A Perceptual Encryption Scheme for Motion JPEG 2000 Standard

Osamu WATANABE†and Takahiro FUKUHARA
Takushoku University
Dept. of Electronics & Computer Systems
†Email: owatanab@es.takushoku-u.ac.jp

Hitoshi KIYA
Tokyo Metropolitan University
Faculty of Info. and Commun. Systems
Email: kiya@sd.tmu.ac.jp

***Abstract*—A new perceptual encryption scheme with an efficient key-management mechanism for the Motion JPEG 2000 based ETC system is proposed in this paper. An ETC system is known as a system that makes image/video communication secure and efficient by using perceptual encryption and image/video compression. Unlike conventional ETC systems, the perceptually encrypted images by the proposed scheme can be efficiently compressed by the Motion JPEG 2000. Moreover, an efficient mechanism to manage a lot of secret keys for video sequences is provided to avoid complex key-management. The experimental results demonstrated that the proposed scheme achieved both acceptable compression performance and enough security for secure image/video communication while remaining compatible with the Motion JPEG 2000 standard.**

I. Introduction

Many methods of protecting compressed multimedia content have been reported with the wide/rapid spread of distributed systems for information processing, such as cloud computing and social networks. Communication and processing for multimedia content is performed over the Internet in these systems. In other words, the content is transmitted/received via insecure telecommunication channels with restricted bandwidth. Therefore, both encryption and compression are necessary to make the communication secure and efficient. In the meantime, the JPEG committee has started to standardize a new work item, which is referred to as *JPEG Privacy & Security* [1], [2]. Secure transmission between network servers that are used in cloud computing and social networks is supposed to be one of the technical requirements of JPEG Privacy & Security.

Image encryption has to be performed anterior to image compression in certain practical scenarios for such distributed systems, e.g., image communication with security/privacy considerations. This framework is known as the *Encryption-then-Compression (ETC)* system [3], [4]. ETC systems have several advantages over traditional solutions called *Compression-then-Encryption (CTE)* systems: one is that image compression can be done more efficiently, another is that it is not necessary for an owner of images to disclose them to network providers. In most cases, such ETC systems adopt a perceptual encryption scheme that is known as an operation for making it difficult to understand images visually and is performed in both spatial and frequency-transformed domains [5]–[9]. This is although most studies on ETC systems have assumed the use of their own compression schemes that had no compatibility with international standards such as JPEG or JPEG 2000 [10]–[14].

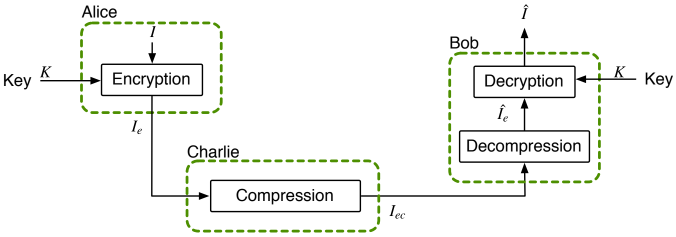


Fig. 1: Image communication with Encryption-then-Compression (ETC) system

The use of an international standard for image compression is one of the essential factors in its wide acceptance considering practical application scenarios for ETC systems.

The Motion JPEG 2000 [15] is well known as the international standard for video compression. The Motion JPEG 2000 adopts intra-frame coding. That is, each frame is encoded as a still image and is encoded into a JPEG 2000 [16] codestream. A new ETC system with JPEG 2000 compliant perceptual encryption schemes have been proposed [17]; however, there is a problem on key-management when we apply this system to the Motion JPEG 2000 based video system because a secret key to encrypt a current frame should be different from the keys that has been used to encrypt previous frames.

In this paper, a new scheme of perceptual encryption for the Motion JPEG 2000 based ETC system is proposed. The proposed scheme is designed for the Motion JPEG 2000 based video system and provides a new key-management mechanism to solve the problem described above.

II. Secure Image Communication Systems with Image Compression

Let us suppose a scenario in which content owner Alice wants to securely and efficiently transmit image I to recipient Bob over insecure and band-restricted communication channel provider Charlie. The concept of ETC systems is described in what follows.

