THE IMPROVED SIGN BIT ENCRYPTION OF MOTION VECTORS FOR H.264/AVC

Yongsheng Wang, Maire O'Neill, Fatih Kurugollu

Center for Secure Information Technologies (CSIT)
Queen's University, Belfast, BT3 9DT, UK
Email: {ywang26, m.oneill, f.kurugollu}@qub.ac.uk

ABSTRACT

In this paper, an improved video encryption method for encrypting the sign bit of motion vectors is proposed based on H.264/AVC, which belongs to selective encryption. This method improves upon previous work involving the sign bit encryption of motion vectors by ensuring the four candidates for the encrypted motion vectors are always located in two orthogonal lines. The improved method can provide a much more effective scrambling effect while keeping the encrypted stream format-compliant and the compression ratio unchanged. The combination of the proposed method with encryption of intra prediction modes can further enhance the scrambling effect, especially for the first few frames which are left clear when only the motion vectors are encrypted.

Index Terms— Video encryption, selective encryption, motion vectors, intra prediction, H.264/AVC, format-compliant

1. INTRODUCTION

With the rapid development of networks and information technology, the illegal use and piracy of video information is now widespread. As a result, video communication applications such as video-on demand, video conferencing and video broadcasting all require security to protect the content.

Due to the unique characteristics of video information, directly applying classical cryptographic algorithms to video streams requires high computational cost and would conflict with the compression efficiency and the syntax of the original video stream. This has spurred on researchers to look for new video encryption methodologies. Selective encryption is one of the most promising technologies to meet the diverse requirements of practical video applications [1] [2]. Its basic concept is to selectively encrypt the important information in a video stream such that if the encrypted information was incorrect, the decoded video would appear as if noise had been added to the original video. In particular, when considering the scenario of handheld devices, like PDAs, cell phones, *etc.*, the computational cost incurred by the encryption procedure will become an important aspect for practical applications. Therefore, selective encryption is much more suitable for such applications.

The latest research on video compression, H.264/AVC [3] (Part 10 of MPEG-4), is widely adopted in various video applications, and can significantly outperform previous compression standards in terms of compression performance [4] [5]. Selective encryption in H.264/AVC has become the focus of recent research. Ahn *et al*. [6] proposed a scrambling method based on the encryption of intra prediction modes, denoted as IPM in this paper. This method is syntax-compliant and can maintain the compression ratio, however, it leaves the motion information clear. Li *et al*. [7] proposed to encrypt the intra and inter prediction modes, the transform coefficients and the sign bit of motion vectors; however, this scheme degrades the compression ratio. Lian *et al*. [8] [9] proposed a further scheme, which also utilized IPM and included a sign bit encryption of nonzero transform coefficients and motion vectors. But this method leads to a relatively high computational cost, since each nonzero coefficient needs one random bit and the number of nonzero coefficients in a frame is very large.

Shi and Bhargava [10] first proposed the random flipping of the sign bit of motion vectors (referred to as EMV in this paper) and the sign bit of transform coefficients. However, as mentioned previously, encrypting the sign bit of coefficients can lead to a relatively higher computational cost. Liu and Li [11] developed a motion vector encryption algorithm by XORing motion vectors with a random number and relocating their positions in the video stream, which decreased the compression ratio. The research in [10][11] is based on MPEG-1. Lian *et al.* [8] first extended the sign bit encryption of motion vectors to H.264/AVC. Kwon *et al*. [12] proposed a scrambling method for H.264/AVC by relocating differential motion vectors and the macroblock data within the same slice. However, this method incurs a longer delay when encoding since the relocation happens in the range of a slice. In this paper, an improved method for encrypting the sign bit of motion vectors based on H.264/AVC is proposed. The improved method has the same computational cost as the previous sign bit encryption of motion vectors and can scramble the video content much more effectively. However, only encrypting the motion vectors will leave the first few frames in the video sequence clear or in a good perceptual quality. Thus, the pro-

posed method can be combined with the method in [6] to further enhance the scrambling effect.

