Cryptanalyzing an Encryption Scheme Based on Blind Source Separation

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

***Abstract*— Recently there is a proposal of using the underdetermined BSS (blind source separation) principle to design image and speech encryption. In this paper, we report a cryptanalysis of this BSS-based encryption scheme and point out that it is not secure against known/chosen-plaintext attack and chosen-ciphertext attack. In addition, we discuss some other security defects of the schemes: 1) it has a low sensitivity to part of the key and to the plaintext; 2) it is weak against a ciphertext-only differential attack; 3) a divide-and-conquer (DAC) attack can be used to break part of the key. We finally analyze the role of BSS in this approach towards cryptographically secure ciphers.**

***Index Terms*— blind source separation (BSS), speech encryption, image encryption, cryptanalysis, known-plaintext attack, chosen-plaintext attack, chosen-ciphertext attack, differential attack, divide-and-conquer (DAC) attack.**

I. Introduction

With the rapid development of multimedia and networking technologies, the security of multimedia data becomes more and more important in many real applications. To fulfill such an increasing demand, during the past two decades many encryption schemes have been proposed for protecting multimedia data, including speech signals, images and videos [1]–[9].

According to the nature of the data, multimedia encryption schemes can be classified into two basic types: analog and digital. Most early schemes were designed to encrypt analog data in various ways: element permuting, signal masking, frequency shuffling, etc., all of which may be exerted in the time domain or the transform domain, or both. However, due to the simplicity of their encryption procedures, almost all analog encryption schemes are not sufficiently secure against cryptographical attacks, especially those modern attacks such as known/chosen-plaintext and chosen-ciphertext attacks [2], [3], [10], [11]. As a comparison, in digital encryption schemes, one can employ many cryptographically strong ciphers, such as DES [12] and AES [13], to achieve a higher level of security. Besides, to achieve a higher efficiency of encryption and to meet some special demands of multimedia encryption (such as format-compliance [14] and perceptual encryption [15]), many specific multimedia encryption schemes have also been developed [4]–[6]. Recent cryptanalysis work [16]–[30] has shown that some multimedia encryption schemes are insecure against cryptographical attacks.

Recently Lin et al. suggested employing blind source separation (BSS) for the purpose of image and speech encryption [31]–[37]. The basic idea is to mix multiple plaintexts (or multiple segments of the same plaintext) with a number of secret key signals, in the hope that an attacker has to solve a hard mathematical problem – the underdetermined BSS problem. In Sec. VII of [37], Lin et al. claimed that this BSS-based cipher "is immune from the attacks such as the ciphertext-only attack, the known-plaintext, and the chosen-plaintext attack", "as long as the intractability of the underdetermined BSS problem is guaranteed by the mixing matrix for encryption".

This paper re-evaluates the security of the BSS-based encryption scheme and points out that it is actually insecure against known/chosen-plaintext attack and chosen-ciphertext attack. In addition, some other security defects are found under the scenarios of ciphertext-only attack, including its low sensitivity to the mixing matrix (part of the secret key) and to the plaintext, and a differential attack can be recommended, which works well when the matrix size is small. Based on the cryptanalytic findings, we further discuss the role of BSS in this approach towards cryptographically secure ciphers.

The rest of this paper is organized as follows. In the next section, a brief introduction is given to the BSS-based encryption scheme. Section III is the main body of this paper, which reports detailed cryptanalysis of the BSS-based encryption scheme. Then, the role of BSS in cryptography is discussed in Sec. IV. Finally, the last section concludes the paper.

II. BSS-Based Encryption

Blind source separation is a technique that tries to recover a set of unobserved sources or signals from some observed mixtures [38]. Given N unobserved signals $\mathbf{s}_1, \cdots, \mathbf{s}_N$ and a mixing matrix $\mathbf{A}$ of size $M \times N$, the BSS problem is to recover $\mathbf{s}_1, \cdots, \mathbf{s}_N$ from M observed signals $\mathbf{x}_1, \cdots, \mathbf{x}_M$, where

$$[\mathbf{x}_1, \cdots, \mathbf{x}_M]^T = \mathbf{A}[\mathbf{s}_1, \cdots, \mathbf{s}_N]^T. \tag{1}$$

When $M \geq N$, blind source separation is possible if $\mathbf{A}$ satisfies some conditions. However, when $M < N$, this is

This research was partially supported by The Hong Kong Polytechnic University's Postdoctoral Fellowships Program under grant no. G-YX63, and by the Alexander von Humboldt Foundation of Germany. 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).

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.

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

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

generally impossible (whatever $\mathbf{A}$ is), thus leading to the difficult underdetermined BSS problem.

In [31]–[37], Lin et al. introduced a number of secret key signals to make the determination of the plaintext signals become an underdetermined BSS problem in the case that the key signals are unknown. Given P input plain-signals $s_1(t), \cdots, s_P(t)$ and Q key signals $k_1(t), \cdots, k_Q(t)$, the encryption procedure is described as follows[1]:

$$\mathbf{x}(t) = [x_1(t), \cdots, x_P(t)]^T = \mathbf{A}\mathbf{s}_k(t), \tag{2}$$

where $\mathbf{x}(t)$ denote P cipher-signals, $\mathbf{s}_k(t) = [s_1(t), \cdots, s_P(t), k_1(t), \cdots, k_Q(t)]^T$, and $\mathbf{A}$ is a $P \times (P+Q)$ mixing matrix whose elements are within $[-1, 1]$. Write $\mathbf{A} = [\mathbf{A}_s, \mathbf{A}_k]$, where $\mathbf{A}_s$ is a $P \times P$ matrix and $\mathbf{A}_k$ is a $P \times Q$ matrix. Then, the encryption procedure can be represented in an equivalent form as

$$\mathbf{x}(t) = \mathbf{A}_s\mathbf{s}(t) + \mathbf{A}_k\mathbf{k}(t), \tag{3}$$

where $\mathbf{s}(t) = [s_1(t), \cdots, s_P(t)]^T$ and $\mathbf{k}(t) = [k_1(t), \cdots, k_Q(t)]^T$. Thus, as long as $\mathbf{A}_s$ is an invertible matrix, one can decrypt $\mathbf{s}(t)$ as follows[2]:

$$\mathbf{s}(t) = \mathbf{A}_s^{-1}\left(\mathbf{x}(t) - \mathbf{A}_k\mathbf{k}(t)\right). \tag{4}$$

Different values of Q was used in Lin et al.'s papers: $Q = 1$ in [31] and $Q = P$ in [32]–[37]. When $Q = P$, Lin et al. further set $\mathbf{A}_s = \mathbf{B}$ and $\mathbf{A}_k = \beta\mathbf{B}$, where $\beta \geq 10$ for image encryption and $\beta \geq 1$ for speech encryption. In this case, the encryption procedure becomes

$$\mathbf{x}(t) = \mathbf{B}\left(\mathbf{s}(t) + \beta\mathbf{k}(t)\right), \tag{5}$$

and the decryption procedure becomes

$$\mathbf{s}(t) = \mathbf{B}^{-1}\mathbf{x}(t) - \beta\mathbf{k}(t). \tag{6}$$

Observing Eq. (3), one can see that the encryption procedure contains two steps:

- *Step 1*: $\mathbf{x}^{(1)}(t) = \mathbf{A}_s\mathbf{s}(t)$;
- *Step 2*: $\mathbf{x}(t) = \mathbf{x}^{(1)}(t) + \mathbf{A}_k\mathbf{k}(t)$.

