

A Framework for Optimizing Nonlinear Collusion Attacks on Fingerprinting Systems

Negar Kiyavash

Department of Electrical and
Computer Engineering
University of Illinois at Urbana-Champaign
Coordinated Science Laboratory
Email: kiyavash@uiuc.edu

Pierre Moulin

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

Abstract— This paper develops a mathematical analysis of the performance of order statistic collusion attacks on Gaussian fingerprinting systems. The attacks considered include the popular memoryless averaging and median attacks as special cases. In this model, the colluders create a noise-free forgery by applying an order statistic mapping to each sample of their individual copies, and next they add a Gaussian noise sequence to form the final forgery. The choice of the mapping may be time-dependent and/or random. The performance of a strategy is evaluated in terms of the resulting probability of error of a correlation focused detector, and in terms of the mean-squared distortion between host and forgery. We prove the surprising fact that all the nonlinear attacks considered result in the same detection performance. Moreover, the linear averaging attack outperforms the other ones in the sense of minimizing mean-squared distortion.

I. INTRODUCTION

Due to the proliferation of digital media on private and public networks, there is a pressing need for reliable digital rights management (DRM) techniques. Digital fingerprinting provides attractive solutions for DRM, as well as for traitor tracing applications. For instance, in copyright protection, a provably reliable digital fingerprinting scheme can deter users from illegally redistributing the digital content. In a generic copyright application, each user is provided with his own individually marked copy of the content. This makes it possible to trace an illegal copy to a traitor. However, some users may collude and mount a strong attack on the scheme by comparing their copies and creating a *forgery* that contains only weak evidence of the presence of their fingerprints.

In our problem setup, the fingerprinting scheme is additive, and the fingerprinting code is taken from a Gaussian ensemble of codes. (This does not mean the code is selected randomly; the random-ensemble technique merely allows us to *prove the existence of a fingerprinting code with the desired properties.*) The detector has access to a forgery as well as to the host signal (non-blind detection) and performs a *binary hypothesis test* to determine whether a user of interest was involved in the forgery. There are two cost functions in this problem: the detector's probability of error, which the colluders want to maximize; and the mean-squared distance between the host and the forgery, which the colluders want to minimize.

A simple but often used attack is linear averaging of the

copies: if there are K colluders, the residual power of each fingerprint is only $\frac{1}{K}$ times the original value. Nonlinear attacks are also popular, e.g., each sample of the forgery could be the median value of the corresponding K samples available to the colluders. Instead of computing the median value, the colluders may perhaps output the minimum, or the maximum, or the midpoint (average of min and max) of their samples. All such attacks may be viewed as *order statistic* attacks as they employ well-established methods from statistics and nonlinear signal processing [1], [2]. In addition to the above operations, the colluders may add noise to further confuse the detector.

The averaging plus noise attack has been studied in several papers where the noise model is assumed to be i.i.d. Gaussian [3], [4] or colored Gaussian [5], as well as in [6] where a worst case noise analysis was conducted. Some special cases of the order statistic attacks, including the median, minimum, maximum, and midpoint attacks, have been studied in [7] using a *noise-free forgery* and some approximations to detector performance analysis.¹ Based on these approximations, it was claimed that minimum and maximum attacks are superior to the median and averaging attacks. In contrast, we prove that for Gaussian fingerprints and in presence of Gaussian noise, the above attacks are all *equivalent* in terms of maximizing the *probability of error* of the detector. Moreover, among these attacks, the averaging attack outperforms all the other attacks in terms of minimizing the *attackers' distortion*. The advantage of the linear averaging attack holds even if the class of unbiased order-statistic filters is extended to include time-varying, signal-dependent, and randomized versions of such filters.

Throughout this paper, we use boldface uppercase letters to denote random vectors, uppercase letters for the components of the vectors, and calligraphic fonts for sets. We use the symbol \mathbb{E} to denote mathematical expectation. For any random vector $\mathbf{X} : \{X_1, \dots, X_K\}$, we denote by $X_{(i)}$ the i -th smallest component of vector \mathbf{X} or the i -th order statistic of vector \mathbf{X} . For instance, $X_{(1)}$ denotes the minimum component of vector \mathbf{X} , and $X_{(K)}$ its maximum component. The symbol

¹In the absence of attacker's noise, the only source of uncertainty at the detector is which coalition created the forgery. The approach used in [7, Eqn (5)] was to use a Gaussian approximation to model this uncertainty.

