OPTIMAL GAUSSIAN FINGERPRINT DECODERS

Pierre Moulin

Beckman Inst., Coord. Sci. Lab and ECE Department
University of Illinois at Urbana-Champaign, USA
Email: moulin@ifp.uiuc.edu

ABSTRACT

This paper proposes codes that achieve the fundamental capacity limits of digital fingerprinting subject to mean-squared distortion constraints on the fingerprint embedder and the colluders. We first show that the traditional method of fingerprint decoding by thresholding correlation statistics falls short of this goal: reliable performance is impossible at code rates greater than some value C_1 that is strictly less than capacity. To bridge the gap to capacity, a more powerful decoding method is needed. The *Maximum Penalized Gaussian Mutual Information* decoder presented here meets this requirement. Finally, a mathematical framework and a capacity expression for fingerprinting of social networks are presented.

Index Terms: Digital fingerprinting, coding, decoding

1. INTRODUCTION

Digital fingerprinting systems can be used for traitor tracing and digital rights management applications. A length-N real-valued signal is to be protected and distributed to M users. K users 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. Their fingerprints were drawn randomly from an i.i.d. (independent and identically distributed) 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 [2–4] for the case of signals defined over finite alphabets, and [5, 6] for the case of real-valued signals. In the latter case, a possible strategy for the colluders is to perform a uniform linear average of their copies and add i.i.d. Gaussian noise. Other strategies involve nonlinear attacks [6–9].

In our model, the decoder returns a list of guilty users. It is assumed that the decoder has access to the host signal (non-blind detection) but knows neither the collusion strategy nor even the number of colluders. However the maximum possible number of colluders is $K_{\max}$, and the decoder knows that value. The cost functions of interest are the probabilities of false positives and false negatives, which should vanish as $N \to \infty$, for any admissible collusion strategy. A simplified version of this problem was analyzed in [6] and a mutual-information game was solved, but no coding system was proposed. An upper bound on fingerprinting capacity was derived in [4], but no coding scheme was given that would achieve this bound in the Euclidean setting.

Research supported by NSF grants CCF 06-35137 and CCF 07-29061.

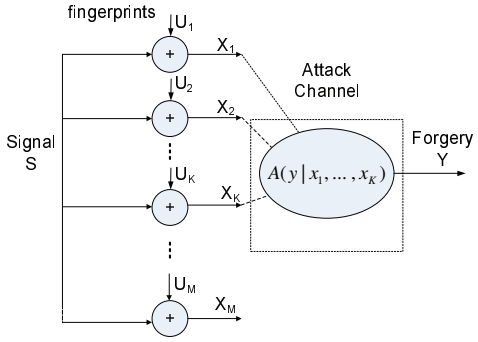

Fig. 1. The fingerprinting process and the attack channel.

Notation: we use boldface uppercase letters to denote random vectors, uppercase letters for the components of the vectors, and calligraphic fonts for sets. The symbol $\mathbb{E}$ denotes mathematical expectation. The restriction of a collection of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ to its components $k \in \mathcal{K}$ is denoted by $\mathbf{x}_{\mathcal{K}} = \{\mathbf{x}_k,\, k \in \mathcal{K}\}$. The symbol $f(N) \sim g(N)$ denotes asymptotic equality: $\lim_{N\to\infty} \frac{f(N)}{g(N)} = 1$. The Gaussian distribution with mean zero and covariance matrix R is denoted by $\mathcal{N}(0, \mathsf{R})$. The $N \times N$ identity matrix is denoted by I_N.

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 $\mathbf{S} = (S(1), \ldots, S(N))$ in $\mathbb{R}^N$, viewed as *deterministic* but *unknown* to the colluders. Fin-

gerprints 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 $m \in \{1, \ldots, M\}$ and $\mathbf{U}_m \in \mathbb{R}^N$ is the fingerprint assigned to user m.

The fingerprints $\mathbf{U}_1, \cdots, \mathbf{U}_M$ form a (N, M) fingerprinting code $\mathcal{C}$. The rate of the code is $R_N = \frac{1}{N} \log M$. In a typical signal fingerprinting application, $N \sim 10^3 - 10^9$ and $M \sim 2 - 10^9$ (not to exceed the number of humans).

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

$$\|\mathbf{U}_m\|^2 = ND_1, \quad \forall m,$$

i.e., the *mean-squared embedding distortion* is equal to D_1. The random ensemble $\mathscr{C}$ is permutation-invariant, i.e., $\mathcal{C} \in \mathscr{C} \Rightarrow \pi\mathcal{C} \in \mathscr{C}$ where π is a permutation of $\{1, \cdots, N\}$, and all $N!$ permutations have the same probability. Moreover, $\mathscr{C}$ is invariant to permutation of the users.