The first step corresponds to a simple matrix-based block cipher, and the second step corresponds to a simple addition-based stream cipher. From a different point of view, the two steps are exchanged as follows:

- *Step 1*: $\mathbf{x}^{(1)}(t) = \mathbf{s}(t) + \mathbf{A}_s^{-1}\mathbf{A}_k\mathbf{k}(t)$;
- *Step 2*: $\mathbf{x}(t) = \mathbf{A}_s\mathbf{x}^{(1)}(t)$.

In any case, the BSS-based encryption scheme is always a product cipher composed by a simple block cipher and a simple stream cipher. In next section, we will show that the two sub-ciphers can be separately broken by known/chosen-plaintext attack and chosen-ciphertext attack.

In Lin et al.'s BSS-based encryption scheme, the key signals $k_1(t), \cdots, k_Q(t)$ are so long as the plain-signals and have to be generated by a pseudo-random number generator (PRNG) with a secret seed I_0, which serves as the secret key. According to the principle of BSS, the decryption process actually does not need any knowledge about the mixing matrix $\mathbf{A}$ to separate (i.e., decrypt) the encrypted plain-signals (i.e., the plaintexts). So, it seems that Lin et al. do not consider $\mathbf{A}$ as part of the secret key. However, we believe that $\mathbf{A}$ should be kept secret like part of the key, due to the following considerations:

- From a cryptographer's point of view, except for the secret key, all details about a cryptosystem are known to attackers. So, if $\mathbf{A}$ is not part of the secret key, it should be considered known to attackers. Unfortunately, in this case the product cipher will be degraded to be a simple stream cipher. Considering $\mathbf{x}^*(t) = \mathbf{A}_s^{-1}\mathbf{x}(t)$ as the equivalent cipher-signal, the encryption procedure becomes
$$\mathbf{x}^*(t) = \mathbf{s}(t) + \mathbf{A}_s^{-1}\mathbf{A}_k\mathbf{k}(t). \tag{7}$$
From the above equation, one can see that the encryption scheme is actually independent of the underdetermined BSS problem.
- As the theoretical basis of the BSS technique, it is assumed that the input signals are mutually independent of each other. Though Lin et al. have shown that the BSS techniques can work well for some images and speeches, it remains doubtful if the BSS technique can still work well to decrypt closely-related input signals (such as an image and its watermarked version, or consecutive frames in a movie). Thus, the knowledge about $\mathbf{A}$ is required to ensure that any plaintexts can be exactly decrypted in any case. This is the reason why we represent the decryption procedure as Eq. (4) without employing any BSS algorithm.
- As will be shown later in Sec. III-A.4, the key signals can be totally circumvented in a ciphertext-only differential attack, so the mixing matrix $\mathbf{A}$ must be kept secret as a second defence to maintain the security.

Thus, in this paper we assume that the secret key consists of both I_0 and $\mathbf{A}$. Note also that adding $\mathbf{A}$ into the secret key will definitely lead to a stronger (at least equivalently strong) cryptosystem compared with the one with a single key I_0. By cryptanalyzing the stronger cryptosystem, the original one is also cryptanalyzed.

In [31]–[35], the BSS-based encryption scheme was mainly designed to encrypt P images simultaneously, where $s_i(t)$ is the t-th pixel in the i-th image. In [36], [37], the encryption scheme was suggested to encrypt a single speech, each frame of which is divided into P segments and $s_i(t)$ is the t-th sample in the i-th segment. This encryption scheme can also be applied to a single image, by dividing it into P blocks of the same size. To facilitate the following discussion, we assume that the encryption scheme is used to encrypt a single plaintext with P segments of equal sizes.

In Sec. VII of [37], it was claimed that the BSS-based encryption scheme is secure against most modern cryptographical attacks, including the ciphertext-only attack, known-plaintext attack, and chosen-plaintext attack. In the next section, we will show that this claim is questionable.

[1] To provide a clearer description of the BSS-based encryption scheme, in this paper we use some notations different from those in Lin et al.'s original papers. For example, in [37], the i-th key signal is denoted by $s_{ni}(t)$, while in this paper we use $k_i(t)$ to emphasize the fact that it is a key signal.

[2] In Lin et al.'s papers, it is said that the decryption procedure was achieved via BSS. However, from the cryptographical point of view, it is more convenient to denote the decryption procedure by Eq. (4).

III. Cryptanalysis

Before introducing the cryptanalytic results, let us see how large the key space is. In [31]–[37], each element of $\mathbf{A}$ is restricted within the interval $[-1,1]$. Thus, assuming that each element in $\mathbf{A}$ has R possible values[3], the number of all possible mixing matrix $\mathbf{A}$ is $R^{P(P+Q)}$. Furthermore, assuming that the bit size of I_0 is L, the size of the whole key space is $R^{P(P+Q)}2^L$. When $Q=P$ and $\mathbf{A}=[\mathbf{B},\beta\mathbf{B}]$, the size of the whole key space is $R^{P^2}2^L$. Later, we will show that the real size of the key space is much smaller than this estimation, due to some essential security defects of the BSS-based encryption scheme. We will also explain why this encryption scheme is not secure against known/chosen-plaintext attack and chosen-ciphertext attack.

A. Ciphertext-Only Attack

1) Divide-and-Conquer (DAC) Attack: Rewriting Eq. (4) in the following form:

$$\mathbf{s}(t)=\hat{\mathbf{A}}\mathbf{x}_k(t), \tag{8}$$

where $\mathbf{x}_k(t)=[x_1(t),\cdots,x_P(t),k_1(t),\cdots,k_Q(t)]^T$ and

$$\hat{\mathbf{A}}=\mathbf{A}_s^{-1}\left[\mathbf{I},-\mathbf{A}_k\right]=\left[\mathbf{A}_s^{-1},-\mathbf{A}_s^{-1}\mathbf{A}_k\right].$$

From the above equation, to recover $s_i(t)$, one only needs to know $\mathbf{k}(t)$ and the i-th row of $\hat{\mathbf{A}}$. In other words, when the BSS-based encryption scheme is used to encrypt P independent plaintexts, the i-th plaintext can be exactly recovered with the knowledge of I_0 and the i-th row of $\hat{\mathbf{A}}$. A similar result can be obtained when P segments of one single plaintext is encrypted with the encryption scheme. This fact means that P rows of $\hat{\mathbf{A}}$ can be separately broken with a divide-and-conquer (DAC) attack. As a result, the size of the key space is reduced to be $PR^{(P+Q)}2^L$. When $Q=P$ and $\mathbf{A}=[\mathbf{B},\beta\mathbf{B}]$, it becomes PR^P2^L.

