ON OPTIMAL COLLUSION STRATEGIES FOR FINGERPRINTING

Negar Kiyavash and Pierre Moulin

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Email: kiyavash@uiuc.edu, moulin@ifp.uiuc.edu

ABSTRACT

We study the theoretical performance of linear and non-linear collusion attacks under the assumptions that orthogonal or regular-simplex fingerprints are used, and that the detector performs a linear correlation test in order to decide whether a user of interest is among the colluders. The colluders create a noise-free forgery by applying a mapping f to their individual copies, and then add a noise sequence $\mathbf{e}$ to form the actual forgery. They seek the mapping f and the distribution of $\mathbf{e}$ that maximize the probability of error of the detector. The performance of mappings such as linear-averaging and interleaving can be compared in this framework. It is also shown that impulsive noise attacks are far more effective than Gaussian attacks.

1. INTRODUCTION

Digital fingerprinting schemes are devised for traitor tracing. In applications such as copyright protection, the goal is to deter users from illegally redistributing the digital content. Each user is provided with his own individually marked copy of the content. Although this makes it possible to trace an illegal copy to a traitor, it also allows for users to collude and form a stronger attack. One form of such attacks is linear averaging, where the colluders average their copies and add noise to create a *forgery*. Averaging reduces the power of each fingerprint and makes the detector's task harder. Another collaborative attack is interleaving, where the colluders form a pre-forgery by contributing samples from their copies and contaminating the pre-forgery with noise. Unlike linear averaging, interleaving is not a linear attack.

The averaging plus noise attack has been studied in numerous works see e.g. [1, 2, 3]. But the noise model is generally assumed to be i.i.d. Gaussian. The interleaving attack has been addressed extensively in [4] where the noise model is still assumed to be Gaussian. One may ask whether Gaussian noise is the most malicious noise the attackers can introduce. This question has not yet been answered in the fingerprinting literature, so it is conceivable that other types of noise are worse.

This research was supported in part by NSF grant CCR 03-25924.

In our problem setup, the fingerprinting scheme is additive, and the fingerprints are chosen from an orthogonal or a regular simplex constellation. The detector has access to the host signal (non-blind detection) and performs a *binary hypothesis test* to verify whether a user of interest is colluding. The cost function in this problem is the detector's probability of error. The main contribution of this paper is to show that under the attack model (1), the best strategy for the colluders is to perform linear averaging of their copies followed by addition of a particular type of impulsive noise. This strategy is far better than strategies that use i.i.d. Gaussian noise or replace the linear averaging by interleaving attacks.

Throughout this paper, we use uppercase letters to denote random variables, lowercase for their individual values, boldface for sequences and vectors, and calligraphic fonts for sets. We use the symbol $\mathbb{E}$ to denote mathematical expectation, and the asymptotic equality notation $f(n) \sim g(n)$ to indicate that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. We also write $f(n) \ll g(n)$ to indicate that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$, and $f(n) \doteq g(n)$ to indicate asymptotic equality on the logarithmic scale: $\ln f(n) \sim \ln g(n)$.

2. PROBLEM STATEMENT

In this section we describe the mathematical setup of the problem, which is also diagrammed in Fig. 1.

2.1. Fingerprint Embedding

The host signal is a sequence $\mathbf{S} = (S(1), \ldots, S(N))$ in $\mathbb{R}^N$, viewed as deterministic but unknown to the colluders. Fingerprints are added to $\mathbf{S}$, and the marked copies of the signal are distributed to L users. Specifically, user j is assigned a marked copy

$$\mathbf{X}_j = \mathbf{S} + \mathbf{Q}_j \qquad j \in \{1, \ldots, L\},$$

where $\mathbf{Q}_j$ denotes the fingerprint assigned to him.

