Linear Combination Collusion Attack and its Application on an Anti-Collusion Fingerprinting

Yongdong Wu
Institute for Infocomm Research, Singapore
wydong@i2r.a-star.edu.sg

***Abstract*— This paper presents a Linear Combination Collusion Attack (LCCA) which is a generalization of the average attack model. LCCA generates a pirated image of good quality but prevents traitors from being identified. As an application example, LCCA is used to attack a fingerprinting scheme published in IEEE Transactions on Signal Processing. The theoretical analysis and experiment results demonstrate that the attack is practical and efficient.**

I. INTRODUCTION

In order to protect the ownership and prevent unauthorized dissemination of digital content, traitor tracing schemes [1]-[7] enable to trace the illegal distributors. In the applications (e.g. pay-TV) of traitor tracing schemes, each legal user has a personal decryption key. The service provider broadcasts the encrypted content such that any legal user can decode the protected content with his personal key. One assumption of traitor tracing schemes is that the user is unwilling to distribute the original content because the size of the content is often very large, but the traitors may disclose pirate keys derived from their personal keys so that the illegal users access the content. In the traitor tracing scheme, at least one traitor is identified from the pirate keys. However, most of the black-box traitor tracing schemes are vulnerable to the general attack addressed in paper [8]. On the other hand, traitor tracing scheme is not viable in protecting images because images are relatively small and easy to be distributed. Luckily, fingerprinting provides one way for image protection. It embeds a special label which identifies the user uniquely. If a label is found in a suspected image, the user with the label is identified as a traitor at a high probability.

However, most of the fingerprinting schemes are prone to average collusion attacks [9]. Such an attack does not consider any specific watermarking scheme given that the probability of implicating an innocent is reasonably low. In the collusion attack, a group of traitors collectively obtains an average of their individually watermarked copies and escapes from being identified. Ergun *et al*. [9] proved that no traitor will be identified with a pirated copy if the number of the traitors is $O(\sqrt{n\ln(n)})$ given that the probability of implicating innocent is low, where n is the size of the cover signal. This result is of more importance in theory than in practice because the number of traitors is too big. For a low-value image, it is probably not worth collecting that many watermarked images. A number of other collusion attacks have been studied, the reader is directed to [10] - [14] for technical details.

However, from the viewpoint of designers, a good fingerprinting scheme should approach to this upper bound as close as possible. Celik *et al.* [15] propose a collusion-resilient watermarking method, wherein the host signal is pre-warped randomly prior to watermarking. Celik remarked that it required substantial computational resources to undo warps.

The traditional orthogonal method [16] is resilient to the average collusion attack with a lot of orthogonal vectors. To reduce the number of orthogonal vectors, Trappe *et al.*[17] proposed an anti-collusion scheme AND-ACC fingerprinting (hereafter referred to as the TWWL scheme)which assigns each user a watermark generated with orthogonal vectors and an anti-collusion codevector. Trappe *et al.* investigated the security of AND-ACC fingerprinting scheme [17] and reported that the scheme is secure against average attack. However, as an extension of the collusion model used in papers [9][17], LCCA enables traitors to create a pirated image of good quality safely.

The reminder of this paper is organized as follows. Section II elaborates the LCCA attack. Section III introduces LCCA application on the TWWL addressed in [17]. Section IV contains the results of the experiments which demonstrate the efficiency of the attack.

II. LCCA ATTACK SCHEME

Because additive embedding method [18] is widely used in watermarking, average attack is used as a main security analysis tool. This section describes LCCA which extends average attack so as to enable k traitors to create a pirate image of good quality safely. For self contained, the average attack is introduced in the following.

A. Average Attack

Trappe *et al.* studied the security of AND-ACC fingerprinting based on the collusion attack model in [9] as

$$\begin{cases} \hat{\mathbf{Y}} = \sum_{i=1}^{k} \lambda_i \mathbf{Y}_i \\ \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1 \\ 0 \le \lambda_i \le 1 \qquad\qquad i = 1, 2, \cdots, k \end{cases} \tag{1}$$

where $\mathbf{Y}_i$ is the legal watermarked image of traitor P_i, $i = 1, 2, \cdots, k$. Trappe *et al.* selected $\lambda_i = 1/k$, and they also noted: "*there may exist cases in which the underlying fingerprints will not necessarily have the same energy, or be independent of each other, and that other choices for* λ_i *might be more appropriate.*" Although Trappe *et al.* noticed

the existence of other collusion attacks, they did not propose an effective collusion attack but average attack. Indeed, Su *et al.* [19] extended the average attack. They noted "*more sophisticated linear temporal filters by allowing β_k (i.e., λ_i in [17]) to take on arbitrary values*". Clearly, their collusion is not right. For example, if $\beta_k = 100$, the traitors will obtain nothing but noise according to Su's attack [19]. Thus, How to select λ_i is very important in the linear attack. In the following, a LCCA model is addressed.