The rest of this paper is arranged as follows. In Section 2, a brief overview of H.264/AVC is presented. The improved sign bit encryption of motion vectors and its combination with IPM [6] are demonstrated in Section 3. In Section 4, experimental results and the performance of the proposed method are given. Finally, conclusions are drawn in Section 5.

2. OVERVIEW OF H.264/AVC

In H.264/AVC, a video sequence is treated picture by picture. Each frame consists of a number of slices, and each slice includes some individual coding units, called macroblocks, each of which contains one 16x16 luminance (Y) array and two corresponding chrominance (Cb and Cr) arrays. A macroblock may be encoded in intra or inter prediction mode.

For a video sequence, the slices in the first frame are always encoded as I slices, where each macroblock is coded in intra prediction mode; the slices in the following frames are often encoded as P or B slices, where each macroblock can be coded in intra or inter prediction mode. The choice of prediction mode is decided by optimizing the distortion.

2.1. Intra Prediction

A macroblock coded in intra prediction mode has two states: intra 4x4 prediction mode and intra 16x16 prediction mode. For high profile, there is a further state, the intra 8x8 prediction mode, which is developed in the later version of H.264/AVC and is not considered in this paper for simplicity.

When the macroblock is coded in intra 4x4 prediction mode, a 16x16 macroblock is partitioned into 16 4x4 blocks, each of which chooses the best intra 4x4 prediction mode from 9 candidate modes to minimize the distortion. In the case of intra 16x16 prediction mode which has four possible modes, the mode optimizing the distortion is adopted as the practical intra 16x16 prediction mode[3][4][5]. The predictions of pixels in a 4x4 block or a macroblock are obtained by linear interpolation of its adjacent pixels as shown in Fig. 1, which can remove the spatial redundancy to reach the aim of compressing the video information.

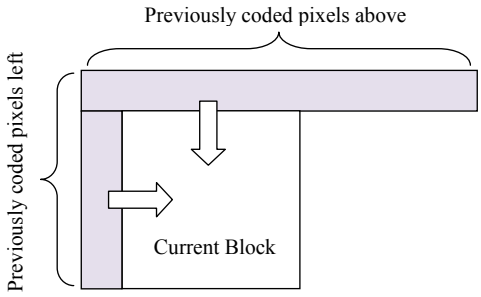


Fig. 1. Intra prediction.

2.2. Inter Prediction and Motion Vectors

In the case of inter prediction, the macroblock may be kept as one 16x16 macroblock partition (covering the whole macroblock), or partitioned into some blocks of size 4x4, 4x8, 8x4, 8x8, 16x8 and 8x16. In a P macroblock, each partitioned block is predicted by a prediction region from one previously coded reference picture; in a B macroblock, each partitioned block is predicted by one or two prediction regions from one or two previously coded reference pictures. As shown in Fig. 2, MB1 is a P macroblock and MB2 is a B macroblock. Inter prediction can compress the video information by eliminating the temporal redundancy.

The offset between the partitioned block and the corresponding prediction block is called a motion vector (MV). Thus, each partitioned block in a P macroblock only has one motion vector; each one in a B macroblock has one or two motion vectors. If the difference between the partitioned block and its prediction region exists, known as the residual data, it will be transformed by the integer DCT, quantized, and then coded in the entropy coding. Because of the high correlation of motion information of neighbouring blocks, only the differential motion vector (DMV) between the current MV and the predictive MV is coded. The predictive MV is calculated from the previous coded DMV [3][4][5].

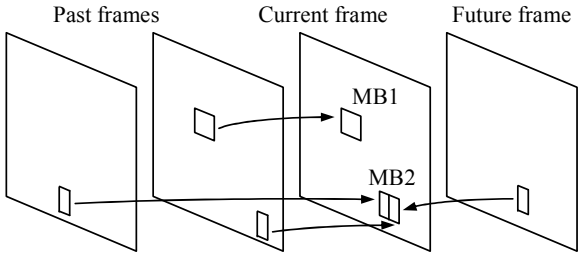


Fig. 2. Inter prediction and motion vectors.

