Available online at www.sciencedirect.com
ScienceDirect
Signal Processing: *Image Communication* 23 (2008) 179–193
SIGNAL PROCESSING: IMAGE COMMUNICATION
www.elsevier.com/locate/image

# AACS-compatible multimedia joint encryption and fingerprinting: Security issues and some solutions

Shih-Wei Sun[a,b], Chun-Shien Lu[a,*], Pao-Chi Chang[b,c]

[a]*Institute of Information Science, Academia Sinica, Taipei 115, Taiwan, ROC*
[b]*Department of Electrical Engineering, National Central University, Chung-Li 320, Taiwan, ROC*
[c]*Department of Communication Engineering, National Central University, Chung-Li 320, Taiwan, ROC*

Received 15 February 2007; received in revised form 7 January 2008; accepted 10 January 2008

**Abstract**

In this paper, a new multimedia joint encryption and fingerprinting (JEF) scheme embedded into the advanced access content system (AACS) is proposed for multimedia transmission over networks. AACS is selected because it has been jointly developed by many famous companies and has been considered as the leading technology in content access control and multimedia distribution. In this framework, many attack points exist and can be exploited to defeat it. Furthermore, multiple attack points can be combined to form multi-point collusion attacks, which also endanger the proposed system. In this paper, we address the security concerns toward AACS-compatible JEF system in its entirety and propose solutions to cope with some security threats. The contributions of this paper include: (i) applying multimedia encryption at different points to resist some attacks points; (ii) proposing rewritable fingerprint embedding (RFE) to deal with some multi-point collusion attacks; (iii) designing a perceptual security spectrum metric (PSSM) to evaluate the degree of security when multiple encryptions are applied. The feasibility of the proposed AACS-compatible JEF method is further demonstrated through simulation results.


*Keywords:* AACS; Collusion attack; (Selective) multimedia encryption; Multimedia fingerprinting; Perceptual security; Rewritable fingerprint embedding

## 1. Introduction

### 1.1. Background

With the development of multimedia technologies and network popularity, the exchange and distribution of multimedia data have become common in daily life. Thus, multimedia protection technologies are indispensable to secure multimedia distribution. Among them, multimedia encryption plays the role of the first line defense [14,17,21,23,24,27]. However, when encrypted multimedia is decrypted, multimedia encryption loses the content protection capability, which is the inherent limitation of encryption techniques. As a result, the study of multimedia watermarking is emergent and plays the role of the second line defense [5,10,7,18,23] by providing passive protection.

One promising application provided by passive protection is traitor tracing. As multimedia sharing is easy to achieve, illegally re-distributed data may

*Corresponding author. Tel.: +886 2 2788 3799x1513.
*E-mail address:* lcs@iis.sinica.edu.tw (C.-S. Lu).

0923-5965/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2008.01.001

be pirated by more than one traitor, which forms a group of traitors, called colluders. This kind of malicious attack is called a collusion attack. The aim of a collusion attack is to reduce the suspicion of colluders, even accusing other innocent users, such that the content owner fails to capture the real colluder with higher probability.

Multimedia fingerprinting techniques [4,9,20,22,25] are developed to deal with collusion attacks. Before transmitting multimedia content from the content owner to the legal users, a multimedia fingerprint should have already been embedded into the multimedia data. The characteristic of a multimedia fingerprint is similar to that of a human being's fingerprint in that both possess uniqueness and singularity. For different users, the corresponding different fingerprints would be embedded. Once a user illegally re-distributes the received multimedia data, the originally embedded multimedia fingerprint can be detected to reveal the traitor. On the other hand, if a group of users perform a collusion attack, some multimedia fingerprints are expected to be detected from the illegally re-distributed copy such that a list of possible colluders are captured to achieve the goal of traitor tracing.

Usually, traitor tracing [22] is expected to:

1. catch one: the target is to maximize the probability of catching at least one colluder, meanwhile, minimizing the probability of catching innocent users;
2. catch many: the goal is to increase the probability of catching more actual colluders at the expense of probably catching innocent users;
3. catch all: in this situation, multimedia fingerprinting is designed to maximize the probability of catching all actual colluders under the constraint that the numbers of innocent users caught is limited.

### 1.2. Related work about joint multimedia encryption and fingerprinting

In this paper, we focus on the dual targets of multimedia content access and traitor tracing. We also investigate the resistance to collusion attacks at a single attack point or multiple attacks points. It is worth noting that resistance to multi-point collusion is relatively unexplored in the literature. The existing joint multimedia encryption and fingerprinting technology, divided into three categories [9], is briefly described as follows.

(a) Transmitter-side encryption and fingerprint embedding: A multimedia plaintext is separately embedded with a user's fingerprint and then encrypted with a global key to form a multimedia ciphertext. However, this scenario incurs some disadvantages: (1) Inefficient bandwidth utilization—since multimedia fingerprint embedding is done at the transmission side, repeat requests of the same copy will waste bandwidth. (2) Insecure encryption—since a single global encryption key is used, if a malicious user eavesdrops on another user's data, then the multimedia plaintext belonging to that user can be obtained. The techniques in [4,20] belong to this category.

(b) Transmitter-side encryption and receiver-side fingerprint embedding: Fingerprint embedding at the receiver side was first proposed in [11] for digital TV. In [3,7], the concept of [11] was applied to digital rights management (DRM) in digital cinema. At the transmitter side, only one global key-based encryption is necessary. This kind of design can save a lot of computation time and bandwidth usage. In this scenario, the receiver acts like a super node in a network, not merely the user end. Thus, multimedia data can be sent to different users via the receiver (super node) for multicasting. At the receiver side, the received multimedia ciphertext can be decrypted according to the global key. Meanwhile, the multimedia fingerprint must be embedded into the multimedia data to generate the fingerprinted multimedia data for each user. A sealed set-top box is necessary in this scenario for joint multimedia decryption and fingerprint embedding. However, the sealed set-top box is still an open problem because the multimedia plaintext is possibly revealed by the set-top box. In addition, if the transmission has a real-time requirement, the total load of decryption and fingerprint embedding gathered at the receiver side (super node) will increase its computational complexity.

(c) Joint fingerprinting and decryption: In order to reduce system complexity and achieve the real-time requirement, Kundur and Karthik [9] proposed a joint fingerprinting and decryption (JFD) method. The idea behind JFD is that the

multimedia ciphertext is partially decrypted such that the un-decrypted parts imitate multimedia fingerprint embedding. This kind of method is conceptually promising, achieving partial multimedia decryption and multimedia fingerprint embedding at the same time. However, the un-decrypted content must satisfy the following two conflicting requirements. On the one hand, the un-decrypted parts should not affect the whole transparency of fingerprinted multimedia data. On the other hand, the un-decrypted parts should preserve meaningful encryption, i.e., the encrypted parts can intrinsically hide their original content.

In this paper, we will present a new joint encryption and fingerprinting (JEF) scheme, which can be incorporated into the advanced access content system (AACS). Most importantly, this method does not encounter the same problems as the above three types of methods. In the next subsection, AACS will be briefly described.

### 1.3. Advanced access content system

The AACS [1] was jointly proposed by many famous companies, including IBM, Intel, Microsoft, Panasonic (Matsushita Electric), Sony, Toshiba, Walt Disney Company, and Warner Bros, for multimedia publishing over HD-DVD and blue-ray disk. AACS is a specification for managing the stored content for the prerecorded optical media (e.g., HD-DVD or blue-ray disk) for consumer usage over PCs and electronic devices. AACS presents a distribution model to improve the functionality and interactivity among the consumers, content providers, aggregators, and device manufacturers. For example, AACS is designed to support the ability to make recordings of content as authorized. In addition, the proven cryptographic methods make AACS flexible enough to interoperate with content protection technologies to enable consumers to access the licensed and protected copies of multimedia, while preventing the unauthorized reproduction and distribution for the prerecorded optical media. Basically, AACS contains four major parts: content owner, licensed replicator, licensing entity for key management, and licensed player at the user end.

### 1.4. Contributions of this paper

AACS is indeed a suitable framework for multimedia content protection, but is currently designed for HD-DVD and blue-ray disk. If the concept of AACS is applied to networked multimedia (e.g., video) transmission, the raised problems need to be dealt with. Therefore, based on the four major elements of AACS, a new multimedia content protection framework for multimedia transmission, as shown in Fig. 1, is proposed in this paper. Specifically, in order to achieve secure transmission of multimedia content, a JEF method is proposed and embedded into AACS for content access control and traitor tracing. The advantage of the proposed AACS-compatible JEF method is that it does not suffer the difficulties of the existing joint multimedia encryption and fingerprinting methods, as have been discussed in Section 1.2.

