RESISTANCE OF ORTHOGONAL GAUSSIAN FINGERPRINTS TO COLLUSION ATTACKS

Z. Jane Wang, Min Wu, Hong Zhao, and K. J. Ray Liu

ECE Department and Institute for Systems Research
University of Maryland
Email: wangzhen, minwu, hzhao, kjrliu@eng.umd.edu

Wade Trappe

WINLAB, ECE Dept.,
Rutgers University
Email: trappe@winlab.rutgers.edu

ABSTRACT

Digital fingerprinting is a means to offer protection to digital data by which fingerprints embedded in the multimedia are capable of identifying unauthorized use of digital content. A powerful attack that can be employed to reduce this tracing capability is collusion. In this paper, we study the collusion resistance of a fingerprinting system employing Gaussian distributed fingerprints and orthogonal modulation. We propose a likelihood-based approach to estimate the number of colluders, and introduce the thresholding detector for colluder identification. We first analyze the collusion resistance of a system to the average attack by considering the probability of a false negative and the probability of a false positive when identifying colluders. Lower and upper bounds for the maximum number of colluders K_{max} are derived. We then show that the detectors are robust to different attacks. We further study different sets of performance criteria.

1. INTRODUCTION

Due to the ease with which digital content can be accessed, retrieved and manipulated, there is a demand for methods to protect digital media and facilitate digital rights management. Digital fingerprinting is one such technique, whereby some unique information, such as a serial number, is embedded in media using watermarking techniques. One powerful class of attacks is *collusion*, whereby a coalition of users combine their different marked copies of the same multimedia content in an attempt to attenuate/remove the trace of any original fingerprint. The fingerprint must therefore survive both standard distortions and collusion attacks by users intending to destroy it. Several methods have been proposed in the literature to embed and hide fingerprints (watermarks) in different media [2, 3, 5]. The spread spectrum watermarking method proposed in [3], where the watermarks have a component-wise Gaussian distribution and are statistically independent, was argued to be highly resistant to collusion attacks [3, 6].

The research on the collusion-resistant fingerprinting systems can be broadly divided into two main directions. One direction focuses on designing collusion-resistant fingerprint codes [1, 9, 10]. The other direction of research is on examining the resistance performance of specific watermarking schemes under different attacks. We are aware of only a few works on the collusion resistance of digital watermarks [4, 6, 7, 8]. Proposing a simple linear collusion attack that consists of adding noise to the average of K independent copies, the authors concluded in [6] that $O(\sqrt{N/\log n})$ independently marked copies are sufficient for an attack to defeat the underlying system with non-negligible probability, when Gaussian watermarks are considered. It was further shown [6] to be optimal: no other watermarking scheme can offer better collusion resistance. These results are also supported by [4]. Stone suggested that the most powerful attack may succeed to defeat uniformly distributed watermarks if as few as one to two dozen independent copies are available [8]. We do not know of any work that provides a precise analysis of the collusion resistance of watermarks when employed with different possible detection schemes. This paper will address this issue. We employ some basic assumptions in this paper:

- We consider independent Gaussian watermarks. Furthermore, we assume that the fingerprints use orthogonal modulation, or at least the correlations among different fingerprints can be ignored.
- A non-blind detection scenario is assumed, meaning that the host signal is available in the detector side. Analysis shows that 2 or 3 independent copies may defeat watermarks under blind scenario.
- The additive distortion is modeled as *iid* Gaussian noise.

We begin, in Section 2, with the problem description and propose an approach to estimate the number of colluders. We then introduce the thresholding detector, and examine the collusion resistance of our fingerprinting system when considering the average attack and the criteria represented by the probabilities of a false positive and a false negative. In Section 4, we further examine other types of collusion and two more sets of performance criteria. We refer the interested readers to [11] for all detailed derivations.

