

Performance Analysis of Physical Signature Authentication

Joseph A. O'Sullivan, *Senior Member, IEEE*, and
Natalia A. Schmid, *Member, IEEE*

Abstract—In physical signature authentication schemes, a candidate signature presented for authentication is compared to a previous signature measurement. Uniqueness of a signature is a result of a random physical process underlying the signature formation. In this correspondence, we model signatures as realizations of random processes with known statistics. Rate functions for the false alarm and miss probabilities are determined. The case where the candidate signature and previous signature measurement are zero-mean, stationary, and Gaussian is covered in detail, yielding quantitative measures of performance for system design. The binary case is also discussed.

Index Terms—Information rates, pattern classification, physical signature authentication, Toeplitz matrices.

I. INTRODUCTION

Random physical processes may be used as sources of signatures for use in authentication. These signatures often arise from sensing realizations of the random configuration of physical particles. The motivating example signature for this correspondence arises from fixing micro-magnetic particles to a substrate using a binding agent [8]. Related examples include recognition based on biometrics [13]–[15], the distribution of trees or other objects in natural scenes, wood grains, and many other realizations of random media and scenes. In each of these cases, the realization (signature) arises from an underlying random physical process.

Loosely speaking, the authentication problem is to determine whether two measurements (a candidate signature and a previous signature measurement) come from the same realization or independent realizations of the physical process. The problem is stated as a hypothesis-testing problem, where under one hypothesis there is a joint distribution of the two measurements and under the null hypothesis the two measurements are independent. In the derivation, we assume that the probability distribution functions for the physical random processes are known. This is a rather stringent requirement for many applications. In fingerprint and other biometric analyses (including face recognition, retinal scans, and hand scans) [13]–[15], the true distribution for the random physical processes is not known. Finding a good model for those processes may be the most challenging part in order to apply the analysis here. In other applications, there may be a good model for the random physical process. For example, with randomly placed objects with random sizes or orientations, a Poisson process may describe the locations, and distributions for the sizes or orientations may be estimated from data.

Manuscript received December 8, 1997; revised September 4, 2000. This work was supported in part by the National Science Foundation under Grant NCR 94-06197. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, June/July 1997.

J. A. O'Sullivan is with the Electronic Systems and Signals Research Laboratory, Department of Electrical Engineering, Washington University, St. Louis, MO 63130 USA (e-mail: jao@ee.wustl.edu).

N. A. Schmid is with the Electronic Systems and Signals Research Laboratory, Department of Electrical Engineering, Washington University, St. Louis, MO 63130 USA. She is now with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: nschmid@ifp.uiuc.edu).

Communicated by R. A. Calderbank, Editor-in-Chief.

Publisher Item Identifier S 0018-9448(01)08589-3.

When the distributions are not known, then nonparametric techniques may be used. In current implementations of systems for authentication of magnetic media, only the correlation coefficient is computed [8]. Suppose that a signature measurement is a realization of a real-valued random process $\{X_k\}$. This is typically collected under controlled conditions and stored for authentication purposes. Suppose that a candidate signature, a realization of a real-valued random process $\{Y_k\}$, is presented and claimed to correspond to a realization of the process $\{X_k\}$. In the absence of a model for the data, if realizations of n consecutive measurements of the processes $\{X_k\}$ and $\{Y_k\}$ are available, the correlation coefficient may be computed as

$$\rho = \frac{\sum_{k=1}^n (X_k - \bar{X})(Y_k - \bar{Y})}{\left[\sum_{k=1}^n (X_k - \bar{X})^2 \sum_{k=1}^n (Y_k - \bar{Y})^2 \right]^{1/2}} \quad (1)$$

where \bar{X} and \bar{Y} are the sample means obtained using the n samples from the processes $\{X_k\}$ and $\{Y_k\}$, respectively. Typically, ρ is compared to a threshold. While good performance is observed empirically, the error rates are generally unknown.

In contrast to this nonparametric approach, we examine performance of an authentication system using the optimal test statistic, the information density. The error-rate functions are determined in terms of the distribution tilted between the joint and the independent distributions. The two extreme cases are of interest. Invoking Stein's lemma, the error rate for the case of a fixed probability of false alarm is the sup-relative entropy rate function between the product and the joint distributions. The complementary error rate function for the fixed probability of detection case is the sup-information rate function. By computing the optimal performance for known distributions, the results in this correspondence, particularly the Gaussian results, provide bounds on the performance and guidelines for system design.

We consider the stationary Gaussian case and a binary case. For the Gaussian case, the power spectra of the random processes determine the performance, but only through a ratio of the signal spectrum to the signal-plus-noise spectrum. The Toeplitz distribution theorem is invoked at several points in the analysis. The rate functions are plotted for an example with low-pass random processes observed in white noise. More detailed proofs are relegated to appendixes. In the binary case, the components of vectors are assumed to be independent and identically distributed (i.i.d.). Closed-form expressions for the rate functions and their plots are obtained.