3. THE IMPROVED SIGN BIT ENCRYPTION OF MOTION VECTORS

3.1. The Improved Method

The sign bit encryption of motion vectors has been integrated into some existing video encryption schemes [7][8][9][10]. For each of two coordinates of an original motion vector MV(x,y), the corresponding sign bit is flipped according to a random bit sequence. As shown in Fig. 3 (a), the encrypted MV could be one of four candidate vectors, A, B, C and MV. When MV is adjacent to the x axis, MV and A are close to B and C, respectively, and may even coincide when MV is on the x axis. Similarly when MV is adjacent to the y axis: MV and B will be close to A and C, respectively. In addition, the adjacent blocks in a picture are often correlated and have similar texture. Thus, when MV is close to the x or y axis, the difference between it and the encrypted MV (one of four

candidate vectors) is possibly too small to result in a good scrambling effect. Therefore, the scrambling effect is significantly decreased in comparison to when MV is in the middle of a quadrant of the coordinates.

With the aim of improving the scrambling effect under such cases, an improved sign bit encryption method is proposed. The basic idea is always to keep the line A-B orthogonal to the line C-MV, as shown in Fig. 3 (b). It can be implemented according to the following pseudo-code:

```
j_rand = generate two random bits;
switch (j_rand)
  case 00: mv_x_encrypted =   mv_x;
           mv_y_encrypted =   mv_y;
  case 01: mv_x_encrypted = - mv_y;
           mv_y_encrypted =   mv_x;
  case 10: mv_x_encrypted =   mv_y;
           mv_y_encrypted = - mv_x;
  case 11: mv_x_encrypted = - mv_x;
           mv_y_encrypted = - mv_y;
```

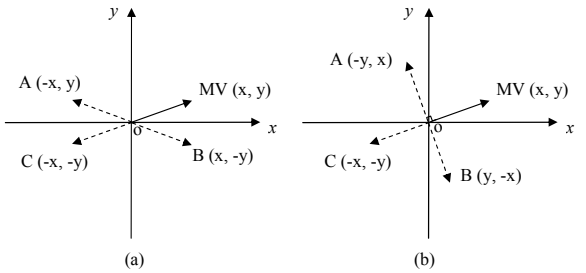
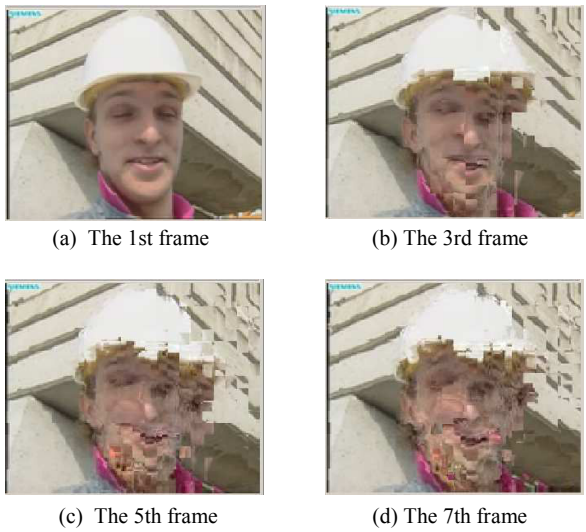


Fig. 3. Illustration of (a) the previous sign bit encryption of motion vectors and (b) the improved method.

3.2. Combination with IPM

The method of encrypting intra prediction modes in [6] is denoted as IPM in this paper. IPM can effectively scramble the I macroblock in the sequence, but leaves the motion information clear. Thus, the proposed improved method can be combined with IPM. This combination can scramble the video sequence much more effectively to provide stronger protection. Since neither technique affects the compression ratio, the combination of the two is also expected to maintain the compression efficiency.

Fig. 4 shows the 1st, 3rd, 5th and 7th frames of 'foreman' encoded in the baseline profile, with QP=18 and only the sign bit of motion vectors encrypted by the proposed method. Here, QP is the quantization parameter, which can adjust the video quality when decoded and affect the compression ratio. Generally, a smaller QP means a higher video quality and lower compression ratio, and vice versa. It is clear that the first few frames in the video sequence are not effectively scrambled. In particular, the first frame is not scrambled since it is an I frame, which does not have motion vectors. In addition, the background of these frames also shows a good perceptual quality.