The L fingerprints form a constellation in $\mathbb{R}^N$. In our setup, the fingerprints are equienergetic, i.e., the constellation lies on the N-dimensional sphere with radius $\sqrt{ND_f}$, where D_f is the embedding distortion per sample. Therefore

$$\|\mathbf{X}_j - \mathbf{S}\|^2 = \|\mathbf{Q}_j\|^2 = ND_f, \quad \forall j.$$

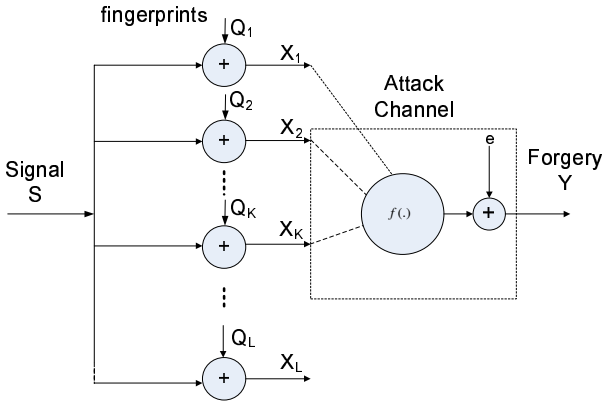

Fig. 1. The fingerprinting process and the attack channel.

For instance, the fingerprints could be orthogonal, or they form a regular simplex [3]. In both cases, $L = N$.

2.2. Attack Model

Throughout this paper, we assume that the colluders know the type of constellation from which their fingerprints were drawn; however the orientation of this constellation is randomized uniformly over the N-dimensional sphere. So the colluders do not know their individual fingerprints. The attacks are of the form

$$\mathbf{Y} = f\left(\mathbf{X}_k, k \in \mathcal{J}\right) + \mathbf{E} \tag{1}$$

where $\mathcal{J}$, the *coalition*, is the index set of the colluding users. The set $\mathcal{J}$ has cardinality $K \leq L$. In this paper, we study the case of large N, with $K \ll L = N$. It is shown in [3] that for $K > \sqrt{N}$ reliable detection is impossible when $\mathbf{E}$ is i.i.d. Gaussian noise.

The mapping $f : \mathbb{R}^{N|\mathcal{J}|} \to \mathbb{R}^N$ is symmetric in its arguments, i.e., any permutation of the index set $\mathcal{J}$ does not change the value of f. We view f as a "noise-free forgery" (to which noise $\mathbf{e}$ is added to form the actual forgery, $\mathbf{Y}$), and the symmetry condition as a *fairness condition*: all members of the coalition incur equal risk. We shall consider two special instances of f satisfying the above fairness condition: the linear averaging forgery, where the colluders average their marked signals; and the interleaving forgery, where each colluder contributes N/K samples of his own copy to form the forgery.

Moreover, in (1) we may view $f\left(\mathbf{X}_k, k \in \mathcal{J}\right)$ as an estimator of the signal $\mathbf{S}$ based on the copies available to the coalition. The noise $\mathbf{E}$ represents an actual degradation of the signal. It is modelled as a length-N vector drawn from a probability distribution function (pdf) $p_{\mathbf{E}}$ with zero mean. The expected squared norm of $\mathbf{E}$ is

$$\int_{\mathbb{R}^N} \|\mathbf{e}\|^2 p_{\mathbf{E}}(\mathbf{e}) d\mathbf{e} = N\sigma^2. \tag{2}$$

The mean-squared distortion of the forgery $\mathbf{Y}$ relative to the host signal $\mathbf{S}$ is given by

$$\mathbb{E}\|\mathbf{Y} - \mathbf{S}\|^2 = ND_c \tag{3}$$

where D_c is the average distortion per sample introduced by the coalition. Under the attack model (1), the total distortion (3) can be decomposed as

$$\mathbb{E}\|\mathbf{Y} - \mathbf{S}\|^2 \quad = \|f\left(\mathbf{X}_k, k \in \mathcal{J}\right) - \mathbf{S}\|^2 + \mathbb{E}\|\mathbf{E}\|^2, \tag{4}$$

and thus $D_c \geq \sigma^2$. The difference

$$D_c - \sigma^2 = \frac{1}{N}\|f\left(\mathbf{X}_k, k \in \mathcal{J}\right) - \mathbf{S}\|^2$$

represents the mean-square estimation error.

2.3. Correlation Detector

The host signal $\mathbf{S}$ is available at the detector and can be subtracted from $\mathbf{Y}$, to form the centered content $\mathbf{Y} - \mathbf{S}$. The detector performs a binary hypothesis test to determine whether a specific user's mark is present in the forgery. We shall call this detector *focused*, because it decides whether a particular user of interest is a colluder [3]. It does not aim at identifying all colluders. The *focused* detector above does not even need to know K, the number of the colluders. (However its performance depends strongly on K.)

