

EXPURGATED GAUSSIAN FINGERPRINTING CODES

Pierre Moulin and Negar Kiyavash

Beckman Inst., Coord. Sci. Lab and ECE Department
University of Illinois at Urbana-Champaign, USA

ABSTRACT

This paper analyzes the performance of collusion attacks on random fingerprinting codes, when the colluders are subject to an almost-sure squared distortion constraint and a list decoder is used. We derive an exact characterization of the type-I and type-II error exponents of the fingerprinting system. A Gaussian ensemble and an expurgated Gaussian ensemble of codes are considered, and the corresponding random-coding exponents are derived. Explicit optimal strategies for the colluders are derived as well.

Index Terms: Digital fingerprinting, random codes, large deviations.

1. INTRODUCTION

Digital fingerprinting systems can be used for traitor tracing or digital rights management applications. A length- N real-valued signal is to be protected and distributed to M users. Some of the users (K of them) may collude and process their copies to create a *forgery* that contains only weak traces of their fingerprints. This problem was first posed by Cox *et al.* [1] who proposed the use of *Gaussian fingerprints* for this purpose. Specifically, their fingerprints were drawn randomly from an i.i.d. Gaussian distribution; the fingerprint code is shared with the decoder but not revealed to the users.

A fundamental question is what are the optimal performance limits for detection of colluders. To make the problem nontrivial, one may assume embedding distortion constraints on the fingerprinter and the colluders. Example of this analysis include [2, 3] for the case of signals defined over finite alphabets, and [4, 5] for the case of real-valued signals. In the latter case, an obvious (but not necessarily optimal) strategy for the colluders is to perform a uniform linear average of their copies and add i.i.d. Gaussian noise; this strategy was examined in the above papers. Possible improvements for the attackers consist of developing (nonlinear) order-statistics attacks [6, 7].

Our study aims at developing a comprehensive detection-theoretic analysis of collusion attacks and identifying an optimal strategy for the colluders. The analysis is rooted in large-deviations theory. Initial results were reported in [8, 9]. In

this paper some of the restrictive assumptions in [8, 9] are relaxed, including the assumption that the noise introduced by the colluders is white and Gaussian. We also use a list decoder that returns a list of guilty users.

We consider two random ensembles of fingerprinting codes. The first one is the same as the one used by Cox [1] and other researchers and is shown to be less performant than the second one, which is an expurgated ensemble (bad codes are eliminated). The decoder has access to a forgery as well as to the host signal (nonblind detection) and performs a *binary hypothesis test* on each user to determine whether that user was involved in the forgery. The cost functions in this problem are the maximum type-I and type-II probabilities of error, which the colluders want to maximize.

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 collection of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, we denote by $\mathbf{x}_{\mathcal{K}} = \{\mathbf{x}_k, k \in \mathcal{K}\}$ the restriction of this collection to its components $k \in \mathcal{K}$. The symbols $f(N) \ll g(N)$ and $f(N) \sim g(N)$ (asymptotic equality) mean that $\lim_{N \rightarrow \infty} \frac{f(N)}{g(N)} = 0$ and $\lim_{N \rightarrow \infty} \frac{f(N)}{g(N)} = 1$, respectively. The symbols $f(N) \doteq g(N)$ and $f(N) \dot{<} g(N)$ denote asymptotic equality and inequality on the exponential scale: $\ln f(N) \sim \ln g(N)$ and $f(N) \ll g(N)$, respectively. The Gaussian distribution with mean zero and covariance matrix R is denoted by $\mathcal{N}(0, R)$.

2. PROBLEM STATEMENT

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

2.1. 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 M users. Specifically, user m is assigned a marked copy $\mathbf{X}_m = \mathbf{S} + \mathbf{U}_m$ where $m \in \{1, \dots, M\}$ and $\mathbf{U}_m \in \mathbb{R}^N$ is the fingerprint assigned to user m .