2) Low Sensitivity to $\mathbf{A}$*:* From the cryptographical point of view, given two distinct keys, even if their difference is the minimal value under the current finite precision, the encryption and decryption results of a good cryptosystem should still be completely different. In other words, this cryptosystem should have a very high sensitivity to the secret key [12]. Unfortunately, the BSS-based encryption scheme does not satisfy this security principle, because the involved matrix computation is not sufficiently sensitive to matrix mismatch. Given two matrices $\mathbf{A}_1=[a_{1;i,j}]$ and $\mathbf{A}_2=[a_{2;i,j}]$ of size $M\times N$, if the maximal difference of all elements is ε, then one can easily deduce that Δx_i, the i-th element of $\Delta\mathbf{x}=\mathbf{A}_1\mathbf{s}(t)-\mathbf{A}_2\mathbf{s}(t)$, satisfies the following inequality:

$$\begin{aligned}|\Delta x_i| &= \left|\sum_{j=1}^{N}(a_{1;i,j}-a_{2;i,j})s_j\right|,\\ &\leq \sum_{j=1}^{N}|a_{1;i,j}-a_{2;i,j}|\cdot|s_j|,\\ &\leq N\varepsilon\max(|\mathbf{s}(t)|),\end{aligned}$$

where $|\mathbf{s}(t)|$ denotes the vector composed of absolutes values of all elements in $\mathbf{s}(t)$, i.e., $|\mathbf{s}(t)|=[\,|s_1(t)|\quad\cdots\quad|s_N(t)|\,]^T$ (the same expression will also be used for other vectors/matrices afterwards without further explanation). As a result, the matrix $\mathbf{A}$ can be approximately guessed under a relatively large finite precision ε, still maintaining an acceptable quality of the recovered plaintexts. Since under the finite precision ε one only needs to guess $\lceil 2/\varepsilon\rceil$ values of each element in $\mathbf{A}$, the size of the key space is significantly reduced from $PR^{(P+Q)}2^L$ to $P\lceil 2/\varepsilon\rceil^{(P+Q)}2^L$, where $\lceil 2/\varepsilon\rceil^{(P+Q)}\ll R^{(P+Q)}$.[4] When $Q=P$ and $\mathbf{A}=[\mathbf{B},\beta\mathbf{B}]$, the size of key space is reduced from PR^P2^L to $P\lceil 2/\varepsilon\rceil^P2^L$.

The above low sensitivity can be easily verified with experiments described as follows:

- *Step 1*: for a randomly-generated key $(\mathbf{A},\mathrm{I}_0)$, calculate the ciphertext $\mathbf{x}(t)$ corresponding to a plaintext $\mathbf{s}(t)$;
- *Step 2*: with another mismatched key $(\mathbf{A}+\varepsilon\mathbf{R},\mathrm{I}_0)$, decrypt $\mathbf{x}(t)$ to get $\tilde{\mathbf{s}}(t)$ – an estimated version of $\mathbf{s}(t)$, where $\varepsilon\in(0,1)$ and $\mathbf{R}$ is a $P\times(P+Q)$ random $(1,-1)$-matrix.

For each value of ε, the second step was repeated for 100 times to get a mean value of the recovery error (measured in MAE – mean absolute error)[5]. Then, one can observe the relationship between the recovery error and the value of ε. Figure 1 shows the experimental results when the plaintexts are a digital image and a speech file, respectively.

The experimental results confirm that a mismatched key can approximately recover the plaintext. Considering that humans have a good capability of correcting errors in viewing images and listening to speech, even relatively large errors may not be able to prevent a human attacker from recognizing the plain-image or plain-speech. Thus, the value of ε may be relatively large. When $P=4$, $\mathbf{A}=[\mathbf{B},\beta\mathbf{B}]$ and $\varepsilon=0.1$, we give two examples of such recognizable plaintexts with relatively large errors in Figs. 2 and 3.

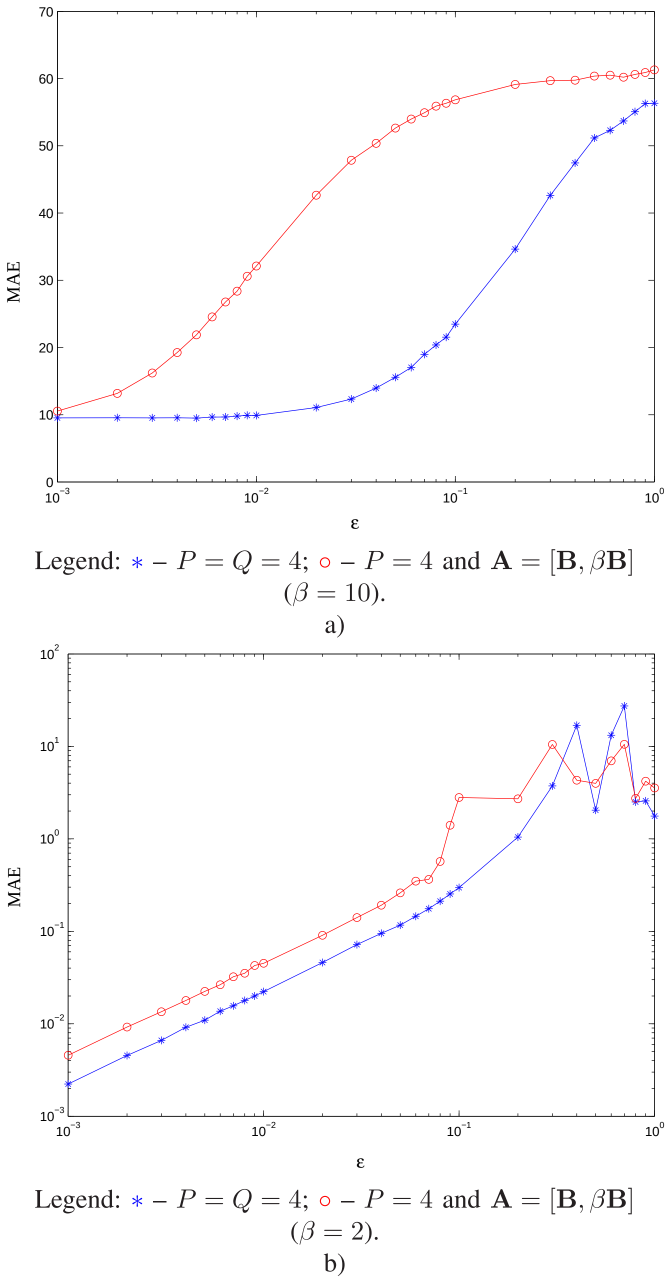
From the above experimental results, we can exhaustively search for an approximate version of $\mathbf{A}$ under the finite precision $\varepsilon=0.01\sim0.1$. Such an approximate version of $\mathbf{A}$ is then used to roughly reveal the plaintext. Since the searching complexity is $O\left(P\lceil 2/\varepsilon\rceil^{P+Q}\right)$, such an exhaustive search is feasible when P,Q is not very large[6]. When $P=2$ and