2.2. Attack Model

Denote by $\mathcal{K} \subseteq \{1, 2, \cdots, M\}$ the *coalition*, i.e., the index set of the colluders. Their coalition has cardinality $K \leq M$. They select a *memoryless collusion channel* $A(y|x_{\mathcal{K}})$. An example is the uniform averaging attack followed by addition of Gaussian noise with mean zero and variance D_2:

$$Y = \frac{1}{K} \sum_{k \in \mathcal{K}} X_k + W \tag{1}$$

where $W \sim \mathcal{N}(0, D_2)$.

The colluders pass their fingerprinted sequences $\mathbf{x}_{\mathcal{K}}$ through the channel A and output a pirated copy, or forgery, $\mathbf{Y} \in \mathbb{R}^N$. (The memoryless assumption can be relaxed and is made solely to simplify the exposition.) The following two constraints on A define a feasible set $\mathscr{A}(D_2)$ of collusion channels.

(A1) *Location-Invariant constraint*:

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

(A2) *Expected Mean-Squared Distortion constraint*:

$$\sigma^2(A) \triangleq \mathbb{E}\left(Y - \frac{1}{K} \sum_{k \in \mathcal{K}} X_k\right)^2 \leq D_2.$$

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, $\frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{X}_k$. The expected mean-squared distortion is at most D_2.

The analysis will show that optimal collusions satisfy the *fairness property*

$$A(y|x_{\pi\mathcal{K}}) = A(y|x_{\mathcal{K}})$$

i.e., all members of the coalition incur the same risk. The attack of (1) is feasible and fair, and so are the nonlinear attacks of [9].

2.3. Decoder

Since the host signal $\mathbf{S}$ is available at the decoder, it can be subtracted from $\mathbf{Y}$ to form the centered data $\mathbf{Y} - \mathbf{S}$. The decoder outputs an estimated coalition

$$\hat{\mathcal{K}} = g_N(\mathbf{y} - \mathbf{s}) \tag{2}$$

where g_N is independent of $\mathbf{s}$. If the decoder returns $\emptyset$, no user is accused. Most decoding rules in the fingerprinting literature are based on thresholding the correlation statistics $\mathbf{u}_m^T(\mathbf{y} - \mathbf{s})$, $1 \leq m \leq M$. As we shall see, *such decoders are always suboptimal*.

2.4. Error Probabilities and Capacity

By our location-invariant assumptions (A1) and (2) on the collusion channel and the decoding regions, the probability of false positives (accuse an innocent user) and the probability of false negatives (fail to catch any colluder):

$$P_{FP}(A) \triangleq Pr[\hat{\mathcal{K}} \setminus \mathcal{K} \neq \emptyset], \quad P_{FN}(A) \triangleq Pr[\hat{\mathcal{K}} \cap \mathcal{K} = \emptyset],$$

are independent of $\mathbf{s}$. To simplify notation, we will thus assume without loss of generality that $\mathbf{s} = 0$. By design of $\mathscr{C}$, these error probabilities are also independent of $\mathcal{K}$.

A fingerprinting code rate R is said to be achievable if there exists a sequence of $(N, 2^{NR})$ codes such that both $\sup_{A \in \mathscr{A}(D_2)} P_{FP}(A)$ and $\sup_{A \in \mathscr{A}(D_2)} P_{FN}(A)$ vanish as $N \to \infty$. Fingerprinting capacity is the supremum of all achievable rates [4]. Hence capacity describes the fundamental limits on the parameters (N, M, K, D_1, D_2) for any reliable fingerprinting system.

3. MUTUAL-INFORMATION GAME

Our coding scheme is related to the solution to the following mutual-information game. The expression for $C(K)$ below is a simple variation on a result by Wang and Moulin [6]. Let $\mathscr{P}_{X_{\mathcal{K}}}(D_1)$ be the set of all pdf's of the product form $p_{X_{\mathcal{K}}}(x_{\mathcal{K}}) = \prod_{k \in \mathcal{K}} p_X(x_k)$ where p_X satisfies $\mathbb{E}X^2 \leq D_1$.