Assume that the detector is focused on user j. Our detector compares the correlation statistic $T(\mathbf{Y})$ with a threshold τ:

$$T(\mathbf{Y}) = \mathbf{Q}_j^T(\mathbf{Y} - \mathbf{S}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5}$$

where H_1 and H_0 respectively denote the "guilty" and "innocent" hypotheses. The decision boundary for this test is a hyperplane normal to the vector $\mathbf{Q}_j$:

$$\Omega = \{\mathbf{Y} \ : \ \mathbf{Q}_j^T(\mathbf{Y} - \mathbf{S}) = \tau\}.$$

For mapping such as linear averaging and interleaving

$$f\left(\mathbf{X}_k, k \in \mathcal{J}\right) = \mathbf{S} + f\left(\mathbf{Q}_k, k \in \mathcal{J}\right). \tag{6}$$

Due to (1) and (6), the pdf's of $T(\mathbf{Y})$ under H_0 and H_1 are translations of each other, with respective means 0 and $\mathbf{Q}_j^T f\left(\mathbf{Q}_k, k \in \mathcal{J}\right)$.

The threshold τ trades off the probabilities of false alarm and miss P_F and P_M. In this paper, the threshold τ is chosen to minimize the probability of error, assuming equal priors on the "guilty" and "innocent" hypothesis.

The coalition wants to design the mapping f and the noise pdf $p_{\mathbf{E}}$ to maximize the probability of error, subject to the distortion constraint D_c.

3. LINEAR AVERAGING ATTACK

In this section we consider the linear averaging attack of a coalition of size K, and optimize the noise pdf $p_{\mathbf{E}}$. The mapping f is given by

$$f\left(\mathbf{X}_k, k \in \mathcal{J}\right) = \frac{1}{|\mathcal{J}|} \sum_{k \in \mathcal{J}} \mathbf{X}_k.$$

The resulting mean-squared estimation error is given by

$$\left\| f\left(\mathbf{X}_k, k \in \mathcal{J}\right) - \mathbf{S} \right\|^2 = \left\| \frac{1}{K} \sum_{k \in \mathcal{J}} \mathbf{Q}_k \right\|^2$$

which is equal to $\frac{N}{K} D_f$ for orthogonal fingerprints, and to $\frac{N}{K}(1 - K/L) D_f$ for simplex fingerprints.

The detector performs the correlation test of (5). It is shown in [3] that if the fingerprints $\mathbf{Q}_j$ are chosen from a regular simplex constellation, then

$$\tau = \frac{L - 2K}{2K(L-1)} N \sim \frac{N}{2K}.$$

For orthogonal fingerprints, we have $\tau = \frac{N}{2K}$ (exactly). Notice that the value of the threshold τ is fixed and does not depend on the noise pdf $p_{\mathbf{E}}$.

The coalition designs $p_{\mathbf{E}}$ such that the detector's probability of error is maximized. Since the orientation of the fingerprint constellation is uniform on the N-dimensional sphere, so is the direction of the vector normal to the decision boundary Ω. The best strategy for the colluders under these circumstances is to choose an isotropic pdf $p_{\mathbf{E}}$. The magnitude r of the noise vector $\mathbf{E}$ has pdf $p_R(r)$. The distortion (2) due to $\mathbf{E}$ therefore takes the form

$$\int_0^\infty r^2 p_R(r) dr = N\sigma^2. \tag{7}$$

The probability of error of the detector can be expanded by conditioning over the radial random variable R:

$$\int_0^\infty P_e(r) p_R(r) dr = P_e, \tag{8}$$

where $P_e(r)$ denotes the error probability conditioned on the event $\|\mathbf{E}\| = r$. The coalition's program is to maximize (8) over the radial pdf p_R subject to the constraint (7). By the fundamental theorem of linear programming, the optimal p_R is a mass distribution with support at two points only. It can be shown (derivations are omitted here) that the first point is at $r = 0$. Therefore the optimal radial pdf for the coalition takes the form

$$p_R(r) = (1 - \epsilon)\delta(0) + \epsilon \delta(r - r_0). \tag{9}$$

The distortion constraint (7) implies that $\epsilon r_0^2 = N\sigma^2$; similarly, from (9), we have $P_e = \epsilon P_e(r_0)$.

