Cryptanalysis of an Image Scrambling Scheme without Bandwidth Expansion

Shujun Li, Chengqing Li, Kwok-Tung Lo, *Member, IEEE* and Guanrong Chen, *Fellow, IEEE*

***Abstract*— Recently, a new image scrambling (i.e., encryption) scheme without bandwidth expansion was proposed based on two-dimensional (2-D) discrete prolate spheroidal sequences (DPSS). A comprehensive cryptanalysis is given here on this image scrambling scheme, showing that it is not sufficiently secure against various cryptographical attacks including ciphertext-only attack, known/chosen-plaintext attack and chosen-ciphertext attack. Detailed cryptanalytic results suggest that the image scrambling scheme can only be used to realize perceptual encryption, but not to provide content protection for digital images.**

***Index Terms*— discrete prolate spheroidal sequence (DPSS), image scrambling, encryption, cryptanalysis, ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, chosen-ciphertext attack, Hadamard matrix, perceptual encryption.**

I. Introduction

Content protection of multimedia data (especially digital images and videos) through encryption has attracted more and more attention due to the rapid development of multimedia and network technologies in past two decades. Various image/video encryption (or scrambling[1]) schemes have been proposed, but some of which have been cryptanalyzed to be insecure. To offer some reasonable background knowledge on the content of this paper, in the following a very brief introduction to some existing image/video encryption schemes will be given. For a more comprehensive survey of the state-of-the-art of this topic, readers are referred to [1]–[6].

The most straightforward idea in image/video encryption is to consider the entire 2-D multimedia data as a 1-D textual bit-stream and then apply any conventional cipher that has been validated in modern cryptography, such as DES, IDEA [7], and AES [8]. This solution is sometime called naive encryption [9]. The major problem of naive encryption lies in the following two aspects: 1) the encryption speed may be too slow; 2) it does not consider the information redundancy existing in the uncompressed images/videos and the syntax structures of the compressed ones. A possible way to overcome these problems is to encrypt part of the given plain-image/video, which is called selective (or partial) encryption. For example, for MPEG videos, only sign bits of the DCT coefficients and the motion vectors can be selected for encryption. Even though partial encryption may not provide a high level of security, it is still useful for realizing perceptual encryption [2], [10] and format-compliant encryption [11], two interesting security requirements of some multimedia applications.

As a frequently-used approach to encrypting images and videos, some schemes were designed by secretly permuting pixels in the plain-image or each frame of the plain-video [12]–[14]. This idea can also be generalized to the transform domain, while in this case encryption is achieved by permuting transform coefficients (and/or nodes for some transforms with a tree-like structure)[2] [17]–[20]. However, a large number of cryptanalysis reports have shown that these permutation-based image/video encryption schemes are not sufficiently secure from a cryptographical point of view [9], [21]–[27]. The main security problems include: 1) the plain-image/video may be partially recovered due to the large information redundancy existing in natural images/videos (under ciphertext-only attack); 2) secret permutations are always insecure against known/chosen-plaintext attack. As a general result of cryptanalysis, secret permutations must be combined with other techniques in order to design a secure image/video encryption scheme.

Another idea in designing image/video encryption schemes is to scramble all the pixels and/or transform coefficients with some multiplicative or additive matrices. This idea can be considered as a generalization of the permutation-only encryption, since secret permutations can be formulated with a permutation matrix as shown in the theoretical models of some permutation-based speech scrambling schemes [15], [16]. Many optical image encryption methods have been developed in this way, by introducing double random phase matrices

Shujun Li was partially supported by the Alexander von Humboldt Foundation of Germany and by The Hong Kong Polytechnic University's Postdoctoral Fellowships Program (under grant no. G-YX63). The work of K.-T. Lo was supported by the Research Grants Council of the Hong Kong SAR Government under Project Number 523206 (PolyU 5232/06E). The authors would like to thank Dr. Van De Ville with the Swiss Federal Institute of Technology Lausanne for sending the Matlab reference codes of the image scrambling scheme to us.

Shujun Li is with the FernUniversität in Hagen, Lehrgebiet Informationstechnik, Universitätsstraße 27, 58084 Hagen, Germany.

Chengqing Li and Guanrong Chen are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon Toon, Hong Kong SAR, P. R. China.

Kwok-Tung Lo is with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, P. R. China.

The corresponding author is Shujun Li. Contact him via his personal web site: http://www.hooklee.com.

[1] The term "scrambling" was used instead of "encryption" by some designers of multimedia encryption schemes, especially by those early designers who intended to encrypt analog signals by "scrambling" them in some way. From a cryptographical point of view, we simply consider "scrambling" as a synonym of "encryption".

[2] Although speech encryption is not the focus of this paper, it is worth mentioning that many speech scrambling schemes were also developed based on this idea working in the transform domain [15], [16].

(keys) [6], [28], [29], which are used to scramble the plain-image in spatial and frequency (Fourier transform) domains, respectively. However, some recent cryptanalysis work [30]–[32] shows that optical image encryption schemes of this kind is not sufficiently secure against known-plaintext attack and chosen-ciphertext attack.

On the other hand, a large number of image encryption schemes were designed by combining different encryption techniques. For example, some image encryption schemes are developed based on the multi-round combination of secret permutations and pixel-value substitutions [33]–[35]. There are also a number of attempts of using chaos to design image/video encryption schemes [2], yet some of which have been cryptanalyzed to be insecure [36]–[42].

This paper focuses on a new image scrambling scheme proposed in [43], which is a 2-D extension of a speech scrambling scheme proposed earlier in [44]. Compared with other existing image scrambling methods, this scheme does not introduce many high-frequency components into the spectrum of the ciphertext, but causes only a negligible expansion of bandwidth. This feature is useful in some real applications, as the cipher-image can be transmitted with a band-limited channel that carries the plain-image. The encryption process is mainly achieved by scrambling low-frequency components of a 2-D DPSS (discrete prolate spheroidal sequences) transform by a multiplicative matrix, which serves as the secret key of the scheme. To further enhance the security, random swapping of some high-frequency components and multiple secret matrices were also suggested, each of which corresponds to the encryption of one single block of the plain-image.

In this paper, we report a thorough investigation on the security of the image scrambling scheme proposed in [43], and point out that it is not sufficiently secure against various cryptographical attacks including ciphertext-only attack, known/chosen-plaintext attack and chosen-ciphertext attack. We also found some other security defects of the scheme when a fixed secret matrix is used to encrypt all blocks of the plain-image. Based on the cryptanalytic results, we conclude that the image scrambling scheme should only be used to realize perceptual encryption, i.e., for degrading the visual quality of the plain-images in a secret manner.

The rest of this paper is organized as follows. First, a description of the image scrambling scheme is given in the next section. Then, Section III focuses on the cryptanalytic findings, with both theoretical and experimental results demonstrated. In Sec. IV, the question of how to use the image scrambling scheme in practice is addressed. Finally, the last section concludes the paper.

II. The Image Scrambling Scheme without Bandwidth Expansion

The image scrambling scheme proposed in [43] is a 2-D extension of Wyner's signal scrambling scheme [44] based on discrete prolate spheroidal sequences (DPSS), which are defined as the normalized eigenvectors of the following real and symmetric matrix:

$$\mathbf{V} = \left[\frac{\sin(2\pi W(m-n))}{\pi(m-n)}\right]_{0\leq m,n\leq N-1}. \tag{1}$$

Denote the DPSS, i.e., the N eigenvectors of $\mathbf{V}$, by $\{\phi_j\}_{j=0}^{N-1}$, where $\phi_j = \begin{bmatrix}\phi_j(0) & \cdots & \phi_j(N-1)\end{bmatrix}^T$, and the corresponding eigenvalues by $\{\lambda_j\}_{j=0}^{N-1}$. It was showed [45] that $\{\phi_j\}_{j=0}^{N-1}$ form an orthonormal basis that spans the subspace of sequences with an energy concentration in a certain band $[-W, W]$. Thus, for any sequence $\boldsymbol{a} = \begin{bmatrix}a(0) & \cdots & a(N-1)\end{bmatrix}^T$, one can use the DPSS to get another sequence, $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_0 & \cdots & \alpha_{N-1}\end{bmatrix}^T$, such that $\boldsymbol{\alpha} = \mathbf{S}\boldsymbol{a}$ (or $\boldsymbol{a} = \mathbf{S}^T\boldsymbol{\alpha}$), where $\mathbf{S} = \begin{bmatrix}\phi_0 & \cdots & \phi_{N-1}\end{bmatrix}^T$. Based on such a DPSS transformation, one can scramble $\boldsymbol{\alpha}$ and then perform an inverse transform as an alternative way of encrypting the original sequence $\boldsymbol{a}$. The encryption process can be described as $\boldsymbol{a}' = \mathbf{S}^T\mathbf{M}\boldsymbol{\alpha} = \mathbf{S}^T\mathbf{M}\mathbf{S}\boldsymbol{a}$, where $\mathbf{M}$ is the secret matrix that scrambles $\boldsymbol{\alpha}$. For such a scrambling scheme, Wyner [44] showed that the bandwidth expansion will be negligible if the smallest eigenvalue corresponding to scrambled coefficients in $\boldsymbol{\alpha}$ is sufficiently large. More precisely, assuming that all coefficients in $\boldsymbol{\alpha}$ are ranked by the eigenvalues in descending order and only the v lowest coefficients[3] $\{\alpha_j\}_{j=0}^{v-1}$ are scrambled, Wyner deduced that the energy concentration of the scrambled sequence $\boldsymbol{a}'$ differs at most by $1-\lambda_{v-1}$ as compared to the concentration of the original sequence $\boldsymbol{a}$. In this case, the secret matrix $\mathbf{M}$ is in the following form:

$$\mathbf{M} = \begin{bmatrix}\mathbf{M}_v & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{N-v}\end{bmatrix}, \tag{2}$$

where $\mathbf{M}_v$ is the sub-matrix scrambling the v lowest coefficients and $\mathbf{I}_{N-v}$ is the $(N-v)\times(N-v)$ identity matrix. One can see that this scrambling scheme is a selective encryption algorithm, since some coefficients in $\boldsymbol{\alpha}$ are left unchanged.