[3] The value of R is determined by the finite precision under which the cryptosystem is realized. For example, if the cryptosystem is implemented with n-bit fixed-point arithmetic, $R=2^n$; if it is implemented with IEEE floating-point arithmetic, $R\approx2^{31}$ (single-precision) or $R\approx2^{63}$ (double-precision) [39], where the sign bit of the floating-point number is always negative.

[4] In this paper, $\lceil x\rceil$ denotes the ceiling function of x, i.e., the minimal integer that is not less than x.

[5] When the plaintext is a digital image with 256 gray scales, we first calibrate each sub-image into the range $\{0,\cdots,255\}$ and then calculate the recovery error of the whole image.

[6] In [31]–[37], small values are used in all examples: $P=2$ or 4 and $Q\leq P$.



Legend: $*$ – $P = Q = 4$; ○ – $P = 4$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$ $(\beta = 10)$.

a)

Legend: $*$ – $P = Q = 4$; ○ – $P = 4$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$ $(\beta = 2)$.

b)

Fig. 1. The experimental relationship between the recovery error and the value of ε: a) the plaintext is a digital image "Lenna" (Fig. 3a); b) the plaintext is a speech file "one.wav" that corresponds to the pronunciation of the English word "one" (from Merriam-Webster Online Dictionary, http://www.m-w.com).

$\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$, we carried out a large number of experiments in the following steps:

- *Step 1*: for a randomly-generated key $(\mathbf{B}, \mathrm{I}_0)$, calculate the ciphertext $\mathbf{x}(t)$ corresponding to a plaintext $\mathbf{s}(t)$;
- *Step 2*: randomly generate a matrix $\mathbf{R}$ (each element in the interval $[-1, 1]$), and then decrypt $\mathbf{x}(t)$ with the guessed key $(\mathbf{R}, \mathrm{I}_0)$ to get $\tilde{\mathbf{s}}(t)$;
- *Step 3*: repeat *Step 2* for r rounds, output the recovered plaintext $\tilde{\mathbf{s}}^*(t)$, every segment of which corresponds to the best recovery performance in all the r rounds;
- *Step 4*: for the i-th segment of $\tilde{\mathbf{s}}^*(t)$, find the corresponding matrix $\mathbf{R}$, and extract its i-th row of its inverse $\mathbf{R}^{-1}$ to form the i-th row of $\tilde{\mathbf{B}}^{-1}$, the inverse of an estimation of the original matrix $\mathbf{B}$.

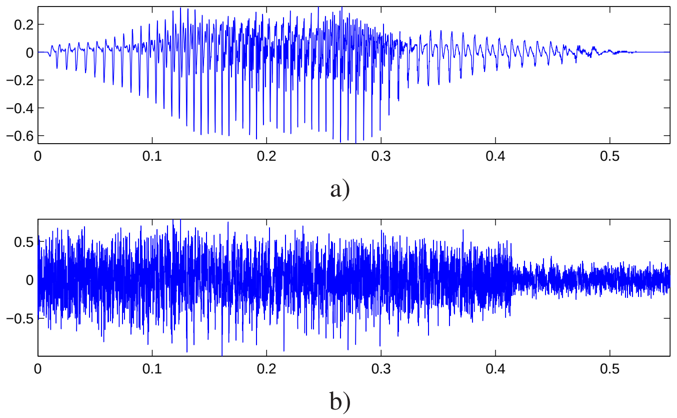


Fig. 2. An example of human capability of recognition against large noises in speech: a) the original plain-speech "one.wav"; b) the recovered speech (MAE=0.164103). For reader's verification, the recovered speech is posted online at http://www.hooklee.com/Papers/Data/BSSE/one_MAE=0.164103.wav.

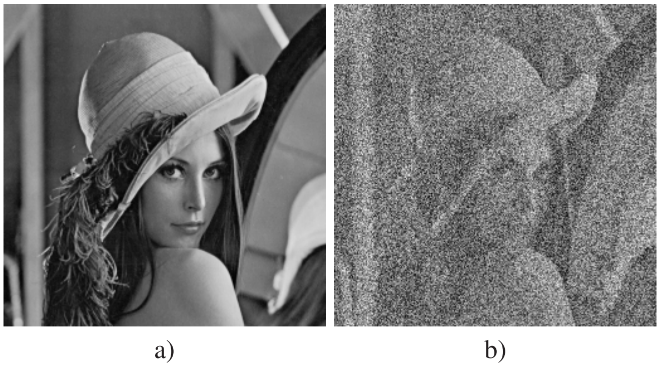


Fig. 3. An example of human capability of recognition against large noises in images: a) the original plain-image "Lenna"; b) the recovered image (MAE=47.6913).

Assuming that the target finite precision is $\varepsilon > 0$, the interval $[-1, 1]$ is divided into $n_\varepsilon = \lceil 2/\varepsilon \rceil$ sub-intervals. Without loss of generality, assume that $2/\varepsilon$ is an integer. Then, each sub-interval is of equal size. Thus, if the element in the random matrix $\mathbf{R}$ has a uniform distribution over $[-1, 1]$, the probability that $|r_{i,j} - a_{i,j}| < \varepsilon$ occurs at least one time in r rounds of experiment is $p(n_\varepsilon, r) = 1 - (1 - 1/n_\varepsilon)^r$, where $r_{i,j}$ and $a_{i,j}$ are the (i, j)-th elements of $\mathbf{R}$ and $\mathbf{A}$, respectively. One can easily deduce that $p(n_\varepsilon, r)$ is an increasing function with respect to r and

$$\begin{aligned} p(n_\varepsilon, n_\varepsilon) > \lim_{n_\varepsilon \to \infty} p(n_\varepsilon, n_\varepsilon) &= 1 - \lim_{n_\varepsilon \to \infty} (1 - 1/n_\varepsilon)^{n_\varepsilon} \\ &= 1 - e^{-1} \approx 0.6321, \end{aligned}$$