A. Image communication with ETC system

Image communication with an ETC system is outlined in Figure 1. Alice encrypts I into I_e then I_e is transmitted to Bob over the channel provided by Charlie. Charlie compresses I_e into I_{ec} according to the band-restrictions of the channel. The received I_{ec} is decompressed and then decrypted. Finally, Bob gets reconstructed image $\hat{I}$. A method of perceptual encryption

978-1-4673-6820-9/15/$31.00 ©2015 IEEE

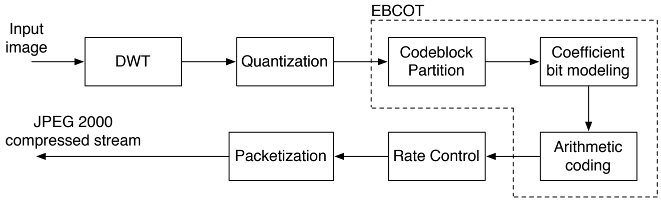


Fig. 2: Block diagram of JPEG 2000 encoder

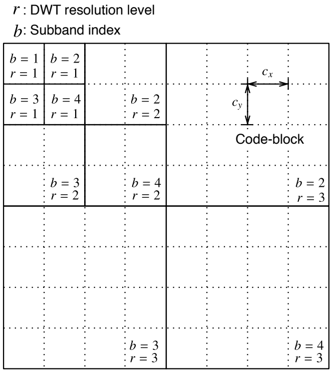


Fig. 3: Definition of code-blocks, subbands, and resolution levels on DWT coefficients (decomposition level $R = 3$)

in an ETC system is used as encryption function $E(\cdot)_k$ instead of an AES-like cipher. Perceptual encryption makes image I difficult to understand visually so that it is possible to decrypt I_{ec} with communication error.

An ETC system using perceptual encryption has three main advantages over a CTE system:

- Alice does not need to disclose image I to Charlie.
- Decryption of I_{ec} with communication error is possible.
- It is possible for Charlie to control the coding rate to maximize network utilization.

B. JPEG 2000 coding

In the Motion JPEG 2000, each video frame is encoded as a still image; a frame is encoded into a JPEG 2000 codestream. Figure 2 is a block diagram of a JPEG 2000 encoder and Figure 3 is an example of an image analyzed with DWT. The R denotes three DWT resolution levels in Fig. 3, and $r = 1, 2, ..., R$ is the index of the resolution level. The b denotes the indices of subbands, where 1, 2, 3, and 4 correspond to LL, HL, LH, and HH. The subbands are divided into $c_x \times c_y$ -sized code-blocks. The default value for c_x and c_y is 64. DWT coefficients in the code-blocks are quantized and then are encoded by embedded block coding with optimized truncation (EBCOT). Then, bit-modeling and arithmetic coding is performed followed by rate control operations. Rate control is used to make the bitstream conform to a target size. Finally, a JPEG 2000 compliant bitstream is generated by adding packet headers, a main header, and other control codes.

III. Proposed perceptual encryption scheme for Motion JPEG 2000

Here, a new perceptual encryption scheme for the Motion JPEG 2000 is proposed, which includes three types of perceptual encryption schemes and key-management mechanism for video sequences. First, the overall procedure for the ETC system with the proposed perceptual encryption schemes is described, and then the perceptual encryption schemes are explained. Finally, a new key-management mechanism for video sequences is described.

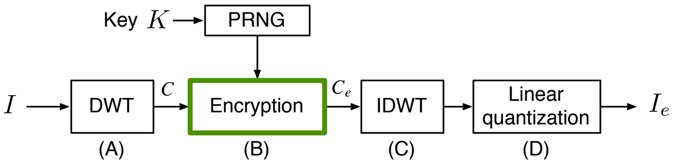
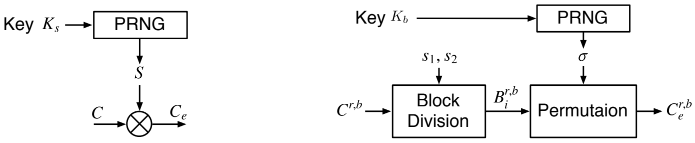


Fig. 4: Procedure of generating a perceptual encrypted image



(a) Sign-scrambling

(b) Block-shuffling

Fig. 5: Procedures of perceptual encryption schemes

A. Procedure for generating perceptual encrypted frames

The procedure to generate the perceptual encrypted frames is outlined in Fig. 4 and is summarized in four steps (A-D). Note that input frame I is supposed to be $X \times Y$ in size.

(A) The JPEG 2000 compliant DWT is applied to I. Then, the DWT coefficients, C, which are $X \times Y$ in size, are obtained.

(B) DWT coefficients C are perceptually encrypted into C_e with two types of encryption schemes based on a pseudo-random number generator (PRNG) with a secret key. Detailed explanations of each of the schemes are described in the next subsection and the key-management mechanism for video sequences is described in III-C.

