ERROR EXPONENTS FOR ONE-BIT WATERMARKING

Tie Liu and Pierre Moulin
{tieliu, moulin}@ifp.uiuc.edu

University of Illinois
Beckman Inst., Coord. Sci. Lab & ECE Dept.
405 N. Mathews Ave., Urbana, IL 61801

ABSTRACT

Quantization index modulation (QIM) is a powerful host-interference rejecting method for data hiding. This paper applies QIM to one-bit watermarking and proposes a simple but powerful watermark detector. We derive lower bounds on the error exponents for the detector under a quadratic distortion constraint for the watermarker and additive white Gaussian noise attacks. These bounds are independent of the host-signal distribution and are substantially better than recently derived bounds for public (blind) spread-spectrum watermarking.

1. INTRODUCTION

Data hiding and one-bit watermarking are two fundamental problems in watermarking research. The data hiding problem has been systematically treated for several years, with quantization index modulation (QIM) [1] being recognized as the most successful method to approach the fundamental limits predicted by information theory [2]. The most fascinating characteristic of QIM is that it can completely reject host interference. Traditional embedding methods such as spread spectrum do not possess that property.

This paper proposes a simple but powerful detector and develops the first quantitative performance analysis of QIM for one-bit watermarking. Specifically, we derive error exponents for binary hypothesis testing and the receiver operating characteristic of the detector. As expected, interference from the host signal can be completely rejected.

2. MATHEMATICAL MODEL

One-bit watermarking can be modeled as in Fig. 1. Let S^n be the length-n host signal to be marked. A watermark may be inserted in S^n, resulting in a marked signal X^n that is made publicly available. We write $X^n = \psi(S^n, M, K^n)$,

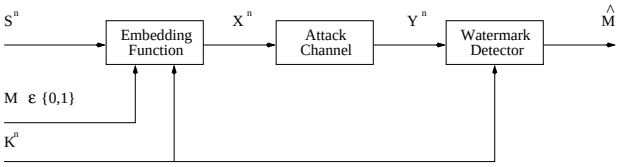

Fig. 1. Mathematical model of one-bit watermarking.

where $M \in \{0, 1\}$ [1] and K^n is a secret key shared with the detector. The embedding function ψ must satisfy a certain distortion constraint. The attacker takes X^n and produces a degraded signal Y^n in an attempt to fool the watermark detector. The detector has access to Y^n and K^n but not to S^n (*public watermarking*) and must decide whether the watermark was embedded or not. In Sec. 3, we shall also consider *private watermarking* where S^n is available to the detector. The detector's output is $\hat{M} = \hat{M}(Y^n, K^n) \in \{0, 1\}$.

We focus on the *additive Gaussian case* where

$$\frac{1}{n}\mathbb{E}\,\|X^n - S^n\|^2 \leq D_1, \tag{1}$$

$$Y^n = X^n + W^n \quad \text{and} \quad W^n \sim \mathcal{N}(0, D_2\mathbf{I}_n). \tag{2}$$

Here, $\mathbb{E}\,(\cdot)$ denotes mathematical expectation, $\|\cdot\|$ denotes l_2 norm, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathcal{N}(\mu, \mathbf{R})$ denotes the Gaussian distribution with mean μ and covariance matrix $\mathbf{R}$. In addition, we assume that

$$\max\{D_1, D_2\} \ll \min_{1\leq i\leq n} \sigma^2_{s,i}, \tag{3}$$

where $\sigma^2_{s,i}$ denotes the variance of the i-th component of S^n. This low-distortion regime is typical of many watermarking problems. We require that

$$\psi(S^n, M = 0, K^n) = S^n, \tag{4}$$

i.e., ψ reproduces the original signal when no watermark is embedded.

THIS WORK WAS SUPPORTED BY NSF GRANTS CCR 00-81268 AND CDA 96-24396.

[1] This is unlike in the data hiding problem [1, 2], where the hidden message rate is R bits per sample: $M \in \{1, 2, \cdots, 2^{nR}\}$.

It is assumed that the watermark detector knows the statistics of the host signal and the attack channel; the detector then implements the optimal likelihood ratio test (LRT). In this context, the performance of the one-bit watermarking system is determined by the embedding function. We evaluate the false-alarm and miss exponents, which are respectively defined as

$$E_{FA} \;=\; \lim_{n\to\infty} -\frac{1}{n}\ln P_{FA}^{(n)} \tag{5}$$

and

$$E_M \;=\; \lim_{n\to\infty} -\frac{1}{n}\ln P_M^{(n)}, \tag{6}$$

when the limits exist. Here, $P_{FA}^{(n)} = Pr\{\hat{M}=1\,|M=0\}$ and $P_M^{(n)} = Pr\{\hat{M}=0\,|M=1\}$.

3. ERROR EXPONENTS FOR SPREAD-SPECTRUM WATERMARKING

Traditional one-bit watermarking systems use spread-spectrum embedding. A game-theoretic methodology to design and embed spread-spectrum watermarks was recently investigated in [3]. The spread-spectrum embedding function takes the additive form

$$X^n \;=\; \psi(S^n, M=1, K^n) \;=\; S^n + P^n, \tag{7}$$

where the watermark P^n is asymptotically $\mathcal{N}(0, D_1\mathbf{I}_n)$, and depends on the secret key K^n. The detector solves the following binary hypothesis testing problem:

$$\begin{cases} M=0: & Y^n = S^n + W^n \\ M=1: & Y^n = P^n + S^n + W^n. \end{cases}$$

Assume that the host signal is a Gaussian random vector with independent and identically distributed components [2]

$$S^n \sim \mathcal{N}(0, \sigma_s^2\mathbf{I}_n). \tag{8}$$

Then, for public watermarking, the LRT is a correlation test:

$$T(Y^n|K^n) = \sum_{i=1}^{n} P_i Y_i \;\underset{\hat{M}=0}{\overset{\hat{M}=1}{\gtrless}}\; n\rho, \tag{9}$$

where $n\rho$ is the threshold of the test. The probabilities of error can be directly calculated as

$$P_{FA}^{(n)} \;=\; \mathcal{Q}\left(\sqrt{\frac{n\rho^2}{\overline{P^2}(\sigma_S^2+D_2)}}\right) \tag{10}$$

[2]The paper [3] treats the more general case of colored host signals and attacks.