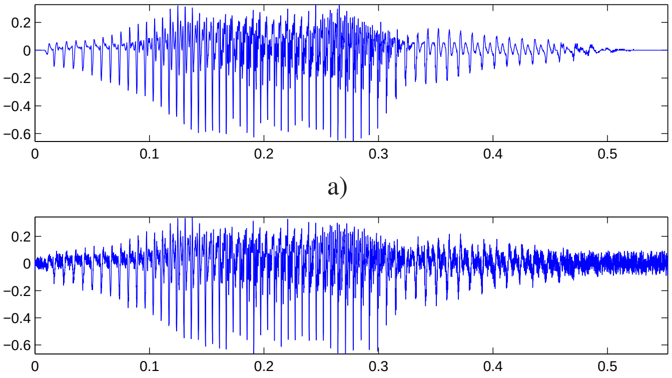
which results in that $p(n_\varepsilon, r) > 1 - e^{-1}$ when $r \geq n_\varepsilon$. In other words, with $r \geq n_\varepsilon$ experiments, it is a high-probability event that at least one $r_{i,j}$ is "equal" to $a_{i,j}$ under the finite precision ε. To get an approximate estimation of the i-th row of $\mathbf{A}$, one can see that $r = O\left(n_\varepsilon^P\right)$ rounds of experiment are needed.

Apparently, the above steps actually simulate the process of a real ciphertext-only attack that tries to reveal the plaintext and to exhaustively guess $\mathbf{B}^{-1}$ (under the assumption that I_0

is known). Note that MAE cannot be calculated to evaluate the recovery performance in a real attack, in which one does not know the plaintext. Fortunately, exploiting the large information redundancy existing in natural images and speech signals, one can resort to use some other measures to reflect the recovery performance of each segment of $\tilde{\mathbf{s}}(t)$. In our experiments, we use a measure called MANE (mean absolute neighboring error), which is defined as follows for the i-th segment of $\tilde{\mathbf{s}}(t)$:

$$\frac{1}{T-2}\sum_{t=2}^{T-1}\frac{|\tilde{s}_i(t)-\tilde{s}_i(t-1)|+|\tilde{s}_i(t)-\tilde{s}_i(t+1)|}{2}, \qquad (9)$$

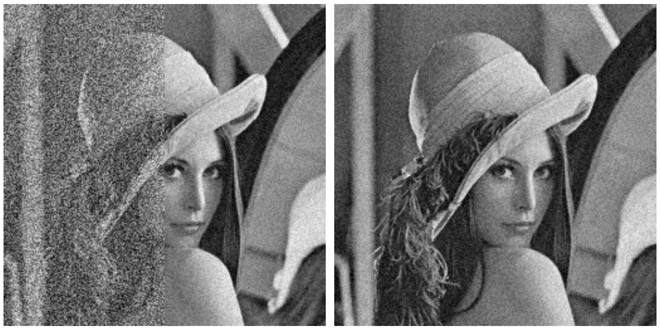
where T denotes the segment length. In Figs. 4 and 5, one recovered plain-speech signals and two recovered plain-images are shown for demonstration. One can see that $r = O(10,000)$ (or $\varepsilon \approx 0.01$) is sufficient to get a good estimation of the plaintext.



a)

b)

Fig. 4. A recovered speech in one 50,000-round experiment of exhaustively guessing $\mathbf{A}$ when $P = 2$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$: a) the original plain-speech “one.wav”; b) the recovered speech (MANE of each segment: 0.0469, 0.0521). For reader’s verification, the recovered speech is posted online at http://www.hooklee.com/Papers/Data/BSSE/one_MANE=0.0469-0.0521.wav.



a) b)

Fig. 5. Two recovered plain-images in experiments of exhaustively guessing $\mathbf{B}$ when $P = 2$ and $\mathbf{A} = [\mathbf{B}, \beta\mathbf{B}]$: a) $r = 1,000$ (MANE of each segment: 39.7491, 14.9373); b) $r = 10,000$ (MANE of each segment: 16.3888, 15.1722).

Note that for 2-D images the above 1-D MANE may be generalized to include more neighboring pixels, thus achieving a more accurate description of the recovery performance. In addition, multiple quality factors can be employed to further enhance the evaluation of the recovery performance.

3) Low Sensitivity to $\mathbf{k}(t)$: Due to the same reason as the low sensitivity to $\mathbf{A}$, one can deduce that the BSS-based encryption scheme is also insensitive to the key signal $\mathbf{k}(t)$. Given two key signals $\mathbf{k}_1(t)$ and $\mathbf{k}_2(t)$, if the maximal difference of all elements is ε, then each element of $|\mathbf{A}_k\mathbf{k}_1(t)-\mathbf{A}_k\mathbf{k}_2(t)|$ is not greater than $Q\max(|\mathbf{A}_k|)\varepsilon = Q\varepsilon$. Since $\mathbf{k}(t)$ itself is not part of the secret key, but generated from I_0, this problem does not have much negative influence on the security of the whole cryptosystem against ciphertext-only attacks.

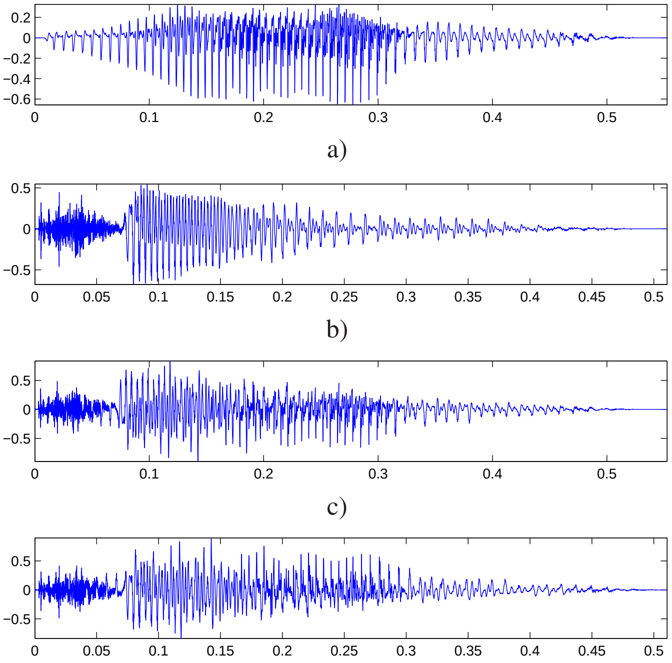
4) Differential Attack: Given two plaintexts $\mathbf{s}^{(1)}(t)$ and $\mathbf{s}^{(2)}(t)$, if they are encrypted with the same key $(\mathbf{A}, \mathrm{I}_0)$, one can get the following formula from Eq. (3):