(C) Encrypted DWT coefficients C_e are transformed to a spatial image by using inverse DWT.

(D) The spatial image is linear-quantized to maintain the signal range of the input image, and then perceptually encrypted frame I_e is obtained.

B. Methods of perceptual encryption

1) Sign-scrambling of DWT coefficients: The sign-scrambling of DWT coefficients is described and the procedure for this is outlined in Fig. 5(a) as a perceptual encryption scheme for the proposed ETC system. Let K_s denote a secret key for encryption and S denote a pseudo-random matrix. The S is generated by using K_s as its seed and $S(i, j)$ with $(1 \le i \le X, 1 \le j \le Y)$, which means an element of S, consists of "1" or "-1". Equation (1) has an example of S as:

$$S = \begin{pmatrix} -1 & 1 & \dots & 1 \\ 1 & -1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & -1 & \dots & 1 \end{pmatrix}. \qquad (1)$$

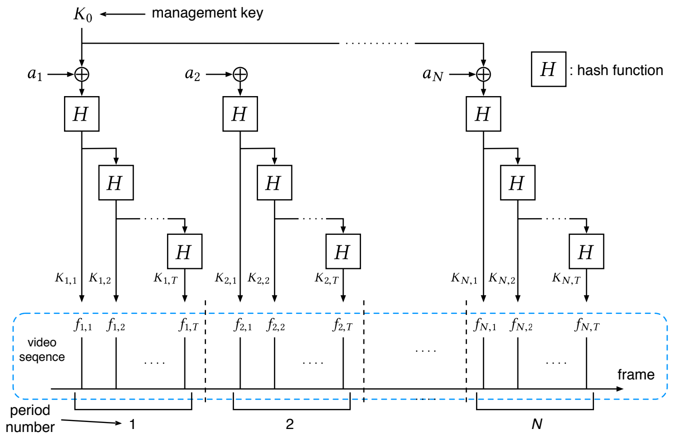


Fig. 6: Generation and assignment of secret keys based on multidimensional hash chain

The DWT coefficients $C(i,j)$ are scrambled by:

$$C_e(i,j) = C(i,j)S(i,j). \tag{2}$$

Note that the ratio of "-1" to "1" can be controlled. If Alice wants to get a fully scrambled image, the ratio should be 1:1.

2) Block-shuffling of DWT coefficients: The block-shuffling of DWT coefficients is described as another perceptual encryption scheme. The procedure for this is given in Fig. 5(b). First, DWT coefficients C are divided into blocks that are $s_1 \times s_2$ in size. Suppose i is the index of the blocks in subband and a $B_i^{r,b}$ is the i-th block in subband b of DWT resolution level r. The permutation, σ, for the block-shuffling is defined as a two-line notation:

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & i & \cdots & n \\ x_1 & x_2 & \cdots & x_i & \cdots & x_n \end{pmatrix}, \tag{3}$$

where the first row represents the elements of i and the second row, which is $(x_1, x_2, \ldots, x_n)$, is a sequence generated by sequentially taking a pseudo-random number between 1 and n, ensuring that there are no repetitions. A block, $B_i^{r,b}$, is set to the position having its index equal to x_i in the block-shuffled subband. Note that a secret key, K_b, is used as a seed for generating the pseudo-random numbers.

3) Combination of sign-scrambling and block-shuffling: Combined scrambling is easily achieved by sign-scrambling followed by the block-shuffling. Let $\hat{C}_e^{r,b}$ denote the output of sign-scrambling in subband b of DWT resolution level r. The $\hat{C}_e^{r,b}$ is input into the following the block-shuffling. Then, the block-shuffled sign-scrambling output, $C_e^{r,b}$, is obtained.

C. Key-management mechanism for encryption of video sequences

Let us consider a scenario in which the proposed perceptual encryption schemes are applied to video sequences for the Motion JPEG 2000 based systems. When a single key is used for all frames, there is no need for the key management mechanism because the number of keys is no more than one. However, the difficulty to estimate the secret key decreases because each encrypted frame mutually has some correlation. To make the ETC system based on the Motion JPEG 2000 more secure, each frame should be encrypted by an individual secret key because the estimation of the keys is more difficult. However, the use of a lot of secret keys results in complicate key management.