$f(N) \ll g(N)$ means that $\lim_{N \rightarrow \infty} \frac{f(N)}{g(N)} = 0$. The Gaussian distribution with mean θ and variance σ^2 is denoted by $\mathcal{N}(\theta, \sigma^2)$.

II. PROBLEM STATEMENT

The mathematical setup of the problem is diagrammed in Fig. 1.

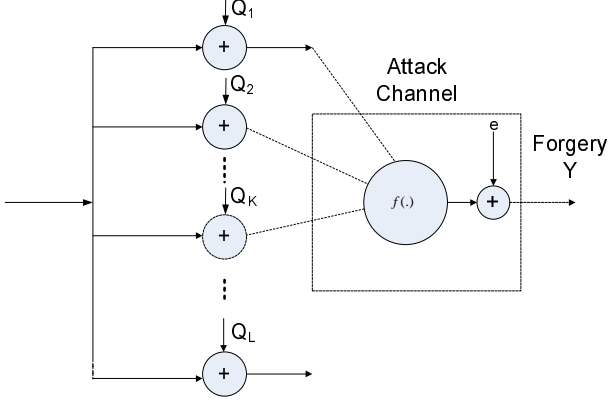


Fig. 1. The fingerprinting process and the attack channel.

A. Fingerprint Generation and Embedding

The host signal is a sequence $\mathbf{S} = (S(1), \dots, S(N))$ 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

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

where the fingerprint $\mathbf{Q}_j = (Q_j(1), \dots, Q_j(N))$ is a N -vector.

The fingerprints $\mathbf{Q}_1, \dots, \mathbf{Q}_L$ form a (N, L) fingerprinting code \mathcal{C} . The code \mathcal{C} is selected from a random ensemble of codes. Each code in this ensemble is made of L i.i.d $\mathcal{N}(0, I_N)$ fingerprints, generated independently of the host signal \mathbf{S} . Therefore,

$$\mathbb{E}[\|\mathbf{X}_j - \mathbf{S}\|^2] = \mathbb{E}[\|\mathbf{Q}_j\|^2] = N, \quad \forall j,$$

where the expectation is taken with respect to the Gaussian ensemble.

B. Attack Model

The attacks are of the form

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

where \mathcal{J} , the *coalition*, is the index set of the colluding users. The coalition has cardinality $K \leq L$. Moreover the noise \mathbf{E} is i.i.d. $\mathcal{N}(0, \sigma^2)$ and is independent of the fingerprints $\{\mathbf{Q}_k\}$ and consequently independent of $\{\mathbf{X}_k\}$.

The mapping $f : \mathbb{R}^{N|\mathcal{J}|} \rightarrow \mathbb{R}^N$ 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}). The requirement (1) represents a *fairness condition*: all members of the coalition incur equal risk. Another requirement of the mapping f is that it satisfies the following *separation condition*:

$$f(\mathbf{X}_k, k \in \mathcal{J}) = f(\mathbf{Q}_k, k \in \mathcal{J}) + \mathbf{S}. \quad (2)$$

The order statistic attacks considered in this paper satisfy both requirements.

If the attackers can retrieve the original signal \mathbf{S} , they will succeed in defeating the detector. It is therefore useful to view $f(\mathbf{X}_k, k \in \mathcal{J})$ as an estimator of the signal \mathbf{S} based on the copies available to the coalition. The mean-squared distortion of the forgery \mathbf{Y} relative to the host signal \mathbf{S} is given by

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

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

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

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

$$D_c - \sigma^2 = \frac{1}{N} \mathbb{E}\|f(\mathbf{X}_k, k \in \mathcal{J}) - \mathbf{S}\|^2 \quad (4)$$

represents the mean-squared estimation error.

C. Order Statistic Attacks

In this work we shall consider a class of attacks that are known as order statistics filters [2]. Let us first define the notion of order statistic and subsequently order statistic attacks.

Definition 1: If random variables X_1, X_2, \dots, X_K are arranged in ascending order of magnitude and then written as

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(K)},$$

we call $X_{(i)}$ the *i*-th order statistic. [1]

To form an order statistic attack, the colluders form a weighted sum of their ordered samples. The m -th sample of user j 's copy is given by

$$X_j(m) = S(m) + Q_j(m), \quad 1 \leq m \leq N.$$

The ordered samples are given by

$$X_{(j)}(m) = S(m) + Q_{(j)}(m), \quad 1 \leq m \leq N. \quad (5)$$

Sample m of the noise-free forgery is given by

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \sum_{i=1}^K a_i X_{(i)}(m). \quad (6)$$

where $X_{(i)}(m)$ is the i -th order statistic of the K -vector $\mathbf{X}(m) = \{X_1(m), \dots, X_K(m)\}$. For the time being, we assume that the vector $\mathbf{a} = \{a_1, \dots, a_K\}$ is deterministic and the same for all samples. With abuse of notation we

shall denote the overall attack for the entire sequence by $f_{\mathbf{a}}(\mathbf{X}_k, k \in \mathcal{J})$.