$$\Delta_{\mathbf{x}}(t) = \mathbf{A}_s\Delta_{\mathbf{s}}(t), \qquad (10)$$

where $\Delta_{\mathbf{x}}(t) = \mathbf{x}^{(1)}(t)-\mathbf{x}^{(2)}(t)$ and $\Delta_{\mathbf{s}}(t) = \mathbf{s}^{(1)}(t)-\mathbf{s}^{(2)}(t)$. Note that $\mathbf{A}_k\mathbf{k}(t)$ disappears from the above equation. This means that from the differential viewpoint only $\mathbf{A}_s$ is the secret key, i.e., I_0 is insignificant in the key. Considering the low sensitivity of the encryption scheme to $\mathbf{A}$, under finite precision ε the key space becomes $O\left(Pn_\varepsilon^P\right)$, and so one might exhaustively search $\mathbf{A}_s$ to recover the plaintext difference as follows:

$$\Delta_{\mathbf{s}}(t) = \mathbf{A}_s^{-1}\Delta_{\mathbf{x}}(t). \qquad (11)$$

From the obtained plaintext difference, one can get a mixed view of the two interested plaintexts, in which both plaintexts may be completely recognizable by humans. Figures 6 and 7 show four plaintext differences of two speech files and two images.



a)

b)

c)

d)

Fig. 6. Differences of two plain-speech files: a) the first speech “one.wav”; b) the second speech “two.wav”; c) the difference speech signal “one”−“two”; d) the difference speech signal “two”−“one”. For readers’ verification, the two difference speech files are posted online at http://www.hooklee.com/Papers/Data/BSSE/one-two.wav and http://www.hooklee.com/Papers/Data/BSSE/two-one.wav.

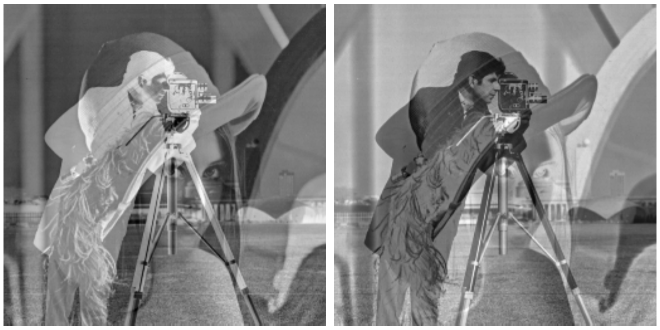


Fig. 7. Differences of two plain-images, “Lenna” and “cameraman”: a) Lenna-cameraman; b) cameraman-Lenna.

Denoting the guessed matrix by $\tilde{\mathbf{A}}_s$, one has

$$\tilde{\Delta}_{\mathbf{s}}(t) = \tilde{\mathbf{A}}_s^{-1} \Delta_{\mathbf{x}}(t) = \tilde{\mathbf{A}}_s^{-1} \mathbf{A}_s \Delta_{\mathbf{s}}(t). \qquad (12)$$

Apparently, if $\tilde{\mathbf{A}}_s \neq \mathbf{A}_s$, the obtained plaintext difference $\tilde{\Delta}_{\mathbf{s}}(t)$ will have an inter-segment mixture, which may make the recognition of the two plaintexts more difficult. Fortunately, when P is relatively small, such an inter-segment mixture may not be too severe to prevent the recognition of the two plaintexts by humans. More importantly, our experiments showed that humans are able to recognize the two plaintexts even when the mismatch between $\tilde{\mathbf{A}}_s$ and $\mathbf{A}_s$ is not very small. When $P = 2$,

$$\mathbf{A}_s = \begin{bmatrix} 0.7123 & -0.4272 \\ 0.1958 & 0.1295 \end{bmatrix}, \tilde{\mathbf{A}}_s = \begin{bmatrix} 0.5914 & 0.9527 \\ 0.5726 & 0.1437 \end{bmatrix}, \qquad (13)$$

a plain difference-image obtained in our experiments is shown in Fig. 8. One can see that both plain-images, “Lenna” and “cameraman”, can still be roughly recognized from such a heavily mixed difference-image. Another obtained plain difference-speech for “one.wav” and “two.wav”, is shown in Fig. 9, from which the two English words (“one” and “two”) are also distinguishable.

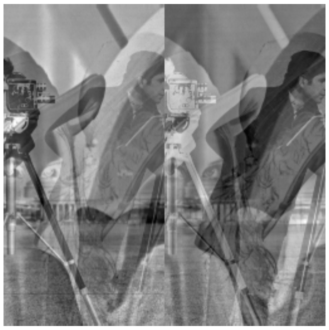


Fig. 8. One obtained plain difference-images with a badly-mismatched key when $P = 2$.

To further show the real performance of the differential attack with a badly-mismatched key, we have also carried out some experiments on the plain-images “Lenna” and “cameraman” when $P = 4$,

$$\mathbf{A}_s = \begin{bmatrix} 0.6444 & -0.2417 & -0.2687 & -0.7043 \\ -0.1770 & 0.2778 & 0.5539 & -0.7035 \\ 0.0539 & -0.6525 & -0.3157 & 0.2781 \\ 0.8404 & 0.5716 & 0.5484 & -0.8133 \end{bmatrix}$$

and

$$\tilde{\mathbf{A}}_s = \begin{bmatrix} 0.9283 & 0.8109 & 0.9567 & 0.4825 \\ 0.1668 & 0.4057 & 0.8549 & 0.3246 \\ 0.3868 & 0.7950 & 0.9863 & 0.0574 \\ 0.4860 & 0.0194 & 0.2507 & 0.2910 \end{bmatrix}.$$

One result is shown in Fig. 10.

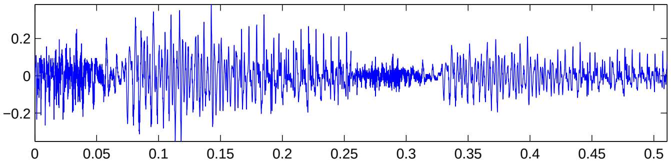


Fig. 9. One obtained plain difference-speech when $\mathbf{A}_s$ and $\tilde{\mathbf{A}}_s$ have a relatively large mismatch. For readers’ verification, this difference-speech is posted online at http://www.hooklee.com/Papers/Data/BSSE/two-one-large-mismatch.wav.

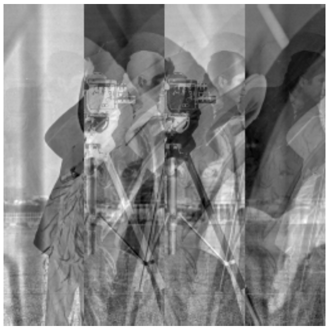


Fig. 10. One obtained plain difference-image with a badly-mismatched key when $P = 4$.