Proposition 3.1 *The values of the maxmin mutual-information games*

$$C(K) = \sup_{p_{X_\mathcal{K}} \in \mathscr{P}_{X_\mathcal{K}}(D_1)} \inf_{A \in \mathscr{A}(D_2)} \frac{1}{K} I(X_\mathcal{K}; Y),$$
$$C_1(K) = \sup_{p_{X_\mathcal{K}} \in \mathscr{P}_{X_\mathcal{K}}(D_1)} \inf_{A \in \mathscr{A}(D_2)} I(X_1; Y)$$

are respectively given by

$$C(K) = \frac{1}{2K} \ln\left(1 + \frac{D_1}{KD_2}\right), \quad (3)$$

$$C_1(K) = \frac{1}{2} \ln\left(1 - \frac{D_1/K^2}{D_1/K + D_2}\right)^{-1} \quad (4)$$

where $C_1(K) \le C(K)$, with equality if and only if $K = 1$. For both games, the supremum over $p_{X_\mathcal{K}}$ is achieved by the product Gaussian pdf with variance D_1, and the infimum over A is achieved by the uniform averaging attack of (1).

Note that even if the decoder knew K, Prop. 3.1 does not imply that the maximum-likelihood decoder tailored to the minimizing channel of (1) performs satisfactorily against all feasible channels. For instance, to show that $C(K)$ is an achievable rate, one must construct a decoder that achieves vanishing error probabilities for any rate below $C(K)$, for all feasible collusion channels. Candidate decoders are examined in the next two sections. Also note that the converse theorem of [4] states that no rate greater than $C(K)$ is achievable. The proof of this theorem was given for finite alphabets but applies to Euclidean alphabets as well.

We now study the performance of rate-R random spherical codes whose 2^{NR} fingerprints are drawn independently and uniformly from the N-dimensional sphere with squared radius ND_1. Two decoders are studied in Secs. 4 and 5 below and achieve rates $C_1(K_{\max})$ and $C(K_{\max})$, respectively.

4. SIMPLE THRESHOLDING DECODER

The normalized empirical correlation coefficient between two sequences $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^N$ is defined as $\rho(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}$.

A simple decoder is the thresholding rule

$$m \in \hat{\mathcal{K}} \quad \Leftrightarrow \quad \rho(\mathbf{x}_m, \mathbf{y}) > \eta \quad (5)$$

with threshold η. Since the code is spherical, (5) is equivalent to a thresholding rule on the unnormalized correlation statistics $\mathbf{x}_m \cdot \mathbf{y}$, $1 \le m \le 2^{NR}$. We choose an arbitrarily small $\epsilon > 0$ and let $\eta = \eta_1(K_{\max}) - \epsilon$, see definition of $\eta_1(\cdot)$ below.

False Negatives. To analyze P_{FN}, first note that the random variables $\frac{1}{N}\mathbf{X}_m \cdot \mathbf{Y}$ (for all $m \in \mathcal{K}$) and $\frac{1}{N}\|\mathbf{Y}\|^2$ converge in probability to their expectations $\frac{D_1}{K}$ and $\frac{D_1}{K} + \sigma^2(A)$, respectively, *for any fair collusion attack, Gaussian or not* [9]. Hence $\rho(\mathbf{X}_m, \mathbf{Y})$ converges in probability to

$$\rho(X_m, Y) = \sqrt{\frac{D_1/K^2}{D_1/K + \sigma^2(A)}} \ge \sqrt{\frac{D_1/K^2}{D_1/K + D_2}} \triangleq \eta_1(K)$$

for any fair attack channel and any $m \in \mathcal{K}$. For general (possibly nonfair) channels $A \in \mathscr{A}(D_2)$, we can show that $\max_{m \in \mathcal{K}} \rho(\mathbf{X}_m, \mathbf{Y})$ converges in probability to $\max_{m \in \mathcal{K}} \rho(X_m, Y) \ge \eta_1(K) \ge \eta + \epsilon$. Hence

$$P_{FN}(A) = Pr\left[\max_{m \in \mathcal{K}} \rho(\mathbf{X}_m, \mathbf{Y}) < \eta\right]$$

vanishes for all rates and for all channels $A \in \mathscr{A}(D_2)$.

False Positives. For any fixed innocent user $m \notin \mathcal{K}$, $\mathbf{X}_m$ and $\mathbf{Y}$ are independent, and Shannon's formula for the volume of a spherical cap yields [11]

$$Pr[\rho(\mathbf{X}_m, \mathbf{Y}) > \eta] \doteq 2^{-nE_{\text{cap}}(\eta)}, \quad m \notin \mathcal{K}$$

where $E_{\text{cap}}(\eta) = -\frac{1}{2}\log(1 - \eta^2) \triangleq \tau$. By the union bound we have

$$\begin{aligned} P_{FP} &= Pr\left[\exists m \notin \mathcal{K} \,:\, \rho(\mathbf{X}_m, \mathbf{Y}) \ge \eta\right] \\ &\le (2^{nR} - K) Pr[\rho(\mathbf{X}, \mathbf{Y}) \ge \eta] \\ &\doteq 2^{-n(E_{\text{cap}}(\eta) - R)} \end{aligned}$$