II. GENERAL RESULTS

Suppose that there are two standard alphabet random processes $\{X_k\}$ and $\{Y_k\}$. Assume that a candidate signature presented for authentication is a realization of the process $\{Y_k\}$ and the previous signature measurement is a realization of the process $\{X_k\}$. Under hypothesis H_1 , the two realizations involve a common underlying physical process, and hence the processes $\{X_k\}$ and $\{Y_k\}$ have a joint distribution function P_{XY} . Under hypothesis H_0 , the two processes are independent with distributions P_X and P_Y that equal the marginals of P_{XY} . Some general results are obtained in this section without specifying exactly how the underlying random physical process is related to the candidate signature and the previous signature measurement. Specific examples are presented in later sections. Denote the distribution functions on the first n samples by $P_{X^n Y^n}$, P_{X^n} , and P_{Y^n} . The notation follows Gray [3]. In particular, $X^n = (X_1, X_2, \dots, X_n)$.

Assumption 1: The pair random process $\{(X_n, Y_n)\}$ is stationary and ergodic under both $P_{X^n Y^n}$ and $P_{X^n} \times P_{Y^n}$. For each n , $P_{X^n Y^n}$ and $P_{X^n} \times P_{Y^n}$ are mutually absolutely continuous.

The optimal test statistic is given by the log-likelihood ratio, which in this case is the information density

$$i_n(X^n, Y^n) = \frac{1}{n} \log \frac{dP_{X^n Y^n}}{dP_{X^n} \times P_{Y^n}}. \quad (2)$$

Definition 1: A stationary process $\{X_n\}$ is said to have the finite-gap information property if there exists an integer K such that

$$I(X_{K+1}; X^- | X^K) < \infty \quad (3)$$

where $X^- = (X_0, X_{-1}, X_{-2}, \dots)$, [3, p. 145].

Theorem 1: Let the random processes $\{X_n\}$ and $\{Y_n\}$ each have the finite-gap information property, and let Assumption 1 hold. Then, under hypothesis H_1 , the information density converges to the sup-information rate

$$\lim_{n \rightarrow \infty} i_n(X^n, Y^n) = \bar{I}(X; Y), \quad \text{a.e. } P_{XY} \quad (4)$$

where

$$\bar{I}(X; Y) = \limsup_{n \rightarrow \infty} E_1 \left\{ \frac{1}{n} \log \frac{dP_{X^n Y^n}}{dP_{X^n} \times P_{Y^n}} \right\} \quad (5)$$

and where E_1 is expectation under the joint distribution of H_1 .

For the proof, see Gray [3, pp. 178–179]. A similar theorem applies to convergence under hypothesis H_0 , but additional assumptions must be made.

Assumption 2: Let $P_{X^n Y^n}^{(K)}$ denote the K th-order Markov approximation to $P_{X^n Y^n}$. For all n , assume

$$P_{X^n Y^n} \gg P_{X^n Y^n}^{(K)} \gg P_{X^n} \times P_{Y^n}.$$

Additionally, assume that there is some K_0 such that for all $K \geq K_0$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \ln \frac{dP_{X^n Y^n}^{(K)}}{dP_{X^n} \times P_{Y^n}} = \eta^{(K)}, \quad \text{a.e. } P_X \times P_Y \quad (6)$$

and

$$\lim_{K \rightarrow \infty} \eta^{(K)} = 0 \quad (7)$$

where

$$\eta^{(K)} = \lim_{n \rightarrow \infty} \frac{1}{n} E_0 \left\{ \ln \frac{dP_{X^n Y^n}^{(K)}}{dP_{X^n} \times P_{Y^n}} \right\}. \quad (8)$$

Theorem 2: Under Assumptions 1 and 2, the information density converges to the negative sup-relative entropy rate between the product and the joint distributions under hypothesis H_0

$$\lim_{n \rightarrow \infty} i_n(X^n, Y^n) = -\bar{D}(P_X \times P_Y, P_{XY}), \quad \text{a.e. } P_X \times P_Y \quad (9)$$

where

$$\bar{D}(P_X \times P_Y, P_{XY}) = \limsup_{n \rightarrow \infty} E_0 \left\{ \frac{1}{n} \log \frac{dP_{X^n} \times P_{Y^n}}{dP_{X^n Y^n}} \right\} \quad (10)$$

and E_0 is expectation under the product distribution of H_0 .

The proof of the theorem is a modification of a proof from Gray [3, pp. 165–167], and is given in Appendix A. The idea behind the proof is that the finite-gap property and absolute continuity between the measures ensure that the limit is finite. A sequence of K th-order Markov processes is used to approximate the joint distribution. The ergodicity of the processes then gives convergence.

These two theorems describe conditions under which the log-likelihood ratio converges to its expected value under either hypothesis with probability one. Under stronger conditions, the rate of convergence is exponential under either hypothesis, with exponent determined by a rate function. We adopt a large-deviations approach, and obtain the rate functions under either hypothesis from the Gartner–Ellis theorem.

Let the rate functions for $n i_n(X^n, Y^n)$ under H_0 and H_1 be $I_0(t)$ and $I_1(t)$, respectively. These rate functions are convex functions that attain their minima at the expected values under the two hypotheses, which are given in the two previous theorems as $-\bar{D}(P_X \times P_Y, P_{XY})$ and $\bar{I}(X; Y)$, respectively.

For the large-deviations results, we follow the approach described by Bucklew [1, pp. 14–19] for his statement of the Gartner–Ellis theorem. Define

$$\phi_{m,n}(s) = \frac{1}{n} \log E_m \{ \exp(s n i_n(X^n, Y^n)) \}, \quad m = 0, 1 \quad (11)$$

where E_1 is expectation under P_{XY} , and E_0 is expectation under $P_X \times P_Y$. Note that $\phi_{1,n}(s) = \phi_{0,n}(s+1)$.