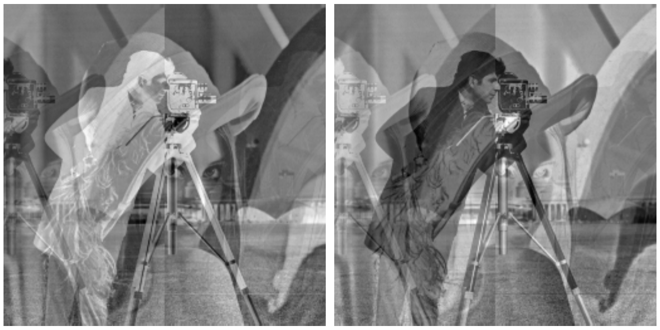


Fig. 11. A visually-optimal result obtained in 100 plain difference-images: a) the optimal difference-image; b) the negative of the optimal difference-image.

In this differential attack, some quality evaluation factors (such as MANE used in Sec. III-A.2) are not suitable to automatically determine the best result in many plain difference-signals, because each segment of an obtained difference-signal is also a natural signal with abundant information redundancy.

Instead, one has to output all obtained difference-signals, and check them with naked eyes or ears to find a perceptually-optimal result with the least inter-segment mixture. Figure 11 shows such a result in 100 plain difference-images when $P=2$ and $\mathbf{A}$ follows Eq. (13). By checking each segment separately and combining the P optimal segments together, one can further get a better result with less inter-segment mixture.

While this differential attack works well for $P=2$ as shown above, it will become infeasible when P is sufficiently large, due to the following reasons: 1) the inter-segment mixture is too severe; 2) the complexity of checking all $O\left(P\lceil 2/\varepsilon\rceil^P\right)$ difference-signals is beyond human's capability.

B. Low Sensitivity to Plaintext

Another cryptographical property required by a good cryptosystem is that the encryption should be very sensitive to plaintext, i.e., the ciphertexts of two plaintexts with a slight difference should be very different [12]. However, this property does not hold for the BSS-based encryption scheme. Given two key signals $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$, if the maximal difference of all elements is ε, then each element of $|\mathbf{A}_s\mathbf{s}_1(t)-\mathbf{A}_s\mathbf{s}_2(t)|$ is not greater than $P\max(|\mathbf{A}_s|)\varepsilon = P\varepsilon$. When the same secret key is used to encrypt two closely-correlated plaintexts, such as a plaintext and its watermarked version, this security defect means that the exposure of one plaintext leads to the revealment of both.

C. Known-Plaintext Attack

In this kind of attack, one can access to a number of plaintexts that are encrypted with the same key. From Eq. (10), with P plaintext differences, one immediately knows that the mixing matrix can be uniquely determined as follows:

$$\mathbf{A}_s = \Delta_{\mathbf{X}}(t)(\Delta_{\mathbf{S}}(t))^{-1}, \tag{14}$$

where $\Delta_{\mathbf{S}}(t)$ and $\Delta_{\mathbf{X}}(t)$ are $P\times P$ matrices, constructed row by row from the P plaintext differences and the corresponding ciphertext differences, respectively. Then, $\mathbf{A}_k\mathbf{k}(t)$ can be further solved from any plaintext and its ciphertext:

$$\mathbf{A}_k\mathbf{k}(t) = \mathbf{x}(t) - \mathbf{A}_s\mathbf{s}(t). \tag{15}$$

Now, $(\mathbf{A}_s, \mathbf{A}_k\mathbf{k}(t))$ can be used to recover other plaintexts encrypted by the same key $(\mathbf{A}, \mathrm{I}_0)$. Note that $\mathbf{A}_k\mathbf{k}(t)$ has a finite length determined by the maximal length of all known plaintexts, so $(\mathbf{A}_s, \mathbf{A}_k\mathbf{k}(t))$ can only recover plaintexts under this finite length.

When $\mathbf{A}=[\mathbf{B}, \beta\mathbf{B}]$, the key signals can also be determined:

$$\mathbf{k}(t) = \frac{\mathbf{s}(t) - \mathbf{B}^{-1}\mathbf{x}(t)}{\beta}. \tag{16}$$

If the PRNG used is not cryptographically strong (such as LFSR [12]), it may be possible to further derive the secret seed I_0, thus completely breaking the BSS-based encryption scheme.

Note that n distinct plaintexts can generate $\binom{n}{2} = n(n-1)/2$ plaintext differences. Solving the inequality $n(n-1)/2 \geq P$, one can get the number of required plaintexts to yield at least P plaintext differences:

$$n \geq \left\lceil \sqrt{P-1/4} + 1/2 \right\rceil \approx \sqrt{P}. \tag{17}$$

D. Chosen-Plaintext/Ciphertext Attack

In chosen-plaintext attack, one can freely choose a number of plaintexts and observe the corresponding ciphertexts, while in chosen-ciphertext attack, one can freely choose a number of ciphertexts and observe the corresponding plaintexts. So, in these attacks, one can choose P plaintext differences easily, which means that the above differential known-plaintext attack still works fine in the same way.

IV. Discussion

As we pointed out in the last section, the BSS-based encryption scheme is always insecure against known/chosen-plaintext attack. So, the secret key cannot be repeatedly used in any case. This means that the encryption scheme has to work like a common stream cipher, by changing the secret key for each distinct plaintext. However, in this case, $\mathbf{k}(t)$ (equivalently, the secret seed I_0) is enough to provide a high level of security, since $\mathbf{k}(t)$ satisfies the cryptographical properties in a perfectly secure one-time-a-pad cipher (see Sec. V.B of [37]). This means that the BSS process becomes excessive.

Even when one wants to add a second function against potential attacks by applying the BSS mixing, the low sensitivity of encryption/decryption to the mixing matrix $\mathbf{A}$ (recall Sec. III-A.2) makes this goal less useful. As a result, in the current encryption design, the BSS model does not play a key role in the security of the scheme. The real core of the encryption scheme actually is the embedded PRNG that generates the key signals for masking the plaintexts.

If one wants to use the BSS-based encryption scheme with a repeatedly used key, some essential modifications have to be made to enhance the security against various attacks. Following the cryptanalysis given in the last section, we suggest adopting two coutermeasures simultaneously: 1) use a sufficiently large P; 2) like the design of most modern block ciphers [12], iterate the BSS-based encryption for many rounds to improve its sensitivity to the secret key and to the plaintext. It is obvious that both countermeasures will significantly influence the encryption/decryption speed of the encryption scheme. It therefore seems doubtful if such an enhanced encryption scheme will have any advantages comparing with other multiple-round block ciphers, especially AES [13] that can be optimized to run at a very high rate on PCs [40].