In [43], the 1-D DPSS was extended to the case of 2-D square passband region as follows:

$$\phi_{j_1,j_2}^{(2D)}(n_1,n_2) = \phi_{j_1}^{(1D)}(n_1)\phi_{j_2}^{(1D)}(n_2), \tag{3}$$

where $0 \leq n_1, j_1 \leq N_1 - 1$ and $0 \leq n_2, j_2 \leq N_2 - 1$. Accordingly, the eigenvalues corresponding to $\phi_{j_1,j_2}^{(2D)}$ are $\lambda_{j_1,j_2}^{(2D)} = \lambda_{j_1}^{(1D)}\lambda_{j_2}^{(1D)}$. Then, by scanning all elements in each $\phi_{j_1,j_2}^{(2D)}$ to form an $N_1N_2\times 1$ vector $\phi_j^{(2D)}$ $(0 \leq j \leq N_1N_2-1)$ and sorting these eigenvectors such that $\lambda_0^{(2D)} \geq \cdots \geq \lambda_{N_1N_2-1}^{(2D)}$, one gets an $N_1N_2 \times N_1N_2$ matrix $\mathbf{S} = \begin{bmatrix}\phi_0^{(2D)} & \cdots & \phi_{N_1N_2-1}^{(2D)}\end{bmatrix}^T$. Now, given an $N_1N_2 \times N_1N_2$ secret matrix $\mathbf{M}$ and an $N_1N_2\times 1$ vector $\boldsymbol{a}$, the 2-D scrambling scheme has the same encryption formula as the 1-D one:

$$\boldsymbol{a}' = \mathbf{S}^T\mathbf{M}\mathbf{S}\boldsymbol{a} = \mathcal{M}\boldsymbol{a}. \tag{4}$$

Each $N_1 \times N_2$ block of a digital image is scanned row by row (or column by column) to form an $N_1N_2 \times 1$ vector, which is then encrypted by using the above equation. After the encryption is done, all elements in the $N_1N_2 \times 1$ vector $\boldsymbol{a}'$ are placed back into the $N_1 \times N_2$ image block in the same scanning order.

[3] Here, "the v lowest coefficients" mean those coefficients corresponding to the v largest eigenvalues, i.e., the v low-frequency components of the 2-D DPSS transform.

Besides scrambling v lowest coefficients in $\boldsymbol{\alpha}$, another encryption operation was also suggested in [43]: swapping these coefficients that correspond to the same eigenvalues. This swapping operation is dependent on the fact that $\lambda^{(2D)}_{j_1,j_2} = \lambda^{(2D)}_{j_2,j_1}$ and is unavailable in the 1-D case. In [43], it was not explicitly mentioned how to perform the swapping operation on valid coefficients. However, if all pairs of coefficients with the same eigenvalue are swapped, then the adversary can swap all of them to totally cancel this encryption operation. So, in this paper, we assume that a secret pseudo-random keystream is used to randomly select some (averagely half of all) pairs of coefficients for swapping.

When the scrambling scheme is exerted on digital images with L gray scales, the input and output have to be calibrated to make the scrambling more efficient. Assuming that one plain-block in the plain-image is J and the corresponding cipher-block is J' (both are $N_1N_2 \times 1$-vectors), the encryption process becomes

$$J' = \operatorname{round}\left(\frac{\mathcal{M}\left(J - \mathcal{L}/2\right)}{\gamma} + \mathcal{L}/2\right), \tag{5}$$

where $\operatorname{round}(\cdot)$ converts the input real number into the nearest integer in $\{0, \cdots, L-1\}$, $\mathcal{L}$ is an $N_1N_2 \times 1$ vector in which each element is L and

$$\gamma = \max_n \left(\sum_{j=0}^{N_1N_2-1} |\mathcal{M}_{n,j}|\right). \tag{6}$$

Accordingly, assuming the recovered image block is $\hat{J}$, the decryption process is carried out as follows:

$$\hat{J} = \operatorname{round}\left(\mathcal{M}^T\left(\gamma\left(J' - \mathcal{L}/2\right)\right) + \mathcal{L}/2\right). \tag{7}$$

Due to the round-off errors existing in the encryption/decryption processes, it is obvious that the original plain-image cannot be exactly recovered in most cases. Another problem is that the use of γ may enlarge the noise added to J'. In Sec. V of [43], experimental results were given to show that γ may not be determined by Eq. (6), and an "optimal" value was found for a set of test images: $\gamma = 3$, in the sense of MAE (mean absolute error) and MSE (mean squared error) in the recovering of the plain-image[4].

As to the choice of the secret sub-matrix $\mathbf{M}_v$ that scrambles the v lowest coefficients in $\boldsymbol{\alpha}$, it was suggested in [43] to derive it from a Hadamard matrix $\mathbf{H}$, which is a $v \times v$ $(-1, 1)$-matrix[5] and its rows and columns are orthogonal [46]. Its inverse matrix is $\mathbf{H}^{-1} = \frac{1}{v}\mathbf{H}^T$. By permuting the rows/columns of $\mathbf{H}$, and/or multiplying some rows/columns by -1, one can get an H-equivalent matrix. In this way, one can get $(v!2^v)^2$ H-equivalent matrices, where some of them are identical. Each H-equivalent matrix $\mathbf{H}^*$ can be scaled to get a secret sub-matrix, $\mathbf{M}_v = \frac{1}{\sqrt{v}}\mathbf{H}^*$.

Generally, the plain-image is much larger than $N_1 \times N_2$, so there are many $N_1 \times N_2$ blocks for encryption. To further enhance the security of the image scrambling scheme, in [43] it was also suggested that one should change the secret matrix for each block, under the control of a cryptographical pseudo-random number generator (PRNG). In this case, the key of the scrambling scheme becomes the seed of the PRNG. To facilitate the following cryptanalysis, we use change_key=1 to denote this encryption configuration and change_key=0 to denote the basic configuration with a fixed secret matrix.

III. Cryptanalysis

As a major part of modern cryptology, cryptanalysis focuses on the security analysis of different kinds of cryptographical algorithms, including encryption schemes (ciphers), hash functions, security protocols, etc. [7]. Generally, the following four types of attacks should be considered when evaluating the security of a cipher:

- *ciphertext-only attack*, in which an attacker can only observe a number of ciphertexts;
- *known-plaintext attack*, in which an attacker can observe a number of plaintexts and the corresponding ciphertexts;
- *chosen-plaintext attack*, in which an attacker can deliberately choose a number of plaintexts and observe the corresponding ciphertexts;
- *chosen-ciphertext attack*, in which an attacker can deliberately choose a number of ciphertexts and observe the corresponding plaintexts.

Among the four attacks, ciphertext-only attack is the simplest one and every cipher should resist this kind of attack. The other three attacks are much more advanced, but become more and more popular in today's digital and networked world. Known-plaintext attack is very common on modern ciphers, since most binary files and data packets transmitted over networks have some fixed segments, such as the leading headers and frequently-used syntax elements. Chosen-plaintext and chosen-ciphertext attacks are possible when an attacker gets a temporary access to the encryption/decryption machine, or can seduce the target user to store some chosen files or transmit some chosen data. An imaginary scenario of chosen-plaintext attack is as follows: 1) Eve sends an interesting photo to Alice; 2) Alice encrypts the photo with her secret key and then forwards it to Bob for sharing; 3) Eve mounts a chosen-plaintext attack after observing the encrypted photo transmitted over the public channel.

If a cipher can resist only ciphertext-only attack, it has to be used very carefully to avoid any possibility of the other three types of attacks. In this section, we report a results of a comprehensive investigation on the security of the image scrambling scheme under study against all the four kinds of attacks. Throughout this section, without loss of generality, we employ the scrambling parameters used in [43, Sec. V] for demonstration: $N_1 = N_2 = N = 8$, $W = 0.25$, $v = 8$, $\gamma = 3$. The secret matrix and the random swapping operations are both controlled by the `rand()` function with a random seed. All the experiments were carried out with Mathwork's

[4] In this paper, we mainly use PSNR to investigate the performance, because PSNR is more popular in the area of image processing than MAE and MSE. To keep a direct comparison with the results in [43], the MAE value is also given for all images measured in this paper with PSNR.

[5] The order of a Hadamard matrix (i.e., the value of v) cannot be an arbitrary value, but can only be 1, 2, or $4n$, where $n \in \mathbb{Z}$. So this key-generate method can be used only when v satisfies this requirement.

Matlab 6.5, based on a series of programs derived from the reference codes[6] used in [43].

A. Ciphertext-Only Attack

The visible encryption performance of the image scrambling scheme is explained first. For a 256×256 plain-image with 256 gray scales (i.e., $L = 256$) shown in Fig. 1, the encryption results when change_key=0 and 1 are given in Figs. 2a and 2b, respectively. It can be seen that some smooth areas in the plain-image is still recognizable after encryption. This problem was also noticed by the authors of [43] and considered as a minor security problem that can be further improved with some other techniques.

Fig. 1. The plain-image "Lenna".

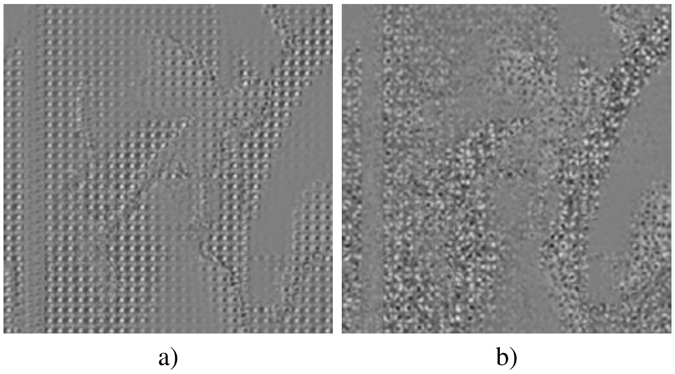


Fig. 2. The encryption results of "Lenna" when a) change_key=0 and b) change_key=1.

In the following, consider how to get more visual information from the cipher-images than that in Fig. 2. There are several different ways to do so.

1) Error-Concealment Based Attack [11]: Since the image scrambling scheme is a selective encryption algorithm, we can try to recover the plain-image from these unencrypted coefficients. This error-concealment based attack (ECA in short) is a common attack on all selective encryption methods. As pointed out in [2], [47], for selective encryption based on any orthogonal transform, there is always some visual information leaking from the unencrypted transform coefficients. Although the corresponding energy of these unencrypted coefficients may be rather small, some important visual information can still be distinguished by human eyes. It is true that 2-D DPSS also form an orthogonal transform, so an attacker can try to carry out an ECA on the image scrambling scheme by setting the v scrambled low coefficients to be some fixed values. For the cipher-images shown in Fig. 2, the broken results are shown in Fig. 3 when the fixed values are 0 and α_v (the lowest unencrypted coefficient). Comparing Figs. 2 and 3, one can see that a rough view of the original plain-image has emerged.