Assumption 3: For each $s \in \mathcal{R}$, the asymptotic log-moment-generating function defined as

$$\bar{\phi}_1(s) = \lim_{n \rightarrow \infty} \phi_{1,n}(s)$$

exists as an extended real number.

Assumption 4: Let $\mathcal{D}_{\bar{\phi}_1} = \{s : \bar{\phi}_1(s) < \infty\}$. The function $\bar{\phi}_1$ is differentiable on $\mathcal{D}_{\bar{\phi}_1}$.

Definition 2: The rate function I_m is the Fenchel–Legendre transform of $\bar{\phi}_m$

$$I_m(t) = \sup_s [st - \bar{\phi}_m(s)], \quad m = 0, 1, t \in \mathcal{R}. \quad (12)$$

Let $\mathcal{D}_{I_m} = \{t \in \mathcal{R} : I_m(t) < \infty\}$.

For notational convenience, we denote the expected values under the two hypotheses as $L_0 = -\bar{D}(P_X \times P_Y, P_{XY})$ and $L_1 = \bar{I}(X; Y)$.

Since $\phi_{0,n}(s+1) = \phi_{1,n}(s)$, this relation is true in the limit, so $\bar{\phi}_0(s+1) = \bar{\phi}_1(s)$. Also, the domain of $\bar{\phi}_0$ is the domain of $\bar{\phi}_1$ shifted, $\mathcal{D}_{\bar{\phi}_0} = \{s+1 : s \in \mathcal{D}_{\bar{\phi}_1}\}$. This implies that the rate function $I_0(t)$ is related to $I_1(t)$ by

$$\begin{aligned} I_1(t) &= \sup_s [st - \bar{\phi}_1(s)] = \sup_s [st - \bar{\phi}_0(s+1)] \\ &= \sup_s [st - t - \bar{\phi}_0(s)] = I_0(t) - t. \end{aligned}$$

From this, we can conclude that $\mathcal{D}_{I_0} = \mathcal{D}_{I_1}$.

Since any rate function equals zero at the mean of the corresponding random variable, $I_0(L_0) = 0$ and $I_1(L_1) = 0$. Then we have $L_0 \in \mathcal{D}_{I_0}$ and $L_1 \in \mathcal{D}_{I_1}$, so $[L_0, L_1] \subset \mathcal{D}_{I_1}$.

Theorem 3 (Gartner–Ellis): Under Assumption 3, if $[a, b] \cap \mathcal{D}_{I_m} \neq \emptyset$, then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_m \{i_n(X^n, Y^n) \in [a, b]\} \leq - \inf_{t \in [a, b]} I_m(t). \quad (13)$$

Under Assumptions 3 and 4, if $(a, b) \subset \mathcal{D}_{I_m}$, then

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_m \{i_n(X^n, Y^n) \in (a, b)\} \geq - \inf_{t \in (a, b)} I_m(t). \quad (14)$$

The proof is given in [1, pp. 14–19].

Consider $[a, b] \subset [L_0, L_1]$. The continuity and convexity of the rate functions on their domains implies that

$$\inf_{t \in [a, b]} I_0(t) = \inf_{t \in (a, b)} I_0(t) = I_0(a) \quad (15)$$

and

$$\inf_{t \in [a, b]} I_1(t) = \inf_{t \in (a, b)} I_1(t) = I_1(b). \quad (16)$$

If the conditions of Theorem 1 and 2 hold, then the exponential rate of a Neyman–Pearson detector can be determined. Suppose that the interval $[-\bar{D}(P_X \times P_Y \| P_{XY}), \bar{I}(X; Y)]$ is partitioned into two disjoint subintervals, that is,

$$[-\bar{D}(P_X \times P_Y \| P_{XY}), \gamma] \cup [\gamma, \bar{I}(X; Y)] \\ = [-\bar{D}(P_X \times P_Y \| P_{XY}), \bar{I}(X; Y)].$$

Define a region

$$A_n = \{(x^n, y^n) : i_n(x^n, y^n) > \gamma\}.$$

Lemma 1 (Stein): Let A_n be an acceptance region for the hypothesis H_1 and assume that the conditions of Theorems 1 and 2 hold. Let $\alpha_n = P_0\{A_n\}$, $\beta_n = P_1\{A_n^c\}$ be the probabilities of error under H_0 and H_1 , respectively. Denote by β_n^c the minimum probability of error under H_1 , under the constraint that $\alpha_n < \epsilon$, $0 < \epsilon < 0.5$. Then

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^c = -\bar{D}(P_X \times P_Y \| P_{XY}).$$

Similarly, under the constraint that $\beta_n < \epsilon$, $0 < \epsilon < 0.5$, the asymptotic rate of the minimum probability of error α_n^c is given by

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n^c = -\bar{I}(X; Y).$$

The proof is similar to the proof of Stein's lemma given in [16, pp. 309–311].