B. Linear Combination Collusion Attack

LCCA extends the average collusion attack [9][17] by removing the unnecessary restraint $0 \leq \lambda_i \leq 1$ from formula (1), and the updated attack model is

$$\begin{cases} \hat{\mathbf{Y}} = \sum_{i=1}^{k} \lambda_i \mathbf{Y}_i \\ \lambda_1 + \lambda_2 + \cdots + \lambda_k = 1 \end{cases} \tag{2}$$

Generally speaking, all the watermarks have almost the same energy. In order that each traitor has the same probability of escaping from being identified, the contribution to the pirated image from any traitor should be almost identical. That is to say, $\mid \lambda_1 \mid = \mid \lambda_2 \mid = \cdots = \mid \lambda_k \mid$. Hence, λ_i is selected to be 1 or -1 in the LCCA of the present paper. Without loss of generality, the LCCA model is

$$\hat{\mathbf{Y}} = -\sum_{i=1}^{r} \mathbf{Y}_i + \sum_{i=r+1}^{2r+1} \mathbf{Y}_i. \tag{3}$$

where $k = 2r + 1$ for some r. Obviously, the challenge for LCCA is how to achieve good fidelity of the pirated image. To quantitatively describe the similarity between the original image $\mathbf{X}$ and the pirated image $\hat{\mathbf{Y}}$, suppose the processing image is 8-bit gray images, and all the independent watermarks have the same energy, calculate the PSNR (peak signal-noise-ratio) as

$$\begin{aligned} \sigma^2 &= \frac{1}{n^2} \parallel \hat{\mathbf{Y}} - \mathbf{X} \parallel^2 = \frac{1}{n^2} \parallel \sum_{i=1}^{k} \lambda_i \mathbf{Y}_i - \mathbf{X} \parallel^2 \\ &= \frac{1}{n^2} \parallel \sum_{i=1}^{k} \alpha \lambda_i \mathbf{W}_i \parallel^2 = \frac{k}{n^2} \parallel \alpha \mathbf{W} \parallel^2 . \\ PSNR &= 10(\lg 255^2 - \lg \sigma^2) \\ &= 10(\lg 255^2 - \lg \frac{1}{n^2} \parallel \alpha \mathbf{W} \parallel^2) - 10 \lg k \\ &= PSNR_0 - 10 \lg k, \end{aligned}$$

where $PSNR_0$ is the PSNR of the original watermarked image. Comparing with PSNR of the original watermarked images, the PSNR of the pirated image is decreased only $10 \lg k$ dB. For instance, if there are three traitors, the PSNR of pirated image is reduced $10 \lg 3$=4.7dB. That is to say, the quality loss of the pirated copy is small. Therefore, the pirated image $\hat{\mathbf{Y}}$ generated from only a few traitors' images is similar to the original image $\mathbf{X}$.

III. LCCA Attack on TWWL

Trappe *et al.* claimed that k traitors can be identified if the AND-ACC $(v, k+1, 1)$ is used to generate the binary codevectors for users. For completeness, the Trappe's scheme (TWWL) is introduced here.

A. TWWL scheme

1) Embedding Watermarks: According to the *Definition 1* and *Theorem 1* in TWWL [17], a binary AND-ACC $(v, k+1, 1)$ is k-resilient. [1] A user P_i, identified with a codevector $\mathbf{b}_i = (b_{i1}, b_{i2}, \cdots, b_{iv})$ of AND-ACC $(v, k+1, 1)$, has a watermark

$$\mathbf{W}_i = \sum_{j=1}^{v} b_{ij} \mathbf{u}_j$$

where $b_{ij} \in \{0, 1\}$ (for simplicity, another case $b_{ij} \in \{-1, 1\}$ in paper [17] is ignored.), all the $\mathbf{u}_j$ $(j = 1, 2, \cdots, v)$ constitute an orthonormal basis. That is to say, $\mathbf{u}_i^T \cdot \mathbf{u}_i$=1, and $\mathbf{u}_i^T \cdot \mathbf{u}_j$=0 if $i \neq j$. According to the additive embedding method [18], the watermarked signal $\mathbf{Y}_i$ for user P_i is

$$\mathbf{Y}_i = \mathbf{X} + \alpha \mathbf{W}_i \tag{4}$$

where $\mathbf{X}$ is a host signal, and α is a public number which is used for perceptibility constraint.