The fingerprints $\mathbf{U}_1, \dots, \mathbf{U}_M$ form a (N, M) fingerprinting code \mathcal{C} . The rate of the code is $R_N = \frac{1}{N} \log M$. In a

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

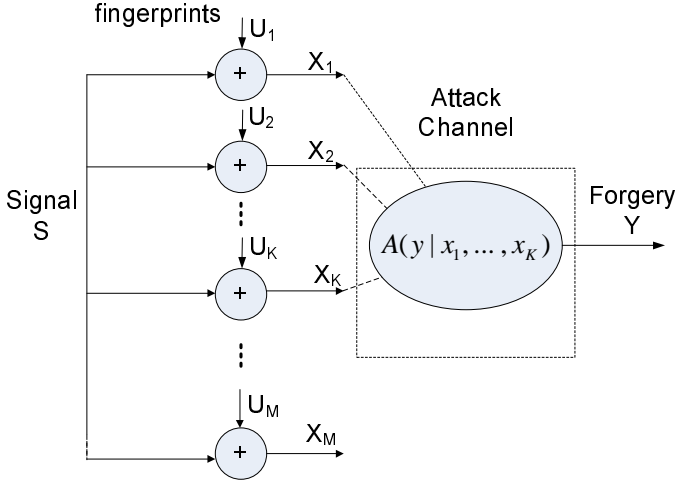


Fig. 1. The fingerprinting process and the attack channel.

typical signal fingerprinting application, $N \sim 10^3 - 10^9$ and $M \sim 2 - 10^9$ (not to exceed the number of humans); therefore we focus our attention on the case of zero-rate codes: $\lim_{N \rightarrow \infty} R_N = 0$.

The code \mathcal{C} is selected independently of \mathbf{S} from a random ensemble of codes, \mathcal{C} , such that

$$\mathbb{E}_{\mathcal{C}}[\|\mathbf{U}_m\|^2] = ND_f, \quad \forall m,$$

i.e., the expected mean-squared distortion is equal to D_f . The random ensembles \mathcal{C} considered in this paper are permutation-invariant, i.e., $\mathcal{C} \in \mathcal{C} \Rightarrow \pi\mathcal{C} \in \mathcal{C}$ where π is a permutation of $\{1, \dots, N\}$, and all $N!$ permutations have the same probability. Moreover, the random ensembles are invariant to permutation of the users.

2.2. Attack Model

Denote by $\mathcal{K} \subseteq \{1, 2, \dots, M\}$ the *coalition*, i.e., the index set of the colluding users. Their coalition has cardinality $K \leq M$. They select a conditional pdf $A(\mathbf{Y}|\mathbf{X}_{\mathcal{K}})$ and draw a pirated copy, or forgery, $\mathbf{Y} \in \mathbb{R}^N$, from that distribution. We write

$$\mathbf{Y} \sim A(\cdot|\mathbf{X}_{\mathcal{K}}). \quad (1)$$

Consider the following constraints on the attack channel A .

(A1) Location-Invariant constraint:

$$A(\mathbf{y}|\mathbf{x}_{\mathcal{K}}) = A(\mathbf{y} - \mathbf{s}|\mathbf{x} - \mathbf{s})_{\mathcal{K}}.$$

(A2) Almost-Sure Mean-Squared Distortion constraint:

$$\|\mathbf{Y} - \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{X}_k\|^2 \leq ND_c.$$

(A3) Fairness constraint:

$$A(\mathbf{y}|\mathbf{x}_{\pi\mathcal{K}}) = A(\mathbf{y}|\mathbf{x}_{\mathcal{K}})$$

for all permutations π of the index set \mathcal{K} .

The model (A1) precludes attacks involving filtering of host signal components. The motivation for this restriction is that it considerably simplifies the mathematical derivation and does not require a statistical model for the host \mathbf{S} . The restriction is relatively mild if embedding is done in a transform domain in which the components of the host \mathbf{S} are approximately independent and are large relative to the embedding distortion. The motivation for (A2) is that distortion is best measured relative to the host \mathbf{S} , but \mathbf{S} is not known to the coalition, so we replace \mathbf{S} by its best linear unbiased estimate, $\frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{X}_k$. Choosing an expected distortion constraint of the form $\mathbb{E}\|\mathbf{Y} - \mathbf{S}\|^2 \leq ND_c$ would allow impulsive noise strategies which blast iid additive noise $\mathcal{N}(0, ND_c)$ with probability $1/N$. Such attacks are extremely effective [10] as they result in zero error exponents. Finally, the fairness condition (A3) will not be imposed but will be shown to hold for optimal attacks: all members of the coalition incur the same risk.