TABLE I: FSIM index of encrypted image I_e

Frame #	Sign	BS 2x2	BS 8x8	BS 10x10	BS (LL)	Mixed
2254	0.449	0.451	0.489	0.507	0.455	0.428
3387	0.564	0.511	0.504	0.520	0.509	0.493
6482	0.569	0.504	0.530	0.525	0.512	0.485
8132	0.373	0.390	0.496	0.518	0.383	0.347
9924	0.529	0.518	0.569	0.571	0.515	0.467
Average	0.497	0.475	0.517	0.528	0.475	0.444

To solve such problems, a key management mechanism using the multidimensional hash chain is proposed. Figure 6 illustrates the generation and assignment of the encryption keys. In the proposed mechanism, a video sequences are divided into N periods in which each period has T frames. With a hash function $H(\cdot)$ and a seed key K_0, the secret key which is used for encrypting the n-th frame is generated as

$$K_{n,1} = H(K_0 + a_n) \qquad (n = 1, 2, \cdots, N) \tag{4}$$
$$K_{n,t+1} = H(K_{n,t}) \qquad (t = 1, 2, \cdots, T) \tag{5}$$

where a_1, a_2,$\cdots$, and a_N, which are all different constants and not secret, are used to change the output values of hash functions. A frame $f_{n,t}$, which is the t-th frame of the n-th period, is encrypted by a secret key $K_{n,t}$. The constants a_1,a_2,$\cdots$, and a_N can be released if K_0 is secret. Therefore, we need to manage only K_0. In addition, in this mechanism, it is possible to decrypt the video sequences partially (from the middle, or only necessary part) by encrypting the divided video sequences. In this mechanism, the number of managed keys can be increased if we want to enhance the safety without any security risk.

IV. Experimental results

This section presents the experimental results obtained from evaluating the performance of the proposed perceptual encryption schemes.

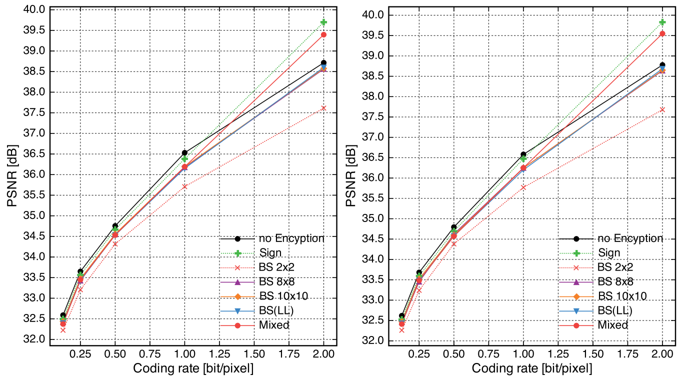
A. Conditions for experiments

Kakadu version 6.3 [18] was used as a codec for the Motion JPEG 2000 standard. The test frames were taken from standard evaluation material (StEM) test data [19]. Each of the test frames consisted of RGB color space, and had a resolution of 4096×1714. The bit-depth of single color components was 12 bits/pixel. Five different frames (#2254, #3387, #6482, #8132, and #9924) were used. Four cases for the proposed encryption schemes were investigated in the experiments.

Sign :	Sign-scrambling with "1"-to-"-1" ratio of 1:1.
BS $n \times n$:	Block-shuffling with $s_1 = s_2 = n$. Two, eight, and ten were used as values for n.
BS (LL):	Block-shuffling with $s_1 = s_2 = 2$ was done for lowest subband LL. Block-shuffling with $s_1 = s_2 = 8$ was done for the others.
Mixed:	Combination of **Sign** and **BS** 8×8.

B. Results and remarks

1) Efficiency of encryption: Perceptual encryption makes an image difficult to understand visually. The feature similarity



(a) Codeblock size 32x32 (b) Codeblock size 64x64

Fig. 7: Comparison of the coding rate-PSNR performance: Each of curves is an average of PSNR values for all test images. There is no significant difference between *no-Encryption* and the others in terms of the value of PSNR without $BS2 \times 2$ case.

(FSIM) index [20] between I and I_e was calculated for all these cases to confirm this property for perceptual encryption we propose. The FSIM indices ranged from zero to one. If a processed image had an FSIM index of less than 0.5, it was very hard to identify the processed image to be the same as the original because an FSIM index of less than 0.5 provided Zhang et al. a very small subjective MOS score that was close to zero [20]. The calculated FSIM indices are summarized in Table I This tableconfirmed that perceptual encryption we propose is sufficiently capable of making an image unrecognizable.