2. A CLASSIFIER APPROACH

Additive embedding is a widely used watermarking scheme. As shown in Figure 1, the content owner has a family of watermarks, denoted by $\{\mathbf{s}_j\}$ and they are fingerprints associated with different users, for distributing marked copies to different users and allowing tracing of pirated copies to their original users. For the j^{th} user, the marked version of the content $\mathbf{y}_j$ is computed by adding the watermark $\mathbf{s}_j$ to the host signal $\mathbf{x}$. Now the observed content $\mathbf{y}$ after the average collusion is

$$\mathbf{y} = \frac{1}{K}\sum_{j \in S_c} \mathbf{y}_j + \mathbf{d} = \frac{1}{K}\sum_{j \in S_c} \mathbf{s}_j + \mathbf{x} + \mathbf{d} \qquad (1)$$

where all vectors have dimension N, K is the number of colluders, and S_c indicates a subset with size K, where $S_c \subseteq [1, ..., n]$ with n be the total number of users. The normally distributed fingerprint $\mathbf{s}_j$ for each user j is assumed to have the equal energy and be orthogonal to each other. The distortion $\mathbf{d}$ is assumed to be an N-dimensional vector following an *iid* $N(0, \sigma_d^2)$ distribution.

0-7803-7663-3/03/$17.00 ©2003 IEEE

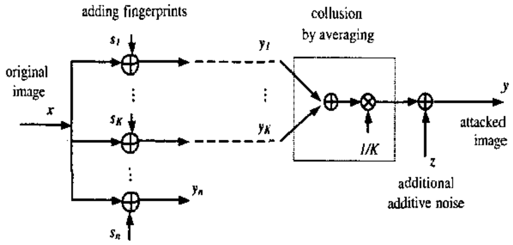


Fig. 1. Model for collusion by averaging.

Here the number of colluders K and the subset S_c are unknown parameters. We have $H_K:\ \mathbf{y}=\frac{1}{K}\sum_{j\in S_c}\mathbf{s}_j+\mathbf{x}+\mathbf{d}$ for $1\le K\le K_m$. To estimate K, we classify an observation $\mathbf{y}$ into one of K_m classes and estimate S_c correspondingly, by the MAP or Bayesian classifier

$$(\hat{K},\hat{S_c}) = arg\max_{K,S_c} p(\mathbf{y}|H_K,S_c)p(H_K)p(S_c|H_K) \qquad (2)$$

where $p(.)$ represents likelihood functions. We choose a noninformative prior such that $p(H_K)p(S_c|H_K)$ is constant and thus can be ignored as long as $|S_c|=K$ is satisfied. Due to the nonblind assumption, the host signal $\mathbf{x}$ is always subtracted from $\mathbf{y}$ for analysis. Because of the orthogonality of basis $\{\mathbf{s}_j\}$, it suffices to consider the correlator vector $\mathbf{T}_N$, with

$$T_N(j)=(\mathbf{y}-\mathbf{x})^T\mathbf{s}_j/\sqrt{\|\mathbf{s}_j\|^2} \qquad (3)$$

for $j=1,...,n$. It is straightforward to show that

$$p(T_N(j)|H_K,S_c)=\begin{cases} N(\frac{\|\mathbf{s}\|}{K},\sigma_d^2), & \text{if } j\in S_c,\\ N(0,\sigma_d^2), & \text{otherwise}\end{cases} \qquad (4)$$

where $\|\mathbf{s}\|=\|\mathbf{s}_j\|$ for all j, and $T_N(j)$ is independent of each other. Now the classifier is equivalent to

$$\begin{aligned}(\hat{K},\hat{S_c}) &= arg\max_{K,S_c} p(\mathbf{y}|H_K,S_c) = arg\max_{K,S_c} p(\mathbf{T}_N|H_K,S_c),\\ thus\ \hat{K} &= arg\max_{K}\{\frac{2\parallel\mathbf{s}\parallel}{K}\sum_{j=1}^{K}T_N^{(j)}-\frac{\parallel\mathbf{s}\parallel^2}{K}\},\\ \hat{S}_c &= \text{the index of } \hat{K} \text{ largest } T_N(j)\text{'s} \qquad (5)\end{aligned}$$

where $T_N^{(j)}$'s are the order statistics of the sample $\mathbf{T}_N$ such that $T_N^{(1)}\ge T_N^{(2)}\ge\cdots\ge T_N^{(n)}$.

