International Conference on Communication Technology and System Design 2011

An Efficient Scalable Video Encryption Scheme for Real time applications.

L.M. Varalakshmi\(^a\), G. Florence Sudha\(^b\), G. Jaikishan\(^c\), a*
Dept. of ECE, Sri Manakula Vinayagar Engineering College, Puducherry.
Dept. of ECE, Pondicherry Engineering College, Puducherry.

Abstract

This work describes an efficient selective encryption scheme for H.264/Scalable Video Coding (SVC). The main feature of this scheme is that it makes use of the characteristics of SVC and fully meets the encryption requirements. The encryption procedures are carried out at the Network Abstraction Layer (NAL) level. The proposed scheme performs encryption in three domains: Intra-Prediction mode (IPM), residual data and motion vector difference values. For enhancement layers, temporal scalability and spatial/SNR scalability are distinguished. Experiments were performed to verify the proposed method using the joint scalable video model (JSVM), and the experimental results show that the proposed method protects the SVC streams effectively and supports full scalability and can guarantee the robustness to transmission errors.

1. Introduction

Multimedia applications involving the transmission of video over communication networks are rapidly increasing in popularity. These applications include multimedia messaging, video telephony, video conferencing, wireless and wired Internet, video streaming, cable and satellite TV broadcasting etc. In general, the communication networks supporting these applications are characterized by a wide variability in throughput, delay and packet loss. Furthermore, a variety of receiving devices with different resources and capabilities are commonly connected to a network. Scalable video coding (SVC) is a highly suitable video transmission and storage system designed to deal with the heterogeneity of the modern communication networks. SVC supports efficient coding of video in such a way that multiple versions of the video signal can be decoded at a range of bitrates, spatial resolutions and/or temporal resolutions or frame rates [1]. The scalability of the SVC format allows easy rate adaptation in any of the scalable dimensions (temporal, spatial and quality) in the compressed domain by removal of parts of the SVC format stream.

* L.M. Varalakshmi. Tel.: 9442067514; fax: 0413-2641136.
E-mail address: varalakshmi_1@yahoo.co.in.
The protection of scalable multimedia data has become increasingly important in recent years [2]. In the past decade, some video encryption algorithms have been reported, most of which are based on MPEG 2 codec and can be classified into two types: complete-encryption algorithm and partial-encryption algorithms [3][4][5]. The first one encrypts raw data or the compressed data directly with traditional or chaotic ciphers. The second type combines encryption with compression and encrypts videos selectively. Such algorithms encrypt the signs of DCT coefficients or motion vectors; some permute DCT coefficients while some algorithms combine encryption process with Variable Length Codes (VLC). These algorithms satisfy real-time requirement and keep file format unchanged. Currently the encryption schemes for SVC are restrictive. The enhancement structure for the scalability is considered in [6]. Base layer encryption could provide enough security of the whole bit-stream because scalable contents are enhanced from the base layer. Y. G. Won et al. [7] proposed an encryption algorithm to protect SVC bit-stream and a method for conditional access control to consume the encrypted SVC bit stream. An encryption method to protect the region of interest (ROI) of SVC was presented in [8]. Recently, a selective encryption scheme for SVC was designed by S. W. Park et al. [9], but the key generation and distribution was not discussed in further detail. So an integrity solution including both encryption method and key management for protection of scalable video coding is still necessary.

An efficient video encryption scheme is presented in this work which encrypts both the Base layer (BL) and Enhancement layer (EL) and maintains the scalability of SVC. This work aims at coding multiple versions of the same video content into a single encoded bit stream by combining all the scalability features and performing selective encryption. The rest of the paper is organized as follows. Section 2 briefly introduces the concept of SVC. The proposed encryption scheme is discussed in detail in Section 3. Experimental results are analyzed in section 4 and concluding remarks are given in section 5.

2. Overview of SVC

The design of SVC allows for spatial, temporal, and quality scalabilities. The video bit stream generated by SVC is commonly structured in layers, consisting of a base layer (BL) and one or more enhancement layers (ELs). Each enhancement layer either improves the resolution (spatially or temporally) or the quality of the video sequence. SVC provides spatial, temporal, and SNR scalability with high coding efficiency. Spatial scalability is achieved by layered coding and the temporal scalability is achieved by hierarchical B picture structure[9]. In the case of SNR scalability, fine granular scalability (FGS) and coarse granular scalability (CGS) are employed. By decoding the base layer, the lowest quality of original video can be obtained. Enhancement layers are added on the base layer to get a better quality [10]. Figure 1 shows the structure of spatial, temporal and SNR scalability in SVC.