Now suppose that for all n , $P_{X^n Y^n}$ has a density function. Following the arguments from [9, Ch. 8], for the hypothesis testing problem, for any number of samples n , the probability density functions minimizing error exponents for the “false alarm” and “miss” probabilities are the tilted density functions

$$q_{s,n}(x^n, y^n) = \frac{p(x^n, y^n) e^{s \log \frac{p(x^n, y^n)}{p(x^n)p(y^n)}}}{e^{n\phi_{1,n}(s)}}, \quad -1 \leq s \leq 0 \quad (17)$$

$$\pi_{s,n}(x^n, y^n) = \frac{p(x^n)p(y^n) e^{s \log \frac{p(x^n, y^n)}{p(x^n)p(y^n)}}}{e^{n\phi_{0,n}(s)}}, \quad 0 \leq s \leq 1 \quad (18)$$

and the error-exponents are

$$(1/n)D(Q_{s,n}, P_{X^n Y^n}) \quad \text{and} \quad (1/n)D(\Pi_{s,n}, P_{X^n} \times P_{Y^n})$$

where $Q_{s,n}$ and $\Pi_{s,n}$ are the tilted distribution functions.

Since the Gartner–Ellis theorem provides conditions for the existence of the large deviation rate functions

$$\lim_{n \rightarrow \infty} \frac{1}{n} D(Q_{s,n}, P_{X^n Y^n}) = I_1(t_{1,s}) \\ \lim_{n \rightarrow \infty} \frac{1}{n} D(\Pi_{s,n}, P_{X^n} \times P_{Y^n}) = I_0(t_{0,s})$$

where

$$t_{1,s} = \frac{d\bar{\phi}_1(s)}{ds} = \lim_{n \rightarrow \infty} E_{q_{s,n}} \{i_n(X^n, Y^n)\} \\ t_{0,s} = \frac{d\bar{\phi}_0(s)}{ds} = \lim_{n \rightarrow \infty} E_{\pi_{s,n}} \{i_n(X^n, Y^n)\}.$$

In the simplest case, the entries of X^n and Y^n are i.i.d. In that case, assuming that $P_{X^1 Y^1}$ has a density function $p(x, y)$, the quantities defined here do not depend on n . In particular, $\phi_{1,n}(s) = \bar{\phi}_1(s)$ for all n , and

$$\bar{\phi}_1(s) = \log \iint p(x, y) e^{s \log \frac{p(x, y)}{p(x)p(y)}} dx dy. \quad (19)$$

The expectations $\bar{I}(X; Y)$ and $\bar{D}(P_X \times P_Y, P_{XY})$ are given directly in terms of the densities, with

$$\bar{D}(P_X \times P_Y, P_{XY}) = \iint p(x)p(y) \log \frac{p(x)p(y)}{p(x, y)} dx dy. \quad (20)$$

III. GAUSSIAN CASE

Often, physical processes are well-modeled by Gaussian distributions, particularly when the data arise from a sum of contributions of a large number of particles. If the particles are fixed, then the sum of their contributions is constant over time. Often the measurement process has noise in it, which is modeled as additive noise in this section. Then the data consist of a linear combination of a Gaussian distributed random signature and Gaussian noise. The binary hypothesis test for this physical signature authentication problem reduces to a decision of whether the realizations of the random signature are identical.

Let the signature process $\{Z_k\}$ and noise processes $\{W_{1k}\}$ and $\{W_{2k}\}$ be zero-mean, stationary Gaussian random processes with real power spectra $S_Z(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$, respectively. Two observations are available

$$X_k = Z_{1k} + W_{1k}, \quad k = 1, 2, \dots, n \quad (21)$$

$$Y_k = Z_{2k} + W_{2k}, \quad k = 1, 2, \dots, n. \quad (22)$$

Assumption 5: $S_Z(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$ are real, bounded, and Riemann integrable, and $S_1(\lambda)$ and $S_2(\lambda)$ are positive for all λ .

The two hypotheses in the detection problem are as follows. Hypothesis H_1 , in which $Z_{2k} = Z_{1k}$ for all k , that is, the two signature process realizations are the same; Hypothesis H_0 , in which Z_2 and Z_1 are i.i.d. realizations of the signature process. Denote the $n \times n$ Toeplitz covariance matrix corresponding to n consecutive observations of a random process, parameterized by the power spectrum $S(\lambda)$, by $\mathbf{K}_n(S)$.

The information density is then

$$i_n(X^n, Y^n) = -\frac{1}{2n} [X^{nT}, Y^{nT}] (\mathbf{R}_1^{-1} - \mathbf{R}_0^{-1}) [X^{nT}, Y^{nT}]^T \\ - \frac{1}{2n} \log \det(\mathbf{R}_0^{-1} \mathbf{R}_1) \quad (23)$$

where \mathbf{R}_1 and \mathbf{R}_0 are the covariance matrices under the hypotheses H_1 and H_0 , respectively,

$$\mathbf{R}_0 = \begin{bmatrix} \mathbf{K}_n(S_Z + S_1) & 0 \\ 0 & \mathbf{K}_n(S_Z + S_2) \end{bmatrix} \\ \mathbf{R}_1 = \begin{bmatrix} \mathbf{K}_n(S_Z + S_1) & \mathbf{K}_n(S_Z) \\ \mathbf{K}_n(S_Z) & \mathbf{K}_n(S_Z + S_2) \end{bmatrix}. \quad (24)$$

Under hypothesis H_1 , X^n and Y^n are correlated since the signature process realizations are identical. Under hypothesis H_0 , X^n and Y^n are independent with the same marginal distributions as under hypothesis H_1 .

Before we place the results in the context of the Gartner–Ellis theorem, we find the asymptotic log-moment-generating functions for the random variable $i_n(X^n, Y^n)$.