2) Motion JPEG 2000 compression performance: Figure 7 plots the results for compression by using the proposed perceptual encryption schemes. Two different code-block sizes, that were 32×32 and 64×64, were used and the results obtained from a PSNR evaluation of the average test frames between reconstructed frame $\hat{I}$ and original frame I are given in Figs. 7(a) and 7(b). The BS 2×2 performed the worst in Figs 7(a) and 7(b); however, better quality in reconstructed frame $\hat{I}$ could be obtained by using larger sizes for s_1 and s_2. We also tested and confirmed that there were no significant differences between *Sign* and the others in terms of the value of PSNR. Using *Mixed* was a good choice because the combination of sign-scrambling and block-shuffling was more secure against brute-force attacks. It was also confirmed that the scalability function of the Motion JPEG 2000 standard, which is known to be an additional feature of the standard, was retained when using the proposed perceptual encryption schemes.

V. Conclusions

This paper proposed a new perceptual encryption schemes with an efficient key-management mechanism for the Motion JPEG 2000 standard. By using multidimensional hash chain, the number of management keys can be increased without any security risk. The experimental results demonstrated that the proposed perceptual encryption schemes achieved both acceptable compression and sufficient security for secure image communication while maintaining the compatibility with the Motion JPEG 2000 standard.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 25730073.

References

[1] "Draft requirements for JPEG Privacy & Security," ISO/IEC JTC 1/SC 29/WG 1 N69030, Jun. 2015.

[2] P. Schelkens, "Image security tools for JPEG standards," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security - IH&MMSec '14*. ACM Press, Jun. 2014, pp. 1–1.

[3] Z. Erkin, A. Piva, S. Katzenbeisser, R. L. Lagendijk, J. Shokrollahi, G. Neven, and M. Barni, "Protection and retrieval of encrypted multimedia content: When cryptography meets signal processing," *EURASIP Journal on Information Security*, vol. 2007, no. 3, 2007.

[4] N. Kalyani G. and S. Milind V., "Article: A survey based on designing an efficient image encryption-then-compression system," *IJCA Proceedings on National Level Technical Conference X-PLORE 2014*, vol. XPLORE2014, pp. 6–8, May 2014, full text available.

[5] I. Ito and H. Kiya, "A new class of image registration for guaranteeing secure data management," in *Proc. ICIP 2008*. IEEE, Oct. 2008, pp. 269–272.

[6] H. Kiya and I. Ito, "Image matching between scrambled images for secure data management," in *Proc. 16th European Signal Process. Conf*, 2008.

[7] I. Ito and H. Kiya, "One-time key based phase scrambling for phase-only correlation between visually protected images," *EURASIP Journal on Information Security*, vol. 2009, p. 3, 2009.

[8] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 118–129, Mar. 2003.

[9] Z. Tang, X. Zhang, and W. Lan, "Efficient image encryption with block shuffling and chaotic map," *Multimedia Tools and Applications*, pp. 1–20, 2014.

[10] M. Johnson, P. Ishwar, V. Prabhakaran, D. Schonberg, and K. Ramchandran, "On compressing encrypted data," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2992–3006, 2004.

[11] W. Liu, W. Zeng, L. Dong, and Q. Yao, "Efficient compression of encrypted grayscale images," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1097–1102, 2010.

[12] X. Zhang, "Lossy compression and iterative reconstruction for encrypted image," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 53–58, 2011.

[13] R. Hu, X. Li, and B. Yang, "A new lossy compression scheme for encrypted gray-scale images," in *Proc. ICASSP 2014*. IEEE, 2014, pp. 7387–7390.

[14] J. Zhou, X. Liu, O. C. Au, and Y. Y. Tang, "Designing an Efficient Image Encryption-Then-Compression System via Prediction Error Clustering and Random Permutation," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 1, pp. 39–50, Jan. 2014.

[15] "Information technology — JPEG 2000 image coding system – Part 3: Motion jpeg 2000," International Standard ISO/IEC IS-15444-3:2007, May 2007.

[16] "Information technology — JPEG 2000 image coding system – Part 1: Core coding system," International Standard ISO/IEC IS-15444-1, Dec. 2000.

[17] O. Watanabe, A. Uchida, T. Fukuhara, and H. Kiya, "An encryption-then-compression system for JPEG 2000 standard," in *Proc. ICASSP 2015*. IEEE, Apr. 2015, pp. 1226–1230.

[18] kakadu software. [Online]. Available: http://www.kakadusoftware.com

[19] Digital Cinema Initiatives, LLC Technology Committee. (2010, Sep.) StEM Access Procedures. [Online]. Available: http://www.dcimovies.com/StEM/

[20] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment." *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–86, Aug. 2011.