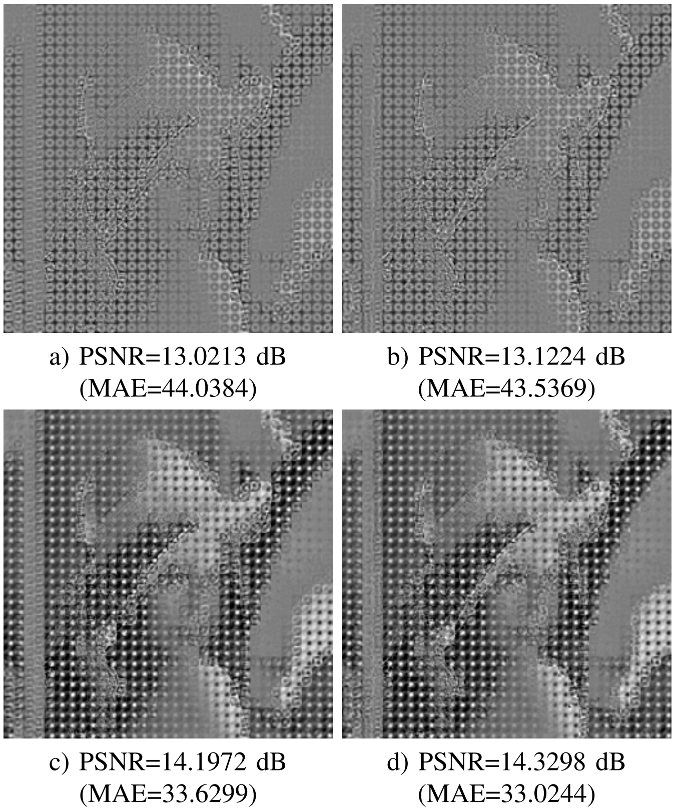


Fig. 3. The breaking performance of ECA on the cipher-images shown in Fig. 2, by setting the $v = 8$ scrambled low coefficients of each block as follows: a) $\alpha_0 = \cdots = \alpha_7 = 0$, change_key=0; b) $\alpha_0 = \cdots = \alpha_7 = 0$, change_key=1; c) $\alpha_0 = \cdots = \alpha_7 = \alpha_8$, change_key=0; d) $\alpha_0 = \cdots = \alpha_7 = \alpha_8$, change_key=1.

The broken results shown in Fig. 3 can be further enhanced by investigating the statistical properties of all the coefficients in α. For the plain-image shown in Fig. 1, we calculated the histograms of all the 2-D DPSS coefficients and those of $\alpha_0 \sim \alpha_9$ are given in Fig. 4. Among the 10 lowest 2-D DPSS coefficients, one can see that the mean values of α_1, α_2, α_3, α_6, α_7 and α_9 are all close to 0, while those of α_0, α_4, α_5 and α_8 are not (see also Fig. 5 for a plot of all the $N^2 = 64$ mean values). Dividing all the mean values by that of α_8, we can get 64 ratios, $\{r_i = E(\alpha_i)/E(\alpha_8)\}_{i=0}^{63}$, as shown in Fig. 6. If these ratios keep approximately unchanged for most natural images, one can use them to statistically optimize the breaking performance of ECA. For 1,200 natural images falling into four different categories, "people", "wild animals", "textures" and "city life and China", we calculated the mean values and variances of $\{r_i\}_{i=0}^{63}$ as shown in Fig. 7. One can see that the mean values of these ratios are really stable for the 1,200 test images (though the variances are not very small for some ones): $r_0 \approx 2.8$, $r_4 \approx r_5 \approx 1.68$,

[6] Courtesy of Dr. Dimitri Van De Ville (the first author of [43]).

$r_1 \approx r_2 \approx r_3 \approx r_6 \approx r_7 \approx 0.$

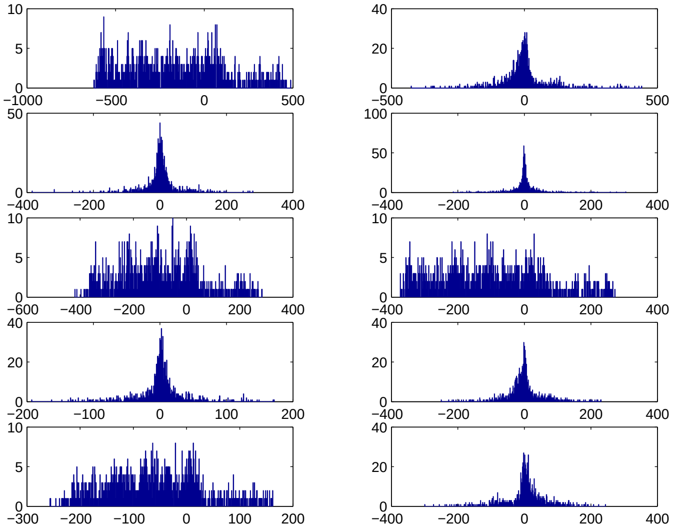


Fig. 4. The histograms of $\alpha_0 \sim \alpha_9$ estimated from all blocks in the plain-image "Lenna" (order: from left to right, from top to bottom).

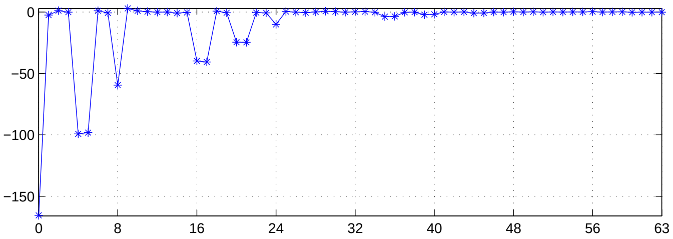


Fig. 5. The mean values of the 64 2-D DPSS coefficients of all blocks in the plain-image "Lenna".

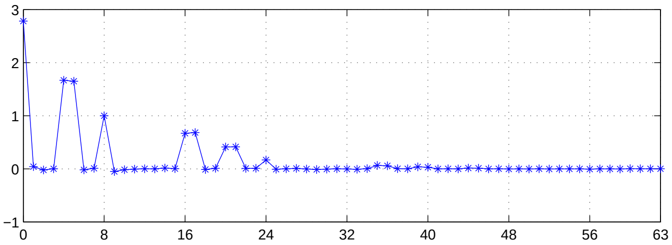
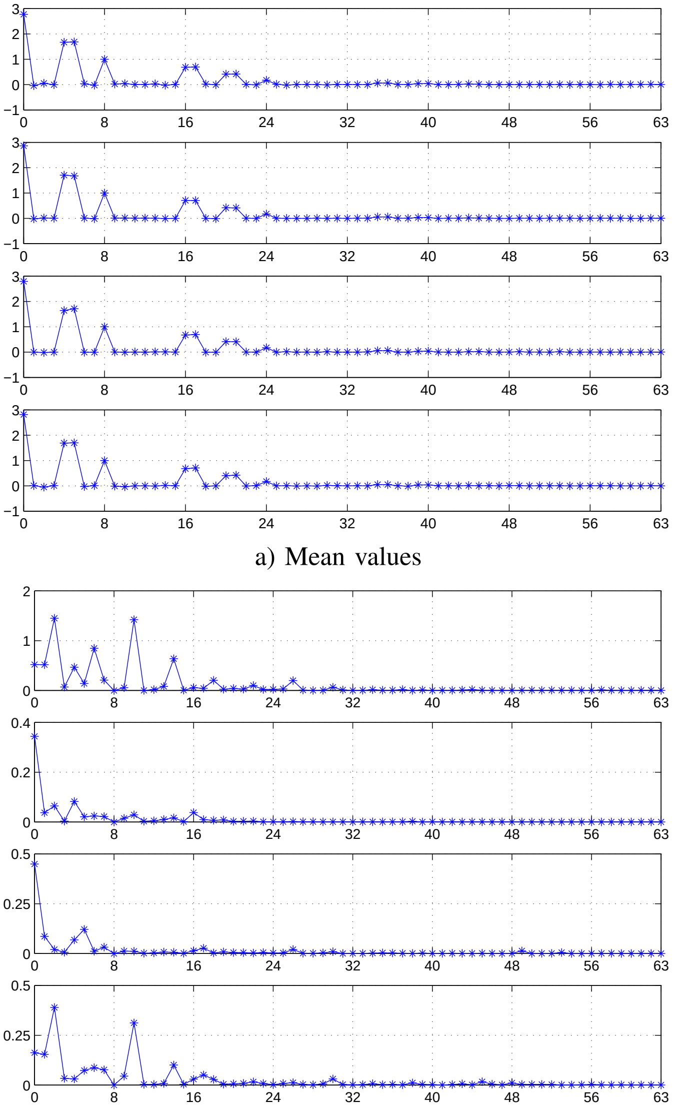


Fig. 6. The ratios of the mean values of the 64 2-D DPSS coefficients to the mean value of α_8 of all blocks in the plain-image "Lenna".

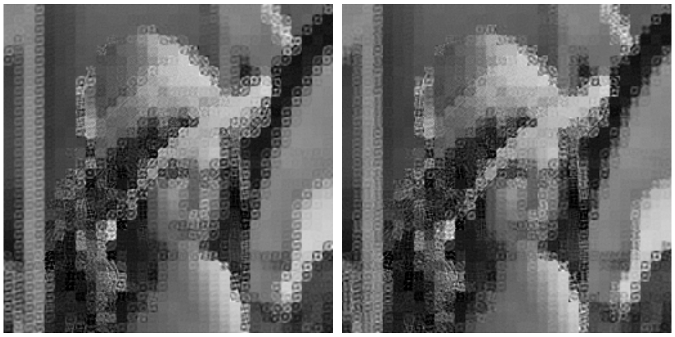


a) Mean values

b) Variances

Fig. 7. The mean values and variances of $\{r_i = E(\alpha_i)/E(\alpha_8)\}_{i=0}^{63}$ of 1,200 test images in four different categories (from top to bottom: "people", "wild animals", "textures", "city life and China").