(a) The 1st frame (b) The 3rd frame

(c) The 5th frame (d) The 7th frame

Fig. 4. The first few decoded frames of 'foreman' encoded in the baseline profile, with QP=18 and only the sign bit of motion vectors encrypted by the proposed method.

4. PERFORMANCE ANALYSIS

In this paper, a stream cipher Rabbit [13], developed as part of the ECRYPT Stream Cipher Project, is adopted to generate the random bit sequence, because it is suitable for software implementation. To date, there have been no effective attacks against this cipher [14]. For convenience, the previous sign-bit encryption of motion vectors [10] and the proposed improved method are denoted as EMV and IEMV, respectively; their combinations with IPM [6] are denoted as IPM+EMV and IPM+IEMV, respectively. Based on the Joint Model (JM) reference software of H.264/AVC, version 17.2 [15], these four schemes were implemented under the baseline and main profiles with CAVLC (Context Adaptive Variable Length Coding) as the entropy coding method. Three standard test videos in QCIF resolution were chosen to evaluate the performance.

4.1. The Perceptual Scrambling Effect

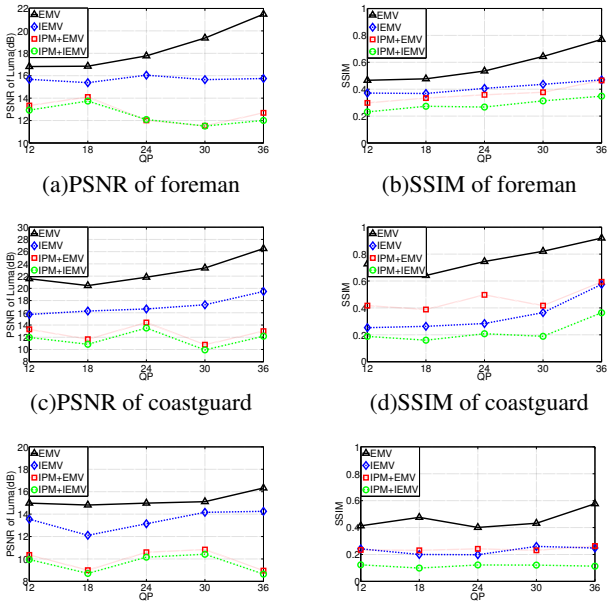
The perceptual quality of a video is a subjective metric, and it is very difficult to practically implement this metric in a subjective way. Peak signal to noise ratio (PSNR) is most widely utilized to give an objective approximation of the perceptual quality. However, it is often criticized for its bad performance. Structural Similarity (SSIM) [16] is a more recently proposed objective metric to measure the video quality. It has been reported that SSIM can perform much closer to the subjective

observation than PSNR [17]. In this section, both PSNR and SSIM are used to measure the perceptual scrambling effect.

4.1.1. *Under the baseline profile*

In the baseline profile, the IPP...P coding sequence is used. For each test video sequence, the first 30 frames under different QPs are encoded and encrypted, and then are decoded without decryption. The perceptual quality of the decoded video without decryption is measured by the average PSNR and SSIM of the first 30 frames. As shown in Fig. 5(a), 5(c) and 5(e), in terms of PSNR, IEMV and IPM+IEMV can much more effectively degrade the perceptual quality than EMV and IPM+EMV, respectively. The results using the SSIM metric, as shown in Fig. 5(b), 5(d) and 5(f), provide the same conclusion.

It is also observed that no matter which metric is used, IPM+IEMV is the best of the four schemes in degrading the perceptual quality. The first few frames of 'foreman' encoded with QP=18 and encrypted using IPM+IEMV are shown in Fig. 6. Compared with Fig. 4, it is shown that IPM+IEMV can effectively scramble the whole frame including the background.