3. DETECTION APPROACHES

In this section, we consider one of the most popular criteria, the probability of a false negative (P_{fn}) and the probability of a false positive (P_{fp}). A detection approach fails if either the detector fails to identify any of the colluders (a false negative) or the detector falsely indicates that an innocent user is a colluder (a false positive) [4, 6]. It is desirable to minimize P_{fn}, with a given P_{fp}. Although it might be interesting to study P_{fn} and P_{fp} for the approach introduced in Section 2 and use $\hat{S}_c$ obtained via (5) to indicate colluders, the approach is not designed to address the desirable goal represented by P_{fn} and P_{fp}, and furthermore it lacks the capability of adjusting parameters to meet a given P_{fp}. Next we introduce the thresholding detector and study its collusion resistance under the average attack.

We employ the traditional correlator $T_N(j)$ and compare it to a threshold h, and report that the $j-th$ fingerprint is present if $T_N(j)$ exceeds h. This simple approach is described as

$$\hat{\mathbf{j}}=\arg_{j=1,...,n}\{T_N(j)\ge h\} \qquad (6)$$

where the set $\hat{\mathbf{j}}$ indicates the indices of colluders, and an empty set means that no user is accused. The threshold h here is determined by such parameters as the document length N, the total number of users n, the number of colluders K, and the WNR.

3.1. Performance Analysis

The threshold h in test (6) is chosen to yield $P_{fp}=\epsilon$, where ϵ is a desired small value. For simplicity, we assume that the number of colluders K is known. We now have

$$\begin{aligned}P_{fp} &= P_r\{\hat{\mathbf{j}}\cap\bar{S}_c\neq\emptyset\}=1-(1-Q(h/\sigma_d))^{n-K} \qquad (7)\\ P_d &= P_r\{\hat{\mathbf{j}}\cap S_c\neq\emptyset\}=1-\left(1-Q\left(\frac{h-\|\mathbf{s}\|/K}{\sigma_d}\right)\right)^K\end{aligned}$$

where $\bar{S}_c$ is the complementary set of S_c. According to (8), we can numerically calculate h to yield $P_{fp}=\epsilon$ with given K, n, and WNR, and then compute the corresponding P_d.

We illustrate the resistance performance using an example, where $WNR=0dB$, $N=10^4$, and $\sigma_d^2=1$. In this example, the system requirements are defined as $P_d\ge 0.8$ and $P_{fp}\le 10^{-3}$. As shown in Figure 2(a) and (b), when the number of users n is on the order of 10^4, the fingerprinting system can resist to up to 28 colluders; when n is set as a small number 75, the system can resist to up to 46 colluders. This behavior can be intuitively explained by the expressions of P_{fp} and P_d in (8). To have an overall understanding of the collusion resistance of this orthogonal fingerprinting scheme, we plot the maximum resistible number of colluders K_{max} as a function of the total number of users n in Figure 3. It is noted that the system can resist to up to n colluders when the total number of users n is less than 60. However, for a system accommodating more than 60 users, its collusion resistance starts to decrease. For a system accommodating more than one thousand users, the number K_{max} is around 28.

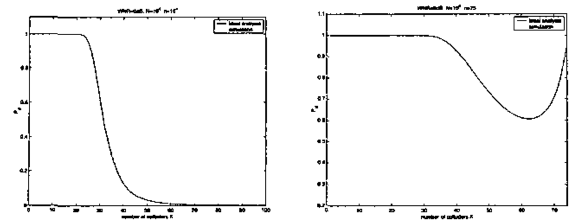
3.2. Lower and Upper Bounds of K_{max}

Since the above analysis is based on numerical computation, we shall study analytic bounds on the maximum number of colluders K_{max} for an orthogonal fingerprinting system employing the thresholding detector.