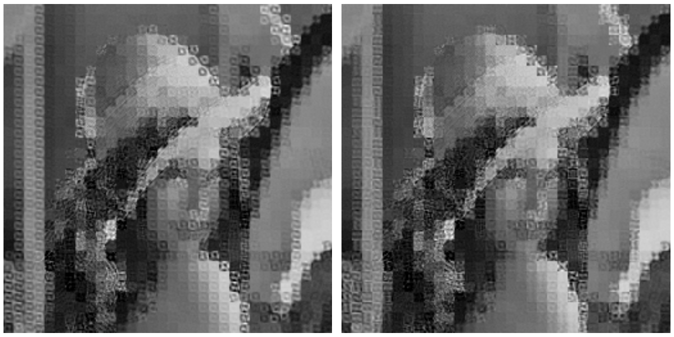
The above fact implies that the following setting of the 8 lowest coefficients is optimal to achieve the best breaking performance of ECA in a statistical sense: $\alpha_0 = 2.8\alpha_8$, $\alpha_4 = \alpha_5 = 1.68\alpha_8$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_6 = \alpha_7 = 0$. For the two cipher-images shown in Fig. 2, the performance of such an optimized ECA is shown in Fig. 8. Considering the variances of r_i shown in Fig. 7b, in a real attack one can further adjust the values of r_1, r_4 and r_5 to get an even better breaking performance. For example, by introducing a small degree of randomness in the values of $r_0 \sim r_7$, we have obtained some results with clearer edges and higher PSNR values (i.e., smaller MAE values). The corresponding results are shown in Fig. 9. Comparing Fig. 9 with Fig. 8, one can easily distinguish the improvements around the edges.



a) PSNR=18.6957 dB (MAE=18.2867)

b) PSNR=19.0843 dB (MAE=17.4056)

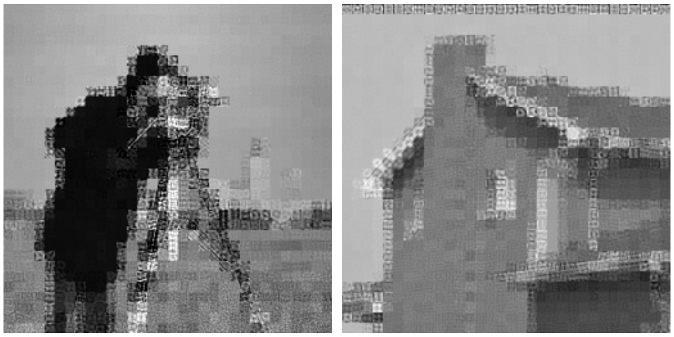
Fig. 8. The optimized ECA of the cipher-images shown in Fig. 2, when $\alpha_0 = 2.8\alpha_8$, $\alpha_4 = \alpha_5 = 1.68\alpha_8$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_6 = \alpha_7 = 0$.



a) PSNR=19.2498 dB (MAE=17.0121)

b) PSNR=19.3390 dB (MAE=16.8824)

Fig. 9. The optimized ECA of the cipher-images shown in Fig. 2, when the value of each α_i is disturbed by a random variable distributed uniformly in $(-0.1, 0.1)$.

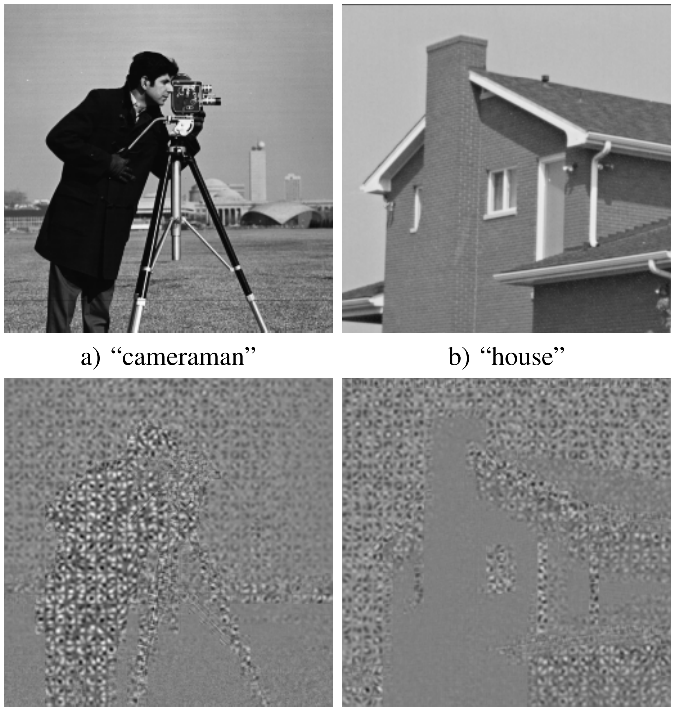


a) PSNR=18.2250 dB (MAE=15.8023)

b) PSNR=20.9006 dB (MAE=12.1457)

Fig. 11. The optimized ECA for the cipher-images corresponding to the two plain-images shown in Fig. 10, when the value of each α_i is disturbed by a random variable distributed uniformly in $(-0.1, 0.1)$.

To further demonstrate the performance of this attack, the results of the optimized ECA for other two plain-images, "cameraman" and "house", are given in Fig. 11.



a) "cameraman"

b) "house"

c) scrambled "cameraman"

d) scrambled "house"

Fig. 10. Two plain-images, "cameraman" and "house", and the corresponding cipher-images when change_key=1.

2) Breaking Random Swapping: When change_key=0, the random swapping coefficients may be exhaustively guessed and then verified by observing the breaking performance of the optimized ECA. When $v = 8$, there are $8(8-1)/2 - 3 = 25$ pairs of coefficients with equal eigenvalues in $\{\alpha_i\}_{i=8}^{63}$, so the complexity of guessing all random coefficients operations is not greater than $O(2^{25})$. Since the 2-D DPSS coefficients $\{\alpha_i\}_{i=32}^{63}$ play a minor role in representing the visual information of an image, one can only guess the random swapping coefficients in $\{\alpha_i\}_{i=8}^{31}$, in which there are only $14 - 3 = 11$ valid pairs of coefficients for possible swapping. In this case, the guessing complexity is reduced to be $O(2^{11})$ and becomes feasible for an attacker to carry out in practice. Because the attacker does not know the original image, he/she has to check each obtained image by naked eyes, trying to find a satisfactory result. Assuming that he/she can examine one image within one minute, a complexity of $O(2^{11})$ means that two days should be enough to finish the work. In the case that a group of people are working together, the attack will be even much easier. Figure 12 shows the broken result when all random swapping operations of $\alpha_8 \sim \alpha_{31}$ are removed. Comparing Fig. 12 with Fig. 8a, one can see that the former contains more recognizable details.

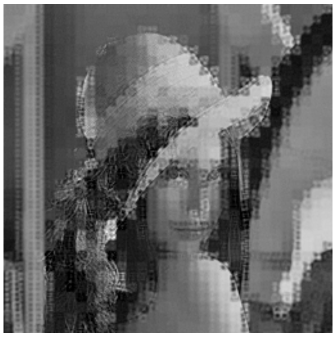


Fig. 12. The breaking performance of the optimized ECA when the random swapping operations of $\alpha_8 \sim \alpha_{31}$ are removed: PSNR=20.4684 dB (MAE=14.2199).

To fix this security defect, one can either enlarge the block size or always set change_key=1. The latter is better since it works for any block size.

3) Insecurity of Hadamard-Based Matrices: When the secret sub-matrix $\mathbf{M}_v$ is generated from a $v \times v$ Hadamard matrix $\mathbf{H}$ as suggested in [43], our experiments showed that the decryption is not sufficiently sensitive to the key mismatch, which is an undesirable property for a good cryptosystem and generally leads to a dramatic reduction of the key space [7]. In our experiments, we exerted some fundamental matrix transformations on the original sub-matrix $\mathbf{M}_v$ so as to get some mismatched matrices, which are then used as a replacement of the original secret matrix $\mathbf{M}$ for decrypting the cipher-image Fig. 2a. Some selected results are given in Fig. 13, from which one can see that many severely mismatched keys

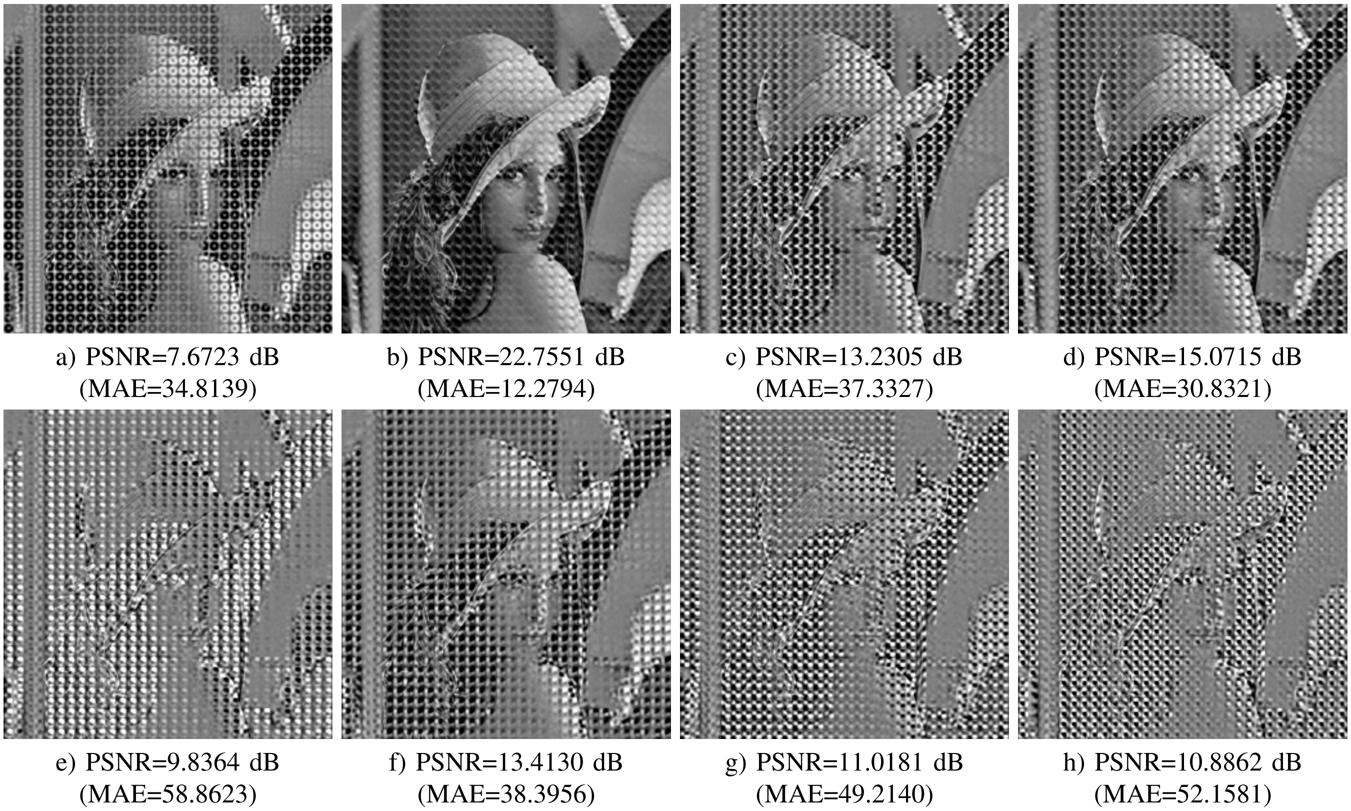

a) PSNR=7.6723 dB (MAE=34.8139)
b) PSNR=22.7551 dB (MAE=12.2794)
c) PSNR=13.2305 dB (MAE=37.3327)
d) PSNR=15.0715 dB (MAE=30.8321)
e) PSNR=9.8364 dB (MAE=58.8623)
f) PSNR=13.4130 dB (MAE=38.3956)
g) PSNR=11.0181 dB (MAE=49.2140)
h) PSNR=10.8862 dB (MAE=52.1581)

Fig. 13. The decryption results (when change_key=0) corresponding to the plain-image "Lenna" with some mismatched keys by processing the sub-matrix $\mathbf{M}_v$ as follows: a) reversing the signs of all elements; b) swapping Rows 1, 8; c) swapping Columns 1, 8; d) swapping Rows 1, 8 and Columns 1, 8; e) reversing the order of all rows; f) reversing the order of all rows and the signs of all elements; g) reversing the order of all columns; h) reversing the order of all rows and that of all columns.

can still recover the plain-image with acceptable qualities.