(a)PSNR of foreman (b)SSIM of foreman

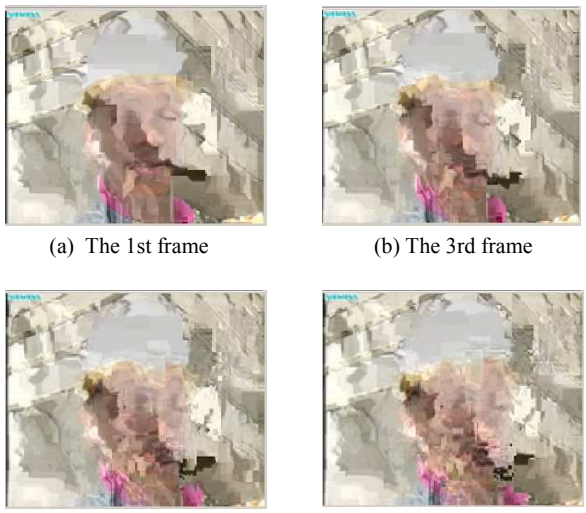
(c)PSNR of coastguard (d)SSIM of coastguard

(e)PSNR of mobile (f)SSIM of mobile

Fig. 5. The perceptual scrambling effect of EMV, IEMV, IPM+EMV and IPM+IEMV, under the baseline profile.

4.1.2. *Under the main profile*

In the main profile, the IBPBP...BP coding sequence is adopted and CAVLC is chosen as the entropy coding method. Again, both PSNR and SSIM are used to measure the scrambling effect of the four schemes. The corresponding results are shown in Fig. 7. It is clear that the same conclusion can be reached as in the baseline profile: IEMV and IPM+IEMV

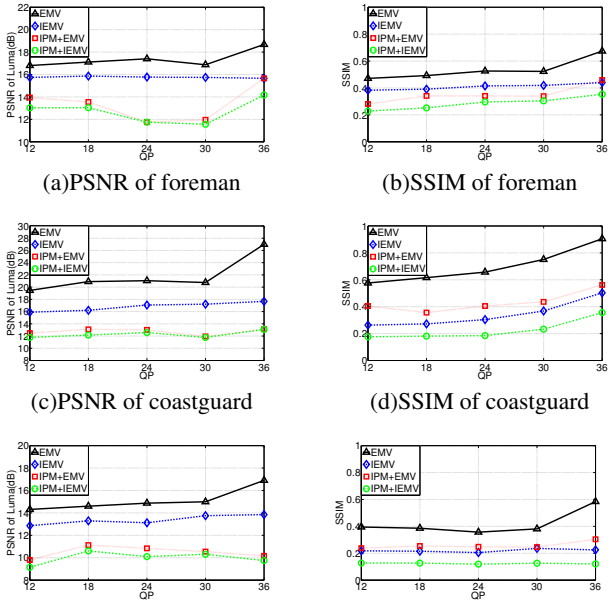


(a) The 1st frame (b) The 3rd frame

(c) The 5th frame (d) The 7th frame

Fig. 6. The first few decoded frames of 'foreman' encoded in the baseline profile, with QP=18 and encrypted by IPM+IEMV.

can much more effectively scramble the perceptual quality than EMV and IPM+EMV, respectively, and IPM+IEMV is the best choice to degrade the perceptual quality.



(a)PSNR of foreman (b)SSIM of foreman

(c)PSNR of coastguard (d)SSIM of coastguard

(e)PSNR of mobile (f)SSIM of mobile

Fig. 7. The perceptual scrambling effect by EMV, IEMV, IPM+EMV and IPM+IEMV, under the main profile.

In addition, in some cases, IEMV appears to perform much better than IPM+EMV in terms of SSIM for some video sequences. However, from practical observation, IPM+EMV can much more effectively scramble the video than IEMV, since IPM+EMV encrypts much more information than EMV.

4.2. Compression Ratio

Encrypting the sign bit of motion vectors or intra prediction modes does not affect the bit length of related syntax elements in the original video stream and therefore, these four schemes do not change the compression ratio. From practical experiments for the three video sequences under different QPs, in the baseline and main profiles, it is observed that the compression ratio is kept unchanged when encrypted by any one of these four schemes.