Setting $\sigma_d^2=1$ for convenience, note that now $\|s\|=\sqrt{\eta N}$ with the WNR $\eta=\|s\|^2/\|d\|^2$. We restate the system requirements as

$$P_{fp}\le\epsilon, \quad P_d\ge\beta, \qquad (8)$$

in which ϵ is a small number and β is close to 1. A key point in determining K_{max} is to figure out the appropriate threshold h in the above equation (8). The assumption that ϵ is small implies that the choice of h can meet the condition $Q(h)<<1/n$. Based on this observation and the two Lemmas, we obtain a lower bound h_L and upper bound h_H of the threshold h. We now proceed to show that a lower and upper bound of the maximum number of



(a) (b)

Fig. 2. Probability of detection P_d as a function of the number of colluders K when apply the thresholding detector in (6). Here $WNR = 0dB$, $N = 10^4$ and $P_{fp} \leq 10^{-3}$. In figure(a) the number of user n is 10^4; in (b) $n = 75$.

colluders K_{max} can be obtained by using the bounds of h. The basic idea is to find a lower bound K_L of K_{max} such that the resulting pair (K_L, h_H) simultaneously satisfies the conditions that the corresponding P_d is larger than but close to the requirement β and P_{fp} is smaller than but close to the requirement ϵ. Similarly, an upper bound K_H is chosen such that the pair (K_H, h_L) results in a P_d, which is smaller than but close to the requirement β, and a P_{fp}, which is larger than but close to the requirement ϵ. A detailed derivation leads to the following collusion resistance:

$$K_{max} \geq \min\{n, K_L\}, \qquad K_L = \sqrt{\frac{\eta N}{\log\left(\frac{n^2}{2\pi\epsilon^2 \log(0.5n^2/\pi)}\right)}}$$

$$K_{max} \leq \min\{n, K_H\}, \qquad K_H = \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1 - \sqrt[\tilde{K}]{1-\beta})} \quad (9)$$

where $Q^{-1}(.)$ represents the inverse Q-function, and $\tilde{K}$ serves as an upper bound of the upper bound K_H: $\tilde{K} = \frac{\sqrt{\eta N}}{h_L - Q^{-1}(1-\sqrt[n]{1-\beta})}$. It is worth mentioning that a tighter lower and upper bound of K_{max} can be obtained by solving the one-dimensional problem $P_d = \beta$ when h_H and h_L are considered, respectively. However, more computational load will be involved and no explicit expressions of K_H and K_L as in (9) be available due to the complex nature of P_d.

We plot the lower and upper bound of K_{max} versus the number of users n, along with the numerical K_{max}, in Figure 3, where $\sigma_d^2 = 1$, $WNR = 0dB$, $N = 10^4$, and the requirements $P_{fp} \leq 10^{-3}$ and $P_d \geq 0.8$. It is noted that the lower and upper bounds are within a factor of 2 of the true value of K_{max}. Some interesting observations are noted from this example. From the attacker point of view, if an attack can only collect up to 20 copies, he/she can never succeed in removing all the traces; however, an attacker is guaranteed to celebrate his/her success if 80 independent copies are available. From the owner (detector) point of view, if the owner has a mean to ensure that a potential attacker has no way to obtain as many as 20 independent copies, the fingerprinting system is claimed to be collusion-free. Meanwhile, in order to maximize the worst case of P_d, the owner should limit the number of independent distributions.

3.3. Simulations

Since the knowledge of K is normally not available in practice, we need to first estimate K before setting a threshold h for the detection process. Our simulations used the following implementation:

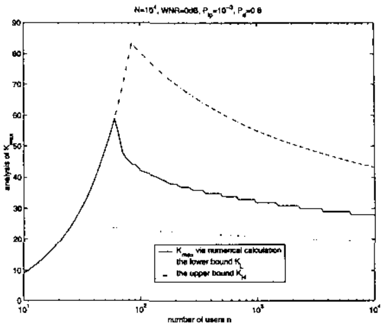


Fig. 3. The lower and upper bound of K_{max} as a function of the number of users n when apply the thresholding detector in (6). Here $WNR = 0dB$, $N = 10^4$, $\epsilon = 10^{-3}$ and $\beta = 0.8$.