In this AACS-compatible JEF method, many attack points exist that can be exploited to defeat it. Furthermore, multiple attack points can be combined to form multi-point collusion attacks, which also endanger the proposed system. In this paper, we examine the security issues of the AACS-compatible JEF system in its entirety and propose solutions to cope with some security threats. The contributions of this paper include: (i) applying multimedia encryption at different points to resist some attacks points; (ii) proposing rewritable

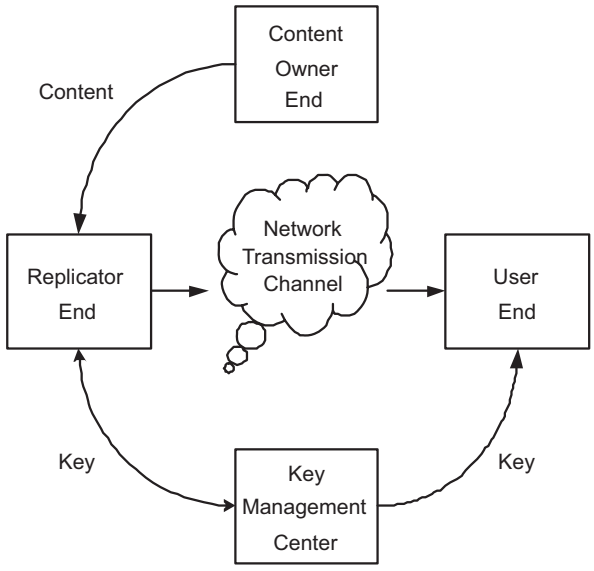


Fig. 1. The conceptual framework for the proposed AACS-based multimedia transmission method.

fingerprint embedding (RFE) to deal with some multi-point collusion attacks; (iii) designing a perceptual security spectrum metric (PSSM) to evaluate the degree of security when multiple encryptions are applied.

The remainder of this paper is organized as follows. In Section 2, the security threats to the proposed method at different single attack points and multiple attack points are first described, and then a description of the system is provided. In Section 3, the proposed RFE method used to cope with the multi-point collusion attacks is described. In Section 4, the perceptual security metric defined in the spectrum domain is presented to measure the security gain of our method because our method will apply more than one encryption. Experimental results are given in Section 5 and conclusions are drawn in Section 6.

## 2. The framework of proposed joint multimedia encryption and fingerprinting method

Based on AACS, the proposed joint multimedia encryption and fingerprinting method contains four major parts: content owner, replicator end, key management center, and user end, as shown in Figs. 2 and 3.

### 2.1. Possible attack points and multi-point collusion attacks, and their countermeasure

Like other security-related systems, there exist some security threats to the proposed framework. In this section, the possible attack points and collusion attacks, as shown in Figs. 2 and 3, will be addressed. Furthermore, solutions to some of these attacks will be described. Basically, we propose to use multimedia encryption to cope with some of the attack points (which will be described in Section 2.2), and propose RFE (which will be discussed in Section 3) to cope with some multi-point collusion attacks.

#### 2.1.1. Single attack points

A single attack point means a basic security threat that is able to endanger the system. The single attack points to be discussed here are labeled as A, B, ..., and H, as shown in Figs. 2 and 3.

(A) Original copy attack: Original copy attack states that the original multimedia data are

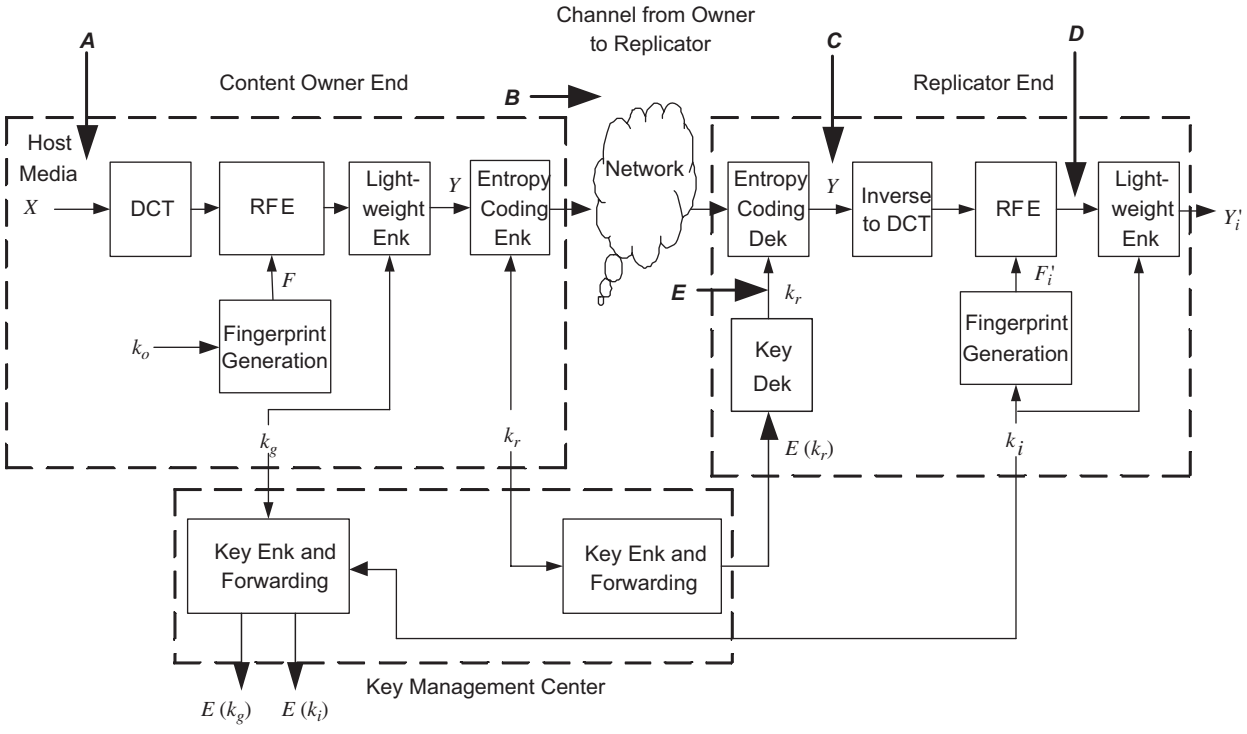


Fig. 2. JEF based on AACS and possible attack points in: (1) content owner end; (2) replicator end; and (3) key management center.

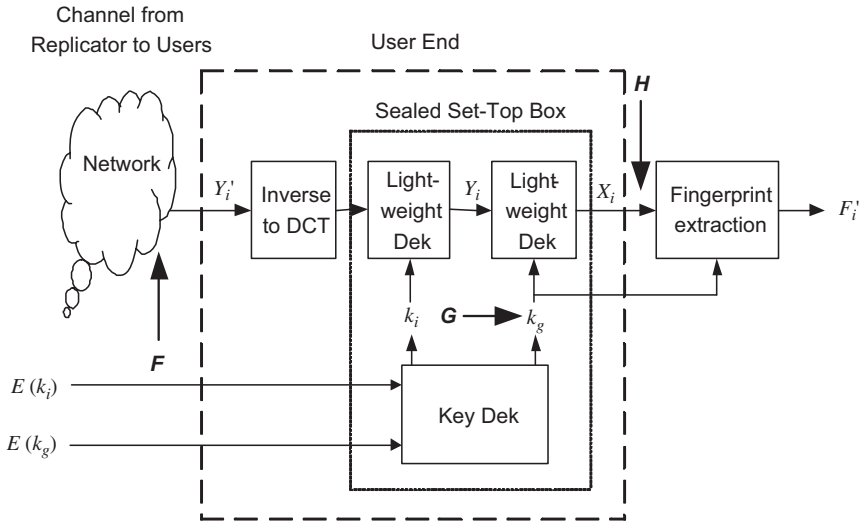


Fig. 3. JEF based on AACS and possible attack points in user end.

illegally re-distributed before any protection operation is imposed, as shown in attack point A of Fig. 2. In general, the original version only belongs to the multimedia content owner. Therefore, the original multimedia data should not be exposed to the risk of being pirated.