2) Tracing Traitors: Trappe *et al.* proposed a traitor tracing method with the orthonormal basis vectors and the codevectors for users. In the tracing approach, although the tracer does not know the original image in advance, he is able to obtain the original image from a watermarked image and the orthonormal basis vectors. Thus, it is reasonable to assume that the tracer knows the original image $\mathbf{X}$ all the time in the tracing process. Furthermore, suppose no noise is added into the watermarked image for the sake of simplicity. Clearly, this simplification increases the robustness of the correlation tracing method proposed in [17]. The tracer calculates the correlation vector $T_N = (T_N(1), T_N(2), \cdots, T_N(v))$ from a suspected image $\hat{\mathbf{Y}}$, where $T_N(j) = (\hat{\mathbf{Y}} - \mathbf{X})^T \mathbf{u}_j / \alpha$. Next the tracer creates a vector Γ=$(\Gamma_1, \Gamma_2, \cdots, \Gamma_v)$,

$$\Gamma_j = \begin{cases} 1 & : \quad T_N(j) > \tau \\ 0 & : \quad else \end{cases} \tag{5}$$

where $\tau \in [0, 1]$ is a predefined threshold value. If the codevector $\mathbf{b}_i$ bitwise-AND Γ is equal to Γ, then user P_i is suspected to be a traitor. Furthermore, if the number of suspected traitors is k or fewer, the suspected traitors are confirmed, otherwise, traitors can not be identified correctly.

B. Resilience to Tracing Traitors

Following the LCCA addressed in Section II, k traitors are able to conspire to create a pirated image. After a pirated

[1] For arbitrary two subsets $\mathbf{U}$ and $\mathbf{V}$, each subset includes k or fewer binary codevectors of AND-ACC $(v, k+1, 1)$, the output of bitwise-AND $\mathbf{U}$'codevectors is distinct from that of bitwise-AND $\mathbf{V}$'codevectors.

image is confiscated, the tracer calculates the correlation vector T_N according to the tracing strategy in [17],

$$T_N(j) = (\hat{\mathbf{Y}} - \mathbf{X})^T \mathbf{u}_j/\alpha = (\sum_{i=1}^{k} \lambda_i \mathbf{W}_i)^T \mathbf{u}_j$$
$$= \sum_{i=1}^{k} \lambda_i b_{ij} = -\sum_{i=1}^{r} b_{ij} + \sum_{i=r+1}^{2r+1} b_{ij}. \quad (6)$$

The abstract in [17] noted: "*We (Trappe et al.) propose a new class of codes, called anti-collusion codes (ACCs), which have the property that the composition of any subset of K or fewer codevectors is unique. Using this property, we can therefore identify groups of K or fewer colluders.*" Unfortunately, this claim does not hold due to the following lemma.

Lemma 1: The probability p_0 of tracing all the traitors is negligible with the original tracing strategy.

Proof: According to Trappe's original tracing strategy, if $T_N(j) > \tau$, then $\Gamma_j = 1$. That is to say, b_{1j} AND b_{2j} AND $\cdots$ AND b_{kj}=1. Thus the tracer's conclusion is $b_{1j} = b_{2j} = \cdots = b_{kj} = 1$. However, according to equation (6), if $T_N(j) > 1$, at least one $b_{ij} \neq 1$ $(i = 1, 2, \cdots, k)$. Therefore, the tracer is hard to identify any traitor. ■

To defend against LCCA, an improvement on Trappe's tracing strategy is that $b_{1j} = b_{2j} = \cdots = b_{kj} = 1$ if and only if $round(T_N(j)) = 1$ where $round(\cdot)$ is a round function.

Denote m to be the number of $T_N(j) = 1$, $j \in [1, v]$. In an AND-ACC $(v, k+1, 1)$, there are $(v^2-v)/(k^2-k)$ codevectors [17]. Because the number of zero elements in each codevector is $k+1$, the output of bitwise-AND k codevectors has at most $k(k+1)$ zero elements, i.e.,

$$m \geq \max(1, v - k(k+1))$$

For example, in a 3-resilient AND-ACC (121,4,1), $n = 1210$ and $m \geq 109$.

Lemma 2: The probability p_0 of tracing all the traitors is negligible with the improved tracing strategy.

