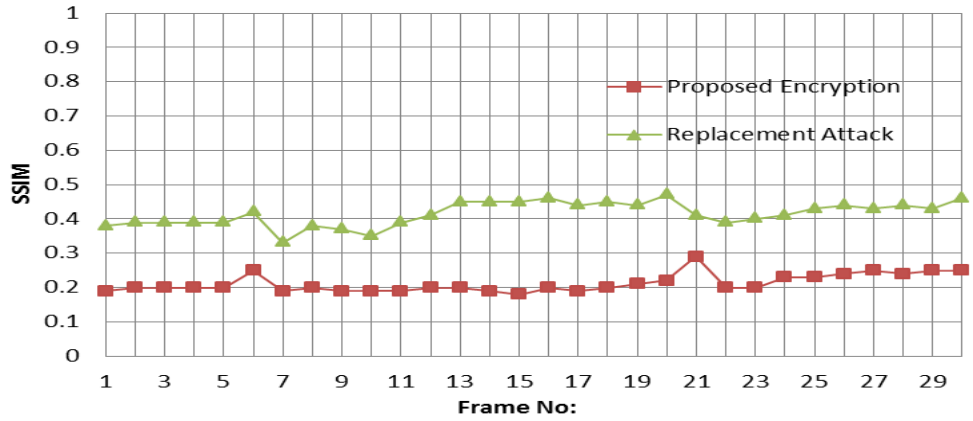

Figure 19: Replacement Attack Soccer frame 100-130 (SSIM Value)

(a) P frame(7th frame) No Scene Transition

(b) P frame(114th frame) Gradual Scene Transition

Figure 20: Security against Replacement Attacks.

The security of the proposed scheme depends on the adapted AES stream cipher. The AES cipher is not vulnerable to any kind of attack and provides good security [31]. Video encryption is vulnerable to attacks such as replacement attacks [32]. In this case, the encrypted code words are replaced by fixed values to enhance the video quality. The replacement attack has been demonstrated on the Soccer video (encrypted by the proposed algorithm) by replacing all the sign bits to positive values and intraprediction modes to 1. Fig.19 shows the SSIM values of the Encrypted Video and the encrypted video with replacement attack. Fig. 5 shows that the perceptual quality could not be regained after the replacement attack. Fig. 20 (b) shows the P frame (114th Frame) with scene transition. The frame is well scrambled as both the intraprediction modes and residual coefficients are encrypted. Fig. 20 (a) (7th frame) shows the frame without scene transition.The results show that even if the intracoded macroblocks are left unencrypted (in case of no scene transition), no information can be retrieved after the replacement attack. Hence, the proposed algorithm can withstand replacement attacks.

4.5 Deactivated Motion Compensation attack

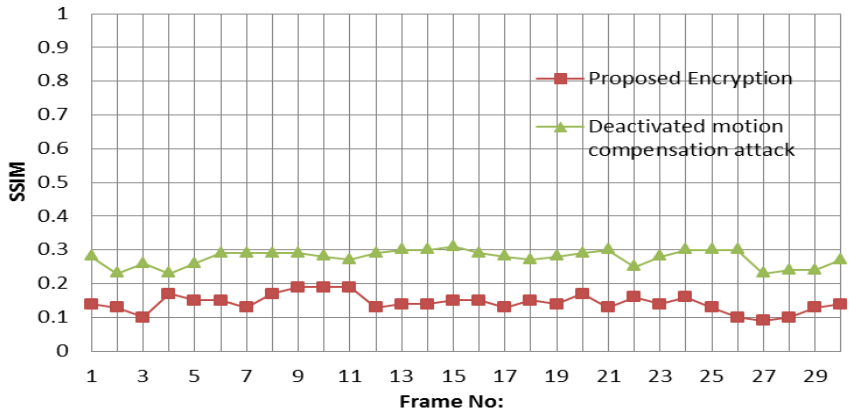

Figure 21: Deactivated Motion Compensation Attack Soccer frame(SSIM Value)

(a) P frame(13th frame) No Scene Transition

(b) P frame(25thth frame) No Scene Transition

Figure 22: Security against Deactivated Motion Compensation Attack

To analyze the fraction of intracoded macroblocks that are not encrypted, deactivated motion compensation attack has been implemented on football video. In this attack, the encrypted video is decoded with deactivated motion compensation i.e. only non-encrypted intra blocks 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 blocks. Fig.21 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.22 (a) & (b) shows the inter frames in which the intracoded macroblocks were left unencrypted. From the figure, it is clear that the unencrypted intracoded macroblocks does not affect the security of the video.

Chapter 5 Conclusion and Future Work

Encryption on the multimedia is essential in both commercial broadcasting and peer to- peer communication. Selective encryption takes relatively small amount of time, compared to the decoding process; the time is not negligible, however. A low overhead yet maximally secure encryption method is required, as not all video streams are of equal value. In order to accomplish this task, a selective encryption based on scene transitions is proposed, which consists of a simple model to detect scene transitions and make a decision for encryption.

5.1 Summary

In this work, it was demonstrated that not all intracoded macroblocks in P and B frames leak information when left unencrypted. Based on the percentage of intracoded macroblock analysis, 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. 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 attacks.