Lemma 2: The asymptotic log-moment-generating functions, defined as

$$\bar{\phi}_m(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \log(E_m[e^{s n i_n(X^n, Y^n)}]), \quad m = 0, 1 \quad (25)$$

are equal to

$$\bar{\phi}_1(s) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - s^2 \hat{f}(\lambda)) d\lambda - \frac{s}{4\pi} \int_{-\pi}^{\pi} \log(1 - \hat{f}(\lambda)) d\lambda \quad (26)$$

and

$$\bar{\phi}_0(s+1) = \bar{\phi}_1(s) \quad (27)$$

where

$$\hat{f}(\lambda) = \frac{S_Z^2(\lambda)}{(S_Z(\lambda) + S_1(\lambda))(S_Z(\lambda) + S_2(\lambda))}. \quad (28)$$

The proof is given in Appendix B and is based on standard results from the asymptotics of Toeplitz matrices (see, for example, [4]–[6]). Note that \hat{f} determines the performance. It is dependent on the signal-to-noise ratio (SNR) of each of the signals, and is a function of frequency. For high SNRs, \hat{f} is close to one. For low SNRs, \hat{f} is close to zero.

Comment 1: Under Assumption 5, then Assumptions 1, 3, and 4 hold, so Theorems 1 and 3 and Lemma 1 are valid for the Gaussian case. While it is difficult to show that Assumption 2 holds, the results stated in Theorem 2 follow from the results of Gray [6].

Comment 2: Under Assumption 5, each of the processes $\{X_n\}$ and $\{Y_n\}$ satisfies the finite-gap information property (see Appendix C for a proof).

Under Assumption 5, we can obtain closed-form expressions for L_0 and L_1 , the asymptotic expectations of the information density under the hypotheses H_0 and H_1 , respectively. L_1 , the mutual information rate, is given by

$$\begin{aligned} L_1 &= \lim_{n \rightarrow \infty} \left\{ -\frac{1}{2n} \text{Tr}[(\mathbf{R}_1^{-1} - \mathbf{R}_0^{-1})\mathbf{R}_1] - \frac{1}{2n} \log \det(\mathbf{R}_0^{-1}\mathbf{R}_1) \right\} \\ &= \lim_{n \rightarrow \infty} -\frac{1}{2n} \log \det(\mathbf{R}_0^{-1}\mathbf{R}_1) \\ &= -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - \hat{f}(\lambda)) d\lambda \end{aligned} \quad (29)$$

and L_0 , the negative sup-relative entropy rate, is given by

$$L_0 = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{2\hat{f}(\lambda)}{1 - \hat{f}(\lambda)} d\lambda - \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - \hat{f}(\lambda)) d\lambda. \quad (30)$$

From Lemma 2, we have explicit expressions for the log-moment-generating functions $\bar{\phi}_m(s)$, $m = 0, 1$ given by (26) and (27). Now we can substitute these expressions into (12) to obtain

$$I_1(t) = \sup_s \left\{ st + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - s^2 \hat{f}(\lambda)) d\lambda + \frac{s}{4\pi} \int_{-\pi}^{\pi} \log(1 - \hat{f}(\lambda)) d\lambda \right\}$$

and

$$I_0(t) = I_1(t) + t.$$

The functions $I_1(t)$ and $I_0(t)$ are convex functions with their minima equal to zero achieved at $t = L_1$ and $t = L_0$, respectively.

IV. EXAMPLE

In this section, we examine the effects of varying two parameters. Suppose that the random signature, modeled as a white Gaussian noise process, is filtered by an ideal low-pass filter. From the form of \hat{f} , if the two observations have different cutoff frequencies, the lower of the two determines performance. Suppose this lower cutoff frequency is a , where $0 < a \leq \pi$. Also, suppose that the two noise levels are constant in the passband of the low-pass filter. Then, the parameter \hat{f} is constant in this band. The two parameters varied are a and this constant value of \hat{f} .

In order to compare performance, the threshold used must vary with the SNR parameter value and with the cutoff frequency. Let $L_1(\hat{f}, a)$ and $L_0(\hat{f}, a)$ be the asymptotic expected values of $i_n(X^n, Y^n)$ under

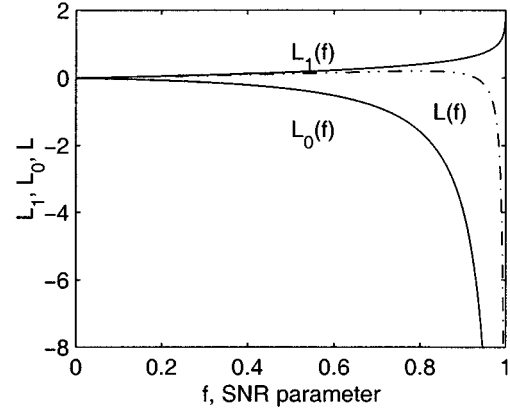


Fig. 1. Expected values L_1 and L_0 as functions of SNR parameter.

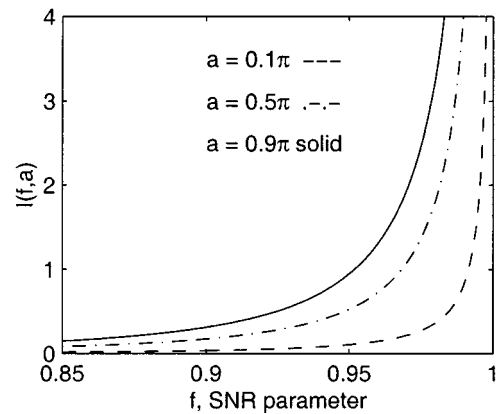


Fig. 2. Rate function I_1 parameterized by $\mathcal{L}(\hat{f})$ and three values of $a = (0.1, 0.5, 0.9)\pi$.