The order statistic attacks include some of the widely used attacks in the fingerprinting literature. For instance:

- Linear averaging attack:

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \frac{1}{K} \sum_{i=1}^K X_i(m)$$

where $a_i = \frac{1}{K}$ $1 \leq i \leq K$.

- Minimum attack:

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \min_{1 \leq i \leq K} X_i(m) = X_{(1)}(m)$$

where $a_1 = 1$, and $a_i = 0$ for $2 \leq i \leq K$.

- Maximum attack: similarly,

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \max_{1 \leq i \leq K} X_i(m) = X_{(K)}(m)$$

where $a_K = 1$ and $a_i = 0$ for $1 \leq i \leq K - 1$.

- Midpoint attack:

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \frac{1}{2}(X_{(1)}(m) + X_{(K)}(m))$$

where $a_1 = a_K = \frac{1}{2}$ and $a_i = 0$ for all $i \neq 1, K$.

- Median attack (for odd K):

$$f_{\mathbf{a}}(X_1, \dots, X_K)(m) = \text{median } X_i(m) = X_{(\frac{K+1}{2})}(m)$$

where $a_{\frac{K+1}{2}} = 1$, and $a_i = 0$ for $i \neq \frac{K+1}{2}$.

All the above estimators are special instances of the order statistic attack of (6).

We require the attack $f_{\mathbf{a}}(\mathbf{X}_k, k \in \mathcal{J})$ to be an unbiased estimator of the signal \mathbf{S} or equivalently,

$$\forall m, \quad \mathbb{E} \left[\sum_{i=1}^K a_i X_{(i)}(m) \right] = S(m)$$

where the expectation is with respect to the random ensemble of fingerprinting codes.

Due to (5), we have

$$\mathbb{E} \left[\sum_{i=1}^K a_i X_{(i)}(m) \right] = \mathbb{E} \left[S(m) \sum_{i=1}^K a_i + \sum_{i=1}^K a_i Q_{(i)}(m) \right]$$

Due to the symmetry of the distribution from which \mathcal{C} is drawn, we have:

$$\mathbb{E}[Q_{(i)}(m)] = -\mathbb{E}[Q_{(K-i+1)}(m)], \quad 1 \leq i \leq K.$$

Subsequently for $f_{\mathbf{a}}(\mathbf{X}_k, k \in \mathcal{J})$ to be unbiased, we must have:

$$\begin{aligned} \sum_{i=1}^K a_i &= 1, \\ a_i &= a_{K-i+1}, \quad 1 \leq i \leq K. \end{aligned} \quad (7)$$

Notice that among the attacks listed above, the minimum and the maximum attacks are biased while the averaging and midpoint attacks are unbiased.

D. Correlation Detector

We shall consider a nonblind scenario where 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 [4]. It does not aim at identifying all colluders. Our focused detector does not need to know K , the number of colluders, or the attack parameters \mathbf{a} and σ^2 . (However its performance generally depends on these quantities.)

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

$$T(\mathbf{Y}) = \mathbf{Q}_j^T (\mathbf{Y} - \mathbf{S}) \begin{matrix} H_1 \\ \geq \tau \\ H_0 \end{matrix} \quad (8)$$

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 :

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

The threshold τ trades off the type-I and type-II probabilities of error. In this paper, we fix

$$\tau = \frac{N}{2K}.$$

E. Detection Performance

Due to (1), the pdf's of $T(\mathbf{Y})$ under H_0 and H_1 , conditioned on the code \mathcal{C} , the coalition \mathcal{J} , user of interest j , and mapping f , are Gaussian with variance σ^2 and means:

$$\mathbb{E}_0[T(\mathbf{Y})|\mathcal{C}, \mathcal{J}, j, f] = \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}), \quad j \notin \mathcal{J} \quad (9)$$

$$\mathbb{E}_1[T(\mathbf{Y})|\mathcal{C}, \mathcal{J}, j, f] = \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}), \quad j \in \mathcal{J}. \quad (10)$$

The conditional variance of $T(\mathbf{Y})$ is equal to $N\sigma^2$ under either hypothesis. Note that the expectation of (9) with respect to the random ensemble of codes is equal to 0 for all f (because the fingerprints in the random ensemble are independent and zero-mean). Moreover, the expectation of (10) with respect to \mathcal{C} is independent of \mathcal{J} and j ; denote this expectation by $\frac{N}{K}\beta(f)$. When f is the linear averaging mapping, $\beta(f) = 1$. For any other choice of f , the value of $\beta(f)$ is to be determined.