5.2 Advantages of Proposed Algorithm

- The proposed encryption algorithm chooses the code words for encryption based on scene transitions, which completely depends on the video content.
- The results and analysis show that the algorithm can provide good scrambling effect with low computational overhead suitable for energy constrained multimedia devices.
- The scene change detection algorithms implemented in certain codecs can be utilized for the purpose of encryption. Thus, making the entire model simple.

5.3 Future work

- Future work aims at designing selective encryption schemes which have low overhead and proficient security.
- Tunable selective encryption algorithms could be developed for balancing computational cost and security.
- The algorithm and accuracy of the scene change detection algorithm for encryption can be improved.

Bibliography

1. E.Richardson, "The H.264 Advanced Video Compression Standard", in 2010 Wiley Publications Ltd.
2. ITU-T. Rec.(ISO/IEC 14496-10):2010, Advanced Video Coding for Generic Audio Visual Services.
3. F. Liu and H. Koenig " A survey of video encryption algorithms" Comput. Security, vol. 29, no. 1, pp.3-15,2010.
4. Y. Shi and H. Sun, Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards. Boca Raton. FL : CRC Press, 2000.
5. A. Puri, X. Chen, and A. Luthra, "Video coding using The H.264/MPEG-4 AVC Compression standard", Signal processing : Image communication, vol. 19, no:9, pp 793-849, oct 2004,
6. Performance evaluation of encryption algorithm for wireless sensor networks Ben Othman, S.ISITC, Univ. of Sousse, and Hammam Sousse, Tunisia.
7. F.Dufaux and T. Ebrahimi "Scrambling for privacy protection in video surveillance systems" IEEE Trans. Circuits syst. Video Technol., vol. 18, pp. 1168-1174, 2008
8. R.Iqbal, S. Shirohammadi and A.El-Saddik "Secured MPEG-21 digital item adaptation for H.264 video" PRoc. ICME, pp. 2181-2184, 2006.
9. 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
10. 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.
11. Z. Shahid, M. Chaumont, and W. Puech, "Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I & P Frames," IEEE Trans. Circuits Syst. Video Technol., vol 21, no. 99, pp. 565–576, May 2011.

12. Z. Shahid, M. Chaumont, and W. Puech, “Fast protection of H.264/AVC by selective encryption of CABAC,” in Proc. IEEE Int. Conf. Multimedia Expo, Jun.–Jul. 2009, pp. 1038–1041
13. 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
14. Tunable Encryption Scheme and Analysis of Fast Selective Encryption for CAVLC and CABAC in H.264/AVC Yongsheng Wang, Student Member, IEEE, Maier O’Neill, Senior Member, IEEE, and Fatih Kurugollu, Senior Member, IEEE, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 23, NO. 9, SEPTEMBER 2013
15. Wei Wang, Micheal Hempel, Dongming Peng, Hongang Wang, Hamid Sharif,”On Energy Efficient Encryption for Video Streaming in Wireless Sensor Networks”, IEEE Transactions on Multimedia.
16. Yingdi Zhao,Li Zhuo “A Content-Based Encryption Scheme for Wireless H.264 Compressed Videos”, Wireless Communications & Signal Processing (WCSP), 2012 International Conference on
17. 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.
18. S. Youm and W. Kim, “Dynamic threshold method for scene change detection”, In Proc. ICME2003, Vol. 2, pp. 337-340, July. 2003.
19. 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.
20. Misra et al : A survey of Multimedia Streaming in wireless sensor networks ,IEEE Communications Survey & tutorial ,Vol. 10 No.4 Fourth Quarter 2008
21. Wireless Multimedia Sensor Networks: Applications and Testbeds By Ian F. Akyildiz, Fellow IEEE, Tommaso Melodia, Member IEEE, and Kaushik R. Chowdhury, Student Member IEEE.
22. Intel Strong ARM RISC Embedded Processors URL: http://www.intel.com/design/strong/datashts/278241.htm.

23. Samsung S3C44B0X RISC Embedded Microprocessor. URL:http://www.samsung.com/products/semiconductor/ mobilesolutions/mobileassp/mobilecomputing/s3c44b0/s3c44b0.htm.
24. 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.
25. Z. He and D. Wu. Resource allocation and performance analysis of wireless video sensors. IEEE Trans. Circuits Syst. Video Technol., 16(5):590–500, May 2006.
26. 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.
27. Multimedia over IP and Wireless Networks: Compression, Networking, and Systems by Mihaela van der Schaar, Philip A Chou.Academeic press,2011.
28. 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.
29. JM Reference Software, ver. 18.5 (2012) [online]. Available http:://iphome.hhi.de/suehring/tml.
30. FIPS 197 (Advanced Encryption Standard),NIST Publications, 2001.
31. NIST Special Publication 800-57, Recommendation for Key Management, 2012
32. 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.
33. Midya, A.; Sengupta, S. "Scene transition based adaptive GOP selection for increasing coding efficiency & resiliency", Informatics, Electronics & Vision (ICIEV), 2012 International Conference on, On page(s): 770 – 773
34. Yao, W.; Rahardja, S. "Dynamic threshold based keyframe detection", Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on, On page(s): 2137 - 2141