The low sensitivity of decryption to key mismatch means that a randomly-generated key may be capable of roughly recovering the plain-image. Figure 14 gives the best recovery result of one experiment, in which 100 randomly-generated keys were used to decrypt the cipher-image Fig. 2a. By testing 100,000 random keys, an estimated probability density function (pdf) of PSNR were obtained as shown in Fig. 15. From this empirical pdf, we can calculate and obtain the probability of PSNR$\geq$14 dB be about 0.017. Thus, it is a high-probability event to get a similar result like Fig. 14 by guessing only $O(100)$ random keys, making the random-guess attack feasible in practice. In some sense, we can say that the size of the Hadamard-based key space is dramatically reduced to be smaller than 100.

Note that this security flaw is not so severe when change_key=1. In this case, each block is encrypted by a different secret matrix. If the number of blocks in a plain-image is sufficiently large, it will be impossible to randomly guess all possible secret matrices to reveal the whole image. Of course, it remains practical for an attacker to guess several selected blocks and roughly recover a small window of the plain-image.

4) Low Sensitivity to Plaintext: The encryption scheme under study has another undesirable property: it is not sufficiently sensitive to plaintext. This feature is a natural result of the low sensitivity of the involved matrix computation to input plaintext. Assume that the additive error in $\boldsymbol{a}$ is bounded by ε. Then, the augmented error in $\boldsymbol{a}' = \mathcal{M}\boldsymbol{a}$ is bounded by $N_1N_2\varepsilon$, since each element of $|\mathcal{M}|$ is not greater than 1. This means that the encryption scheme under study is not suitable to be used for encrypting a number of similar plain-images (such as an image and its watermarked counterpart) with the same key; otherwise, the exposure of one plain-image leads to a rough revealment of all the plain-images.

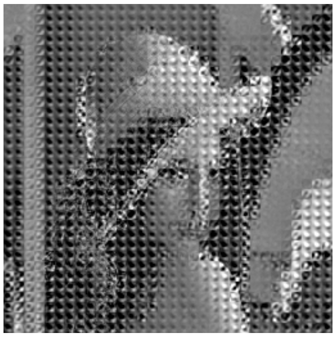


Fig. 14. The best recovery result of the plain-image "Lenna" decrypted with 100 randomly-generated keys: PSNR=16.0590 dB (MAE=27.9677).

B. Known-Plaintext Attack

In known-plaintext attack, one can get a number of plain-images and the corresponding cipher-images. According to Eq. (5), the encryption matrix $\mathcal{M}$ can be derived as follows when N_1N_2 known plain-blocks form an invertible $N_1N_2 \times N_1N_2$

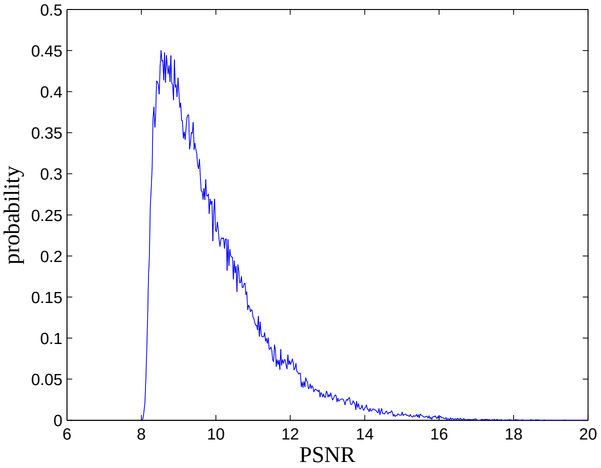


Fig. 15. The empirical pdf of PSNR of the decrypted result of the cipher-image Fig. 2a, estimated from 100,000 random keys.

matrix $(\boldsymbol{J}-\mathcal{L}/2)$:

$$\mathcal{M}=\gamma(\boldsymbol{J}'+\Delta_{\boldsymbol{J}'}-\mathcal{L}/2)(\boldsymbol{J}-\mathcal{L}/2)^{-1}, \qquad (8)$$

where $\boldsymbol{J}'$ is the $N_1N_2\times N_1N_2$ cipher-matrix corresponding to $\boldsymbol{J}$, and $\Delta_{\boldsymbol{J}'}$ denotes the error matrix induced by the $\mathrm{round}(\cdot)$ function. By ignoring the error matrix $\Delta_{\boldsymbol{J}'}$, $\mathcal{M}$ can be estimated by the following equation:

$$\widetilde{\mathcal{M}}=\gamma(\boldsymbol{J}'-\mathcal{L}/2)(\boldsymbol{J}-\mathcal{L}/2)^{-1}, \qquad (9)$$

and the estimation error is

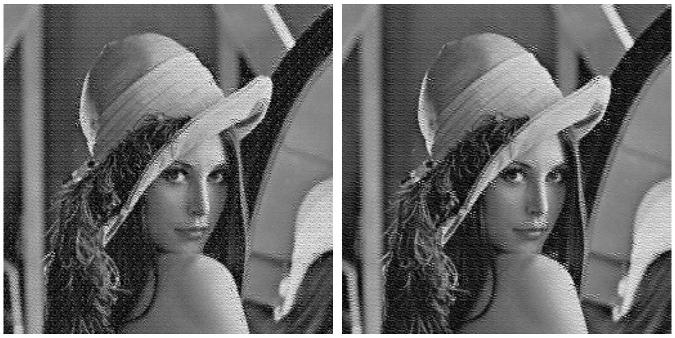
$$\Delta_{\mathcal{M}}=\widetilde{\mathcal{M}}-\mathcal{M}=\gamma\Delta_{\boldsymbol{J}'}(\boldsymbol{J}-\mathcal{L}/2)^{-1}. \qquad (10)$$

Then, $\widetilde{\mathcal{M}}^T$ can be used as a replacement of $\mathcal{M}^T$ for decryption.

It is not easy to theoretically analyzes the relationship between $\boldsymbol{J}$ and $\Delta_{\mathcal{M}}$ (i.e., the relationship between $\boldsymbol{J}$ and the decryption performance of $\widetilde{\mathcal{M}}$), so we carried out a large number of experiments to investigate the real decryption performance of $\widetilde{\mathcal{M}}^T$ by choosing some sets of N_1N_2 plain-blocks to form $\boldsymbol{J}$. In the following, we report our experimental results for two different cases according to the two values of change_key.

1) change_key=0: In this case, all blocks of a plain-image are encrypted with the same matrix $\mathcal{M}$, so generally one known plain-image is enough for an attacker to choose many sets of N_1N_2 plain-blocks, some of which may correspond to a good estimation of $\mathcal{M}$ (i.e., to an acceptable recovery performance of any given plain-image).

When the known plain-image is "Lenna" (Fig. 1), the best results of decrypting the cipher-image of "Lenna" (Fig. 2a) in two separate attacks are given in Fig. 16, with 1,000 and 10,000 sets of N_1N_2 blocks[7], respectively. One can see that the decryption performance is good enough to reveal almost all visual information in the plain-image.

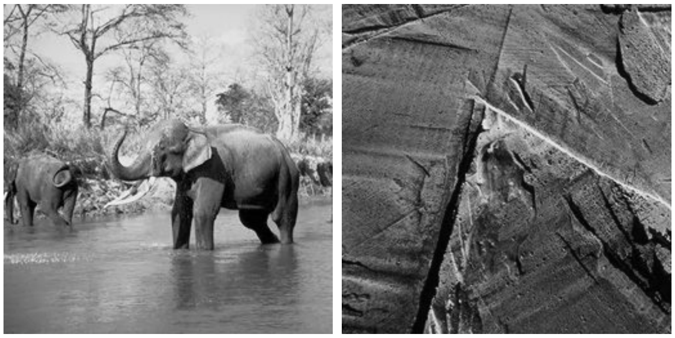


a) PSNR=19.2223 dB (MAE=19.7005)

b) PSNR=20.4598 dB (MAE=16.5549)

Fig. 16. The best decryption results in two experiments of known-plaintext attack with the known plain-image "Lenna" (when change_key=0): a) 1,000 sets of $N_1N_2=64$ plain-blocks; b) 10,000 sets of $N_1N_2=64$ plain-blocks.

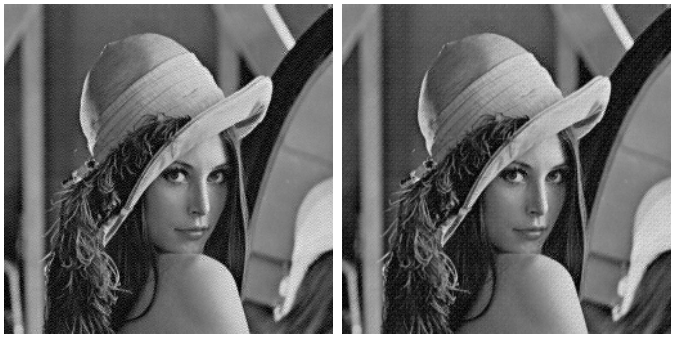
Further experiments showed that some known plain-images can even yield a much better performance than "Lenna". When the known plain-image "Lenna" is replaced by the two images shown in Fig. 17, respectively, the decryption results of the cipher-image Fig. 2a are given in Fig. 18. It can be seen that the decryption performance is nearly perfect. Although it is hard to give a theoretical explanation, this phenomenon can be intuitively explained: "Lenna" contains less blocks with large variances than the other two images.



a)

b)

Fig. 17. Other two images for testing the performance of known-plaintext attack when change_key=0.



a) PSNR=27.7240 dB (MAE=7.4056)

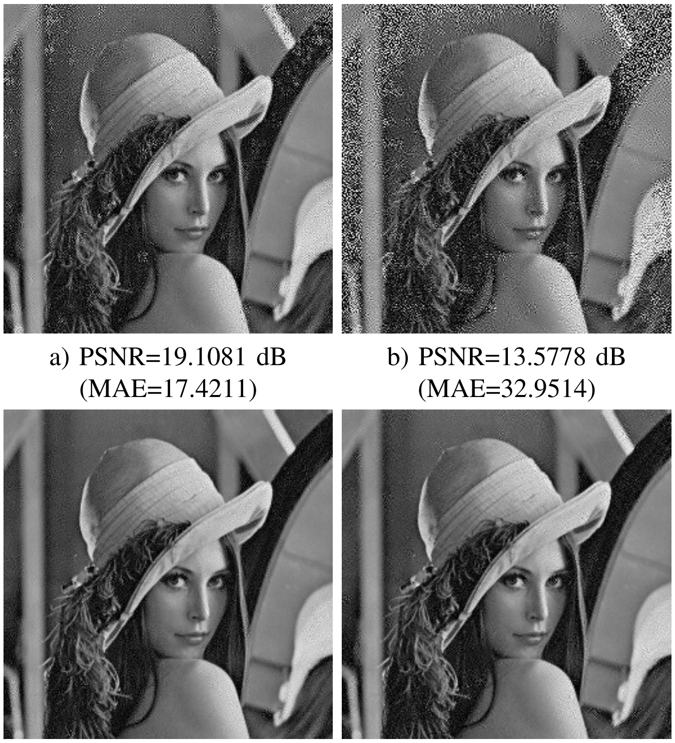
b) PSNR=29.8383 dB (MAE=5.6989)

Fig. 18. The decryption results of known-plaintext attack when when Figs. 17a and 17b serve as the known plain-image, respectively, where change_key=0 and 1,000 sets of $N_1N_2=64$ plain-blocks are processed.