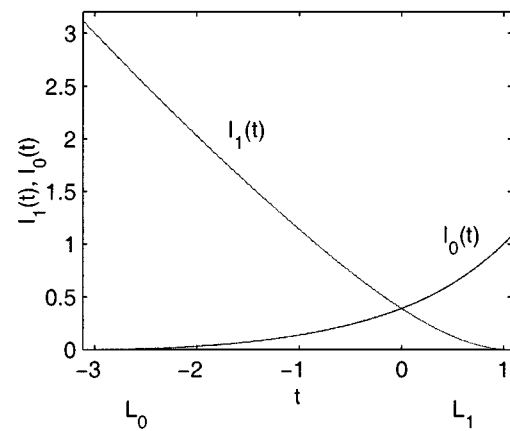


Fig. 3. Rate functions I_1 and I_0 for parameters $a = 0.9\pi$ and $\hat{f} = 0.7$.

hypotheses H_1 and H_0 , respectively, as functions of \hat{f} and a . From (29) and (30), these functions are linear in a . Shown in Fig. 1 are plots of the values of L_1 and L_0 versus \hat{f} for a specific value of $a = 0.5\pi$, and a plot of $\mathcal{L}(\hat{f}) = L_1 - 0.1(L_1 - L_0)$. We use $\mathcal{L}(\hat{f})$ as our threshold. Fig. 2 then shows three plots of the rate function I_1 evaluated at $\mathcal{L}(\hat{f})$ versus \hat{f} , for $a = 0.1\pi, 0.5\pi$, and 0.9π . Note that the rate function increases with a and with \hat{f} , as expected. Also, note that the performance approaches being perfect as \hat{f} increases to one. Plots of $I_1(t)$ and $I_0(t)$ as functions of the parameter t for the case $a = 0.9\pi$ and $\hat{f} = 0.7$ are presented in Fig. 3.

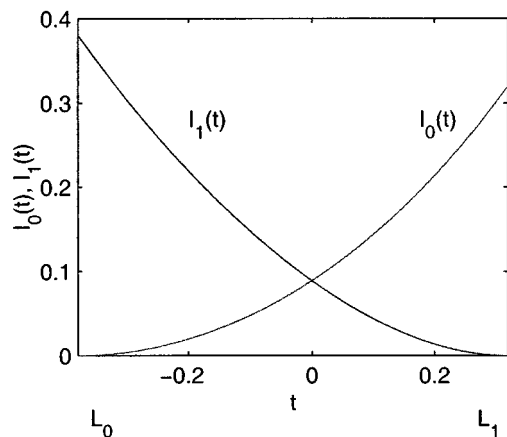


Fig. 4. Rate functions I_1 and I_0 for $p = 0.1$.

V. BINARY CASE

For a binary case, we assume the same additive model as in the Gaussian case. Thus, the two available observations are given by (21) and (22) with X_k, Y_k, Z_{ik} , and W_{ik} , $i = 1, 2$ being binary-valued, with all additions performed modulo 2. For Bernoulli $(1/2)$ entries of Z_{ik} and Bernoulli (p) entries of W_{ik} , we obtain

$$i_n(X^n, Y^n) = \frac{1}{n} \sum_{k=1}^n (1 + \Phi(X_k = Y_k) \log q + \Phi(X_k \neq Y_k) \log(1 - q)) \quad (31)$$

where $q = p^2 + (1 - p)^2$, and $\Phi(\cdot)$ is an indicator function that equals one if its argument is true. In this case,

$$\bar{I}(X; Y) = 1 - H(q)$$

and

$$\bar{D}(P_X \times P_Y, P_{XY}) = -1/2 \log(2q(1 - q)).$$

The asymptotic log-moment-generating function is given by the following expressions:

$$\bar{\phi}_1(s) = s + \log\{q^{s+1} + (1 - q)^{s+1}\}$$

under hypothesis H_1 and

$$\bar{\phi}_0(s) = (s - 1) + \log\{q^s + (1 - q)^s\}$$

under the hypothesis H_0 .

Similar to the Gaussian example, we find the explicit expressions for the large deviation rate functions under both hypotheses

$$I_1(t) = s_1 t - s_1 - \log\{q^{s_1+1} + (1 - q)^{s_1+1}\} \quad (32)$$

where s_1 is the optimal parameter s given by

$$s_1 = -1 + \frac{1}{\log\left(\frac{1-q}{q}\right)} \log \frac{(t - 1 - \log q)}{\log(1 - q) - (t - 1)}$$

and $I_0(t) = I_1(t) + t$. Plots of the rate functions for the case $p = 0.1$ are given in Fig. 4.

VI. CONCLUSION

This correspondence describes a framework for determining the performance of physical signature authentication based on likelihood models. The hypothesis-testing approach yields the information density as the test statistic for deciding if two realizations of a random process are independent. Under the hypothesis that the realizations

are dependent, the information density converges to the sup mutual information rate between the observations. Under the independence hypothesis, the information density converges to the negative sup-relative entropy rate between the product and the joint distributions. Conditions are given under which the convergence rates are exponential, and rate functions are derived. In the Gaussian case, the data are modeled as signature plus noise, the power spectra of the signature and noise processes determine the rate functions, and explicit formulas for log-moment-generating functions and rate functions are derived. Plots of the rate functions are shown when the signal processes are low-pass. A simple binary example is also discussed.