where $\mathbf{X} \sim \mathcal{N}(0, D_1 \, \mathsf{I}_N)$ is independent of $\mathbf{Y}$. Hence P_{FP} vanishes for $R < E_{\text{cap}}(\eta)$. Moreover

$$E_{\text{cap}}(\eta) \uparrow E_{\text{cap}}(\eta_1(K_{\max})) = C_1(K_{\max})$$

as $\epsilon \to 0$. Since ϵ can be chosen arbitrarily small, P_{FP} vanishes for $R < C_1(K_{\max})$. It may also be shown that P_{FP} tends to 1 for $R > C_1(K_{\max})$, hence reliable decoding is impossible at those rates.

5. JOINT FINGERPRINT DECODER

At rates $R > C_1(K_{\max})$, the number of fingerprints that satisfy $\rho(\mathbf{X}_m, \mathbf{Y}) > \eta_1(K_{\max})$ is of the order of $2^{N(R - C_1(K_{\max}))}$, and all these fingerprints would trigger false alarms for the thresholding decoder of Sec. 4. Also note they are strongly correlated with the fingerprints of the guilty users.

The key idea to improve decoding performance is to perform a *joint decision* on the guilt of a candidate coalition *instead of separate decisions* on its individual members. The problem is in many ways analogous to the problem of decoding for the multiple-access channel [10], as further developed in [4]. The developments below are based on these concepts.

5.1. Empirical Gaussian Mutual Information

The mutual information (m.i.) between two Gaussian random variables X and Y with normalized correlation coefficient ρ is given by $I(X; Y) = -\frac{1}{2}\log(1 - \rho^2)$ [10]. If X_k, $k \in \mathcal{K}$, are iid $\mathcal{N}(0, D_1)$ and $Y = \sum_k a_k X_k + W$ where $\sum_k a_k = 1$ and $W \sim \mathcal{N}(0, D_2)$ is independent of $\{X_k\}$, the m.i. between $X_\mathcal{K}$ and Y is given by

$$I_G(X_\mathcal{K}; Y) = -\frac{1}{2} \log\left(1 - \sum_{k \in \mathcal{K}} \rho^2(X_k, Y)\right). \quad (6)$$

If $a_k \equiv \frac{1}{K}$ then (6) does coincide with $K\,C(K)$ in (3). If $X_{\mathcal{K}}, Y$ are non-Gaussian, then (6) will be termed "Gaussian m.i." instead of m.i. Now we define the *empirical Gaussian m.i.* between two sequences $\mathbf{x}_{\mathcal{K}} \in \mathbb{R}^{N\times\mathcal{K}}$ and $\mathbf{y} \in \mathbb{R}^N$ as

$$\hat{I}_G(\mathbf{x}_{\mathcal{K}};\mathbf{y}) \triangleq -\frac{1}{2}\log\left(1-\sum_{k\in\mathcal{K}}\rho^2(\mathbf{x}_k,\mathbf{y})\right). \qquad (7)$$

5.2. Decoder

Define $\mathscr{K}(\epsilon)$ as the set of $\mathcal{K}$ such that the fingerprints $\mathbf{x}_k$, $k \in \mathcal{K}$ have absolute normalized correlation at most equal to ϵ:

$$\frac{1}{nD_1}\,|\mathbf{x}_k \cdot \mathbf{x}_l| \le \epsilon, \quad \forall k \ne l \in \mathcal{K}. \qquad (8)$$

By convention, the empty coalition $\emptyset$ is an element of $\mathscr{K}(\epsilon)$.

For our random spherical codes, for any fixed $\mathcal{K}$ and arbitrarily small ϵ, it follows from the weak law of large numbers that (8) holds with probability approaching 1 as $N \to \infty$.

Let $\tau = C(K_{\max}) - \epsilon$. Our proposed decoder outputs $\hat{\mathcal{K}}$ that maximizes the *maximum penalized Gaussian mutual information* (MPGMI) criterion

$$\widetilde{MPGMI}(\mathcal{K}) = \hat{I}_G(\mathbf{x}_{\mathcal{K}};\mathbf{y}) - K\tau \qquad (9)$$

over all $\mathcal{K} \in \mathscr{K}(\epsilon)$. By convention, $\widetilde{MPMI}(\emptyset) \triangleq 0$. Roughly speaking, the MPGMI decoder of (9) favors fingerprints that are strongly correlated with the forgery, but (i) these fingerprints must be nearly uncorrelated since $\mathcal{K} \in \mathscr{K}(\epsilon)$ and (ii) large coalitions are penalized linearly in the coalition size K. These two conditions are key to ensure vanishing P_{FP}.