Assumption 1: The number of users satisfies $L \leq e^{\sqrt{N}}$.
Assumption 2: The number of colluders satisfies $K \ll \sqrt{N}$.

In Sec. V, we prove under the above restrictions on L and K , there exists a ‘‘good code’’ \mathcal{C} from the random ensemble with the following property. Given an arbitrarily small $\epsilon > 0$ and $N > N_0(\epsilon)$, \mathcal{C} satisfies

$$\begin{aligned} \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}) &\leq N\epsilon, & j \notin \mathcal{J}, \\ \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}) &\geq N \left[\frac{\beta(f)}{K} - \epsilon \right], & j \in \mathcal{J}, \end{aligned}$$

i.e., the conditional means (9) and (10) are close to their expectation, *uniformly over all j and \mathcal{J}* . Therefore the type-I and type-II probabilities of error are upper bounded as follows:

$$P_I(\mathcal{C}, f) \leq P_I^* \triangleq \mathcal{Q}\left(\frac{\sqrt{N}}{2\sigma K}(1 - \epsilon)\right) \quad (11)$$

$$P_{II}(\mathcal{C}, f) \leq P_{II}^*(f) \triangleq \mathcal{Q}\left(\frac{\sqrt{N}}{2\sigma K}(2\beta(f) - 1 - \epsilon)\right) \quad (12)$$

where $\mathcal{Q}(t) = (2\pi)^{-1/2} \int_{-\infty}^t \exp\{-x^2/2\} dx$ is the tail probability of a normally distributed random variable.

Remark. One may wish that randomly drawing a code from the Gaussian ensemble be good enough. However, the probability of drawing a bad code may exceed P_I^* and/or P_{II}^* . We can prove the existence of a good code but cannot propose a construction.

In Sec. V, we guarantee the existence of good codes no matter how complicated the colluders' strategy is.

III. OPTIMAL ORDER STATISTICS ATTACKS

In this section we shall consider the problem of finding the optimal order statistic attack. An ideal attack would have the property that introduces the least distortion while maximizing the detector's probability of error.

A. Distortion

The first criterion can be expressed as follows:

$$\min_{\mathbf{a}} \left\{ \psi_1(\mathbf{a}) \triangleq \frac{1}{N} \mathbb{E}[\|f_{\mathbf{a}}(\mathbf{X}_k, k \in \mathcal{J}) - \mathbf{S}\|^2] \right\}. \quad (13)$$

In light of (4), the attackers' mean-squared estimation error is

$$\frac{1}{N} \mathbb{E} \left[\sum_{m=1}^N (f_{\mathbf{a}}(X_k(m), k \in \mathcal{J}) - S(m))^2 \right]. \quad (14)$$

Recall from (7) that $f_{\mathbf{a}}(\mathbf{X}_k, k \in \mathcal{J})$ is an unbiased estimator of the signal \mathbf{S} . Therefore, (14) can be thought of as the time-averaged variance of this estimator. This reduces the problem to identifying \mathbf{a} that minimizes the variance of an unbiased estimator of a signal corrupted by Gaussian noise. The sample mean is the best unbiased estimator for the mean of a signal in Gaussian i.i.d. noise, which implies that the optimal choice for the coefficients $\{a_i\}$ is

$$a_i = \frac{1}{K} \quad 1 \leq i \leq K.$$

So with respect to the distortion criterion, linear averaging is optimal.

