JOINT MULTIMEDIA FINGERPRINTING AND ENCRYPTION: SECURITY ISSUES AND SOME SOLUTIONS

Shih-Wei Sun[1,2]*, Chun-Shien Lu*[1,*]*, and Pao-Chi Chang*[2,3]

[1]Institute of Information Science, Academia Sinica, Taipei, Taiwan 115, ROC
[2]Dept. Electrical Engineering, National Central Univ., Chung-Li, Taiwan 320, ROC
[3]Dept. Communication Engineering, National Central Univ., Chung-Li, Taiwan 320, ROC

ABSTRACT

In this paper, a new multimedia joint fingerprinting and encryption (JFE) scheme embedded into the advanced access content system (AACS) is proposed. Like other security-related systems, there exist some security threats to the proposed framework. To cope with these difficulties, the contributions of this paper include: (i) we apply multimedia encryption at different points to resist some attacks points; and (ii) we propose rewritable fingerprint embedding (RFE) to deal with some multi-point collusion attacks. Experimental results are provided to demonstrate the proposed AACS-compatible JFE method.

1. INTRODUCTION

In multimedia security, encryption plays the first line of defense for secure multimedia transmission. However, it is known that the decrypted data loses the imposed protection capability and may be illegally distributed. In order to persistently preserve the capability of multimedia protection, fingerprinting is further applied to give traitor tracing. In this paper, we study the dual goals of multimedia content access control and traitor tracing. The emerging joint multimedia encryption and fingerprinting technology, divided into three categories [5], 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 [3, 7]. This scenario incurs some disadvantages: (1) Inefficient bandwidth utilization: since multimedia fingerprint embedding is done at the transmission side, repeat request of the same copy will waste bandwidth; (2) Insecure encryption: since a single global encryption key is used, if a malicious user eavesdrops other user's data, then the multimedia plaintext belonging to that user can be obtained.

b) Transmitter-side Encryption and Receiver-side Fingerprint Embedding: Fingerprint embedding at the receiver-side was first proposed in [6] for digital TV and then applied to digital right management (DRM) of digital cinema [2]. At the transmitter side, only 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 plays 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 could be decrypted according to the global key. Meanwhile, the user fingerprint should be embedded into the multimedia data to generate the fingerprinted multimedia data for each user. However, the total load of decryption and fingerprint embedding gathered at the receiver side (super node) will increase computational complexity.

c) Joint Fingerprinting and Decryption: In order to reduce system complexity and achieve real-time requirement, Kundur and Karthik [5] proposed a joint fingerprinting and decryption method. The idea is that the multimedia ciphertext is partially decrypted such that the un-decrypted parts imitate fingerprinted multimedia. This kind of method is conceptually promising, achieving multimedia partial decryption and multimedia fingerprint embedding at the same time. However, the un-decrypted content must satisfy 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 fingerprinting and encryption (JFE) scheme, which can be incorporated with the advanced access content system (AACS). AACS [1] is a leading technology proposed by many famous companies. Basically, AACS contains four major parts: content owner end, licensed replicator end, licensing entity for key management, and licensed player at the user end. The contributions of our AACS-compatible JFE method include: (i) we discuss the security threats to the proposed method and present solutions to cope with some of them; (ii) we apply multimedia encryption at different points to resist some

*Corresponding author: Dr. C. S. Lu (lcs@iis.sinica.edu.tw)

attack points; (iii) we propose rewritable fingerprint embedding (RFE) to deal with some multi-point collusion attacks.

2. THE FRAMEWORK OF PROPOSED JOINT MULTIMEDIA FINGERPRINTING AND ENCRYPTION METHOD

Based on AACS, the proposed joint multimedia fingerprinting and encryption method contains four major parts, including content owner end, replicator end, key management center, and user end, as shown in Fig. 1.

2.1. Attack Points and Multi-point Collusion Attacks

Like other security-related systems, there exist some security threats, as discussed below, to the proposed framework.