(B) Snooping attack: The attacker can eavesdrop on the network link between the multimedia content owner end and replicator end, as shown in the attack point B of Fig. 2. If the multimedia is transmitted in the plaintext form, the eavesdropped multimedia data can be illegally re-distributed without any effort. Therefore, for secure multimedia transmission, the bitstream should be encrypted before transmission.

Two different encryptions performed at different bitstream domains are proposed in this paper. If $k_g$ can be obtained (at the attack point G), and is used at the attack point B, then the multimedia plaintext still cannot be obtained due to the protection of the second encryption based on $k_r$.

(C) Content owner fingerprinted replicator back-end attack: The back-end attack here means that the administrator at the replicator end reveals the multimedia data after decryption using $k_r$. In the proposed framework, due to the protection of the first encryption based on the global key $k_g$, multimedia plaintext still cannot be generated at the replicator end.

(D) User fingerprinted replicator back-end attack: The attacker plays the role of the system administrator at the replicator end and illegally distributes the fingerprinted copy generated using $k_i$. At this attack point, multimedia plaintext still cannot be revealed because the multimedia is still in the encryption domain according to $k_g$.

(E) Replicator key back-end attack: When the second encryption key $k_r$ is accessible at the replicator end, the attacker (or replicator) can use it and collude with other attackers at different attack points. This shall be further discussed in the next subsection.

(F) Encrypted copy attack: The encrypted multimedia data transmitted from the replicator end to the user end might be eavesdropped on by the attacker over the network link for collusion. Because the transmitted multimedia data are kept in the encrypted form, the collusion attack for encrypted multimedia data cannot obtain the multimedia plaintext.

(G) Set-top box attack: If the set-top box is not safe enough such that the keys $k_i$ and $k_g$ can be derived from the set-top box, the security of the whole system is suddenly degraded. The global key $k_g$ may be sent to the replicator end for obtaining the multimedia plaintext (at the attack point B).

(H) Decrypted copy attack: When the decrypted multimedia is transformed to the plaintext form, it will be collected for use in a collusion

attack. In particular, the collusion attack at this attack point has been widely discussed [4,9,20,22,25].

#### 2.1.2. Multi-point collusion attacks

A multi-point collusion attack is a combination of more than one single attack point. In this paper, the most intuitive collusion attacks generated from different attack points, producing meaningful plaintext of multimedia content, are addressed and possible solutions are presented. The combinations including point A are not discussed here because it is assumed that the owner should undoubtedly preserve his/her multimedia plaintext without being leaked. In addition, the attack point H is also not included for multi-point collusion because it serves as the conventional collusion point that will be separately discussed. The multi-point collusion attacks considered here are described as follows.

(C) + (G) collusion attack: The global encryption key $k_g$, derived and revealed from the set-top box, is sent to the replicator end for obtaining the multimedia plaintext. At the content owner end, the multimedia is encrypted twice by using the encryption keys $k_g$ and $k_r$, respectively. Since the replicator can obtain $k_r$ legally, if the global key $k_g$ is illegally obtained from the user end and colluded with $k_r$, then the multimedia can be completely decrypted to the plaintext form. In the proposed JEF scheme, if (C) + (G) collusion attack happens, then the rewritable fingerprint embedded at the content owner end still can be detected. Therefore, the replicator end can be recognized as being involved in the collusion attack, such that the catch one requirement is satisfied.

(D) + (G) collusion attack: The attacker plays the role of the system administrator at the replicator end (at point D) to illegally distribute the fingerprinted copy generated using $k_i$ and collude with the key $k_g$ obtained from point G. Once this attack happens, an innocent user can be framed as the colluder because the user fingerprint $k_i$ can be detected. This kind of malicious attack still cannot be avoided. We also note that this situation is still an open problem in biometric identification as well [13].

(C) + (H) collusion attack: The multimedia data without user fingerprint embedded can be obtained from C at the replicator end. On the other end, the decrypted multimedia data can be obtained from H at the user end. When both are available, the copy-and-paste [8,15,16,18] attack can be implemented. The decrypted part at H can be replaced with that at C to generate a decrypted copy with fake user fingerprint embedded. In this paper, RFE performed at the content owner end is used to deal with this type of collusion attack. Once a (C) + (H) collusion attack happens, the fingerprint embedded at the content owner end can still be detected. Therefore, the replicator end can be recognized as being involved in the collusion attack, such that the catch one requirement is satisfied.

(B) + (E) + (G) collusion attack: The attacker can eavesdrop on the multimedia bitstreams from point B, and collude with $k_r$ from the replicator end at point E and $k_g$ from the user end at point G to successfully decrypt the eavesdropped multimedia bitstream. Thus, the decrypted multimedia bitstream without user fingerprint embedded is obtained. However, the RFE performed at the content owner end according to $k_o$ still can be detected. Based on this, although not all colluders participating in this collusion attack at the attack points B, E, and G can be caught, the replicator end at point G can be recognized as being involved in the collusion attack such that the catch one requirement is satisfied.

### 2.2. System overview

At the content owner end, as shown in Fig. 2, the content owner end plays the role of multimedia content provider (e.g., Walt Disney or Warner Bros). Let either the images or videos be the host media. Before transmission, multimedia data must be compressed in order to save bandwidth. Since the conventional discrete cosine transform (DCT)-based image/video compression standards (e.g., JPEG, MPEG, H.26x), perform compression in DCT domain, the proposed JEF method is also conducted in DCT domain. In order to embed fingerprints, the widely applied digital watermarking technique, spread spectrum (SS) watermarking [5], is adopted. In addition, a novel concept of fingerprint embedding called "RFE" is proposed for dealing with multi-point collusion attacks. RFE at the owner end aims at embedding the rewritable multimedia fingerprints according to the content owner key $k_o$ for owner identification. The design of RFE will be later described in Section 3. RFE can be currently thought of as a general digital watermark embedding scheme designed to embed the content owner key ($k_o$)-based fingerprint $F$ into the

multimedia to form the content owner fingerprinted multimedia copy $Y$. In order to encrypt the multimedia data effectively and fast, a light-weight encryption scheme aimed at encrypting the AC signs of DCT coefficients according to the global key $k_g$ is adopted. In addition, $k_r$ is applied for encryption in the entropy coding domain. The function of double-encryption is mainly used for secure transmission at different attack points. If $k_g$ can be obtained at the attack point G and is used at the attack point B, then the multimedia plaintext still cannot be obtained due to the protection of the second encryption based on $k_r$.

At the multimedia replicator end, as shown in Fig. 2, the received data should be decrypted in the entropy coding domain by using the replicator end key, $k_r$. After transforming back to the DCT domain, the multimedia fingerprint $F_i'$, generated by the user identification key $k_i$, will be embedded into the multimedia data according to RFE for user identification. Meanwhile, the multimedia data are light-weight encrypted again according to the user identification key $k_i$ before transmission out of the replicator end. Therefore, each user will obtain a different version $Y_i'$ of multimedia data.

The encryption keys, $k_g$, $k_r$, and $k_i$, will be sent to the key management center for storage and distribution. The key management center is designed to manage the keys collected from different ends. In addition, the key management center here should be highly trustworthy in that he/she will not wrongly transmit or reveal keys. These keys should be encrypted before transmission and decrypted after being received. Finally, at the user end, the received encrypted keys $E(k_g)$ and $E(k_i)$ can be decrypted in a sealed set-top box to decrypt the received multimedia stream.

Finally, if the multimedia plaintext $X_i$ is, subsequently, illegally re-distributed at the user end, the multimedia fingerprint $F_i'$ can be extracted from the revealed multimedia data to achieve the goal of traitor tracing.

## 3. Rewritable fingerprint embedding

The goal of RFE is to deal with the collusion attacks discussed in Section 2.1.2. RFE should embed fingerprints $F$ and $F_i'$ generated by $k_o$ and $k_i$, respectively, both at the content owner end and replicator end. The RFE for content owner fingerprint $F$ at the content owner end can be overwritten by the user multimedia fingerprint $F_i'$ at the replicator end. In this study, robustness is not the major concern. Therefore, we derive analytic bounds of the embedded signals to achieve the highest transparency.