1. Estimate the number of colluders K via (5).
2. Determine the threshold h correspondingly to yield a desired P_{fp}, according to (8). The threshold h is only a function of $\hat{K}$ when N, WNR and n are given.
3. Apply the thresholding test statistic described in (6).

We compare the simulation results with the ideal performance in Figure 2(a) and (b). When K is estimated based on simulated observations, the resulting P_d always decreases with increasing K. A good match is observed over the non-increasing part of the ideal case (when K is small). Mis-match is noted over the increasing part of the ideal case (when K is close to n), since K is under-estimated in this situation due to the increasing overlap between the two Gaussian distributions $N(0, \sigma_d^2)$ and $N(\|s\|/K, \sigma_d^2)$ as K increases. However, estimating K does not significantly affect the results of K_{max}, compared with that of the ideal performance analysis, since only the non-increasing part (also the matched part) of the ideal case in the P_d versus K curve is evaluated to decide K_{max}.

4. EXTENSIONS

In this section, we consider three nonlinear attacks suggested by Stone in [8]. We show in [11] that different attacks provide very close performance as long as the powers of the composite observations satisfy

$$E\{\| \mathbf{y}_g \|^2\} = E\{\| \mathbf{y}_{mean} \|^2\} \triangleq \xi_0 \qquad (10)$$

where $g(.)$ represents the attack operation. Note that the power of the observation indicates the level of MSE introduced to the host signal. The above fact that, from the detector point of view, different attacks provide close performance suggests that with the same MSE distortion allowed, the average attack is most efficient from the attacker point of view. This is because from the detector point of view, there exists better detection schemes than the thresholding detector given a specific attack except the average attack.

Different goals arise under different situations, and there are other possible sets of performance measures. These measures provide different balance between capturing colluders and accusing innocents. We consider two new sets of performance criteria and study the thresholding detector under the average attack.

Case 1: Capture More This set of performance criteria consists of the expected fraction of colluders that are successfully captured, denoted as r_c, and the expected fraction of innocent users that are falsely placed under suspicion, denoted as r_i. Here the major concern is to catch more colluders, possibly at a cost of accusing more innocents. The system requirements are represented as

$$r_i = Q(h/\sigma_d) \leq \alpha_i; \quad r_c = Q\left(\frac{h- \parallel s \parallel /K}{\sigma_d}\right) \geq \alpha_c. \quad (11)$$

We obtain the following

$$\begin{aligned} h &= Q^{-1}(\alpha_i)\sigma_d, \\ K_{max} &= \frac{\sqrt{\eta N}}{Q^{-1}(\alpha_i) - Q^{-1}(\alpha_c)}. \end{aligned} \quad (12)$$

It is interesting to note that the threshold h is a constant value determined by α_i, and K_{max} is not affected by the total number of users n. If placing a larger fraction of innocents into suspicion is allowed, the system can resist to more colluders.

Case 2: Capture All This set of performance criteria consists of the efficiency rate R, which describes the amount of expected innocents accused per colluder, and the probability of capturing all K colluders, referred as P_d. Here the goal is to capture all colluders with a high probability. The tradeoff between capturing colluders and placing innocents under suspicion is through the adjustment of the efficiency rate R. The system requirements are expressed as

$$R = \frac{(n-K)Q(h/\sigma_d)}{KQ(\frac{h-\|s\|/K}{\sigma_d})} \leq \alpha; \quad P_d = Q\left(\frac{h- \parallel s \parallel /K}{\sigma_d}\right)^K \geq \beta. \quad (13)$$

We may find lower and upper bounds for K_{max} under this criteria, and an example is given in Figure 4.

The analysis in this section reveals that the maximum number of colluders allowed is on the same order under two different sets of criteria. Basically, a few dozen of colluders could break down the Gaussian fingerprinting system using orthogonal modulation by generating a new composite copy such that the identification of the original fingerprints will unlikely be successful.