B. Probability of Error

The second goal of the attackers is to maximize the probability of error. Under Assumptions 1 and 2, (11) and (12) hold for any good code. Hence we replace the problem of maximizing the error probability of the detector with the problem

of minimizing the expectation of (10) over the ensemble of Gaussian codes. This expectation is given by

$$d(\mathbf{a}) = \mathbb{E} [\mathbf{Q}_j^T f_{\mathbf{a}}(\mathbf{Q}_k, k \in \mathcal{J})].$$

The optimization problem for the normalized distance $\frac{d(\mathbf{a})}{N}$ is therefore stated as:

$$\min_{\mathbf{a}} \left\{ \psi_2(\mathbf{a}) \triangleq \frac{1}{N} \mathbb{E}[\mathbf{Q}_j^T f_{\mathbf{a}}(\mathbf{Q}_k, k \in \mathcal{J})] \right\}. \quad (15)$$

Because the fingerprints in the random ensemble are i.i.d, we have

$$\begin{aligned} \psi_2(\mathbf{a}) &= \mathbb{E} \left[Q_j(m) \sum_{i=1}^K a_i Q_{(i)}(m) \right], \quad \forall m \\ &= \mathbb{E} \left[Q_j \sum_{i=1}^K a_i Q_{(i)} \right]. \end{aligned}$$

Moreover, the expectation is independent of j . Therefore,

$$\begin{aligned} \psi_2(\mathbf{a}) &= \frac{1}{K} \mathbb{E} \left[\sum_{j=1}^K Q_j \sum_{i=1}^K a_i Q_{(i)} \right] \\ &= \frac{1}{K} \mathbb{E} \left[\sum_{j=1}^K Q_{(j)} \sum_{i=1}^K a_i Q_{(i)} \right] \\ &= \frac{1}{K} \sum_{i=1}^K a_i \mathbb{E} \left[\sum_{j=1}^K Q_{(j)} Q_{(i)} \right]. \end{aligned} \quad (16)$$

Now to gain some insight on the quantity $\mathbb{E}[\sum_{j=1}^K Q_{(j)} Q_{(i)}]$, let us consider the following example from [8].

Example 1: Let us consider a simple case where $Q_{(j)}$, $j \in \{1, 2, 3\}$ are the order statistics of a Gaussian vector of length $K = 3$. Table I shows the correlation matrix of $\mathbf{Q} = \{Q_{(1)}, Q_{(2)}, Q_{(3)}\}$. Notice that the expectation in (16) is the columnwise sum of the entries of Table I, which is remarkably equal to 1.

	$Q_{(1)}$	$Q_{(2)}$	$Q_{(3)}$
$Q_{(1)}$	1.27566	0.27566	-0.55132
$Q_{(1)}$	0.27566	0.44867	0.27566
$Q_{(1)}$	-0.55132	0.27566	1.27566

TABLE I
CORRELATION MATRIX OF ORDER STATISTICS OF GAUSSIAN VECTOR OF LENGTH $K = 3$.

Remarkably, the numerical observation of Example 1 applies to order statistics of any zero mean Gaussian vector. This result is stated in the following theorem and is proved by means of an auxiliary estimation problem.

Theorem 1: Let $Q_{(k)}$, $1 \leq k \leq K$ denote the k -th order statistic of K i.i.d. Gaussian random variables with mean zero and unit variance. Then

$$\sum_{j=1}^K \mathbb{E}[Q_{(i)}Q_{(j)}] = 1, \quad i = 1, \dots, K. \quad (17)$$

Proof: By assumption, the random variables Q_k , $1 \leq k \leq K$ are i.i.d. $\mathcal{N}(\theta, 1)$, with $\theta = 0$. Now let us consider the auxiliary problem of estimating θ from these K observations. Consider the estimator

$$\hat{\theta} = \sum_{i=1}^K a_i Q_{(i)} \quad (18)$$

which is linear in $\{Q_{(i)}\}$. The MSE of this estimator is $\mathbb{E}[(\sum_i a_i Q_{(i)})^2]$.

Let us find \mathbf{a} that minimizes this MSE, under the restriction that the linear estimator is unbiased, i.e (7) holds. Consider the Lagrangian

$$\mathbb{E}[(\sum_i a_i Q_{(i)})^2] + \lambda(\sum_i a_i - 1). \quad (19)$$

Taking its partial derivatives with respect to the a_i 's, we have:

$$2 \sum_j a_j \mathbb{E}[Q_{(i)}Q_{(j)}] + \lambda = 0, \quad 1 \leq i \leq K. \quad (20)$$

On the other hand, we know that (i) the sample mean is the minimum variance unbiased (MVUB) estimator of θ and (ii) the sample mean is of the form (18) with $a_i = \frac{1}{K}$. Thus $a_i = \frac{1}{K}$ must solve (20), and we obtain

$$2 \sum_j \frac{1}{K} \mathbb{E}[Q_{(i)}Q_{(j)}] + \lambda = 0, \quad 1 \leq i \leq K, \quad (21)$$

$$\Rightarrow \sum_i \sum_j \mathbb{E}[Q_{(i)}Q_{(j)}] + \lambda \frac{K^2}{2} = 0. \quad (22)$$