The results rely on models for the data. For many anticipated applications, the derivation of these models may be the most challenging aspect. For example, in biometric applications, stochastic models for the biometric features being measured are required. In magnetic medium authentication, explicit stochastic micromagnetic models for the medium must be used. Typically, present implementations use suboptimal, nonparametric test statistics such as the correlation coefficient between the candidate signature presented for authentication and the previous signature measurement [8]. Much work remains in quantifying the performance loss in using a suboptimal approach.

APPENDIX A

PROOF OF THEOREM 2

For the proof of the theorem, let $P_{X^n Y^n}^{(K)}$ be the K th-order Markov approximation to the stationary distribution $P_{X^n Y^n}$, and, thus, a K -step Markov source.

From the statement of the theorem, $P_{X^n Y^n}^{(K)} \gg P_{X^n} \times P_{Y^n}$, for all n . This ensures that the density

$$f_{X^n Y^n}(X^n, Y^n) = \frac{dP_{X^n} \times P_{Y^n}}{dP_{X^n Y^n}^{(K)}}$$

is well-defined and by [3, Theorem 8.2.1] has a limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log f_{X^n Y^n}(X^n, Y^n) = H_{P_X \times P_Y \| P_{XY}^{(K)}}(X_0, Y_0 | X_{-1}^-, Y_{-1}^-), \quad \text{a.e. } P_X \times P_Y$$

where

$$X_{-1}^- = (X_{-1}, X_{-2}, \dots) \quad \text{and} \quad Y_{-1}^- = (Y_{-1}, Y_{-2}, \dots)$$

and $H_{(\cdot|\cdot)}(\cdot)$ is a notation for the relative entropy between two distributions.

If $i_n(X^n, Y^n)$ is finite for any n , the chain rule for densities is applied

$$\begin{aligned} i_n(X^n, Y^n) &= -\frac{1}{n} \log f_{X^n Y^n}(X^n, Y^n) - \frac{1}{n} \log \frac{dP_{X^n Y^n}^{(K)}}{dP_{X^n Y^n}} \\ &\stackrel{n \rightarrow \infty}{\rightarrow} -H_{P_X \times P_Y \| P_{XY}^{(K)}}(X_0, Y_0 | X_{-1}^-, Y_{-1}^-) - \eta^{(K)}, \\ &\quad \text{a.e. } P_X \times P_Y \\ &= -\bar{H}_{P_X \times P_Y \| P_{XY}^{(K)}}(X, Y) - \eta^{(K)}. \end{aligned}$$

Since the last expression is a constant, the constant must also be the limit of the expected value of $i_n(X^n, Y^n)$ under the product distribution.

APPENDIX B

PROOF OF LEMMA 2

Consider the log-moment-generating function of the random variable $i_n(X^n, Y^n)$

$$\phi_{m,n}(s) = \frac{1}{n} \log E_m[e^{s i_n(X^n, Y^n)}], \quad m = 0, 1. \quad (33)$$

Substituting (23) into (33) and taking the expectation, one obtains

$$\phi_{m,n}(s) = -\frac{1}{2n} \log \det[s\mathbf{R}_m(\mathbf{R}_1^{-1} - \mathbf{R}_0^{-1}) + \mathbf{I}] - \frac{s}{2n} \log \det[\mathbf{R}_1\mathbf{R}_0^{-1}], \quad m = 0, 1 \quad (34)$$

where \mathbf{R}_1 and \mathbf{R}_0 are given by (24).

Under hypothesis H_1

$$\begin{aligned} \phi_{1,n}(s) &= -\frac{1}{2n} \log \det[(s+1)\mathbf{I} - s\mathbf{R}_1\mathbf{R}_0^{-1}] \\ &\quad - \frac{s}{2n} \log \det[\mathbf{R}_1\mathbf{R}_0^{-1}] \\ &= -\frac{1}{2n} \log \det[\mathbf{I}_n - s^2\mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_1)]^{-1} \\ &\quad \cdot \mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_2)]^{-1}] \\ &\quad - \frac{s}{2n} \log \det[\mathbf{I}_n - \mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_1)]^{-1} \\ &\quad \cdot \mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_2)]^{-1}]. \end{aligned}$$

Consider the product of the matrices

$$\mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_1)]^{-1}\mathbf{K}_n(S_Z)[\mathbf{K}_n(S_Z) + \mathbf{K}_n(S_2)]^{-1}. \quad (35)$$

To analyze eigenvalues of the matrix product in (35), we first apply triangular factorization to the matrix $\mathbf{K}_n(S_Z)$ given by

$$\mathbf{K}_n(S_Z) = \mathbf{K}_n^{1/2}(S_Z)\mathbf{K}_n^{\dagger/2}(S_Z)$$

where $\mathbf{K}_n^{1/2}(S_Z)$ is a lower-triangular matrix and $\mathbf{K}_n^{\dagger/2}(S_Z)$ is its conjugate transpose. Then (35) can be rewritten as

$$\mathbf{K}_n^{1/2}(S_Z)[\mathbf{I}_n + \mathbf{K}_n^{-1/2}(S_Z)\mathbf{K}_n(S_1)\mathbf{K}_n^{-\dagger/2}(S_Z)]^{-1} \cdot [\mathbf{I}_n + \mathbf{K}_n^{-1/2}(S_Z)\mathbf{K}_n(S_2)\mathbf{K}_n^{-\dagger/2}(S_Z)]^{-1}\mathbf{K}_n^{-1/2}(S_Z).$$