A special case of (1) that will be of interest is ¹

$$\mathbf{Y} = f_N(\mathbf{X}_{\mathcal{K}}) + \mathbf{E} \quad (2)$$

where the noise \mathbf{E} is independent of $\mathbf{X}_{\mathcal{K}}$ and uniformly distributed on the $N - K$ dimensional centered sphere with radius $\sqrt{N\sigma_e^2}$ contained in the $N - K$ dimensional hyperplane orthogonal to the coalition vectors $\mathbf{X}_k, k \in \mathcal{K}$. The mapping $f_N : \mathbb{R}^{N|\mathcal{K}|} \rightarrow \mathbb{R}^N$ may be viewed as an estimator of \mathbf{S} based on $\mathbf{X}_{\mathcal{K}}$, or as a “noise-free forgery”. (Note that if the attackers can retrieve the original signal \mathbf{S} , they will succeed in defeating the decoder.) The location-invariant condition on the attack channel implies that the mapping f_N is location-invariant as well:

$$f_N(\mathbf{X}_{\mathcal{K}}) = f_N(\mathbf{U}_{\mathcal{K}}) + \mathbf{S}. \quad (3)$$

This condition is satisfied by various nonlinear mappings including order-statistics mappings [8]. A special case is uniform linear averaging,

$$\bar{\mathbf{X}} = f_N(\mathbf{X}_{\mathcal{K}}) = \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{X}_k. \quad (4)$$

The model studied in [9] was of the form (2), but \mathbf{E} was AWGN with mean zero and variance σ_e^2 .

When the attack is of the form (2), we need $\sigma_e^2 \leq D_c$ to satisfy the distortion constraint (A2). Hence σ_e^2 is maximized (made equal to D_c) by letting f_N be the uniform linear averaging mapping of (4).

¹When $K \geq 2$, the model (2) is more restrictive than (1), even if \mathbf{E} is unconstrained. For instance a mixture of models of the form (2) (3) satisfies (1) and (A1) but generally not (2).

2.3. Decoder

We study the nonblind scenario where the host signal \mathbf{S} is available at the decoder and can be subtracted from \mathbf{Y} , to form the centered data $\mathbf{Y} - \mathbf{S}$. The decoder computes a *guilt index* for each user m and returns the list of users whose guilt index exceeds a certain threshold $N\tau$. The guilt index adopted here is the correlation statistic

$$T_m(\mathbf{y} - \mathbf{s}) = \mathbf{u}_m^T(\mathbf{y} - \mathbf{s}) = \mathbf{u}_m^T[f_N(\mathbf{u}_K) + \mathbf{e}]. \quad (5)$$

The decoder (5) does not know the channel A used by the colluders or even the exact number K of colluders. However the decoder knows K_{\max} , the maximum number of colluders. For any given user m , the possible error events are a false positive (incorrectly declaring the user to be guilty) or a false negative (incorrectly declaring the user to be innocent). Denote by $\Lambda_m = \{\mathbf{y} \in \mathbb{R}^N : \mathbf{u}_m \cdot (\mathbf{y} - \mathbf{s}) > \tau\}$ the acceptance region for user m and by

$$P_I(m, \mathcal{K}, A) = \int_{\mathbb{R}^{K \times N}} d\mathbf{u}_K p_{\mathbf{U}_K}(\mathbf{u}_K) \int_{\Lambda_m} d\mathbf{y} A(\mathbf{y} | (\mathbf{u} + \mathbf{s})_{\mathcal{K}}),$$

$$P_{II}(m, \mathcal{K}, A) = \int_{\mathbb{R}^{K \times N}} d\mathbf{u}_K p_{\mathbf{U}_K}(\mathbf{u}_K) \int_{\Lambda_m^c} d\mathbf{y} A(\mathbf{y} | (\mathbf{u} + \mathbf{s})_{\mathcal{K}})$$

the corresponding type-I and type-II error probabilities.

By our location-invariant assumption on the attack channel and the decoding regions, these probabilities are independent of \mathbf{s} .