Based on the SS watermarking technique [5], fingerprint embedding is accomplished by

$$y_b = x_b(1 + \alpha \cdot f_b), \tag{1}$$

where $y_b$ is the $b$th stego data of $Y$, $x_b$ is the $b$th cover data of $X$, $\alpha$ is the scaling factor of fingerprint embedding at the content owner end, and $f_b$ is the $b$th fingerprint bit of $F$. In this paper, the embedded fingerprint is a bipolar sequence.

Since both fingerprints, i.e., content owner fingerprint and user fingerprint, are sequentially embedded at the same positions, there are four states describing the change of embedded fingerprint bits, as shown in Fig. 4. In Fig. 4, $f_b'$ denotes the $b$th replicator end fingerprint bit. As a result, the scaling factors of user fingerprint embedding are denoted as $\alpha'_{(+1,-1)}$, $\alpha'_{(+1,+1)}$, $\alpha'_{(-1,+1)}$, and $\alpha'_{(-1,-1)}$ at the replicator end. In the following, we will describe how these scaling factors can be defined to satisfy robust fingerprint extraction in a non-blind watermarking scenario, which is considered reasonable in multimedia fingerprinting [20,22].

Let us first consider the case of $\alpha'_{(+1,-1)}$, i.e., $f_b = +1$ and $f_b' = -1$. At the replicator end, the user fingerprint embedding, similar to Eq. (1), is defined as

$$y_b' = y_b(1 + \alpha'_{(+1,-1)} \cdot f_b'), \tag{2}$$

where $y_b'$ is the $b$th stego data corresponding to the $b$th cover data $y_b$. By substituting $f_b = +1$, $f_b' = -1$, and Eq. (1) into Eq. (2), one gets

$$y_b' = x_b(1 + \alpha - \alpha'_{(+1,-1)} - \alpha \cdot \alpha'_{(+1,-1)}). \tag{3}$$

If $f_b' = -1$ is expected to be successfully extracted under non-blind detection, then $y_b' < x_b$ is required to be achieved. As a result, Eq. (3) can be rewritten as

$$x_b > x_b(1 + \alpha - \alpha'_{(+1,-1)} - \alpha \cdot \alpha'_{(+1,-1)}). \tag{4}$$

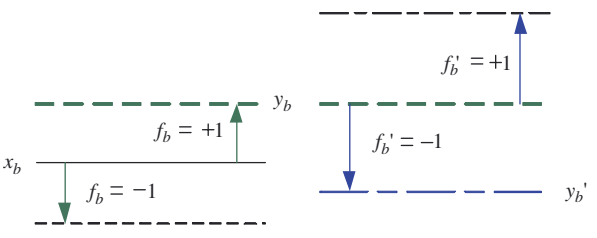


Fig. 4. An example of rewritable fingerprint embedding (RFE).

One can further derive to obtain

$$\alpha'_{(+1,-1)} > \frac{\alpha}{(1+\alpha)}. \tag{5}$$

Similar derivations can be derived for the remaining three cases of scaling factors as

$$\alpha'_{(-1,+1)} > \frac{\alpha}{(1-\alpha)} \quad \text{for } f_b = -1,\ f'_b = +1, \tag{6}$$

$$\alpha'_{(-1,-1)} > \frac{\alpha}{(\alpha-1)} \quad \text{for } f_b = -1,\ f'_b = -1, \tag{7}$$

and

$$\alpha'_{(+1,+1)} > \frac{-\alpha}{(1+\alpha)} \quad \text{for } f_b = +1,\ f'_b = +1. \tag{8}$$

However, the prior knowledge of the fingerprint state change is unknown either at the content owner end or at the replicator end. On the contrary, a global parameter of $\alpha'$ should be determined and sent to the replicator end for user fingerprint embedding. According to Eqs. (5)–(8), the lower bound of the scaling factor of embedding at the replicator end, $\alpha'$, is defined as

$$\begin{aligned}\alpha' &> \max\{\alpha'_{(+1,-1)}, \alpha'_{(-1,+1)}, \alpha'_{(-1,-1)}, \alpha'_{(+1,+1)}\} \\ &= \max\left\{\frac{\alpha}{(1+\alpha)}, \frac{\alpha}{(1-\alpha)}, \frac{\alpha}{(\alpha-1)}, \frac{-\alpha}{(1+\alpha)}\right\}.\end{aligned} \tag{9}$$

Since $0<\alpha'<1$ and $0<\alpha<1$ hold, one can derive

$$\alpha' > \frac{\alpha}{(1-\alpha)}. \tag{10}$$

The derived relationship between the two embedding factors, $\alpha$ and $\alpha'$, will be verified in the experimental results.

## 4. Spectrum perceptual security metric

In this section, a PSSM for measuring the perceptual security of the JEF scheme in the spectrum domain is proposed. For most multimedia security systems, the perceptual security is evaluated in the spatial domain [5,12,21,24,26] because human eyes perceive the image or video in the spatial domain. However, transmission or processing for securing multimedia data generally operates in the transform (spectrum) domain. If perceptual security should be measured, inverse transform back to the spatial domain is necessary but increases computational overhead [19] for JPEG or MPEG compressed multimedia. Therefore, developing a perceptual security metric in the spectrum domain is necessary.

Let $x(t)$ be the original multimedia, $y(t)$ be the encrypted format, $X(f)$ be the original spectrum, and $Y(f)$ be the encrypted spectrum. The spectrum difference $\mathrm{diff}(f) = X(f) - Y(f)$ represents the distortion in spectrum domain caused by encryption. In the proposed JEF scheme, Shi and Bhargava's method [14] is applied for light-encryption, where only the selected signs in DCT domain are modified. As a result, the spectrum energies of $X(f)$ and $Y(f)$ can be, respectively, represented as

$$\mathrm{Eng}(X(f)) = \int X^2(f) \cdot \mathrm{d}f \tag{11}$$

and

$$\mathrm{Eng}(Y(f)) = \int Y^2(f) \cdot \mathrm{d}f, \tag{12}$$

where $f$ is the frequency subband. For sign modification-based encryption, the maximum spectrum difference at the subband $f$ is

$$\begin{aligned}\mathrm{diff}(f) &= X(f) - Y(f) = X(f) - (-X(f)) \\ &= 2 \cdot X(f) = 2 \cdot |X(f)| \cdot (\mathrm{sign}(X(f))).\end{aligned} \tag{13}$$

The smaller $|X(f)|$ is, the less distortion is found in spectrum subband $f$. Finally, the spectrum security metric can be defined as

$$\mathrm{Sec}_\phi = \int W_f \cdot \mathrm{diff}(f), \tag{14}$$

which represents the degree of security at the point $\phi$. In Eq. (14), $W_f$ is the weighting factor for the difference spectrum subband $f$. It is determined according to the reciprocals of the luminance mask derived in [2], called just noticeable difference (JND), based on the human visual system.

Thus, the gain of PSSM defined between any two points, $\phi_1$ and $\phi_2$, is

$$\mathrm{GSec} = \frac{\mathrm{Sec}_{\phi_1}}{\mathrm{Sec}_{\phi_2}} \begin{cases} = 1 & \text{no security gain,} \\ > 1 & \text{security increasing,} \\ < 1 & \text{security decreasing.} \end{cases} \tag{15}$$

From different points of a multimedia security system, the gain of GSec can reveal whether a certain stage of processing increases the security or not.

## 5. Experimental results

In the experiments, several standard images, including Baboon, Lena, and Pepper, with size of $512 \times 512$ were used for JEF. AES [6] is selected for encryption with the encryption unit of 128 bits. In

order to select at least 128 signs of DCT coefficients for blockwise-encryption, an image was divided into blocks of size $16 \times 16$. In the experiments, the 128 largest DCT AC coefficients in a $16 \times 16$ block were selected for encryption. The size of fingerprints embedded using $k_o$ and $k_i$ was 64 bits. There was 1 fingerprint bit embedded in the (1,2)th subband of a DCT block.

In the following, we will show (A) the persistent encryption in our system; (B) the verification of the fingerprint embedding factors; (C) the PSSM measurement results; (D) the resistance to the collusion attack at the single attack point H; and (E) the resistance to the multi-point collusion attacks.