Now

$$\sum_i \sum_j \mathbb{E}[Q_{(i)}Q_{(j)}] = \mathbb{E}[\|Q\|^2] = \sum_i \sum_j \mathbb{E}[Q_i Q_j] = K.$$

Replacing in (21), we obtain $\lambda = -\frac{2}{K}$. Then (20) yields

$$\sum_j \mathbb{E}[Q_{(i)}Q_{(j)}] = 1, \quad \forall i,$$

which proves the claim. \blacksquare

Corollary 1: Any unbiased order statistic estimator of Section II-C, is a solution to (16). The value of the minimum is $\psi_2(\mathbf{a}) = 1/K$.

Proof: Applying successively (16), (17) and (7), we have $\psi_2(\mathbf{a}) = \frac{1}{K} \sum_{i=1}^K a_i \mathbb{E}[\sum_{j=1}^K Q_{(j)}Q_{(i)}] = \frac{1}{K} \sum_{i=1}^K a_i = \frac{1}{K}$. \blacksquare

Therefore any choice of \mathbf{a} satisfying (7) results in the same probability of the error at the detector.

In conclusion, the attackers' ideal is to choose \mathbf{a} such that the probability of the error at the detector is maximized while the distortion is minimized. The corresponding cost functions are $\psi_1(\mathbf{a})$ and $\psi_2(\mathbf{a})$ in (13) and (15). While any unbiased estimator is optimal in terms of optimizing $\psi_2(\mathbf{a})$, only the linear averaging attack with $\mathbf{a}_i = \frac{1}{K}$ for all i optimizes $\psi_1(\mathbf{a})$.

IV. TIME-VARYING, RANDOM MAPPINGS

In Section III we restricted our attention to time-invariant, deterministic mappings, where the weight vector \mathbf{a} is fixed for all samples of the attack. In this section, we show that the colluders gain no advantage by choosing a time-varying and/or random strategy. The colluders choose an arbitrary value of \mathbf{a} for each sample (either deterministically or randomly). They construct a noise-free forgery whose samples are given by:

$$f_{\mathbf{a}(m)}(X_k(m), k \in \mathcal{J}) \quad (23)$$

Let $\mathbf{a}^{1:N} = (\mathbf{a}(1), \dots, \mathbf{a}(N))$.

The results of Sections II-D and III extends as follows. In terms of distortion, the linear averaging estimator ($a_i(m) = \frac{1}{K}$ for all i, m) outperforms all other unbiased estimators, including randomized, time-varying estimators. In terms of error probability, for a fixed $\mathbf{a}^{1:N}$, the expectation of (10) with respect to the random ensemble of codes can be evaluated as in Sec. III-B. The colluders seek $\mathbf{a}^{1:N}$ that minimizes

$$\psi_1(\mathbf{a}^{1:N}) \triangleq \frac{1}{N} \sum_{m=1}^N \mathbb{E}[Q_j(m)(f_{\mathbf{a}(m)}(Q_k(m), k \in \mathcal{J}))].$$

Since $\{Q_j(m), 1 \leq j \leq L, 1 \leq m \leq N\}$ are i.i.d., the minimum is achieved by some $\mathbf{a}^{1:N}$ with constant components – precisely the problem treated in Sec. III-B. Here again randomization of $\mathbf{a}^{1:N}$ does not help the colluders' performance.

V. EXISTENCE OF GOOD CODES

This section proves the existence of good codes, for any strategy (time-varying and/or randomized) that the colluders may select. Given a (N, L) fingerprinting code \mathcal{C} , a colluder's mapping $f = (f(1), \dots, f(N))$, a coalition \mathcal{J} , and a user of interest j , the conditional means of the test statistic $T(\mathbf{Y})$ are given by

$$\mu_0(\mathcal{C}, \mathcal{J}, j, f) = \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}) \quad j \notin \mathcal{J}, \quad (24)$$

$$\mu_1(\mathcal{C}, \mathcal{J}, j, f) = \mathbf{Q}_j^T f(\mathbf{Q}_k, k \in \mathcal{J}) \quad j \in \mathcal{J} \quad (25)$$

under H_0 and H_1 , respectively. Define the random variables

$$Z_{0,f} = Q_j f(Q_k, k \in \mathcal{J}) \quad (j \notin \mathcal{J}),$$

$$Z_{1,f} = Q_j f(Q_k, k \in \mathcal{J}) \quad (j \in \mathcal{J}).$$

Therefore (24) and (25) take the form

$$\mu_i(\mathcal{C}, \mathcal{J}, j, f) = \sum_{m=1}^N Z_{i,f(m)}, \quad i = 0, 1.$$