By design of \mathcal{C} , $P_I(m, \mathcal{K}, A)$ is independent of $m \notin \mathcal{K}$. The type-I and type-II error probabilities are given by

$$\begin{aligned} P_I(\mathcal{K}, A) &= Pr[\exists m \notin \mathcal{K} : T_m(\mathbf{Y} - \mathbf{s}) > N\tau] \\ &\leq (M - K) P_I(m, \mathcal{K}, A) \quad \forall m \notin \mathcal{K}, \end{aligned} \quad (6)$$

(incorrectly accusing an innocent user), and

$$\begin{aligned} P_{II}(\mathcal{K}, A) &= Pr[\forall m \in \mathcal{K} : T_m(\mathbf{Y} - \mathbf{s}) \leq N\tau] \\ &\leq \min_{m \in \mathcal{K}} P_{II}(m, \mathcal{K}, A) \end{aligned} \quad (7)$$

(missing all members of the coalition). The inequality (6) is due to the union bound. Detection is said to be reliable if (6) and (7) are small enough. We shall impose the following requirement on the type-I error probability:

$$\max_{\mathcal{K}} \sup_A P_I(\mathcal{K}, A) \stackrel{\cdot}{\leq} \exp\{-N\lambda_I(K_{\max})\} \quad (8)$$

where $\lambda_I(K_{\max})$ is a constraint on the exponent of P_I . The maximum over \mathcal{K} and A indicates that we do not allow innocent users to be accused for any possible choice of \mathcal{K} and A . (A weaker requirement, not recommended here, would be to replace the maximum over \mathcal{K} with an average.) We would like to find the largest possible value of $\lambda_{II}(K_{\max})$ such that

$$\max_{\mathcal{K}} \sup_A P_{II}(\mathcal{K}, A) \stackrel{\cdot}{\leq} \exp\{-N\lambda_{II}(K_{\max})\}. \quad (9)$$

Associated with the random ensemble \mathcal{C} is a curve of λ_{II} vs λ_I , indexed by the threshold τ .

3. WORST ATTACK

Lemma 1 *For any permutation-invariant random ensemble \mathcal{C} and the list decoder (5), there is no loss of optimality for the colluders in choosing a strongly exchangeable attack channel: $A(\mathbf{y} | \mathbf{x}_K) = A(\pi\mathbf{y} | \pi\mathbf{x}_K)$ for every permutation π of the index set $\{1, 2, \dots, N\}$. In the special case of Model (2), there is no loss of optimality in choosing a memoryless mapping*

$$f_N(\mathbf{x}_K) = \{f(x_K(1)), \dots, f(x_K(N))\} \quad (10)$$

for some $f : \mathbb{R}^K \rightarrow \mathbb{R}$.

Given a vector-valued sequence $\mathbf{z} \in \mathbb{R}^{d \times N}$, denote by $R_{\mathbf{z}} = \frac{1}{N} \mathbf{z} \mathbf{z}^T$ the $d \times d$ empirical correlation matrix for \mathbf{z} .

Given \mathbf{x}_K and \mathbf{y} , define the $N - K$ dimensional sphere $\Sigma(\mathbf{x}_K, \mathbf{y})$ whose elements \mathbf{y}' have the following property: $(\mathbf{x}_K, \mathbf{y}')$ have the same empirical correlation matrix as $(\mathbf{x}_K, \mathbf{y})$. The sphere $\Sigma(\mathbf{x}_K, \mathbf{y})$ is centered at $\hat{\mathbf{s}}_K(\mathbf{x}_K, \mathbf{y}) \in \mathbb{R}^N$ which is the orthogonal projection of \mathbf{y} onto the K -dimensional subspace of \mathbb{R}^N spanned by $\mathbf{x}_k, k \in \mathcal{K}$. This subspace is orthogonal to the $N - K$ dimensional subspace in which the sphere $\Sigma(\mathbf{x}_K, \mathbf{y})$ is embedded. The radius of the sphere is

$$\|\mathbf{y} - \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{x}_k\| = \|\mathbf{y}' - \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{x}_k\|, \quad \forall \mathbf{y}' \in \Sigma(\mathbf{x}_K, \mathbf{y}). \quad (11)$$