### 5.1. Persistent encryption

The encryption results generated at the different stages using different keys are shown in Fig. 5. In Figs. 5(a)–(c), the visual qualities of the encrypted images show that the transmitted images in the proposed JEF scheme are always kept in the encryption domain at the different points of the system. Fig. 5(a) shows the image that has been decrypted at the replicator end using $k_r$ but is still encrypted using $k_g$. Fig. 5(b) shows the result with the user fingerprint $F_i'$ embedded and another encryption based on $k_i$ applied. The images in Figs. 5(a) and (b) show that, at the replicator end, the multimedia is restricted to being processed in the encryption domain. Fig. 5(c) shows the result that is processed in the sealed set-top-box, as shown in Fig. 3. In Fig. 5(d), the decrypted multimedia with user fingerprint $F_i'$ embedded shows that the visual quality is acceptable. The PSNR between Fig. 5(d) and its corresponding original image is 41.78 dB. The above results indicate that our method can achieve secure multimedia transmission because the transmitted images are persistently kept in the encryption domain at the different points of the system.

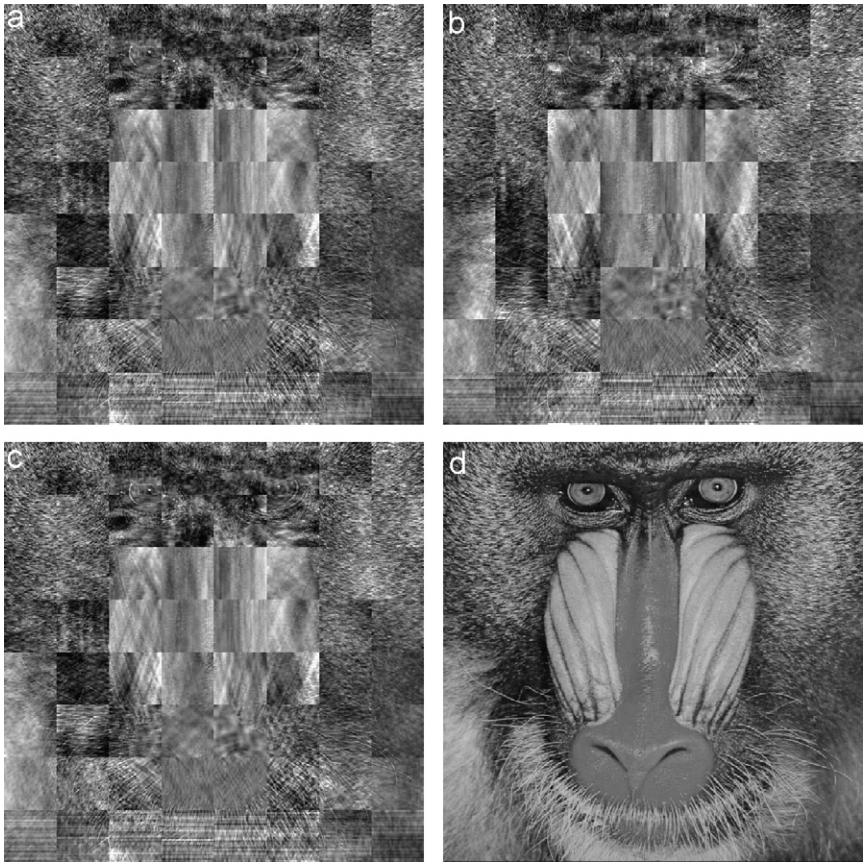


Fig. 5. Visual quality comparison of Baboon: (a) the image encrypted by $k_g$; (b) the image encrypted by $k_g + k_i$; (c) the image decrypted by $k_i$; and (d) the image decrypted by $k_i + k_g$ (PSNR = 41.78 dB).

### 5.2. Verification of fingerprint embedding factors for RFE

Simulation results are shown to verify the derived fingerprint embedding factors for the RFE method. In the simulation, the length of an original signal is set to 4 and its signal strength is set to $[6, 7, 4, 2]$. Let the rewritable fingerprint sequence, $f$, at the content owner end be $[+1, -1, -1, +1]$ and let the user fingerprint sequence, $f'$, at the replicator end be $[-1, +1, -1, +1]$. The pairs of elements in $f$ and $f'$, $(f_b, f'_b)$, constitute a set of fingerprint state changes: $\{(+1, -1), (-1, +1), (-1, -1), (+1, +1)\}$. Here, $\alpha$ is set to 0.1 and $\varepsilon$ is set to 0.0001. Hereafter, $f''$ denotes the extracted fingerprint, $x$ represents the cover data, $y$ represents the watermarked data with fingerprint embedded at the content owner end, and $y'$ represents the watermarked data with user fingerprint embedded at the replicator end.

#### 5.2.1. State change $(+1, -1)$

The first test is to verify the fingerprint bit change from $+1$ to $-1$ for Eq. (5). Based on Eq. (5), the embedding scaling factor is set to $\alpha' = \alpha'_{(+1,-1)} = \alpha/(1+\alpha) + \varepsilon$ for the case of state change $(+1, -1)$. In Fig. 6(a), the first elements of $y'$ and $x$ are very close, indicating very high transparency. This is because a proper embedding scaling factor derived from Eq. (5) is used to correctly reflect the state change of fingerprint bit. In Fig. 6(b), the first fingerprint bits of $f'$ and $f''$ are the same, representing that $f$ is rewritten by $f'$ at that bit. However, some other fingerprint bits of $f''$ are not identical to those of $f'$, representing the fact that $f$ cannot be completely rewritten by $f'$ if Eq. (5) is adopted.

#### 5.2.2. State change $(-1, +1)$

The second test is to verify the fingerprint bit change from $-1$ to $+1$ for Eq. (6). Based on Eq. (6), the embedding scaling factor is set to $\alpha' = \alpha'_{(-1,+1)} = \alpha/(1-\alpha) + \varepsilon$ for the case of state change $(-1, +1)$. In Fig. 7(a), the second elements of $y'$ and $x$ are very close, indicating very high transparency. This is because a proper embedding scaling factor derived from Eq. (6) is used to correctly reflect the state change of fingerprint bit. In Fig. 7(b), the second fingerprint bits of $f'$ and $f''$ are the same, representing that $f$ is rewritten by $f'$ at that bit. Furthermore, other fingerprint bits of $f''$ are also identical to those of $f'$, representing that $f$ is able to be completely rewritten by $f'$ if Eq. (6) is adopted.

#### 5.2.3. State change $(-1, -1)$

The third test is to verify the fingerprint bit change from $-1$ to $-1$ for Eq. (7). Based on Eq. (7), the embedding scaling factor is set to $\alpha' = \alpha'_{(-1,-1)} = \alpha/(\alpha-1) + \varepsilon$ for the case of state change $(-1, -1)$. In Fig. 8(a), the third elements of $y'$ and $x$ are very close, indicating very high transparency. This is because a proper embedding scaling factor derived from Eq. (7) is used to correctly reflect the state change of fingerprint bit. In Fig. 8(b), the third fingerprint bits of $f'$ and $f''$ are the same,

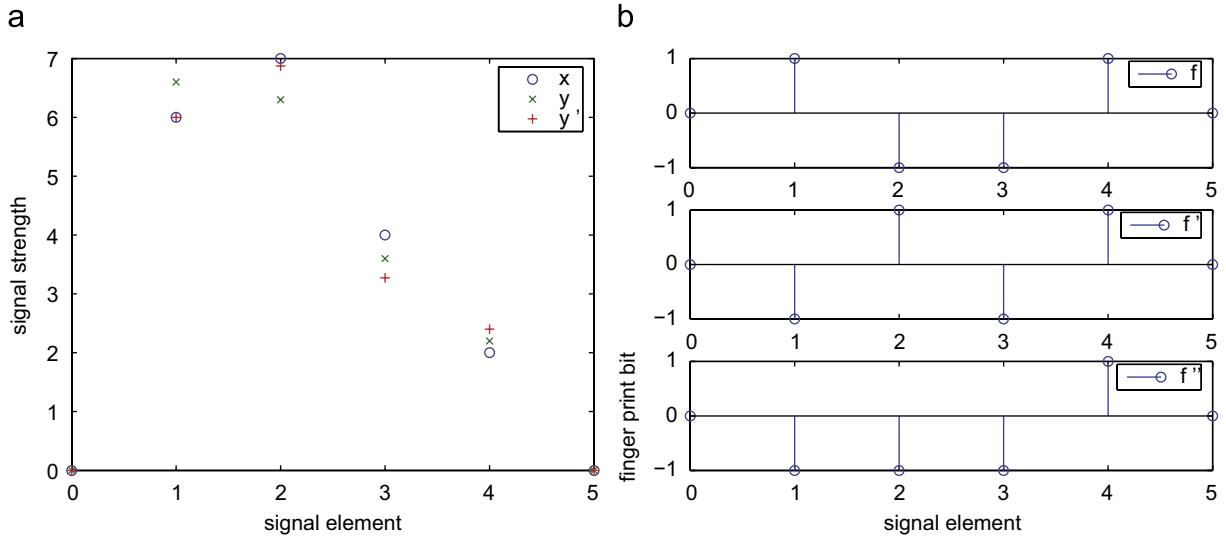


Fig. 6. Verification of $\alpha' = \alpha/(1+\alpha)$: (a) change of signal strength after fingerprint embedding; (b) fingerprint bit detection.