[7] To ensure the invertibility of the formed matrix $\boldsymbol{J}-\mathcal{L}/2$ and to increase the attacking efficiency, in our experiments we first ranked all valid blocks by their variances and then randomly chose N_1N_2 blocks from the 100 ones with larger variances for attacking. A similar but slightly different measure was also used for the experiments given in next sub-subsection when change_key=1.

2) change_key=1: In this case, each block of a plain-image corresponds to a distinctive encryption matrix $\mathcal{M}_{i,j}$, so one plain-image is not capable of supporting the known-plaintext attack. Instead, $m \geq N_1N_2$ plain-images should be known such that for each block, N_1N_2 plain-blocks (one in each plain-image) can be chosen to estimate each $\mathcal{M}_{i,j}$. When $m = 200$, for the four different categories of natural images used in the last subsection, "wild animals", "people", "textures", and "city life and China", we tested the decryption performances of the known-plaintext attack with 500 valid sets of N_1N_2 plain-blocks for each $\mathcal{M}_{i,j}$ (note that there are $\binom{m}{N_1N_2}$ possible sets, but some may not be valid). The decryption results are shown in Fig. 19. The different performances of different sets can be qualitatively explained as follows: for a given block-position (i,j), the breaking performance will be better, if the corresponding blocks of all known plain-images are less correlated. For example, for the category "people", most plain-images have a relatively smooth background, so blocks from different plain-images may be closely correlated or even be almost identical. This will significantly reduce the breaking performance. On the other hand, for the category "textures", generally there does not exist a smooth background, so blocks from different plain-images are relatively less correlated.

By choosing more valid sets of N_1N_2 plain-blocks for each encryption matrix $\mathcal{M}_{i,j}$ or employing some noise-filtering methods, the performance can be further improved.



a) PSNR=19.1081 dB (MAE=17.4211)

b) PSNR=13.5778 dB (MAE=32.9514)

c) PSNR=28.6367 dB (MAE=6.3547)

d) PSNR=27.3269 dB (MAE=7.2946)

Fig. 19. The decryption results of known-plaintext attack when change_key=1, with $m = 200$ known plain-images lying in one of the following four different categories of natural images: a) "wild animals", b) "people", c) "textures", d) "city life and China". For each value of (i,j), 500 valid sets of $N_1N_2 = 64$ plain-blocks are chosen to estimate each $\mathcal{M}_{i,j}$.

C. Chosen-Plaintext Attack

Compared with known-plaintext attack, in chosen-plaintext attack one can freely choose some plain-blocks to optimize the breaking performance. Now let us choose $\boldsymbol{J} - \mathcal{L}/2 = s\mathbf{I}$, where $\mathbf{I}$ denotes the $N_1N_2 \times N_1N_2$ identity matrix. Then, Eq. (10) can be simplified as follows:

$$\Delta_{\mathcal{M}} = \gamma\Delta_{\boldsymbol{J}'}(s\mathbf{I})^{-1} = \frac{\gamma\Delta_{\boldsymbol{J}'}}{s}. \tag{11}$$

On the range of each element in $\Delta_{\boldsymbol{J}'}$, we have the following proposition.

Proposition 1: When $\gamma \geq 1$, $|\Delta_{\boldsymbol{J}'}(i,j)| \leq 1$, $\forall(i,j)$.

Proof: Observing Eq. (8), one can deduce that

$$\boldsymbol{J}' + \Delta_{\boldsymbol{J}'} - \mathcal{L}/2 = \frac{\mathcal{M}(\boldsymbol{J} - \mathcal{L}/2)}{\gamma} = \frac{s\mathcal{M}}{\gamma}.$$

Since $\mathcal{M}$ is an orthogonal matrix, we have $-1 \leq \mathcal{M}(i,j) \leq 1$, so that

$$\frac{-|s|}{\gamma} + L/2 \leq \boldsymbol{J}'(i,j) + \Delta_{\boldsymbol{J}'}(i,j) \leq \frac{|s|}{\gamma} + L/2.$$

Considering $|s| \leq L/2$ and $\gamma \geq 1$, we have

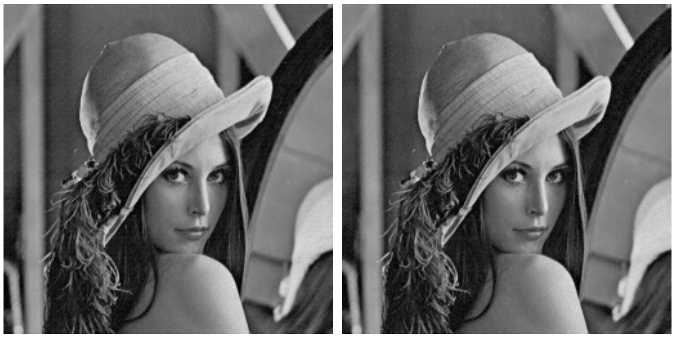
$$\boldsymbol{J}'(i,j) + \Delta_{\boldsymbol{J}'}(i,j) \in [0, L],$$

which immediately leads to $|\Delta_{\boldsymbol{J}'}(i,j)| \leq 1$ and proves the proposition[8]. ■

Then, from Proposition 1, one can get

$$|\Delta_{\mathcal{M}}(i,j)| = \left|\frac{\gamma\Delta_{\boldsymbol{J}'}(i,j)}{s}\right| \leq \frac{\gamma}{|s|}, \tag{12}$$

which means that the best breaking performance is reached when $|s|$ is maximized, i.e., $s = -L/2$ when $\boldsymbol{J} = (1 - \mathbf{I})L/2$. With this chosen value of s, some experiments have been made to confirm this theoretical result, as shown in Fig. 20. Note that this attack needs only N_1N_2 plain-blocks when change_key=0 and N_1N_2 plain-images when change_key=1.



a) PSNR=37.7884 dB (MAE=2.3627)

b) PSNR=37.5721 dB (MAE=2.4110)

Fig. 20. The decryption results of chosen-plaintext attack when $s = -L/2 = -128$ for the following two cases: a) change_key=0; b) change_key=1.

[8] When $\gamma \geq \frac{L/2}{L/2-1} = 1 + 2/(L-2)$, this proposition has a stronger result: $|\Delta_{\boldsymbol{J}'}(i,j)| \leq 1/2$, since $\boldsymbol{J}'(i,j) + \Delta_{\boldsymbol{J}'}(i,j) \in [0, L-1]$. But the weak result is already sufficient to support the following analysis on the choice of $|s|$.

D. Chosen-Ciphertext Attack

In this attack, one can choose cipher-images instead of plain-images, so the target for reconstruction changes from $\mathcal{M}$ to its transpose matrix $\mathcal{M}^T$. When N_1N_2 cipher-blocks form an invertible $N_1N_2 \times N_1N_2$ matrix $\boldsymbol{J}' - \mathcal{L}/2$, one can get the following equation from Eq. (7):

$$\mathcal{M}^T = \frac{(\hat{\boldsymbol{J}} - \mathcal{L}/2 + \Delta_{\hat{\boldsymbol{J}}})(\boldsymbol{J}' - \mathcal{L}/2)^{-1}}{\gamma}. \tag{13}$$

Removing the quantization error $\Delta_{\hat{\boldsymbol{J}}}$, one has

$$\widetilde{\mathcal{M}^T} = \frac{(\hat{\boldsymbol{J}} - \mathcal{L}/2)(\boldsymbol{J}' - \mathcal{L}/2)^{-1}}{\gamma}, \tag{14}$$

and

$$\Delta_{\mathcal{M}^T} = \widetilde{\mathcal{M}^T} - \mathcal{M}^T = \frac{\Delta_{\hat{\boldsymbol{J}}}(\boldsymbol{J}' - \mathcal{L}/2)^{-1}}{\gamma}. \tag{15}$$

Similarly, choosing $\boldsymbol{J}' - L/2 = s\mathbf{I}$, one further gets

$$\Delta_{\mathcal{M}^T} = \frac{\Delta_{\hat{\boldsymbol{J}}}(s\mathbf{I})^{-1}}{\gamma} = \frac{\Delta_{\hat{\boldsymbol{J}}}}{s\gamma}. \tag{16}$$

On the range of each element in $\Delta_{\hat{\boldsymbol{J}}}$, we have the following proposition.

Proposition 2: $|\Delta_{\hat{\boldsymbol{J}}}(i,j)| \leq 1/2$, $\forall(i,j)$ if and only if

$$-\frac{L+1}{2\gamma\mathcal{M}^T(i,j)} \leq s \leq \frac{L-1}{2\gamma\mathcal{M}^T(i,j)}. \tag{17}$$

Proof: From Eq. (7), one has

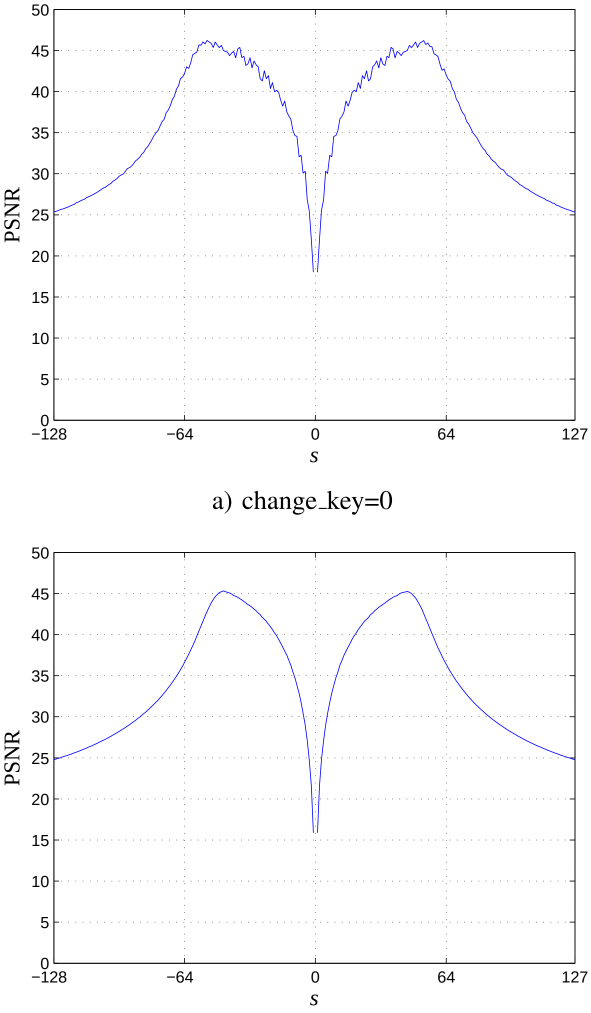
$$\hat{\boldsymbol{J}} + \Delta_{\hat{\boldsymbol{J}}} = \mathcal{M}^T(\gamma(\boldsymbol{J}' - \mathcal{L}/2)) + \mathcal{L}/2 = s\gamma\mathcal{M}^T + \mathcal{L}/2.$$