It may also be shown that for the correlator decoder (5), there is no loss of optimality for the colluders in selecting $A(\cdot | \mathbf{x}_K)$ that is uniform over each sphere $\Sigma(\mathbf{x}_K, \mathbf{y})$. Furthermore, P_I and P_{II} are maximized by $A(\cdot | \mathbf{x}_K)$ that is uniform over a single sphere $\Sigma(\mathbf{x}_K, \mathbf{y})$. That is, there is no loss of optimality for the colluders in choosing \mathbf{Y} according to the additive-noise model (2), where $f_N(\mathbf{x}_K) \in \text{span}\{\mathbf{x}_k, k \in \mathcal{K}\}$. Moreover, by application of Lemma 1, there is no loss of optimality in choosing a memoryless mapping f_N . We still need to identify the worst f in (10). Denote by \mathcal{F} a compact set of feasible f , such that $\|f_N(\mathbf{x}_K) - \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{x}_k\|^2 \leq N(D_c - \sigma_e^2)$.

4. GAUSSIAN ENSEMBLE

A possible choice for \mathcal{C} is the Gaussian ensemble, in which random codes $\mathcal{C} = \{\mathbf{U}_m, 1 \leq m \leq M\}$ are obtained by drawing fingerprint components $U_m(n)$ i.i.d. $\mathcal{N}(0, D_f)$.

We now analyze detection performance for the random ensemble \mathcal{C} under any attack mapping $f \in \mathcal{F}$ selected by the colluders. For simplicity of the exposition, we only report results in the case where \mathbf{E} iid $\mathcal{N}(0, \sigma_e^2)$; the error exponents when \mathbf{E} is uniformly distributed on the sphere of radius $\sqrt{N\sigma_e^2}$ are slightly larger.

First we define the five random variables

$$\begin{aligned} Z_{0,f} &= U_m f(U_{\mathcal{K}}) \quad (m \notin \mathcal{K}) \\ Z_{1,f} &= U_m f(U_{\mathcal{K}}) \quad (m \in \mathcal{K}) \\ W &= U_m E \\ U_{0,f} &= Z_{0,f} + W \\ U_{1,f} &= Z_{1,f} + W. \end{aligned}$$

By our assumptions on f and \mathcal{C} , the distributions of these random variables do not depend on m and \mathcal{K} . In particular, $\text{Var}(W) = D_f \sigma^2$ and

$$\mathbb{E}[Z_{0,f}] = \mathbb{E}[U_{0,f}] = \mathbb{E}[W] = 0.$$

Finally, note that $Z_{0,f}$, $Z_{1,f}$, W , $U_{0,f}$ and $U_{1,f}$ are non-Gaussian, even if the mapping f is linear.

Lemma 2 $\mathbb{E}[Z_{1,f}] = \mathbb{E}[U_{1,f}] = D_f/K$ for all $f \in \mathcal{F}$.

Proof: The random variables $U_m f(U_{\mathcal{K}})$, $m \in \mathcal{K}$, are identically distributed. Therefore $\mathbb{E}[Z_{1,f}] = \mathbb{E}[\bar{U} f(U_{\mathcal{K}})]$ where $\bar{U} = \frac{1}{K} \sum_{m \in \mathcal{K}} U_m$. By the location-invariant property of f , we have $f(U_{\mathcal{K}}) = f((U - \bar{U})_{\mathcal{K}}) + \bar{U}$. Since $U_m, m \in \mathcal{K}$, are iid Gaussian, the mean \bar{U} is independent of the centered random variables $(U - \bar{U})_{\mathcal{K}}$ and therefore of any function of them. Hence

$$\mathbb{E}[\bar{U} f(U_{\mathcal{K}})] = \mathbb{E}[\bar{U}^2] = D_f/K$$

which proves the claim. \square

Since the test statistic $T_m(\mathbf{Y} - \mathbf{s})$ in (5) is a sum of N i.i.d. random variables $U_{0,f}(n)$ and $U_{1,f}(n)$ under hypotheses H_0 and H_1 , respectively, we immediately obtain Proposition 1 below, where $\Lambda_{U_{0,f}}^*(\cdot)$ and $\Lambda_{U_{1,f}}^*(\cdot)$ are the large-deviations functions for the random variables $U_{0,f}$ and $U_{1,f}$, respectively.