2.1.1. Possible Attack Points

The possible attack points existing in our proposed framework are shown in Fig. 1 and briefly summarized as follows. They include (A) original copy attack; (B) snooping attack; (C) content owner fingerprinted replicator back end attack; (D) user fingerprinted replicator back end attack; (E) replicator key back end attack; (F) encrypted copy attack; (G) set-top box attack; and (H) decrypted copy attack. In particular, the collusion attacks at the attack point H have been widely discussed recently [3, 7, 8].

2.1.2. Multi-point Collusion Attacks

A multi-point collusion attack is a combination of more than one single attack points. In this paper, the most intuitive collusion attacks, producing meaningful plaintext of multimedia content, are addressed and possible solution 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 property. 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.

2.2. System Description

At the content owner end, in order to embed fingerprints, the widely applied digital watermarking technique, spread spectrum (SS) watermarking [4], is adopted. In addition, a novel concept of fingerprint embedding called "rewritable fingerprint embedding (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. In order to encrypt the multimedia data effectively and fast, a light-weight encryption scheme aiming at encrypting the AC signs of DCT coefficients according to the global key k_g is adopted. In addition, another key, k_r, is applied for encryption in the entropy coding domain. The function of twice-encryption is mainly used for achieving 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, the received data should be decrypted in the entropy coding domain by using the replicator end key, k_r. After transforming from the decrypted entropy coding domain back to the DCT domain, the multimedia fingerprint F_i', generated by the user identifying key k_i, will be embedded into the multimedia data according to RFE for user identification. Meanwhile, the multimedia data is light-weight encrypted again according to the user identification key k_i. 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 store and distribution. 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.

Once the multimedia plaintext X_i is 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

RFE should embed both fingerprints F and F_i' generated by k_o and k_i, respectively, at the content owner end and replicator end. The content owner fingerprint F at the content owner end can be overwritten by the user multimedia fingerprint F_i' at the replicator end. The extraction of either of them can be used to cope with some of the multi-point collusion attacks. In this study, we derive analytic bound of the embedded fingerprints to achieve the highest transparency.

Based on the spread spectrum watermarking technique [4], 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, α 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. 2. In Fig. 2, f_b' denotes the b-th replicator end fingerprint bit. As a result, the scaling factor of user fingerprint embedding are denoted as

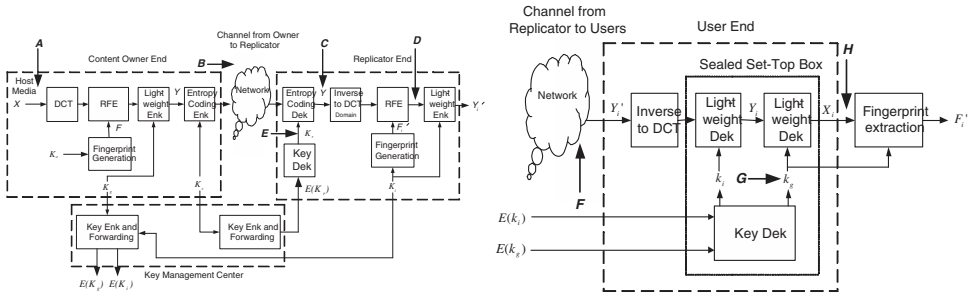

Figure 1: AACS-compatible multimedia joint fingerprinting and encryption method.

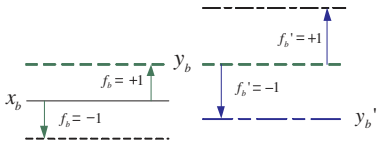

Figure 2: An example of rewritable fingerprint embedding.

$\alpha'_{(+1,-1)}$, $\alpha'_{(+1,+1)}$, $\alpha'_{(-1,+1)}$, 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 [7, 8].

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), we have:

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

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:

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

We can further derive to obtain:

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

The similar derivations can be derived for the remaining three cases of scaling factors as:

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

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

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

However, the prior knowledge of the fingerprint state change cannot be obtained neither at the content owner end nor at the replicator end. On the contrary, a global parameter of α' should be determined and sent to the replicator end for user fingerprint embedding. According to Eqs.(5)$\sim$(8), the lower bound of the scaling factor of embedding at the replicator end, α', is defined as:

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

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

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

4. EXPERIMENTAL RESULTS

Some common images of size 512×512 were used for joint fingerprinting and encryption. AES was selected for performing 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 16×16. In the experiments, the 128 largest DCT AC coefficients in a 16×16 block were selected for encryption. The size of fingerprints embedded using k_o and k_u was 64 bits. There was 1 fingerprint bit embedded in the $(1, 2)$-th subband of a DCT block. In addition, α was set as 0.1 and α' was set as 0.12 to the satisfy the RFE requirements.