![Fig. 1. Spatial, Temporal and SNR scalability.](image)

One of the fundamental issues in the development of SVC was its integration into the existing H.264/AVC standard. Largely responsible for the successful integration was the conceptually clear structure of H.264/AVC, which distinguishes between a coding layer (VCL, video coding layer, and non-VCL, non video coding layer) and a network abstraction layer (NAL). The VCL is responsible for creating a coded representation of the moving
pictures, while the NAL formats these data and provides header information in a simple and effective fashion. VCL data are organized into NAL units, which start with a one byte header. Most important is the type of a NAL unit (NUT) that is inferred from the NAL unit header. In SVC, NAL is the smallest unit through which spatial, temporal and SNR scalability can be defined. In order to make the encryption scheme applicable in bit-stream extraction process, the encryption should be applied at the NAL level. Depending on the NUT (NAL unit type) the subsequent data are interpreted. An SVC NAL unit header extension is specified (Fig. 2(b)), which contains valuable information about the NAL unit content, such as the PID (priority id), the DID (dependency id), the QID (quality id) and the TID (temporal id).

![Fig.2. SVC extension structure of NAL unit](image)

3. Proposed work

From the above analysis of the structure of SVC, it is seen that the base layer of SVC bit-stream is more important than the enhancement layers because the information in the base layer is the basis for enhancement layers. An efficient video encryption scheme is proposed in this work, according to the encryption requirements that encrypts different parts of base and enhancement layers to satisfy the trade-off between security and computation cost.

3.1 Base layer Encryption

As the base layer should be protected with the highest level of security, Intra prediction modes(4x4 and 16x16), MV and Residue data are encrypted.

3.1.1. Fixed length Encryption for Intra (4x4) Prediction Mode

In intra-prediction coding, the intra-prediction mode changes with the block size. There are 16 modes for 4×4 luma block, 4 modes for 16×16 luma block and 4 modes for 8×8 chroma block, respectively. In intra mode, a prediction block is formed based on previously encoded and reconstructed blocks and is subtracted from the current block prior to encoding. The encoder typically selects the prediction mode for each block that minimizes the difference between prediction block and the block to be encoded. The choice of intra prediction mode for each 4x4 block must be signaled to the decoder and Intra 4 x 4 prediction mode is encoded with 3 bits fixed length code by utilization of a flag bit. When the flag bit is 1, the encoder does not send the prediction mode and no scrambling is done in this case, but when the flag bit is 0, intra prediction modes are modified by using the stream cipher to directly encrypt the 3 bits prediction mode.

3.1.2. Exponential-Golomb Code Encryption for Intra(16 x16) Prediction Mode(IPM) and Motion Vectors (MV)

The intra prediction modes (4 modes) for 16x16 luma block are encoded with Exp-Golomb codes. This kind of codeword is composed of R zeros, one 9 ‘1’-bit and R bits of information (Y). Here, the intra-prediction mode is X=2AR+Y-1 and the encryption process is shown in Fig. 3. That is, X is firstly encoded into a variable-length code with Exp-Golomb coding, and then only the information part Y is encrypted into Z with a stream cipher. This encryption algorithm realizes encryption and variable-length coding at the same time and keeps the length of the codeword unchanged. Encrypting the IPMs alone does not provide efficient security in inter frames (P-frames and B-frames). Therefore, the proposed scheme encrypts the MVs in the inter-frame using the stream cipher as the motion vectors are also encoded with Exp-Golomb codes.
3.1.3. Texture (Residual Data) Encryption

The texture of the IPMs and the MVs are compressed using DCT transform and entropy coding. The entropy coding uses powerful compression methods like Context-based Adaptive Variable Length Coding (CAVLC) and Context-based Adaptive Binary Arithmetic coding (CABAC). In the base-layer encryption, however, only CAVLC is considered as it supports all profiles. During CAVLC encoding, parameters as the number of coefficients, trailing ones (coeff_token), the sign of each T1, the levels of the remaining non-zero coefficients, the total number of zeros before the last coefficient and each run of zeros are encoded respectively [10]. Among them, the sign of each T1 and the levels of the remaining non-zero coefficients are more sensitive to the understandability of the video.