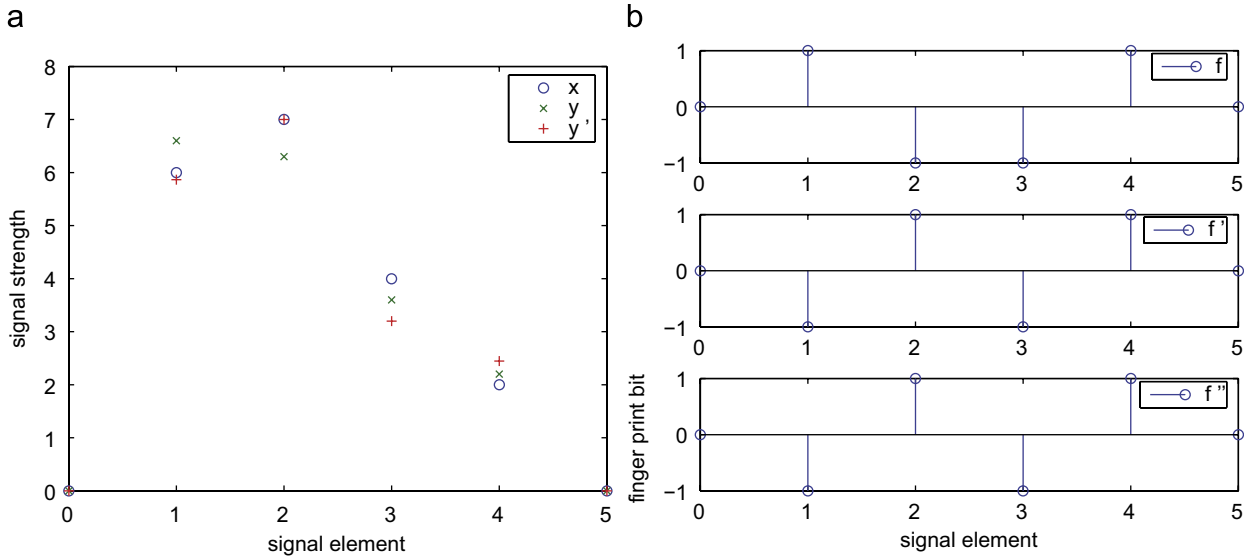


Fig. 7. Verification of $\alpha' = \alpha/(1-\alpha)$: (a) change of signal strength after fingerprint embedding; (b) fingerprint bit detection.

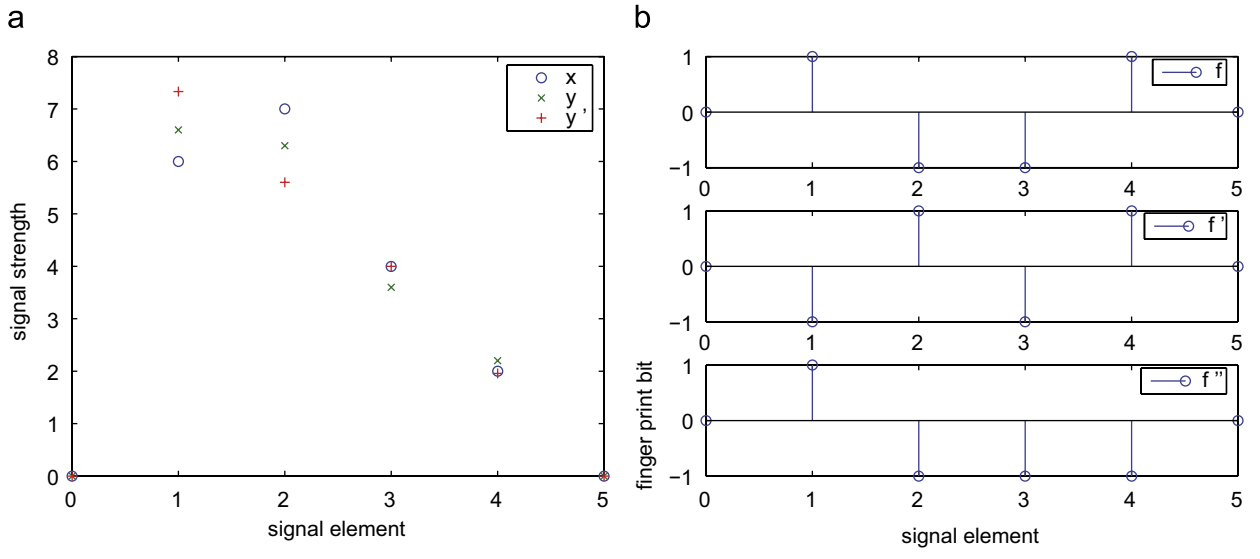


Fig. 8. Verification of $\alpha' = \alpha/(1-\alpha)$: (a) change of signal strength after fingerprint embedding; (b) fingerprint bit detection.

representing that $f$ is rewritten by $f'$ at that bit. However, some other fingerprint bits of $f''$ are not identical to those of $f'$, representing that $f$ cannot be completely rewritten by $f'$ if Eq. (7) is adopted.

*5.2.4. State change* $(+1,+1)$

The fourth test is to verify the fingerprint bit change from $+1$ to $+1$ for Eq. (8). Based on Eq. (8), the embedding scaling factor is set to $\alpha' = \alpha'_{(+1,+1)} = -\alpha/(1+\alpha) + \varepsilon$ for the case of state change $(+1,+1)$. In Fig. 9(a), the fourth elements of $y'$ and $x$ are very close, indicating very high transparency. This is because a proper embedding scaling factor derived from Eq. (8) is used to correctly reflect the state change of fingerprint bit. In Fig. 9(b), the fourth fingerprint bits of $f'$ and $f''$ are the same, representing that $f$ is rewritten by $f'$ at that bit. However, some other fingerprint bits of $f''$ are not identical to those of $f'$, representing that $f$ cannot be completely rewritten by $f'$ if Eq. (8) is adopted.

The values of $\alpha' + \varepsilon$ used for the above simulations are shown in Table 1. Since $0 < \alpha' < 1$ is used as

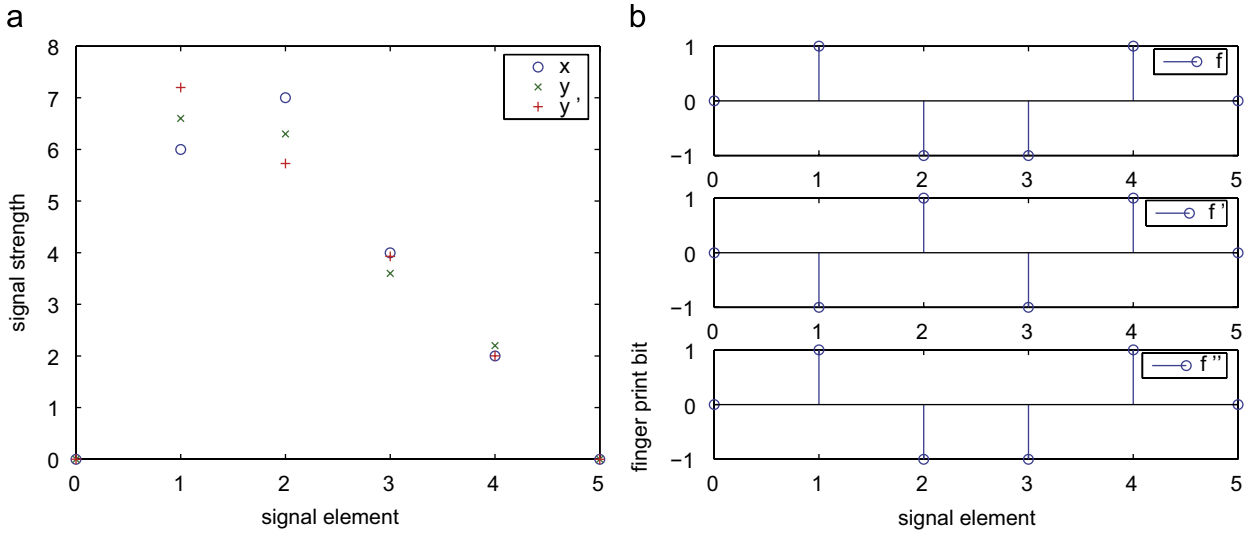


Fig. 9. Verification of $\alpha' = \alpha/(1-\alpha)$: (a) change of signal strength after fingerprint embedding; (b) fingerprint bit detection.