Note that the distributions of $Z_{0,f}$ and $Z_{1,f}$ depend on f but not on j and \mathcal{J} (by construction of the random ensemble of codes \mathcal{C}). Observe that $\mathbb{E}_{\mathcal{C}}[Z_{0,f}] = 0$, and let

$$\mu_{Z,1}(f) \triangleq \frac{1}{N} \sum_m \mathbb{E}_{\mathcal{C}}[Z_{1,f(m)}].$$

The conditional type-I and type-II error probabilities are given by

$$\begin{aligned} P_I(\mathcal{C}, \mathcal{J}, j, f) &= P_0[T(\mathbf{Y}) \geq \tau \mid \mathcal{C}, \mathcal{J}, j, f] \\ &= \mathcal{Q}\left(\frac{\tau - \mu_0(\mathcal{C}, \mathcal{J}, j, f)}{\sigma\sqrt{N}}\right) \\ P_{II}(\mathcal{C}, \mathcal{J}, j, f) &= P_1[T(\mathbf{Y}) \leq \tau \mid \mathcal{C}, \mathcal{J}, j, f] \\ &= \mathcal{Q}\left(\frac{\mu_1(\mathcal{C}, \mathcal{J}, j, f) - \tau}{\sigma\sqrt{N}}\right). \end{aligned}$$

The maximum (over \mathcal{J}, j) probabilities of error are given by

$$\begin{aligned} P_I(\mathcal{C}, f) &= \max_{\mathcal{J}, j} P_I(\mathcal{C}, \mathcal{J}, j, f) \\ P_{II}(\mathcal{C}, f) &= \max_{\mathcal{J}, j} P_{II}(\mathcal{C}, \mathcal{J}, j, f). \end{aligned}$$

Given an arbitrarily small $\epsilon > 0$ and N larger than some $N_0(\epsilon)$, we ask whether there exists a code \mathcal{C} such that

$$\begin{aligned} \mu_0(\mathcal{C}, \mathcal{J}, j, f) &\leq N\epsilon, \\ \mu_1(\mathcal{C}, \mathcal{J}, j, f) &\geq N[\mu_{Z,1}(f) - \epsilon] \end{aligned} \quad (26)$$

uniformly over \mathcal{J}, j, f . The worst-case type-I and type-II error probabilities for such a code would satisfy

$$\begin{aligned} P_I(\mathcal{C}, f) &\leq \mathcal{Q}\left(\frac{\tau - N\epsilon}{\sigma\sqrt{N}}\right) \quad (\text{indep. of } f), \\ P_{II}(\mathcal{C}, f) &\leq \mathcal{Q}\left(\frac{N[\mu_{Z,1}(f) - \epsilon] - \tau}{\sigma\sqrt{N}}\right). \end{aligned}$$

To determine whether such a code exists, define the random variables

$$Z_0^* = \max_{f \in \mathcal{F}} Z_{0,f} \quad \text{and} \quad Z_1^* = \min_{f \in \mathcal{F}} [Z_{1,f} - \mu_{Z,1}(f)]$$

and consider the associated large-deviations functions [?]

$$\phi_0^*(\omega) = \sup_{t > 0} (t\omega - \ln \mathbb{E} e^{tZ_0^*}) \quad (27)$$

$$\phi_1^*(\omega) = \sup_{t < 0} (t\omega - \ln \mathbb{E} e^{tZ_1^*}). \quad (28)$$

Due to Assumptions 1 and 2 on L and K , the number of possible coalitions is a subexponential function of N :

$$\binom{L}{K} \leq L^K \leq e^{K\sqrt{N}} \Rightarrow \ln \binom{L}{K} \ll N.$$

Then

$$\begin{aligned} P_0(\epsilon) &\triangleq \Pr_{\mathcal{C}} \left[\max_{\mathcal{J}, j} \max_{f \in \mathcal{F}^N} \mu_0(\mathcal{C}, \mathcal{J}, j, f) > N\epsilon \right] \\ &= \Pr_{\mathcal{C}} \left[\exists (\mathcal{J}, j) : \max_{f \in \mathcal{F}^N} \mu_0(\mathcal{C}, \mathcal{J}, j, f) > N\epsilon \right] \\ &\stackrel{(a)}{\leq} \binom{L}{K+1} \max_{\mathcal{J}, j} \Pr_{\mathcal{C}} \left[\max_{f \in \mathcal{F}^N} \mu_0(\mathcal{C}, \mathcal{J}, j, f) > N\epsilon \right] \\ &= \binom{L}{K+1} \Pr \left[\max_{\{f(m)\} \in \mathcal{F}^N} \sum_{m=1}^N Z_{0,f(m)}(m) > N\epsilon \right] \end{aligned}$$