Proof: According to the improved tracing strategy, if $T_N(j) = 1$, $b_{1j} = b_{2j} = \cdots = b_{kj} = 1$. However, there are a lot of solutions to any equation $T_N(j) = 1$. Let n_j to be the number of solutions to equation $T_N(j) = 1$. Rewrite formula (6) as

$$-\sum_{i=1}^{r} b_{ij} + \sum_{i=r+1}^{2r+1} b_{ij} = 1 > \tau \quad (7)$$

Obviously, $n_j \geq 2$. Since only one out of n_j resolutions can be used for identifying the traitor group, the tracing probability $p_0 < (n_j^{-1})^m < 2^{-m}$. For example, the probability $p_0 < 2^{-109}$ if the above AND-ACC (121,4,1) is employed for 1210 users. Therefore, the probability p_0 is negligible with the improved tracing scheme. ■

Of course, the tracer may select other strategies to trace the traitors. However, the tracing probability is still very small. For instance, the tracer may select m' equations out of *m* equations (7), where m' is the number of 1s in the output of bitwise-AND k codeverctors. This improved tracing strategy has the success probability $1/\binom{m}{m'}$. Furthermore, the cunning traitors may improve their attack by starting block-wise collusion such that the tracing probability is reduced. In summary, the AND-ACC fingerprinting [17] is merely resilient to average collusion attack, but vulnerable to LCCA.

IV. Experiments

In the experiment in paper [17], Trappe *et al.* constructed an anti-collusion code (16,4,1) AND-ACC code to trace up to three traitors out of 20 users. The codevectors are illustrated in the following matrix C whose column is a codevector for one user. i.e., the first column is the codevector for user P_1, the second column is the codevector for user P_2, and so on.

$$C = \begin{pmatrix} 00000 & 11111 & 11111 & 11111 \\ 01111 & 00001 & 11111 & 11111 \\ 01111 & 11110 & 00011 & 11111 \\ 01111 & 11111 & 11100 & 00111 \\ 10111 & 01110 & 11101 & 11011 \\ 10111 & 10111 & 01110 & 11101 \\ 10111 & 11011 & 10111 & 01110 \\ 11011 & 01111 & 01111 & 10110 \\ 11011 & 11011 & 11010 & 11011 \\ 11011 & 11101 & 10101 & 11101 \\ 11101 & 01111 & 11011 & 01101 \\ 11101 & 10111 & 10111 & 10011 \\ 11101 & 11100 & 11110 & 11110 \\ 11110 & 10111 & 11001 & 11110 \\ 11110 & 11010 & 11111 & 10101 \\ 11110 & 11101 & 01111 & 01011 \end{pmatrix}$$

As the experiment in paper [17], the host signal is a gray image of 512×512 shown in Figure 1. The orthonormal basis vectors are generated by applying the Matlab 5.3 uniform distribution random function with mean value 0. In addition, $\alpha = 1.0$ and $\tau = 0.5$ are selected.

Fig. 1. Original gray image of size 512×512.

As the experiment in the paper [17], three users P_1, P_4, P_8 are selected as traitors. They produce a pirated copy from their watermarked image $\mathbf{Y}_1$, $\mathbf{Y}_4$, and $\mathbf{Y}_8$, as $\hat{\mathbf{Y}} = -\mathbf{Y}_1 + \mathbf{Y}_4 + \mathbf{Y}_8$.

The pirated image $\hat{\mathbf{Y}}$ is shown in Figure 2. With reference to the original image in Figure 1, the PSNR of the pirated image

is 36.1dB. Taubman *et al* [20] said that good reconstructed image typically has $PSNR \geq 30$dB. That is to say, the pirated image is of high quality. To identify the traitors, the tracer calculates the correlation values $T_N(j) = -b_{1j} + b_{4j} + b_{8j}$ for all $j = 1, 2, \cdots, v = 16$. The results of $T_N(j)$ are {1, 1, 2, 2, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1}. Therefore, the vector Γ= {1111 1101 0100 0101 }. As a result, the suspected traitor set Φ = {0000 0000 0000 0000 0000}. In other words, no traitor is identified.

Fig. 2. The pirated image generated by three traitors P_1, P_4 and P_8 with LCCA. Its PSNR is 36.1dB.

To measure the quality of the pirated image, repeat the LCCA with different orthogonal basis vectors. The result is shown in Figure 3. The PSNR loss is less than 1.27dB. This experimental PSNR loss is less than the theoretical value (4.7dB) in subsection II-B because the watermarks are not independent in nature. Figure 3 indicates that the pirated images are of good quality.