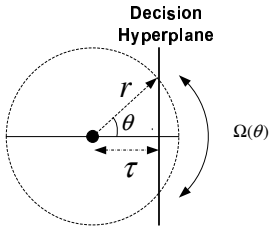

Fig. 2. Decision boundary τ and norm r of noise vector $\mathbb{E}$.

Conditioned on $R = r$ and given the threshold τ, the decision boundary Ω of (5) cuts a spherical cap away from the sphere of radius $R = r$. Figure 2 shows the decision boundary and the corresponding spherical cap. The half angle corresponding to the spherical cap is denoted by θ, and $\cos\theta = \frac{\tau}{r}$. Owing to the isotropic nature of the noise, we have $P_e(r) = \frac{\Omega(\theta)}{\Omega(\pi)}$ [5], where $\Omega(\theta)$ is the area of the spherical cap in N dimensions corresponding to the half angle θ. The probability of error is thus given by

$$P_e = \epsilon \frac{\Omega(\theta)}{\Omega(\pi)} = \frac{N\sigma^2}{r_0^2} \frac{\Omega(\theta)}{\Omega(\pi)}.$$

with $\cos\theta = \frac{\tau}{r_0}$, where

$$\frac{\Omega(\theta)}{\Omega(\pi)} = \frac{\sin^N \theta}{\sqrt{2\pi N} \cos\theta \sin\theta} \left[1 + O(1/N)\right].$$

Hence we have

$$P_e \sim N \frac{\sigma^2}{r_0^2} \frac{\sin^{N-1}\theta}{\sqrt{2\pi N}\cos\theta}. \tag{10}$$

The right side of (10) is maximized for a choice of

$$r_0 = \sqrt{N}\tau \sim \frac{N^{3/2}}{2K}.$$

The corresponding maximum value of the probability of error is

$$P_e \sim \frac{4\sigma^2}{\sqrt{2\pi e}} \frac{K^2}{N^2}. \tag{11}$$

Compare with the case of i.i.d Gaussian noise $\mathbf{e} = \mathcal{N}(0, \sigma^2 I_N)$. Then P_e decays *exponentially* with N/K^2 [3]:

$$P_e \doteq \exp\left\{ -\frac{N}{8\sigma^2 K^2} \right\}.$$

Thus, the coalition can form a significantly stronger attack by choosing impulsive noise according to (9), while incurring the same distortion $D_c = \sigma^2 + \frac{1}{K} D_f$.

4. INTERLEAVING ATTACK

Consider the following nonlinear attack. For notational simplicity, assume that N/K is an integer. The K colluders design a partition $\Lambda_k, k \in \mathcal{J}$ of the set $\{1, 2, \cdots, N\}$, where each Λ_k contains exactly N/K samples, and the partition Λ is selected randomly and uniformly over the set of all such partitions. The n-th sample of the noise-free forgery $f(\mathbf{X}_k, k \in \mathcal{J})$ in (1) is equal to X_{kn} whenever $n \in \Lambda_k$. In other words, user k's sample is simply copied onto the noise-free forgery. Therefore f is an interleaving operator. The mean-squared error of this estimator (averaged over all partitions Λ) is equal to D_f. (As expected this estimator is neither better nor worse than any individual copy $\mathbf{X}_k$). Also we have

$$\mathbf{Y} - \mathbf{S} \quad = \quad f(\mathbf{Q}_k, k \in \mathcal{J}) + \mathbf{E}.$$

The n-th output sample of f is equal to Q_{kn} whenever $n \in \Lambda_k$.

Given the noise pdf $p_{\mathbf{E}}$, the performance of the test (5) depends on f only via the distance between the means of the "guilty" and "innocent" conditional distributions. (This distance was equal to $2\tau = N/K$ for the linear averaging attack of Section 3.) The most confusing case for the detector is derived below. Given a size-K coalition $\mathcal{J}$, either j or some other user j' is part of it. Similarly to [3], denote by $\mathbf{F}$ and $\mathbf{F}'$ the corresponding worst-case (closest) noise-free forgeries; the n-th sample of their difference is given by

$$F_n - F'_n = \begin{cases} Q_{jn} - Q_{j'n} & : \ n \in \Lambda_j = \Lambda_{j'} \\ 0 & : \ \text{else.} \end{cases}$$