![Fig. 3. Encryption process for IPM and MV](image)

It is preferred to encrypt only the signs of T1 and the levels of the remaining non-zero coefficients. The encoding/encryption process is shown in Fig. 4, for which the decoding/decryption process is symmetric.

3.2. Encryption of Enhancement layers

The proposed scheme applies a light-weighted encryption on enhancement tier. S. W. Park et al. has shown experimentally [9] that the residual data domain occupies most parts in the spatial scalability and the SNR scalability layer, while both motion vectors and residual data are important components in the temporal scalability layer.

3.2.1. Temporal scalability layer

In the temporal scalability layer, SVC adopts hierarchical B prediction structure and both MV and residue data should be encrypted to gain high security. The encryption process is described in Fig. 5.
### 3.2.2 Spatial and SNR scalability layer

Spatial scalability is achieved by an oversampled pyramid approach. SNR scalability can be considered as a special case of spatial scalability for which the picture sizes of base and enhancement layer are identical. The sign of residual data alone is encrypted using a stream cipher.

### 3.3 Key Generation

In the proposed encryption scheme, encryption is carried out at the NAL level. In order to maintain security and scalability of SVC, the layer-based encryption operations can be controlled by different keys. The key used by stream cipher is generated by using the NAL_unit_type, dependency_id, temporal_id, and quality_level for each NAL unit. In this work, two spatial layers (spatial 0, spatial 1), two temporal levels (temporal 0, temporal 1), and two quality layers (SNR 0, SNR 1) are considered. If a user wants to access $s$ spatial layers, $t$ temporal layers, and $q$ SNR layers, a set of the keys is needed to decrypt the encrypted NAL units [7]. NAL unit key is represented as Key (spatial layer, temporal level, SNR layer). The number of keys needed to encrypt SVC bit stream can be written as shown in equation (1)

$$\text{Key}_{\text{Total}} = \sum_{s=1}^{\text{NS}} (NQ_s \times NT_s),$$  

(1)

where $\text{NS}$ is the number of spatial layers, $NQ_s$ is the number of SNR layers in $s$-th spatial layer, and $NT_s$ is the number of temporal levels in $s$-th spatial layer. SNR and temporal scalabilities are related with corresponding spatial layers. The key set to access the SVC bit stream with $s$ spatial layer, $t$ temporal level, and $q$ SNR layer can be expressed as shown in equation (2)

$$\text{KEYSET}_{s,t,q} = \left\{ \text{Key}(l,m,n) \bigg| \begin{array}{l} 0 \leq l \leq s, \\
0 \leq m \leq \min(t,NT_s), \\
\text{if } l = s \text{ then } 0 \leq n \leq q, \text{ else } 0 \leq n \leq NQ_s \end{array} \right\},$$

(2)

### 4. Experimental Analysis

The proposed encryption scheme is implemented in JSVM software (version 9.13.1) [11]. Test sequences Bus, Foreman and Mobile are used for simulation. The sequence is encoded by 2 spatial layers (CIF, QCIF), 2 temporal levels (15fps, 30fps) and 2 SNR layers (base SNR layer, enhancement SNR layer). The encryption results of the test video is shown in Fig. 6.
In the proposed scheme, both the predicted information and the residue data are encrypted using stream cipher RC4, which makes the video unintelligible. This is seen in Fig. 6 which shows that the encryption scheme offers high perceptual security. QCIF video is represented with CAVLC and CIF uses CABAC for the entropy coding. And, domains of each layer are encrypted using different keys extracted from the stream cipher.