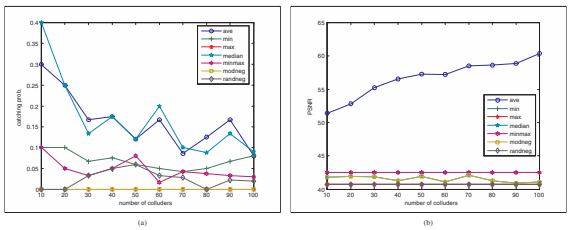

Figure 3: Single collusion attack at point H for Lena.

4.1. Encryption Results

The encryption results generated at the different stages using different keys were visually examined. Basically, the visual qualities of encrypted images show that the transmitted images in the proposed JFE scheme are persistently kept in the encryption domain at the different points of the system. We have also developed a spectrum perceptual security metric to measure the change of encryption security and find that the security gain is almost unchanged.

4.2. Resistance to Single Collusion Attack

The used collusion attacks are 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. Unsurprisingly, the results of resistance to the single collusion attack (attack point H) show that the trend of catching probability reduces with the increase of the number of colluders (Fig. 3(a)). The catching probability is calculated as the number of copies detected to have bit error rate (BER) less than a threshold over the total number of colluders. In Fig. 3(b), we show the PSNR values of colluded images under different number of colluders and different collusion attacks.

4.3. Multi-point Collusion Attack

The results of resistance to multi-point collusion attack are described as follows. For (C)+(G) collusion attack, although the key is revealed from the set-box at the point G and colluded with the replicator at the 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 at the replicator end. For (C)+(H) collusion attack, the decrypted DCT signs can be available from H and the amplitudes without user 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 (B)+(E)+(G) collusion attack, the multimedia stream can be eavesdropped from the content owner end at point B, colluded with the replicator end at point E and with the user end at point G to obtain un-fingerprinted copy. However, the content owner fingerprint F can still be detected with $BER = 0.00$. Thus, the replicator can be determined as one of the colluders to satisfy the requirement of catching at least one colluder at the replicator end.

5. CONCLUSION

A new joint multimedia fingerprinting and encryption method is proposed and incorporated with advanced access content system for content access control and traitor tracing. 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 rewritable fingerprint embedding to cope with some multi-point collusion attacks. Although all the security leaks in the proposed framework have not been completely solved, we hope the raised security issues and solutions can provide directions in developing a new multimedia security system in the future.

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.

6. REFERENCES

[1] [AACS] *http://www.aacsla.com/specifications/specs091/AACS_Spec_Prerecorded_0.91.pdf*.

[2] J. Bloom, "Security and Rights Management in Digital Cinema," *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing*, Vol. 4, pp. 712-715, 2003.

[3] D. Boneh and J. Shaw, "Collusion-secure Fingerprinting for Digital Data," *IEEE Trans. Inform. Theory*, Vol. 44, pp. 1897-1905, 1998.

[4] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. Image Processing*, Vol. 6, No. 12, 1997.

[5] D. Kundur and K. Karthik, "Video Fingerprinting and Encryption Principles for Digital Rights Management," *Proceedings of the IEEE*, Vol. 92, No. 6, pp. 918-932, 2004.

[6] B. M. Macq and J. J. Quisquater, "Cryptology for digital TV broadcasting," *Proceedings of the IEEE*, Vol. 83, No. 6, pp. 944-957, 1995.

[7] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu, "Anti-collusion Fingerprinting for Multimedia," *IEEE Trans. Signal Processing*, Vol. 51, pp. 1069-1087, 2003.

[8] M. Wu, W. Trappe, Z. J. Wang, and K. J. R. Liu, "Collusion-Resistant Fingerprinting for Multimedia," *IEEE Signal Processing Magazine*, pp. 15-27, 2003.