4.3. Security

The security of EMV, IEMV, IPM+EMV and IPM+IEMV relies on the security of the chosen stream cipher for the random bit generator. For this reason, Rabbit [13], which to date has no known weaknesses, has been adopted in this work.

5. CONCLUSION

In this paper, an improved encryption method for encrypting the sign bit of motion vectors, IEMV, is proposed for H.264/AVC video encoding. This method improves upon the previous sign bit encryption of motion vectors, EMV, by ensuring the four candidates for the encrypted motion vectors are always located in two orthogonal lines. Experiments under the baseline and main profiles show that IEMV can achieve a better scrambling effect than EMV, while keeping the compression ratio. Since encryption of only the sign bit of motion vectors will leave the first few frames clear or in a good perceptual state. It is suggested that the proposed method should be combined with IPM [6] to much more effectively degrade the perceptual quality. Experimental results under the baseline and main profiles support that this combination works well and that it does not affect the compression ratio.

6. REFERENCES

[1] T. Lookabaugh and D. C. Sicker, "Selective Encryption for Consumer Applications," *IEEE Commun. Mag.*, vol.42, no.5, pp.124-129, 2004.

[2] T. Stutz and A. Uhl, "Survey of H.264 AVC/SVC encryption," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 325–339, 2011.

[3] ITU-T Rec, H.264 ISO/IEC 14496-10, "Advanced Video Coding for Generic Audio-visual Service," Mar. 2010.

[4] I. Richardson, *The H.264 Advanced Video Compression Standard*, 2nd ed. John Wiley & Sons, 2010.

[5] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of The H. 264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, 2003.

[6] J. Ahn, H. Shim, B. Jeon and I. Choi, "Digital Video Scrambling Method Using Intra Prediction Mode," *PCM2004, Springer, LNCS*, vol. 3333, pp. 386-393.

[7] Y. Li, L. Liang, Z. Su and J. Jiang, "A New Video Encryption Algorithm for H.264," in *Proc. 5th IEEE Int. Conf. Info., Commun. & Signal Process., 2005*, pp. 1121-1124.

[8] S. Lian, Z. Liu, Z. Ren, and Z. Wang, "Selective Video Encryption Based on Advanced Video Coding," *PCM 2005, Springer, LNCS*, vol. 3768, pp. 281–290.

[9] S. Lian, Z. Liu, Z. Ren and H. Wang, "Secure Advanced Video Coding Based on Selective Encryption Algorithms," *IEEE Trans. Consumer Electron.*, vol. 52, no. 2, pp. 621-629, 2006.

[10] C. Shi and B. Bhargava, "An Efficient MPEG Video Encryption Algorithm," in *Proc. 17th IEEE Symp. Reliable Distributed Systems, 1998*, pp. 381–386.

[11] Z. Liu and X. Li, "Motion Vector Encryption in Multimedia Streaming," in *Proc. 10th IEEE Int. Conf. Multimedia Modelling, 2004*, pp. 64–71.

[12] S. Kwon, W. Choi and B. Jeon, "Digital Video Scrambling Using Motion Vector and Slice Relocation," in *Proc. 2nd Int. Conf. Image Analysis & Recognition, Springer, LNCS, 2005*, vol. 3656, pp. 207-214.

[13] M. Boesgaard, M. Vesterager, T. Christensen and E. Zenner, "The Stream Cipher Rabbit," available via `http://www.ecrypt.eu.org/stream/p3ciphers/rabbit/rabbit_p3.pdf`, 2011.

[14] S. Babbage, C. Canniere, A. Canteaut, C. Cid, H. Gilbert, T. Johansson, M. Parker, B. Preneel, V. Rijmen, and M. Robshaw, "The eSTREAM Portfolio," available via `http://www.ecrypt.eu.org/stream/portfolio_revision1.pdf`, 2011.

[15] JM reference software, ver. 17.2, `http://iphome.hhi.de/suehring/tml`, Apr. 2011.

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

[17] Z. Wang and A.C. Bovik, "Mean Squared Error: Love It or Leave It? A New Look at Signal Fidelity Measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, 2009.