Proposition 1 For the Gaussian ensemble \mathcal{C} , the error probabilities satisfy the following upper bounds for all $f \in \mathcal{F}$:

$$\begin{aligned} P_I(\mathcal{K}, f) &\leq \exp\left\{-N \Lambda_{U_{0,f}}^*(\tau)\right\} \\ P_{II}(\mathcal{K}, f) &\leq \exp\left\{-N \Lambda_{U_{1,f}}^*\left(\frac{D_f}{K} - \tau\right)\right\} \end{aligned}$$

where $0 \leq \tau \leq D_f/K_{\max}$. Moreover these bounds are tight in the exponent as $N \rightarrow \infty$.

Proposition 2 In the limit as $N \gg K_{\max} \geq K \rightarrow \infty$, we have the asymptotic equalities

$$\begin{aligned} \Lambda_{U_{0,f}}^*(\tau) &\sim \frac{\tau^2}{2 \text{Var}(W)} = \frac{\tau^2}{2 D_f \sigma_e^2} \leq \frac{D_f}{2 \sigma_e^2 K_{\max}^2}, \\ \Lambda_{U_{1,f}}^*\left(\frac{D_f}{K} - \tau\right) &\sim \frac{1}{2 D_f \sigma_e^2} \left(\frac{D_f}{K} - \tau\right)^2 \leq \frac{D_f}{2 \sigma_e^2 K_{\max}^2}. \end{aligned}$$

Proof. The arguments of the large-deviations functions $\Lambda_{U_{0,f}}^*$ and $\Lambda_{U_{1,f}}^*$ above vanish as $K \rightarrow \infty$. The asymptotic equality results from a quadratic expansion of these functions around zero. \square

Prop. 1 states that the error exponents depend on the mapping f selected by the colluders. However, as indicated by Prop. 2, that exponential dependency vanishes for large K . We conclude this section with Prop. 3 which establishes a fundamental relationship between N , M and K_{\max} , guaranteeing reliable detection for the Gaussian ensemble \mathcal{C} .

Proposition 3 For the Gaussian ensemble \mathcal{C} , reliable detection is guaranteed provided that $K_{\max}^2 \ln M \ll N$.

Proof: immediate consequence of Props. 1, 2, and (6), (7).

5. EXPURGATED GAUSSIAN ENSEMBLE

The problem with the Gaussian ensemble \mathcal{C} of Sec. 4 is that error probability (which is obtained by averaging over all codes in \mathcal{C}) may be dominated by bad codes. This is a standard problem for the design of low-rate codes, for which performance is dictated by minimum-distance considerations, and the bad codes are the ones with poor minimum distance [12]. Improvements can be obtained by using expurgation, i.e., removing bad codes from the random ensemble.

We apply this principle to our fingerprinting problem and show that performance can indeed be improved by expurgating the Gaussian random ensemble. We assume here that $K_{\max}^3 \ln M \ll N$.

To gain some insight, we first consider the case $M = K + 1$ with fixed f and \mathcal{K} (the simplest possible extension of the two-codeword problem of [12, Ch. 5]). For a typical selection of \mathcal{C} , we have

$$\|\mathbf{U}_m\|^2 - N D_f = O(N^{1/2}) \quad \forall m,$$

$$\mathbf{U}_m^T f_N(\mathbf{U}_{\mathcal{K}}) = O(N^{1/2}) \ll N \tau$$

when $m \notin \mathcal{K}$, and

$$\mathbf{U}_m^T f_N(\mathbf{U}_{\mathcal{K}}) - N \frac{D_f}{K} = O(N^{1/2}) \ll N \tau$$

when $m \in \mathcal{K}$. Thus

$$\begin{aligned} P_I &\doteq \Pr[\mathbf{U}_m^T \mathbf{E} > N \tau], \\ P_{II} &\doteq \Pr[\mathbf{U}_m^T \mathbf{E} < -N(\frac{D_f}{K} - \tau)]. \end{aligned}$$

The goal now is to design a random ensemble \mathcal{C} of codes that have the above properties for much larger M and for any mapping $f \in \mathcal{F}$ and coalition \mathcal{K} .