### 4.1 Security Analysis

The security of the proposed algorithm is checked against ciphertext-only attack and known-plaintext attack. Ciphertext-only attack is the most difficult attack since the cryptanalyst has access only to the encrypted data. For a video frame of size 176 × 144, the number of luminance macroblocks would be 11×9, the computational complexity of breaking the first step would be 511^11*9*10 where 10 specify the number of XOR operations per block. Hence the overall cost of XOR and permutation makes breaking the video file practically infeasible, making the proposed scheme robust to ciphertext-only attack. In the case of known plaintext attack, the unauthorized user has the original video, the corresponding encrypted video and the encryption algorithm. Renewing of the key has been done at periodic intervals and so it is secure against known-plaintext. Meanwhile, the scheme is also secure against exhaustive attack, since it is almost impossible to completely get plaintext which requires about 2^128 attacks.
4.2 Peak Signal-to Noise ratio (PSNR)

PSNR which is an objective measure to test the quality of the reconstructed video is given in Table 1 for different test videos (Bus, Foreman and Mobile). As shown in Table 1, the PSNR value of the encrypted video is lower than that of the original video. Thus, the objective measure as well as the perceptual measure shows the strength of the proposed encryption scheme.

Table 1. PSNR results of test videos.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Format</th>
<th>Bus</th>
<th>Foreman</th>
<th>Mobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original video</td>
<td></td>
<td>Y 27.57</td>
<td>28.94</td>
<td>26.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U 36.43</td>
<td>35.42</td>
<td>30.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V 37.69</td>
<td>36.75</td>
<td>31.87</td>
</tr>
<tr>
<td>Encrypted video (BL alone)</td>
<td></td>
<td>Y 21.56</td>
<td>23.02</td>
<td>20.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U 29.63</td>
<td>28.24</td>
<td>24.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V 31.98</td>
<td>29.72</td>
<td>25.67</td>
</tr>
<tr>
<td>Encrypted video (BL + EL)</td>
<td></td>
<td>Y 9.24</td>
<td>10.48</td>
<td>9.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>U 13.47</td>
<td>12.78</td>
<td>10.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V 14.51</td>
<td>13.87</td>
<td>11.62</td>
</tr>
</tbody>
</table>

4.3 Computational Complexity.

The encrypted data (IPM, MV and texture) volume is less compared with the whole volume of video. Hence the operation time ratio between encryption and encoding process is very small as shown in Table 2. Time ratio is a parameter which measures encryption time with respect to encoding time and decryption with respect to decoding time.

Table 2. Time ratio of the proposed encryption scheme.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Format</th>
<th>Encryption/Compression (%)</th>
<th>Decryption/Decompression(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>QCIF</td>
<td>0.72</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.83</td>
<td>0.45</td>
</tr>
<tr>
<td>Foreman</td>
<td>QCIF</td>
<td>0.91</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.96</td>
<td>0.43</td>
</tr>
<tr>
<td>Mobile</td>
<td>QCIF</td>
<td>0.50</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.81</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Table 2 shows that the encryption/decryption operation does not affect the compression/decompression process much, because the time ratio is very less. The proposed encryption scheme thus has low computational complexity, making it suitable for real-time streaming of the video content.

4.4 Bit Overhead Analysis.
In the encryption process, the Intra prediction mode and Motion vectors encryption does not cause overhead, but texture encryption causes bit overhead through CAVLC. The average increase in bit rate due to encryption is shown in Table 3. Bit overhead is calculated using the equation given in (3).

\[
\text{Bit overhead} = \frac{\text{Difference between Encrypted and Encoded bits}}{\text{Encoded bits alone}} \quad (3)
\]

Table 3. Bit overhead of the proposed encryption scheme.

<table>
<thead>
<tr>
<th>Video Sequence</th>
<th>Format</th>
<th>Bit Overhead (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>QCIF</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.012</td>
</tr>
<tr>
<td>Foreman</td>
<td>QCIF</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.009</td>
</tr>
<tr>
<td>Mobile</td>
<td>QCIF</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>CIF</td>
<td>0.007</td>
</tr>
</tbody>
</table>

As seen in table 3 the bit overhead for different test video is very less and it will not affect the transmission as the encrypted bit stream has little difference from the encoded bit stream.

5. Conclusion.

In this paper, an efficient encryption scheme for Scalable Video Coding (SVC) is proposed. For Base layer, intra-prediction mode, motion vectors and residua data are encrypted, while enhancement layers are encrypted based on the type of scalability and the most prominent data of that type. Experimental results show that the proposed algorithms satisfy SVC encryption requirements by providing high security, low computation cost as well as robustness to transmission errors. The proposed scheme is also computationally efficient by encrypting domains selectively according to each layer, offers security through the use of mutually different keys and is format compliant by utilizing the H.264/SVC structure.

References