Note that $|\Delta_{\hat{\boldsymbol{J}}}(i,j)| \leq 1/2$ if and only if $-1/2 \leq \hat{\boldsymbol{J}}(i,j) + \Delta_{\hat{\boldsymbol{J}}}(i,j) \leq (L-1) + 1/2$, which is equivalent to $-1/2 \leq s\gamma\mathcal{M}^T(i,j) + L/2 \leq L - 1/2$. Solving the two inequalities proves the proposition. ■

From the above proposition, when $-\frac{L+1}{2\gamma\max(\mathcal{M}^T)} \leq s \leq \frac{L-1}{2\gamma\max(\mathcal{M}^T)}$, one has

$$|\Delta_{\mathcal{M}^T}(i,j)| = \left|\frac{\Delta_{\hat{\boldsymbol{J}}}(i,j)}{s\gamma}\right| \leq \frac{1}{2|s|\gamma}. \tag{18}$$

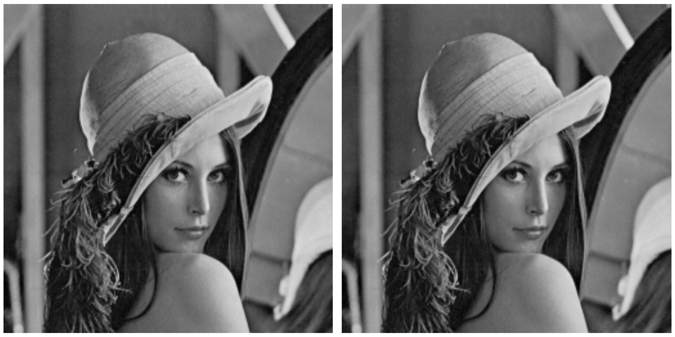
When $s > \frac{L-1}{2\gamma\max(\mathcal{M}^T)}$ or $s < -\frac{L+1}{2\gamma\max(\mathcal{M}^T)}$, it is not easy to directly estimate the range of $|\Delta_{\mathcal{M}^T}(i,j)|$, but it is expected to be greater than $\frac{1}{2|s|\gamma}$. For a randomly generated key, Fig. 21 gives the experimental relationship between the value of s and the breaking performance of the chosen-ciphertext attack, from which one can see that the best decryption performance is achieved when $|s| \approx 47$ (see Fig. 22 for the decryption results when $s = -47$). Due to the symmetry of the curve shown in Fig. 21, in a real attack one can try all positive (or negative) values of s to determine an optimal value as the outcome of the cryptanalysis. This means that the attack needs no more than $N_1N_2L/2$ cipher-blocks when change_key=0 and $N_1N_2L/2$ cipher-images when change_key=1. Note that in most cases it is actually sufficient to achieve a nearly optimal result by fixing the value of $|s|$ around 50. In this case, N_1N_2 cipher-blocks/images are enough to support this attack.



a) change_key=0

b) change_key=1

Fig. 21. The experimental relationship between the value of s and the breaking performance of chosen-ciphertext attack (measured by PSNR), when the test plain-image is "Lenna".



a) PSNR=45.5619 dB (MAE=0.9937)

b) PSNR=45.2878 dB (MAE=1.0094)

Fig. 22. The decryption results of chosen-ciphertext attack when $s = -47$ and the test plain-image is "Lenna": a) change_key=0; b) change_key=1.

IV. Discussion

Recalling the cryptanalysis given in the last section, one can see that one essential reason for all the attacks to be successful is the low sensitivity of the decryption to key mismatch. Actually, this feature is not unique for the original 2-D DPSS basis set shown in Fig. 1 of [43]. We also tested some other basis sets and similar results have been obtained (but with some differences in the details, such as the histograms of $\alpha_0 \sim \alpha_9$ shown in Fig. 4). This implies that the low sensitivity to key mismatch is a common feature of most (if not all) orthogonal transforms[9]. It can be explained by the low sensitivity of matrix computation to small quantization errors and the marvelous capability of human eyes to resist noises in natural images.

In Tables I and II, we give a summary of all the cryptanalytic results obtained through the last section. It is clear that the image scrambling scheme under study is not secure against all the four types of attacks, although it does not suffer from two security flaws, i.e., random swapping breaking and insecurity of Hadamard-based key, when change_key=1.

TABLE I
The computational complexity of ciphertext-only attacks reported in this paper, where N_b denotes the number of $N_1 \times N_2$ blocks in the plain-image.

	change_key=0	change_key=1
error-concealment based attack	O(1)	
random swapping breaking	$O(2^{11})$	$O(2^{11N_b})$
insecurity of Hadamard-based key	$O(0.03^{-1})$	$O(0.03^{-N_b})$

TABLE II
The numbers of plaintexts/ciphertexts needed in known/chosen-plaintext and chosen-ciphertext attacks.

	change_key=0 (blocks)	change_key=1 (images)
known-plaintext attack	$O(N_1N_2)$	
chosen-plaintext attack	N_1N_2	
chosen-ciphertext attack	$\leq N_1N_2L/2$	

From the experimental results given in the last section, the breaking performance of the four attacks can be ranked as follows (from the best to the worst): chosen-ciphertext attack > chosen-plaintext attack > known-plaintext attack > ciphertext-only attack (ECA). For the worst attack, the error-concealment based attack, only a low-resolution view of the plain-image can be successfully recovered, and most high-resolution details are lost (see Fig. 8). As a result, we have the following recommendations on how to apply this particular image scrambling scheme in real applications:

- Use it ONLY for the purpose of perceptual encryption.
- NEVER use the same key to encrypt more than one plain-image[10]. Or, NEVER repeatedly use the same key for more than one plain-image if known/chosen-plaintext or chosen-ciphertext attack is likely.
- ALWAYS set change_key=1 if the secret matrix is generated from a Hadamard matrix.

Perceptual encryption is a technique of multimedia encryption that is used to degrade the perceptible quality of multimedia data, under the control of a secret key and a quality-degradation factor [2], [10]. Here, the secret key is used to avoid any illegal attempt of reconstructing the multimedia data in a higher quality, and the quality-degradation factor determines the degradation degree induced by the perceptual encryption. Apparently, for the image scrambling scheme under study, the degradation on the visual quality of the plain-image should not be measured by the cipher-image, but by the recovered plain-image via the optimized ECA discussed in Sec. III-A.1.

Despite the above security problems and limitations, this particular image scrambling scheme has some advantages in realizing a **lossy** perceptual encryption scheme, i.e., an encryption scheme that works well with any lossy compression algorithm. This is mainly because this scrambling scheme does not incur significant bandwidth expansion, which is generally not the case for many other image encryption schemes. Our experiments have shown that the encryption has only negligible influence on the compression efficiency of a standard JPEG algorithm, as expected from the bandwidth preservation feature. Another important factor is that the decryption is not very sensitive to errors in cipher-images, due to the same reason that the encryption is not very sensitive to plaintext as discussed in Sec. III-A.4. Figure 23 gives the decryption results when the cipher-image Fig. 2b is compressed by the standard JPEG algorithm with the parameter “Quality” equal to $40 \sim 90$, respectively. One can see that the lossy compression really leads to a lossy decryption result, but the recovery performance remains acceptable as long as the compression ratio is not very high.

V. Conclusion

This paper presents a comprehensive investigation on the security of an image scrambling scheme recently proposed in [43]. As a result, it has been found that this image scrambling scheme is not sufficiently secure against various types of attacks: ciphertext-only attack, known-plaintext attack, chosen-plaintext attack, chosen-ciphertext attack. Other two major security flaws have also been pointed out when a fixed secret matrix is used to encrypt the whole image. Based on the cryptanalytic results, it is concluded that this image scrambling scheme can only be used for (lossless or lossy) perceptual encryption, instead of providing a full protection of all (or most) visual information in the plain-image.

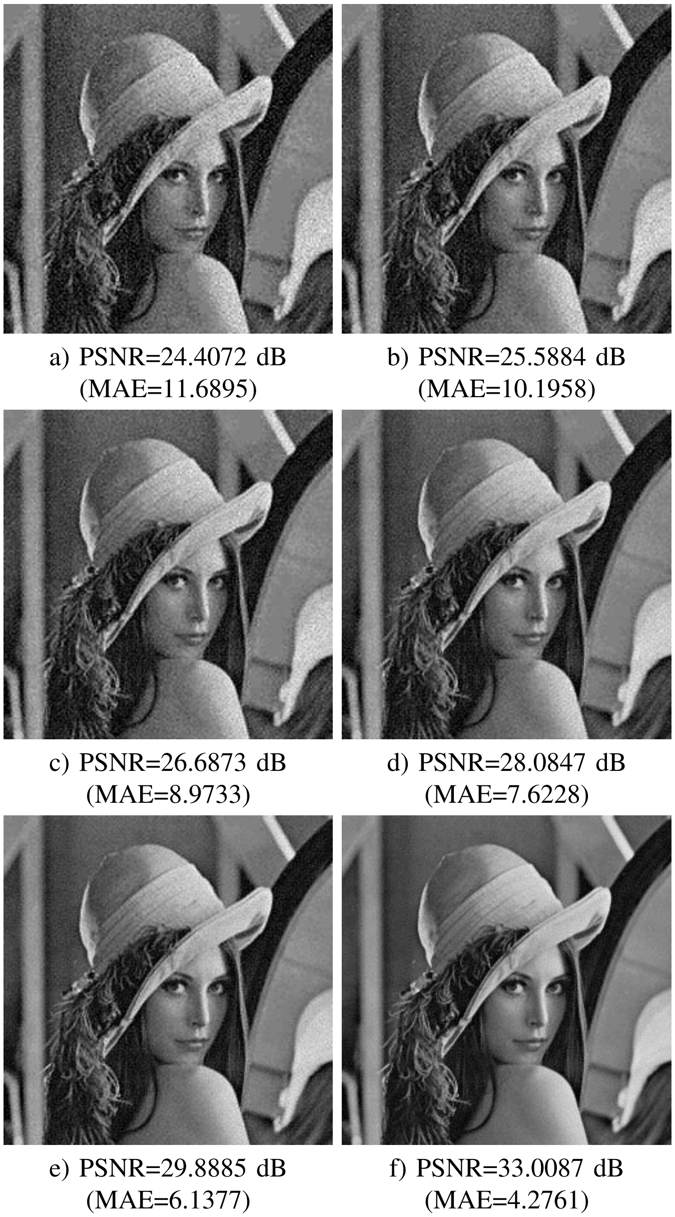
References

[1] B. Furht, D. Socek, and A. M. Eskicioglu, “Fundamentals of multimedia encryption techniques,” in *Multimedia Security Handbook*, B. Furht and D. Kirovski, Eds. Boca Raton, Florida: CRC Press, 2004, ch. 3, pp. 93–132.