$$\begin{aligned} &= \binom{L}{K+1} \Pr \left[\sum_{m=1}^N \max_{f \in \mathcal{F}} Z_{0,f}(m) > N\epsilon \right] \\ &= \binom{L}{K+1} \Pr \left[\sum_{m=1}^N Z_0^*(m) > N\epsilon \right] \\ &\stackrel{(b)}{\leq} e^{(K+1)\sqrt{N}} \exp\{-N\phi_0^*(\epsilon)\} \\ &\rightarrow 0 \quad \text{as } N \rightarrow \infty \end{aligned}$$

where the inequality (a) follows from the union bound and the fact that $j \notin \mathcal{J}$, and (b) follows from the large-deviations bound for Z_0^* . Similarly,

$$\begin{aligned} P_1(\epsilon) &\triangleq \Pr_{\mathcal{C}} \left[\min_{\mathcal{J}, j} \min_{f \in \mathcal{F}^N} [\mu_1(\mathcal{C}, \mathcal{J}, j, f) - \mu_{Z,1}(f)] < -N\epsilon \right] \\ &\leq \binom{L}{K} \Pr \left[\sum_{m=1}^N \min_{f \in \mathcal{F}^N} [Z_{1,f(m)}(m) - \mu_{Z,1}(f)] < -N\epsilon \right] \\ &= \binom{L}{K} \Pr \left[\sum_{m=1}^N Z_1^*(m) < -N\epsilon \right] \\ &\leq e^{K\sqrt{N}} \exp\{-N\phi_1^*(\epsilon)\} \rightarrow 0 \quad \text{as } N \rightarrow \infty. \end{aligned}$$

Since $P_0(\epsilon) + P_1(\epsilon) < 1$ for N large enough, there exists a code \mathcal{C} that satisfies (26), which proves the claim. \square

VI. DISCUSSION

We have studied order statistic attacks (possibly time-varying or stochastic) plus Gaussian noise for Gaussian fingerprints. The attacks are evaluated both in terms of the resulting probability of error and the mean-squared distortion between the host signal and the forgery. While all unbiased order statistic attacks considered result in the same detection performance, the linear averaging attack outperforms all the other attacks in terms of minimizing mean-squared distortion.

We have also shown that if the total number of users $L \leq \exp\sqrt{N}$ and the number of colluders $K \ll \sqrt{N}$, where N is the length of the host signal, good codes exist. These codes guarantee that probability of error at the detector does not exceed $\mathcal{Q}\left(\frac{\sqrt{N}}{2\sigma K}(1 - \epsilon)\right)$, for any $\epsilon > 0$ and $N > N_0(\epsilon)$.

REFERENCES

- [1] H. A. David, *Order Statistics*, 2nd ed. Wiley, 1981.
- [2] A. Bovik, T. Huang, and D. Munson, "A generalization of median filtering using linear combinations of order statistics," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-31 (6), pp. 1342–1350, December 1983.
- [3] Z. Wang, M. Wu, H. Zhao, W. Trappe, , and K. Liu, "Collusion resistance of multimedia fingerprinting using orthogonal modulation," *IEEE Trans. on Image Proc.*, vol. 14, no. 6, pp. 804–821, 2005.
- [4] N. Kiyavash and P. Moulin, "Regular simplex fingerprints and their optimality properties." in *International Workshop on Digital Watermarking, IWDW*, Siena, Italy, sep. 2005, pp. 97–109.
- [5] P. Moulin and A. Briassouli, "The Gaussian fingerprinting game," *Conference on Information Sciences and Systems, CISS'02*, March 2002.
- [6] N. Kiyavash and P. Moulin, "On optimal collusion strategies for fingerprinting," in *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, 2006.
- [7] H. Zhao, M. Wu, Z. Wang, and K. J. R. Liu, "Forensic analysis of nonlinear collusion attacks for multimedia fingerprinting," *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 646–661, May 2005.
- [8] A. Bovik, "Nonlinear filtering using linear combinations of order statistics," Master's thesis, University of Illinois at Urbana-Champaign, 1982.