Table 1
The $\alpha'$'s for different RFE states

| Fingerprint state $(f_b, f_b')$ | $\alpha'$ | $\alpha' + \varepsilon$ |
|---|---|---|
| $(+1,-1)$ | $\frac{\alpha}{(1+\alpha)}$ | 0.0910 |
| $(-1,+1)$ | $\frac{\alpha}{(1-\alpha)}$ | 0.1112 |
| $(-1,-1)$ | $\frac{\alpha}{(\alpha-1)}$ | −0.1110 |
| $(+1,+1)$ | $\frac{-\alpha}{(1+\alpha)}$ | −0.0908 |

the scaling factor for embedding, the $\alpha'$ values derived from $\alpha/(\alpha-1)$ and $-\alpha/(1+\alpha)$ do not satisfy the requirement. In addition, the simulation also shows that $\alpha/(1+\alpha)$ does not satisfy the requirement of rewriting the owner's fingerprint at the replicator end. As a result, in order to achieve the requirement of RFE, the lower bound of $\alpha'$ has to be set to $\alpha/(1-\alpha)$, which is identical to the analytic results discussed in Section 3.

### 5.2.5. Transparency and false alarm in fingerprint embedding

Based on the above verification, $\alpha$ is set to 0.1 and $\alpha' + \varepsilon$ is set to 0.12 for fingerprint embedding in these experiments. The PSNR values between a cover image and its watermarked version are shown in Table 2. In this table, $X$ means the cover image, and $Y$ means the image at the content owner end that is embedded with the rewritable fingerprint $F$ and encrypted with $k_g$. In addition, $Y_i'$ means the image that is first decrypted with $k_r$, and then embedded with the user fingerprint $F_i'$, which overwrites $F$, and finally encrypted with $k_i$. For the convenience of comparison, although $Y$ and $Y_i'$ are in the encryption domain, the corresponding keys are utilized to decrypt them to the plaintext domain to measure the PSNRs. It can be observed from the results that when the owner fingerprint is embedded, the PSNRs between $X$ and $Y$ range from 40 to 46 dB. When user fingerprint $F_i'$ is embedded, the PSNR values between $X$ and $Y_i'$ fall within the range of 38–42 dB.

Table 2
Transparency

| Images | PSNR $(Y, X)$ | PSNR $(Y_i', X)$ |
|---|---|---|
| Baboon | 46.07 | 41.78 |
| Lena | 42.62 | 37.70 |
| Pepper | 40.97 | 39.15 |

Table 3
Fingerprint detection (FD) with content owner key and user key

| Images | BER FD of $Y$ with $k_o$ | BER FD of $Y_i'$ with $k_i$ | BER FD of $Y_i'$ with $k_o$ | BER FD of $Y$ with $k_i$ |
|---|---|---|---|---|
| Baboon | 0.00 | 0.00 | 0.45 | 0.52 |
| Lena | 0.00 | 0.00 | 0.44 | 0.50 |
| Pepper | 0.00 | 0.00 | 0.47 | 0.55 |

The false alarm verification is shown in Table 3. If the correct key is used, the bit error rate (BER) of

the fingerprint, respectively, extracted from $Y$ and $Y'_i$ is zero. On the other hand, if $k_o$ is used to detect the fingerprint from $Y'_i$, the obtained BER approximates 0.5, which implies that the rewritable fingerprint $F$ is actually overwritten at the replicator end and satisfies the goal of designing RFE. In addition, if $k_i$ is used to detect the fingerprint from $Y$, the obtained BER approximates 0.5, which represents the fact that the user's fingerprint $F$ and the content owner's fingerprint $F'_i$ are uncorrelated.

### 5.3. PSSM measurement

The spectrum security metric $\mathrm{Sec}_\phi$ is first calculated in a blockwise manner and then accumulated for the whole image. Fig. 10 shows the spectrum security measurement, which ranges from $6.5 \times 10^5$ to $8.5 \times 10^5$, for different user keys. In addition, the security gain $\mathrm{GSec} = \mathrm{Sec}_{\phi_1}/\mathrm{Sec}_{\phi_2}$ calculated between the replicator end and content owner end is shown in Fig. 11, where $\phi_1$ denotes the user fingerprint embedding point for the $i$th user at the replicator end and $\phi_2$ denotes the content owner fingerprint embedding point at the content owner end. It can be observed from Fig. 11 that the values of GSec are kept at about 1, which means that, as long as the multimedia data are kept in the encryption domain, its security is kept nearly unchanged no matter how many encryptions and decryptions are performed within our framework.

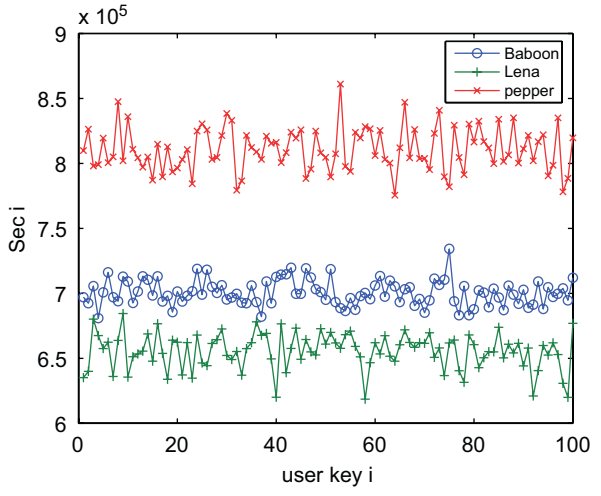


Fig. 10. The spectrum security metric from different users.

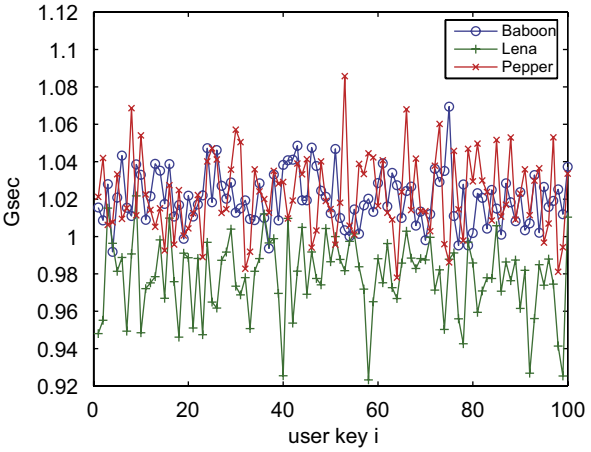


Fig. 11. Gain of PSSM.

### 5.4. Single collusion attack

Because collusion at the attack point H is not the focus of this paper, we only adopt a conventional orthogonal fingerprint embedding technique [5] to evaluate the JEF method under several common collusion attacks [22]. They are abbreviated as follows: ave: average attack; min: minimum attack; max: maximum attack; median: median attack; minmax: MinMax attack; modneg: modified negative attack; rendneg: randomized negative attack. The results of resistance to the single collusion attack (attack point H) are shown in Fig. 12 for different images. Basically, they show that the trend of catching probability reduces with an increase in the number of colluders. The catching probability is calculated as the number of copies detected to have BER less than a threshold over the total number of colluders. However, it is found that different thresholds actually lead to different catching probabilities; however, the trend of catching probability vs. number of colluders is similar.