5.3. Achievable Rates

The analysis of P_{FP} and P_{FN} is considerably more involved than it was for the decoder of Sec. 4. We refer the reader to [4] for a closely related analysis in the case of finite alphabets. Some key steps of the derivations are outlined below.

First, the random variable $\hat{I}_G(\mathbf{x}_{\mathcal{K}};\mathbf{y})$ converges in probability to its expectation, $I_G(X_{\mathcal{K}};Y) \ge K\,C(K)$, as $N \to \infty$. This property is used to establish that P_{FN} vanishes as $N \to \infty$, for all feasible collusion channels.

Second, P_{FP} is upper bounded by applying the union bound to all possible false-positive error events: $\hat{\mathcal{K}} = \mathcal{A} \cup \mathcal{B}$ where $\mathcal{A} \ne \emptyset$ and $\mathcal{B}$ are sets of innocent and guilty users, respectively:

$$P_{FP} \le \sum_{\mathcal{B}\subseteq\mathcal{K}}\sum_{|\mathcal{A}|\ge 1} 2^{N|\mathcal{A}|R}\,Pr[\hat{I}_G(\mathbf{X}_{\mathcal{A}};\mathbf{Y}\mathbf{X}_{\mathcal{B}}) > |\mathcal{A}|\tau].$$

It may be shown that each probability in the right side vanishes as $2^{-N|\mathcal{A}|\tau}$. Hence P_{FP} vanishes for all $R < \tau$. The maximum possible value of τ is $C(K_{\max})$. Hence reliable decoding is possible at all rates below $C(K_{\max})$.

Also observe that $C_1(K) \sim C(K) \sim \frac{D_1}{2(\ln 2)K^2 D_2}$ as $K \to \infty$. Hence, for large K, the simple decoder is nearly as good as the more complex joint decoder.

6. SOCIAL NETWORKS

Assuming that every coalition is possible, there are $\binom{2^{NR}}{K} \approx 2^{NKR}$ coalitions of size K. However colluders are usually acquaintances, i.e., they are part of some social network. It is interesting to study how this constraint affects capacity. We assume a simple social network model in which relations are represented by a graph with 2^{NR} nodes (one for each user) and edge connectivity bounded by some constant c at each node. That is, each user has up to c acquaintances. The number of possible coalitions is then upper bounded by $c!\,2^{NR}$ which is considerably smaller than 2^{NKR} for $K \ge 2$ and $NKR \gg \log c!$. Assuming the social network is known to the decoder, the capacity analysis can be easily revisited to account for that constraint. The set $\mathscr{K}(\epsilon)$ of feasible coalitions for the MPGMI decoder of (9) has size of the order of $c!\,2^{NR}$. The union bound used in the proof of the direct coding theorem involves only $c!\,2^{NR}$ terms instead of 2^{NKR}. As a consequence, the maximum achievable rate is $K_{\max}$ times larger than before. It may also be shown that no rate larger than $K_{\max}\,C(K_{\max})$ is achievable, using a simple modification of the converse theorem of [4]. Hence capacity is $K_{\max}$ times the value of $C(K_{\max})$ given by (3), which is rather substantial. We emphasize that the social network is assumed to be known to the decoder in this analysis.

7. REFERENCES

[1] I. J. Cox, J. Killian, F. T. Leighton and T. Shamoon, "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 Info. Theory*, Yokohama, Japan, p. 191, July 2003.

[4] P. Moulin, "Universal Fingerprinting: Capacity and Random Coding Exponents," *preprint*, Jan. 2008, revised Sep. 2008. Short version in *Proc. ISIT*, Toronto, Canada, July 2008.

[5] J. Kilian, F. T. Leighton, L. R. Matheson, T. G. Shamoon, R. E. Tarjan, and F. Zane, "Resistance of digital watermarks to collusive attacks," *Proc. ISIT*, p. 271, Cambridge, MA, 1998.

[6] Y. Wang and P. Moulin, "Capacity and Optimal Collusion Attack Channels for Gaussian Fingerprinting Games," *Proc. SPIE*, San Jose, CA, Jan. 2007.

[7] H. S. Stone, "Analysis of Attacks on Image Watermarks With Randomized Coefficients," *NEC TR 96-045*, Princeton, NJ, 1996.

[8] 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.

[9] 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.

[10] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley.

[11] C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," *Bell Systems Tech. J.*, Vol. 38, pp. 611—656, 1959.