Finally, it is worth mentioning that the original BSS-based encryption scheme can be used to realize **lossy** decryption, an interesting feature that may be useful in some real applications[7]. This feature means that an encryption scheme can still (perhaps roughly) recover the plaintext even when there are some errors in the ciphertexts. An typical use of this feature is that the ciphertext can be compressed with some lossy algorithms to save the required storage in local computers or

[7] Another possible scheme is a matrix-based image scrambling system proposed in [41], as pointed out in [30].

the channel bandwidth for transmission. For the BSS-based encryption scheme, the lossy decryption feature is ensured by low sensitivity of decryption to ciphertext, which is due to the same reason of the low sensitivity of encryption to plaintext (recall Sec. III-B). However, one should keep in mind that the lossy decryption feature is induced by the low sensitivity to plaintext/ciphertext, so there is a tradeoff between this feature and the security.

V. Conclusion

This paper has analyzed the security of an image/speech encryption scheme based on BSS mixing technique proposed in [31]–[37]. It has been shown that this BSS-based encryption scheme suffers from several security defects, including its vulnerability to a ciphertext-only differential attack, known/chosen-plaintext attack and chosen-ciphertext attack. It remains to see how the BSS-based technique can be further improved for constructing cryptographically strong ciphers.

References

[1] H. J. Beker and F. C. Piper, *Secure Speech Communications*. London: Academic, 1985.

[2] I. J. Kumar, "Cryptology of speech signal," in *Cryptology: System Identification and Key-Clustering*. Laguna Hills, California: Aegean Park Press, 1997, ch. 6.

[3] R. K. Nichols and P. C. Lekkas, "Speech cryptology," in *Wireless Security: Models, Threats, and Solutions*. New York: McGraw-Hill, 2002, ch. 6, pp. 253–327.

[4] 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 LLC, 2004, ch. 3, pp. 93–132.

[5] 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 LLC, 2004, ch. 4, pp. 133–167, preprint is available at http://www.hooklee.com/pub.html.

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

[7] B. Furht, E. Muharemagic, and D. Socek, *Multimedia Encryption and Watermarking*. Springer, 2005.

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

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

[10] M. G. Kuhn, "Analysis for the nagravision video scrambling method," Online document, available at http://www.cl.cam.ac.uk/~mgk25, 1998.

[11] S. Li, C. Li, G. Chen, N. G. Bourbakis, and K.-T. Lo, "A general quantitative cryptanalysis of permutation-only multimedia ciphers against plaintext attacks," accepted by Signal Processing: Image Communication, in press, doi: 10.1016/j.image.2008.01.003, an early preprint available online at http://eprint.iacr.org/2004/374, 2008.

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

[13] 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.

[14] 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.

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

[16] M. Bertilsson, E. F. Brickell, and I. Ingemarson, "Cryptanalysis of video encryption based on space-filling curves," in *Advances in Cryptology - EUROCRYPT'89 Proceedings*, ser. Lecture Notes in Computer Science, vol. 434. Berlin / Heidelberg: Berlin, 1989, pp. 403–411.

[17] 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.

[18] L. Qiao, K. Nahrstedt, and M.-C. Tam, "Is MPEG encryption by using random list instead of ZigZag order secure?" in *Proc. IEEE Int. Symposium on Consumer Electronics (ISCE'97)*, 1997, pp. 226–229.

[19] 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.

[20] 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.

[21] A. M. Youssef and S. E. Tavares, "Comments on the security of fast encryption algorithm for multimedia (FEA-M)," *IEEE Trans. Consumer Electron.*, vol. 49, no. 1, pp. 168–170, 2003.

[22] S. Li and K.-T. Lo, "Security problems with improper implementations of improved FEA-M," *J. Systems and Software*, vol. 80, no. 5, pp. 791–794, 2007.

[23] 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.

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

[25] 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 Proceedings, Part III*, ser. Lecture Notes in Computer Science, vol. 3333. Berlin / Heidelberg: Springer-Verlag, 2004, pp. 418–425.

[26] 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.

[27] 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.

[28] 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.

[29] S. Li, C. Li, G. Chen, and K.-T. Lo, "Cryptanalysis of RCES/RSES image encryption scheme," accepted by *J. Systems and Software*, in press, doi: 10.1016/j.jss.2007.07.037, preprint available online at http://eprint.iacr.org/2004/376, 2008.

[30] S. Li, C. Li, K.-T. Lo, and G. Chen, "Cryptanalysis of an image scrambling scheme without bandwidth expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 338–349, 2008.

[31] Q.-H. Lin and F.-L. Yin, "Blind source separation applied to image cryptosystems with dual encryption," *Electronics Letters*, vol. 38, no. 19, pp. 1092–1094, September 2002.

[32] Q. Lin and F. Yin, "Image cryptosystems based on blind source separation," in *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing (ICNNSP'2003)*, vol. 2. IEEE, 2003, pp. 1366–1369.

[33] Q.-H. Lin, F.-L. Yin, and Y.-R. Zheng, "Secure image communication using blind source separation," in *Proceedings of the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication (CASSET'2004)*, vol. 1. IEEE, 2004, pp. 261–264.

[34] Q. Lin, F. Yin, and H. Liang, "Blind source separation-based encryption of images and speeches," in *Advances in Neural Networks - ISNN 2005 Proceedings, Part II*, ser. Lecture Notes in Computer Science, vol. 3497. Berlin / Heidelberg: Springer-Verlag, 2005, pp. 544–549.

[35] Q.-H. Lin, F.-L. Yin, and H.-L. Liang, "A fast decryption algorithm for BSS-based image encryption," in *Advances in Neural Networks - ISNN 2006 Proceedings, Part III*, ser. Lecture Notes in Computer Science, vol. 3973. Berlin / Heidelberg: Springer-Verlag, 2006, pp. 318–325.

[36] Q.-H. Lin, F.-L. Yin, T.-M. Mei, and H. Liang, "A speech encryption algorithm based on blind source separation," in *Proceedings of the 2004 International Conference on Communications, Circuits and Systems (ICCCAS'2004)*, vol. 2. IEEE, 2004, pp. 1013–1017.

[37] ——, "A blind source separation based method for speech encryption," *IEEE Trans. Circuits Syst. I*, vol. 53, no. 6, pp. 1320–1328, June 2006.

[38] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[39] IEEE Computer Society, "IEEE standard for binary floating-point arithmetic," ANSI/IEEE Std. 754-1985, 1985.

[40] B. Gladman, "AES and combined encryption/authentication modes," online document, available at http://fp.gladman.plus.com/AES/index.htm, 2006.

[41] 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.