Note that the products of matrices

$$\mathbf{K}_n^{-1/2}(S_Z)\mathbf{K}_n(S_i)\mathbf{K}_n^{-\dagger/2}(S_Z), \quad i = 1, 2$$

in the square brackets, are Hermitian matrices. After substituting the above expression into the last term of $\phi_{1,n}(s)$, we obtain

$$-\frac{s}{2n} \log \det[\mathbf{I}_n - [\mathbf{I}_n + \mathbf{K}_n^{-1/2}(S_Z)\mathbf{K}_n(S_1)\mathbf{K}_n^{-\dagger/2}(S_Z)]^{-1} \cdot [\mathbf{I}_n + \mathbf{K}_n^{-1/2}(S_Z)\mathbf{K}_n(S_2)\mathbf{K}_n^{-\dagger/2}(S_Z)]^{-1}]. \quad (36)$$

By [6, Theorems 4.3, 4.4] of Gray and Assumption 5, the matrix in the outer square brackets in (36) is asymptotically equivalent to the matrix

$$\mathbf{C}_n \left(1 - \frac{S_Z^2}{(S_Z + S_1)(S_Z + S_2)} \right)$$

where \mathbf{C}_n is the notation for a circulant matrix.

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(-\frac{1}{2n} \log \det[(s+1)\mathbf{I} - s\mathbf{R}_1\mathbf{R}_0^{-1}] - \frac{s}{2n} \log \det[\mathbf{R}_1\mathbf{R}_0^{-1}] \right) \\ = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - s^2 \hat{f}(\lambda)) d\lambda - \frac{s}{4\pi} \int_{-\pi}^{\pi} \log(1 - \hat{f}(\lambda)) d\lambda. \end{aligned}$$

APPENDIX C PROOF OF LEMMA 3

One can easily verify that under Assumption 5, the pair random process $\{(X_n, Y_n)\}$ is stationary and ergodic with $P_{X^n Y^n}$ and $P_{X^n} \times P_{Y^n}$ being mutually absolutely continuous for each n , and that Assumptions 3 and 4 hold. In this appendix, we show that the finite-gap information property is satisfied for the Gaussian stationary processes $\{X_n\}$ and $\{Y_n\}$.

By the property of mutual information

$$I(X_{k+1}; X^- | X^k) = -h(X_{k+1} | X^-, X^k) + h(X_{k+1} | X^k). \quad (37)$$

The first term in (37) is the differential entropy rate for the stationary Gaussian process and is equal to

$$h(X_{k+1} | X^-, X^k) = \frac{1}{2} \log 2\pi e + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(S_Z(\lambda) + S_1(\lambda)) d\lambda.$$

The last term in (37) is bounded as

$$\begin{aligned} h(X_{k+1} | X^k) &\leq h(X_{k+1}) \\ &= \frac{1}{2} + \frac{1}{2} \log \left(\int_{-\pi}^{\pi} (S_Z(\lambda) + S_1(\lambda)) d\lambda \right). \end{aligned} \quad (38)$$

Therefore, the conditional mutual information (37) is finite for every $k \geq 1$, and the finite-gap information property is satisfied.

REFERENCES

- [1] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation*. New York: Wiley, 1990.
- [2] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*. New York: Springer-Verlag, 1985.
- [3] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
- [4] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. New York: Chelsea, 1984.
- [5] R. M. Gray, "On the asymptotic eigenvalue distribution of toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 767–800, 1972.
- [6] —, "Toeplitz and Circulant Matrices: A Review," Report, Elec. Eng. Dept., Stanford Univ., Stanford, CA 94305, 1993.
- [7] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752–772, May 1993.
- [8] J. R. Hoinville, R. S. Indeck, and M. W. Muller, "Spatial noise phenomena of longitudinal magnetic recording media," *IEEE Trans. Magn.*, vol. 28, pp. 3398–3406, 1992.
- [9] R. E. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.
- [10] R. S. Indeck and E. Glavinias, "Fingerprinting magnetic media," *IEEE Trans. Magn.*, vol. 29, pp. 4095–4097, Nov. 1993.
- [11] D. Agrawal, "Magnetic recording system design to reduce medium noise through signal precompensation," Master's thesis, Washington Univ., Dept. Elec. Eng., St. Louis, MO, Aug. 1995.
- [12] J. A. O'Sullivan, D. Agrawal, R. S. Indeck, and M. W. Muller, "Write-read-write signal precompensation techniques for magnetic recording," in *Coding and Signal Processing for Information Storage*. Philadelphia, PA, Oct. 1995, pp. 39–47.
- [13] "Special issue on automated biometric systems," *Proc. IEEE*, vol. 85, pp. 1341–1516, Sept. 1997.
- [14] A. Jain et al., Ed., *Biometrics: Personal Identification in Networked Society*. Norwell, MA: Kluwer, 1999.
- [15] "Biometrics: the Future of Identification," *IEEE Computer*, vol. 33, no. 2, pp. 46–81, 2000.
- [16] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.