The conditional expectation of $T(\mathbf{Y})$ given that coalition $\mathcal{J}$ and partition Λ were used is given by $\mathbb{E}[T(\mathbf{Y})|j \in \mathcal{J}, \Lambda]$ when $j \in \mathcal{J}$ (i.e., when user j is guilty). Likewise, the conditional expectation when $j \notin \mathcal{J}$ (j is innocent) is denoted by $\mathbb{E}[T(\mathbf{Y})|j \notin \mathcal{J}, \Lambda]$. The difference of the two conditional expectations is given by

$$\begin{aligned} &\mathbb{E}[T(\mathbf{Y})|j \in \mathcal{J}, \Lambda] - \mathbb{E}[T(\mathbf{Y})|j \notin \mathcal{J}, \Lambda] \\ &= \quad (\mathbf{F} - \mathbf{F}')^T \mathbf{Q}_j \\ &= \quad \sum_{n \in \Lambda_j} Q_{jn}^2 - \sum_{n \in \Lambda_{j'}} Q_{j'n} Q_{jn} \end{aligned}$$

Integrating out the uniformly distributed random variable Λ and taking into account the fact that $\|\mathbf{Q}_j\|^2 = ND_f$ and $\mathbf{Q}_j \mathbf{Q}_k^T = 0$ (for orthogonal fingerprints), we obtain

$$\mathbb{E}[T(\mathbf{Y})|j \in \mathcal{J}] - \mathbb{E}[T(\mathbf{Y})|j \notin \mathcal{J}] = \frac{N}{K}. \qquad (12)$$

When simplex fingerprints are used, the right-hand side is multiplied by a factor of $\frac{L}{L-1}$ and thus remains essentially unchanged.

Notice that the distance (12) between the two means is the same that was obtained when the noise-free forgery was linear averaging. For any given $p_{\mathbf{E}}$, the performance of the test is therefore the same.

Now recall that the mean-squared error for the "interleaving estimator" f is D_f, i.e., K times larger than that of the linear-average estimator of Sec. 3. Hence if the interleaving attack is to yield the same detection performance as the linear averaging attack, it must introduce excess distortion $D_f(1 - \frac{1}{K})$, no matter what noise pdf $p_{\mathbf{E}}$ is used. In conclusion, the averaging attack outperforms the interleaving attack.

5. DISCUSSION

Our attack model (1) may be viewed as host signal estimation (using an estimator f) followed by the addition of a noise vector $\mathbf{E}$ independent of the input. Our assumptions have allowed us to characterize the worst-case noise pdf $p_{\mathbf{E}}$ for the correlation detector. The worst noise is impulsive, and the performance of the detector is dramatically worse than that obtained under i.i.d. Gaussian noise. Also, for any choice of $p_{\mathbf{E}}$, the detection performance depends on f via the distance between two worst-case conditional means at the detector. The distortion D_c due to the attack is the sum of the distortion due to $\mathbf{E}$ and the mean-squared estimation error. The linear averaging estimator is nearly ideal in this respect. Some improvements may be obtained if a statistical model for $\mathbf{S}$ is available to the coalition and they design an optimal estimator of $\mathbf{S}$ based on these statistics. However, if $\mathbf{S}$ has large entropy (e.g., i.i.d. Gaussian process whose variance is much larger than D_f), these improvements are marginal for large K and the noise term $\mathbf{E}$ dominates the estimation error.

6. REFERENCES

[1] P. Moulin and A. Briassouli. The Gaussian fingerprinting game. *Conference on Information Sciences and Systems, CISS'02*, March 2002.

[2] Z. Wang, M. Wu, H. Zhao, W. Trappe, and K.J.R. Liu. Collusion resistance of multimedia fingerprinting using orthogonal modulation. *IEEE Trans. on Image Proc.*, 14(6):804–821, 2005.

[3] N. Kiyavash and P. Moulin. Regular simplex fingerprints and their optimality properties. In *Proc. International Workshop on Digital Watermarking*, pages 97–109, 2005. Siena, Italy.

[4] Shan He and Min Wu. Performance study on multimedia fingerprinting employing traceability codes. In *Proc. International Workshop on Digital Watermarking*, pages 84–96, 2005. Siena, Italy.

[5] C.E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell System Technical Journal*, 38(3):611–657, 1959.