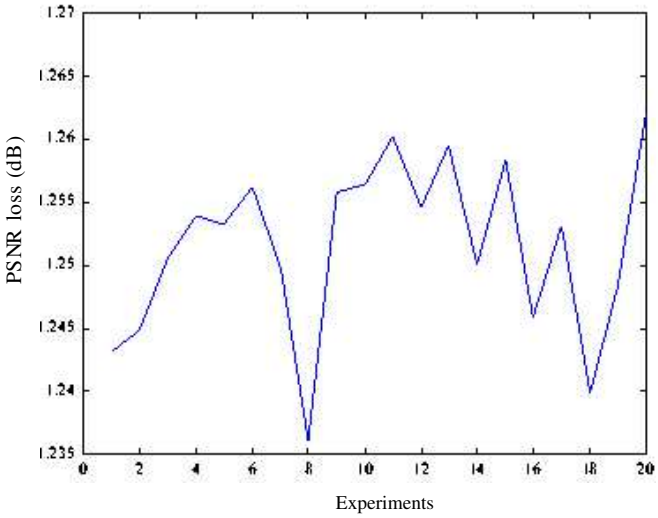

Fig. 3. PSNR loss of pirated image created by three traitors P_1, P_4 and P_8. The PSNR loss is the difference between the PSNRs of the pirated image and the average PSNR of watermarked images.

V. Conclusion

The present paper proposes a LCCA attack which extends the conventional average attack so as to create a pirated image from the linear combination of the traitors' images. Furthermore, LCCA is employed to investigate the security of an anti-collusion fingerprinting [17]. Although the AND-ACC fingerprinting is resilient to the average attack, it is still vulnerable. The experiments illustrate that the pirated images are of good quality but no traitor is identified.

References

[1] D. Boneh and M. Franklin, "An efficient public key traitor tracing scheme," Crypto99, Lecture Notes in Computer Science (LNCS) 1666, pp 338-353, 1999.

[2] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," CRYPTO'95, LNCS 963, pp.452-465, 1995.

[3] B. Chor,A. Fiat and M. Naor, "Tracing traitors," Crypto94, LNCS 839, pp. 257-270, 1994.

[4] M. Naor and B. Pinkas, "Threshold traitor tracing," Crypto98, LNCS 1462,pp. 502-517,1998.

[5] B. Pfitzmann, "Trails of traced traitors," Information Hiding Workshop, LNCS 1174, pp. 49-64, 1996.

[6] Y. Dodis and N. Fazio, "Public key trace and revoke scheme secure against adaptive chosen ciphertext attack," Publick key conference 2003, pp.100-115, 2003.

[7] D. R. Stinson and R. Wei, "Combinatorial properties and constructions of traceability schemes and frameproof codes," *SIAM J. on Discrete Math*, 11(1):41- 53,1998.

[8] Jeff Jianxin Yan and Yongdong Wu, "An attack on a traitor Tracing Scheme," eprint 2001/067.

[9] Funda Ergun, Joe Kilian and Ravi Kumar, "A note on the limits of collusion-resistant watermarks," Advances in Cryptology - EUROCRYPT'99, LNCS 1592, pp. 140-149, 1999.

[10] G. C. Langelaar, R. L. Lagendijk and J. Biemond, "Removing spatial spread spectrum watermarks by nonlinear filtering," 9th European Signal Processing Conference pp. 2281-2284, 1998.

[11] Z. Jane Wang, Min Wu, Hong Zhao, K. J. Ray Liu and Wade Trappe, "Resistance of orthogonal Gaussian fingerprints to collusion attacks," ICASSP (2003) IV724-727.

[12] Yongdong Wu and Robert Deng, "Adaptive collusion attack to a block oriented watermarking scheme," International Conference on Information and Communications Security 2003, LNCS 2836, pp.238-248.

[13] Darko Kirovski, and Fabien A.P. Petitcolas,"Replacement Attack on Arbitrary Watermarking Systems," 2002 ACM Workshop on Digital Rights Management.

[14] Hong Zhao, Min Wu, Z. Jane Wang and K. J. Ray Liu, "Nonlinear Collusion Attacks on Independent Fingerprints For Multimedia," ICASSP (2003), V664-667 1.

[15] M. Celik, G. Sharma and A.M. Tekalp, "Collusion-resilient fingerprinting using random pre-warping," IEEE Int. Conf. On Image Processing 2003.

[16] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[17] W. Trappe, M. Wu, Z. Jane Wang and K. J. Ray Liu, "Anti-collusion fingerprinting for Multimedia," *IEEE Trans. on Signal Processing*, 51(4):1 069-1087, 2003.

[18] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. on Image Processing*, 6(12):1673-1687, 1997.

[19] Karen Su, Deepa Kundur, and Dimitrios Hatzinakos, "Statistical Invisibility for Collusion-resistant Digital Video Watermarking,," IEEE Transactions on Multimedia, `http://www.srcf.ucam.org/~ks349/research.php3`

[20] D. S. Taubman and M. W. Marcellin, *JPEG2000 - Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, pp.6, 2001.