[2] S. Li, G. Chen, and X. Zheng, “Chaos-based encryption for digital images and videos,” in *Multimedia Security Handbook*, B. Furht and D. Kirovski, Eds. Boca Raton, Florida: CRC Press, 2004, ch. 4, pp. 133–167. [Online]. Available: http://www.hooklee.com/pub.html

[9] As an interesting comparison, there is another similar (but with different reason) phenomenon [2], [47]: selective encryption working with any orthogonal transform cannot conceal all visual information of the plain-image.

[10] To do so, a key-management system is generally needed to generate a secret key for each plain-image [7].



a) PSNR=24.4072 dB (MAE=11.6895)

b) PSNR=25.5884 dB (MAE=10.1958)

c) PSNR=26.6873 dB (MAE=8.9733)

d) PSNR=28.0847 dB (MAE=7.6228)

e) PSNR=29.8885 dB (MAE=6.1377)

f) PSNR=33.0087 dB (MAE=4.2761)

Fig. 23. The lossy decryption results when the cipher-image Fig. 2b is compressed by the standard JPEG algorithm: a) Quality=40; b) Quality=50; c) Quality=60; d) Quality=70; e) Quality=80; f) Quality=90.

[3] A. Uhl and A. Pommer, *Image and Video Encryption: From Digital Rights Management to Secured Personal Communication*. Boston: Springer Science + Business Media Inc., 2005.

[4] B. Furht, E. Muharemagic, and D. Socek, Eds., *Multimedia Encryption and Watermarking*. New York: Springer, 2005.

[5] W. Zeng, H. Yu, and C.-Y. Lin, Eds., *Multimedia Security Technologies for Digital Rights Management*. Orlando, Florida: Academic Press, Inc., 2006.

[6] B. Javidi, *Optical and Digital Techniques for Information Security*. New York: Springer Science + Business Media Inc., 2005.

[7] B. Schneier, *Applied Cryptography – Protocols, Algorithms, and Souce Code in C*, 2nd ed. New York: John Wiley & Sons, Inc., 1996.

[8] National Institute of Standards and Technology (US), "Specification for the advanced encryption standard (AES)," Federal Information Processing Standards Publication 197 (FIPS PUB 197), November 2001.

[9] L. Qiao, "Multimedia security and copyright protection," Ph.D. dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 1998.

[10] S. Li, G. Chen, A. Cheung, B. Bhargava, and K.-T. Lo, "On the design of perceptual MPEG-video encryption algorithms," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 2, pp. 214–223, 2007.

[11] J. Wen, M. Severa, W. Zeng, M. H. Luttrell, and W. Jin, "A format-compliant configurable encryption framework for access control of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 545–557, 2002.

[12] Y. Matias and A. Shamir, "A video scramFbing technique based on space filling curve (extended abstract)," in *Advances in Cryptology – Crypto'87*, ser. Lecture Notes in Computer Science, vol. 293, 1987, pp. 398–417.

[13] N. Bourbakis and C. Alexopoulos, "Picture data encryption using scan patterns," *Pattern Recognition*, vol. 25, no. 6, pp. 567–581, 1992.

[14] A. Kudelski, "Method for scrambling and unscrambling a video signal," U.S. Patent 5375168, 1994.

[15] S. Sridharan, E. Dawson, and B. Goldburg, "Fast Fourier transform based speech encryption system," *IEE Proc. I – Comm., Speech & Vision*, vol. 138, no. 3, pp. 215–223, 1991.

[16] B. Goldburg, S. Sridharan, and E. Dawson, "Design and cryptanalysis of transform-based analog speech scramblers," *IEEE J. Select. Areas Commun.*, vol. 11, no. 5, pp. 735–744, 1993.

[17] L. Tang, "Methods for encrypting and decrypting MPEG video data efficiently," in *Proc. 4th ACM Int. Conference on Multimedia*, 1996, pp. 219–229.

[18] K.-L. Chung and L.-C. Chang, "Large encrypting binary images with higher security," *Pattern Recognition Letters*, vol. 19, no. 5-6, pp. 461–468, 1998.

[19] S. U. Shin, K. S. Sim, and K. H. Rhee, "A secrecy scheme for MPEG video data using the joint of compression and encryption," in *Information Security: Second Int. Workshop (ISW'99) Proc.*, ser. Lecture Notes in Computer Science, vol. 1729, 1999, pp. 191–201.

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

[21] M. Bertilsson, E. F. Brickell, and I. Ingemarson, "Cryptanalysis of video encryption based on space-filling curves," in *Advances in Cryptology – EuroCrypt'88*, ser. Lecture Notes in Computer Science, vol. 434, 1989, pp. 403–411.

[22] J.-K. Jan and Y.-M. Tseng, "On the security of image encryption method," *Information Processing Letters*, vol. 60, no. 5, pp. 261–265, 1996.

[23] M. G. Kuhn, "Analysis for the nagravision video scrambling method," 1998. [Online]. Available: http://www.cl.cam.ac.uk/~mgk25/nagra.pdf

[24] H. C. H. Cheng, "Partial encryption for image and video communication." Master's thesis, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, Fall 1998.

[25] T. Uehara and R. Safavi-Naini, "Chosen DCT coefficients attack on MPEG encryption schemes," in *Proc. IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM'2000)*, 2000, pp. 316–319.

[26] C.-C. Chang and T.-X. Yu, "Cryptanalysis of an encryption scheme for binary images," *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1847–1852, 2002.

[27] S. Li, C. Li, G. Chen, N. G. Bourbakis, and K.-T. Lo, "A general cryptanalysis of permutation-only multimedia encryption algorithms," accepted by *Signal Processing: Image Communication* in January 2008, DOI: 10.1016/j.image.2008.01.003, also available in IACR's Cryptology ePrint Archive as Report 2004/374, 2008. [Online]. Available: http://eprint.iacr.org/2004/374

[28] P. Refregier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," *Optics Letters*, vol. 20, no. 7, pp. 767–769, 1995.

[29] X. Tan, O. Matoba, T. Shimura, K. Kuroda, and B. Javidi, "Secure optical storage that uses fully phase encryption," *Applied Optics*, vol. 39, no. 35, pp. 6689–6694, 2000.

[30] A. Carnicer, M. Montes-Usategui, S. Arcos, and I. Juvells, "Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys," *Optics Letters*, vol. 30, no. 13, pp. 1644–1646, 2005.

[31] X. Peng, P. Zhang, H. Wei, and B. Yu, "Known-plaintext attack on optical encryption based on double random phase keys," *Optics Letters*, vol. 31, no. 8, pp. 1044–1046, 2006.

[32] ——, "Known-plaintext attack on double random encoding encryption technique," *Acta Physica Sinica*, vol. 55, no. 3, pp. 1130–1136, 2006.

[33] J. Scharinger, "Fast encryption of image data using chaotic Kolmogorov flows," *J. Electronic Imaging*, vol. 7, no. 2, pp. 318–325, 1998.

[34] J. Fridrich, "Symmetric ciphers based on two-dimensional chaotic maps," *Int. J. Bifurcation and Chaos*, vol. 8, no. 6, pp. 1259–1284, 1998.

[35] Y. Mao, G. Chen, and C. K. Chui, “A symmetric image encryption scheme based on 3D chaotic cat maps,” *Chaos, Solitons & Fractals*, vol. 21, no. 3, pp. 749–761, 2004.

[36] S. Li and X. Zheng, “Cryptanalysis of a chaotic image encryption method,” in *Proc. IEEE Int. Symposium on Circuits and Systems*, vol. II, 2002, pp. 708–711.

[37] ——, “On the security of an image encryption method,” in *Proc. IEEE Int. Conference on Image Processing*, vol. 2, 2002, pp. 925–928.

[38] C. Li, S. Li, D. Zhang, and G. Chen, “Cryptanalysis of a chaotic neural network based multimedia encryption scheme,” in *Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia, Tokyo, Japan, November 30 - December 3, 2004. Proceedings, Part III*, ser. Lecture Notes in Computer Science, vol. 3333. Springer-Verlag, 2004, pp. 418–425.

[39] S. Li, C. Li, G. Chen, and X. Mou, “Cryptanalysis of the RCES/RSES image encryption scheme,” accepted by *J. Syst. Software* in July 2007, DOI: 10.1016/j.jss.2007.07.037, also available in IACR’s Cryptology ePrint Archive as Report 2004/376, 2004. [Online]. Available: http://eprint.iacr.org/2004/376

[40] C. Li, X. Li, S. Li, and G. Chen, “Cryptanalysis of a multistage encryption system,” in *Proc. IEEE Int. Symposium on Circuits and Systems*, 2005, pp. 880–883.

[41] C. Li, S. Li, G. Chen, G. Chen, and L. Hu, “Cryptanalysis of a new signal security system for multimedia data transmission,” *EURASIP J. Applied Signal Processing*, vol. 2005, no. 8, pp. 1277–1288, 2005.

[42] C. Li, S. Li, D.-C. Lou, and D. Zhang, “On the security of the Yen-Guo’s domino signal encryption algorithm (DSEA),” *J. Systems and Software*, vol. 79, no. 2, pp. 253–258, 2006.

[43] D. V. D. Ville, W. Philips, R. V. de Walle, and I. Lemanhieu, “Image scrambling without bandwidth expansion,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 892–897, 2004.

[44] A. D. Wyner, “An analog scrambling scheme which does not expand bandwidth, Part I: Discrete time,” *IEEE Trans. Inform. Theory*, vol. 25, no. 3, pp. 261–274, 1979.

[45] D. Slepian, “Prolate spheroidal wave functions, Fourier analysis, and uncertainty–V: The discrete case,” *Bell Syst. Tech. J.*, vol. 57, no. 5, pp. 1371–1430, 1978.

[46] E. W. Weisstein, “Hadamard matrix,” From MathWorld–A Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/HadamardMatrix.html

[47] C.-P. Wu and C.-C. J. Kuo, “Design of integrated multimedia compression and encryption systems,” *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 828–839, 2005.