5. CONCLUSION

In this paper, we investigated how many independently marked copies of the same multimedia content is required for an attacker to thwart a fingerprinting system. We studied the collusion resistance of a fingerprinting system to the average attack when considering the performance criteria represented by P_{fp} and P_{np}. We derived lower and upper bounds of the maximum number of colluders K_{max}. Using the upper bound, an attacker can know how many independent copies are required to guarantee the success of a collusion attack; on the other hand, an owner will benefit from these bounds in designing a fingerprinting system. Our work was further extended to different attacks and performance criteria. From the detector point of view, the thresholding detector is robust to different attacks, since different attacks yield very close performance as long as the levels of MSE distortion introduced by different attacks are the same. And it seems that attacks based on a few dozen independent copies will confound a fingerprinting system accommodating as many as ten thousand users. This observation suggests that the number of independently marked copies of the same content that can be distributed should be determined by many concerns, such as the system requirements, and the cost of obtaining multiple independent copies. Furthermore, it suggests that tracing colluders via fingerprints should work in concert with other operations, for example, suspecting a user leads the owner to more closely monitor that user and further gather other evidences.

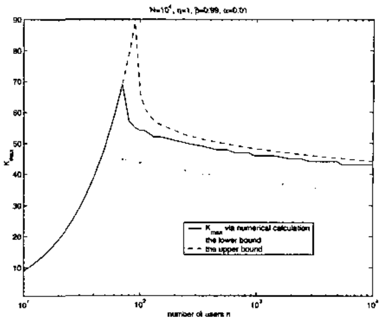


Fig. 4. Resistance performance of the orthogonal fingerprinting system under the criteria R and P_d. Here $N = 10^5$, $\eta = 1$, $\alpha = 0.01$ and $P_d = 0.99$.

6. REFERENCES

[1] D. Boneh and J. Shaw, "Collusion-Secure Fingerprinting for Digital Data", *IEEE Trans. on Information Theory*, Vol. 44, pp. 1897-1905, 1998.

[2] I. Cox, J. Bloom and M. Miller, *Digital Watermarking: Principles & Practice*, Morgan Kaufman Publishers, 2001.

[3] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Trans. on Image Proc.*, vol. 6, no. 12, pp. 1673-87, 1997.

[4] F. Ergun, J. Kilian and R. Kumar, "A Note on the limits of Collusion-Resistant Watermarks", in *Eurocrypt'99*, pp. 140-49, 1999.

[5] F. Hartung and M. Kutter, "Multimedia Watermarking Techniqures", *Pro. of IEEE*, Vol. 87, pp. 1079-1107, July 1999.

[6] J. Kilian, T. Leighton, L. Matheson, T. Shamoon, R. Tarjan, and F. Zane, "Resistance of Digital Watermarks to Collusive Attacks", *Proc. IEEE International Symposium on Information Theory*, pp. 271, Aug. 1998.

[7] J. Su, J. Eggers, and B. Girod, "Capacity of Digital Watermarks Subjected to An Optimal Collusion Attack", *Pro. EUSIPCO 2000*, Vol. 4, Sep. 2000.

[8] H. Stone, "Analysis of Attacks on Image Watermarks with Randomized Coefficients", Tech. Rep. 96-045, NEC Research Institute, Princeton, NJ, 1996.

[9] W. Trappe, M. Wu, Z. Jane Wang and K. J. R. Liu, "Anti-Collusion Fingerprinting for Multimedia", *to appear IEEE Trans. on Signal Processing, Special issue on Signal Processing for Data Hiding in Digital Media & Secure Content Delivery*, Feb. 2003.

[10] Y. Yacobi, "Improved Boneh-Shaw Content Fingerprinting", CT-RSA 2001, LNCS 2020, pp. 378-91, Springer-Verlag Berlin Heidelberg, 2001.

[11] Z. Jane Wang, M. Wu, H. Zhao, W. Trappe, and K. Ray Liu, "Collusion Resistance of Multimedia Fingerprinting Using Orthogonal Modulation",*submitted to IEEE Trans. on Image Processing*, Nov. 2002.