### 5.5. Multi-point collusion attack

The results of resistance to multi-point collusion attacks are described as follows. For the (C) + (G) collusion attack, although the key is revealed from the set-box at point G and colluded with the replicator at point C to generate a un-fingerprinted copy, the content owner fingerprint $F$ can still be detected with BER = 0.00 to satisfy the condition of catching at least one colluder. For the (C) + (H) collusion attack, the decrypted DCT signs can be available from H and the amplitudes without user

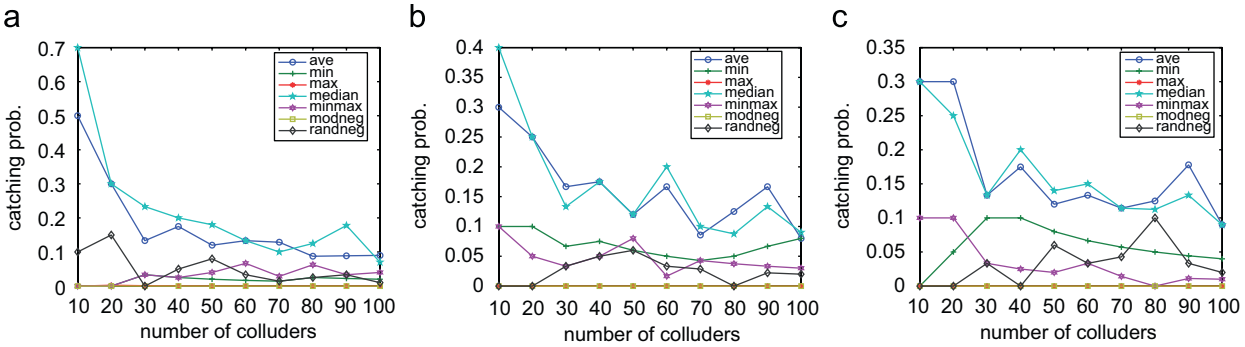


Fig. 12. Resistance to the single collusion attack at point H (catching probability vs. number of colluders) for the images Baboon (a), Lena (b), and Pepper (c).

fingerprint embedded can be obtained from C. Both can be exploited to create a un-fingerprinted copy. However, the content owner fingerprint $F$ can still be detected with BER = 0.00 to satisfy the condition of catching one colluder at the replicator end. For the (B) + (E) + (G) collusion attack, the multimedia stream can be eavesdropped on from the content owner end at point B, $k_r$ can be revealed from the replicator end at point E, and $k_g$ can be revealed from the user end at point G. By collecting the three pieces of information, the eavesdropped multimedia can be successfully decrypted. However, the content owner fingerprint $F$ can still be detected with BER = 0.00. Thus, the replicator can be determined as the colluder to satisfy the requirement of catch one colluder at the replicator end.

## 6. Conclusion

In this paper, a new joint multimedia encryption and fingerprinting method is proposed and incorporated with AACS for content access control and traitor tracing over networks. Unavoidably, there exist some security threats to the proposed framework. We discuss the possible attack points and multi-point collusion attacks, and propose partial solutions to these attacks. Specifically, we propose to use multimedia encryption to cope with some of the attack points and propose RFE to cope with some multi-point collusion attacks. Although all the security leaks cannot be guaranteed to be completely solved in the proposed framework, it is hoped that the raised security issues and solutions can provide directions in developing a secure and practical multimedia transmission system in the future. Compared with the existing joint multimedia encryption and fingerprinting methods, the advantage of this AACS-compatible JEF method is that it does not suffer the same difficulties, as described previously.

## Acknowledgment

This research was supported by the National Science Council under NSC Grants NSC 94-2422-H-001-007 and NSC 95-2422-H-001-008.

## References

[1] [AACS] ⟨http://www.aacsla.com/home⟩.
[2] A.J. Ahumada, H.A. Peterson, Luminance-model-based DCT quantization for color image compression, in: Proceedings of the SPIE, vol. 1666, 1992, pp. 365–374.
[3] J. Bloom, Security and rights management in digital cinema, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, 2003, pp. 712–715.
[4] D. Boneh, J. Shaw, Collusion-secure fingerprinting for digital data, IEEE Trans. Inform. Theory 44 (September 1998) 1897–1905.
[5] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Trans. Image Process. 6 (12) (December 1997) 1673–1687.
[6] J. Daemen, V. Rijmen, The Rijndael block cipher, AES Proposal: Rijndael, 1999, pp. 1–45.
[7] F. Hartung, B. Girod, Digital watermarking of MPEG-2 coded video in the bitstream domain, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, 1997, pp. 2621–2624.
[8] D. Kirovski, F.A. Petitolas, Blind pattern matching attack on watermarking systems, IEEE Trans. Signal Process. 51 (April 2003) 1045–1053.
[9] D. Kundur, K. Karthik, Video fingerprinting and encryption principles for digital rights management, Proc. IEEE 92 (6) (2004) 918–932.
[10] C.S. Lu, S.W. Sun, C.Y. Hsu, P.C. Chang, Media Hash-dependent image watermarking resilient against both

geometric attacks and estimation attacks based on false positive-oriented detection, IEEE Trans. Multimedia 8 (4) (August 2006) 668–685.
[11] B.M. Macq, J.J. Quisquater, Cryptology for digital TV broadcasting, Proc. IEEE 83 (6) (June 1995) 944–957.
[12] Y. Mao, M. Wu, Security evaluation for communication-friendly multimedia encryption, in: Proceedings of the IEEE International Conference on Image Processing, vol. 1, 2004, pp. 569–572.
[13] N.K. Ratha, J.H. Connell, R.M. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM Systems J. 40 (3) (2001) 614–634.
[14] C. Shi, B. Bhargava, A fast MPEG video encryption algorithm, in: Proceedings of the ACM International Conference on Multimedia, 1998, pp. 81–88.
[15] K. Su, D. Kundur, D. Hatzinakos, Statistical invisibility for collusion-resistant digital video watermarking, IEEE Trans. Multimedia 7 (1) (February 2005) 43–51.
[16] K. Su, D. Kundur, D. Hatzinakos, Spatially localized image-dependent watermarking for statistical invisibility and collusion resistance, IEEE Trans. Multimedia 7 (1) (February 2005) 52–66.
[17] S.W. Sun, J.R. Chen, C.S. Lu, P.C. Chang, K.C. Fan, Motion-embedded residual error for packet loss recovery of video transmission and encryption, in: Proceedings of the IS&T/SPIE: Visual Communications and Image Processing (EI127), vol. 6077, 2006.
[18] M.D. Swanson, B. Zhu, A.T. Tewifk, Multiresolution scene-based video watermarking using perceptual models, IEEE J. Select. Areas Comm. 16 (May 1998) 540–550.
[19] L. Tang, Methods for encrypting and decrypting MPEG video data efficiently, in: Proceedings of the ACM International Conference on Multimedia, 1997, pp. 219–229.
[20] W. Trappe, M. Wu, Z.J. Wang, K.J.R. Liu, Anti-collusion fingerprinting for multimedia, IEEE Trans. Signal Process. 51 (April 2003) 1069–1087.
[21] J. Wen, M. Severa, W. Zeng, M.H. Luttrell, W. Jin, A format-compliant configurable encryption framework for access control of video, IEEE Trans. Circuits Systems Video Technol. 12 (6) (2002) 545–557.
[22] M. Wu, W. Trappe, Z.J. Wang, K.J.R. Liu, Collusion-resistant fingerprinting for multimedia, IEEE Signal Process. Magazine (March 2003) 15–27.
[23] X. Xu, S. Dexter, A.M. Eskicioglu, A hybrid scheme of encryption and watermarking, in: IS&T/SPIE Symposium on Electronic Imaging 2004, Security, Steganography, and Watermarking of Multimedia Contents VI Conference, vol. 5306, 2004, pp. 725–736.
[24] W. Zeng, S. Lei, Efficient frequency domain selective scrambling of digital video, IEEE Trans. Multimedia 5 (1) (2003) 118–129.
[25] H. Zhao, K.J.R. Liu, Fingerprint multicast in secure video streaming, IEEE Trans. Image Process. 15 (1) (January 2006) 12–28.
[26] B.B. Zhu, M.D. Swanson, A.H. Tewfik, When seeing isn't believing [multimedia authentication technologies], IEEE Signal Process. Magazine 21 (2) (2004) 40–49.
[27] B.B. Zhu, C. Yuan, Y. Wang, S. Li, Scalable protection for MPEG-4 fine granularity scalability, IEEE Trans. Multimedia 7 (2) (2005) 222–233.