To this end, fix a positive sequence ϵ_N such that

$$\sqrt{\frac{K \ln M}{N}} \ll \epsilon_N \ll \frac{1}{K}$$

and define \mathcal{C} as the random ensemble of codes \mathcal{C} such that

$$\max_m \|\mathbf{U}_m\|^2 - N D_f < N \epsilon_N \quad (12)$$

$$\max_{\mathcal{K}} \max_{m \notin \mathcal{K}} \sup_{f \in \mathcal{F}} \mathbf{U}_m^T f_N(\mathbf{U}_{\mathcal{K}}) < N \epsilon_N \quad (13)$$

$$\min_{\mathcal{K}} \min_{m \in \mathcal{K}} \inf_{f \in \mathcal{F}} \mathbf{U}_m^T f_N(\mathbf{U}_{\mathcal{K}}) > N \left(\frac{D_f}{K} - \epsilon_N\right). \quad (14)$$

Lemma 3 The probability that a code \mathcal{C} drawn from the iid Gaussian ensemble also belongs to \mathcal{C} tends to 1 as $N \rightarrow \infty$.

By application of this lemma, the following procedure may be used to draw a code from \mathcal{C} : draw \mathcal{C} from the iid Gaussian ensemble and verify whether \mathcal{C} satisfies (12), (13) and (14). If yes (this happens with high probability), keep \mathcal{C} . If not, repeat the above procedure until \mathcal{C} satisfies the above conditions.

Sketch of Proof of Lemma 3. Denote by \mathcal{E}_0 , \mathcal{E}_1 and \mathcal{E}_2 the events (12), (13), and (14), respectively. By the union bound, we have

$$Pr[\mathcal{E}_0^c] \leq MPr \left[\left| \sum_{n=1}^N U(n)^2 - ND_f \right| > N\epsilon_N \right] \quad (15)$$

$$Pr[\mathcal{E}_1^c] \leq \binom{M}{K+1} Pr \left[\sup_{f \in \mathcal{F}} \sum_{n=1}^N Z_{0,f}(n) > N\epsilon_N \right] \quad (16)$$

$$Pr[\mathcal{E}_2^c] \leq \binom{M}{K} Pr \left[\inf_{f \in \mathcal{F}} \sum_{n=1}^N Z_{1,f}(n) < N \left(\frac{D_f}{K} - \epsilon_N \right) \right] \quad (17)$$

where $\binom{M}{K} \leq M^K = \exp\{K \ln M\} \dot{<} \exp\{N\epsilon_N^2\}$. We have

$$Pr \left[\left| \sum_{n=1}^N U(n)^2 - ND_f \right| > N\epsilon_N \right] \leq \exp \left\{ -\frac{N\epsilon_N^2}{4D_f^2} \right\},$$

therefore $Pr[\mathcal{E}_0^c]$ vanishes exponentially with N . By application of Sanov's theorem [11], we prove that

$$\begin{aligned} & Pr \left[\sup_{f \in \mathcal{F}} \sum_{n=1}^N Z_{0,f}(n) > N\epsilon_N \right] \\ & \leq \exp \left\{ -\frac{N\epsilon_N^2}{2 \sup_{f \in \mathcal{F}} \text{Var}(Z_{0,f})} \right\}, \end{aligned} \quad (18)$$

$$\begin{aligned} & Pr \left[\inf_{f \in \mathcal{F}} \sum_{n=1}^N Z_{1,f}(n) < N \left(\frac{D_f}{K} - \epsilon_N \right) \right] \\ & \leq \exp \left\{ -\frac{N\epsilon_N^2}{2 \sup_{f \in \mathcal{F}} \text{Var}(Z_{1,f})} \right\}. \end{aligned} \quad (19)$$

Therefore (18) and (19) vanish exponentially with N , and so do (16) and (17). \square

Using Shannon's formula for spherical caps [13], we derive the following result.

Proposition 4 *The type-I and type-II error probabilities for the expurgated Gaussian ensemble \mathcal{C} are given by*

$$P_I(\mathcal{K}, f) \doteq \exp \left\{ -NE_{\text{cap}} \left(\frac{\tau}{\sigma_e} \right) \right\}, \quad (20)$$

$$P_{II}(\mathcal{K}, f) \doteq \exp \left\{ -NE_{\text{cap}} \left(\frac{D_f}{\sigma_e K} - \frac{\tau}{\sigma_e} \right) \right\}, \quad (21)$$

for all \mathcal{K} and $f \in \mathcal{F}$, where

$$E_{\text{cap}}(\rho) = -\frac{1}{2} \ln(1 - \rho^2), \quad 0 \leq \rho \leq 1. \quad (22)$$

Since $\tau < D_f/K_{\text{max}}$ tends to 0 for increasing K_{max} and $E_{\text{cap}}(\rho) \sim \rho^2/2$ as $\rho \rightarrow 0$, the exponents above coincide with those derived for the Gaussian ensemble in Prop. 2 in the limit as $K_{\text{max}} \rightarrow \infty$.

Proposition 5 *The error exponents in (20) and (21) are minimized by the uniform linear averaging f with $K = K_{\text{max}}$; then $\sigma_e^2 = D_c$, and*

$$\lambda_I(K_{\text{max}}) = E_{\text{cap}} \left(\frac{\tau}{D_c^{1/2}} \right)$$

$$\lambda_{II}(K_{\text{max}}) = E_{\text{cap}} \left(\frac{D_f}{D_c^{1/2} K_{\text{max}}} - \frac{\tau}{D_c^{1/2}} \right).$$

Proof. Given K , the detection bounds (20) and (21) depend on f via σ_e . As discussed below (4), uniform linear averaging with $K = K_{\text{max}}$ simultaneously maximizes K and σ_e , and therefore minimizes the error exponents (20) and (21). \square

The error exponents in (20) and (21) are uniformly better than those obtained by drawing codes from the Gaussian ensemble.

6. REFERENCES

- [1] I. J. Cox, J. Killian, F. T. Leighton and T. Shamoan, "Secure Spread Spectrum Watermarking for Multimedia," *IEEE T-IP*, Vol. 6, pp. 1673–1687, Dec. 1997. (Also NEC Tech. Rep. 95-10, 1995).
- [2] P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE T-IT*, Vol. 49, No. 3, pp. 563–593, 2003.
- [3] A. Somekh-Baruch and N. Merhav, "On the Capacity Game of Private Fingerprinting Systems Under Collusion Attacks," *Proc. IEEE Int. Symp. on Information Theory*, Yokohama, Japan, p. 191, July 2003.
- [4] J. Kilian, F. T. Leighton, L. R. Matheson, T. G. Shamoan, R. E. Tarjan, and F. Zane, "Resistance of digital watermarks to collusive attacks," *Proc. ISIT*, p. 271, Cambridge, MA, 1998.
- [5] P. Moulin and A. Briassouli, "The Gaussian Fingerprinting Game," *Proc. CISS'02*, Princeton, NJ, March 2002.
- [6] H. S. Stone, "Analysis of Attacks on Image Watermarks With Randomized Coefficients," *NEC TR 96-045*, Princeton, NJ, 1996.
- [7] H. Zhao, M. Wu, Z. Wang and K. J. R. Liu, "Forensic Analysis of Nonlinear Collusion Attacks for Multimedia Fingerprinting," *IEEE T-IP*, Vol. 14, No. 5, pp. 646–661, May 2005.
- [8] N. Kiyavash and P. Moulin, "A Framework for Optimizing Nonlinear Collusion Attacks on Fingerprinting Systems," *Proc. Conf. on Information Systems and Science*, Princeton, NJ, March 2006.
- [9] P. Moulin and N. Kiyavash, "Performance of Random Fingerprinting Codes Under Arbitrary Nonlinear Collusion Attacks," *Proc. ICASSP*, Hawaii, Apr. 2007.
- [10] N. Kiyavash and P. Moulin, "On Optimal Collusion Strategies for Fingerprinting," *Proc. ICASSP*, Toulouse, France, May 2006.
- [11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer-Verlag, New York, 1998.
- [12] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, NY, 1968.
- [13] C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," *Bell Systems Tech. J.*, Vol. 38, pp. 611–656, 1959.