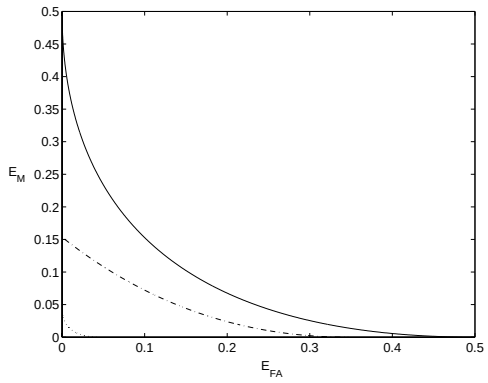

Fig. 2. Receiver operating characteristics for spread-spectrum embedding and QIM embedding. The solid line and the dotted line give the possible tradeoff between error exponents for private and public spread-spectrum embedding, respectively. The dashed line represents lower bounds on the error exponents for (public) QIM embedding. Numerical values are based on $D_1 = D_2 = \sigma_s^2/10$.

$$P_M^{(n)} \;=\; \mathcal{Q}\left(\sqrt{\frac{n(\overline{P^2}-\rho)^2}{\overline{P^2}(\sigma_S^2+D_2)}}\right), \tag{11}$$

where $\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$ and $\overline{P^2} = \frac{1}{n}\sum_{i=1}^n P_i^2$, which converges to D_1 almost surely by the strong law of large numbers. Using the property

$$\lim_{n\to\infty} -\frac{1}{n}\ln\mathcal{Q}(\sqrt{nx}) = \frac{1}{2}x, \quad x>0,$$

we obtain the error exponents as

$$E_{FA} \;=\; \frac{\rho^2}{2D_1(\sigma_S^2+D_2)} \text{ and } E_M \;=\; \frac{(D_1-\rho)^2}{2D_1(\sigma_S^2+D_2)} \tag{12}$$

for any $0 \le \rho \le D_1$.

For private watermarking, the LRT becomes

$$T(Y^n|K^n,S^n) = \sum_{i=1}^{n} P_i(Y_i - S_i) \;\underset{\hat{M}=0}{\overset{\hat{M}=1}{\gtrless}}\; n\rho, \tag{13}$$

and the error exponents can be similarly calculated as

$$E_{FA} = \frac{\rho^2}{2D_1D_2} \quad \text{and} \quad E_M = \frac{(D_1-\rho)^2}{2D_1D_2}. \tag{14}$$

The error exponents for both cases are shown in Fig. 2 (see solid line and dotted line parameterized by ρ). The performance for public watermarking is much worse than that for private watermarking due to the host interference, which is extremely strong in the low-distortion regime.

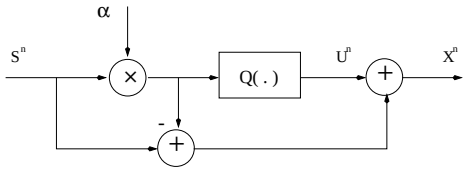

Fig. 3. The QIM embedding function $(M = 1)$.

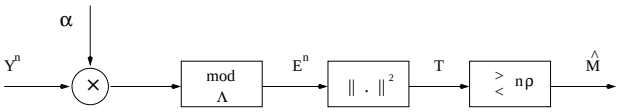

Fig. 4. Watermark detector: $\hat{M} \in \{0, 1\}$.

4. ERROR EXPONENTS FOR QIM WATERMARKING

4.1. Embedding Function and Watermark Detector

Fig. 3 shows the QIM embedding function which has been previously used in data hiding and communication problems [1, 4]. The host signal S^n is scaled by $\alpha \in (0, 1]$ and then quantized using an n-dimensional "good" lattice D_1-quantizer Q [4, 5]:

$$U^n = Q(\alpha S^n) = \alpha S^n + Z^n. \tag{15}$$

The basic Voronoi cell $\mathcal{V}_0$ associated with the lattice Λ is "nearly spherical". We have $\frac{1}{n}\mathbb{E}\,\|Z^n\|^2 = D_1$. Under the low-distortion assumption (3), Z^n is uniformly distributed within the reflective image of $\mathcal{V}_0$, i.e.,

$$Z^n \sim \mathbb{U}(-\mathcal{V}_0). \tag{16}$$

Here, we do not need to assume the host signal is Gaussian as we did in (8) for spread-spectrum embedding, as long as (3) holds. The marked signal X^n is obtained as

$$X^n \;=\; \psi(S^n, M = 1, K^n) \;=\; U^n + (1 - \alpha)S^n, \tag{17}$$

where the lattice Λ may depend on the secret key K^n. Inserting (15) into (17), we have

$$\begin{aligned} \psi(S^n, M = 1, K^n) &= S^n + Z^n \\ &= \tfrac{1}{\alpha}U^n + (1 - \tfrac{1}{\alpha})Z^n. \end{aligned} \tag{18}$$

In the data hiding problem, the choice $\alpha = \frac{D_1}{D_1+D_2}$ achieves the embedding capacity [4]. As we will see, in one-bit watermarking, α can also be judiciously chosen to optimize the appropriate performance measure, i.e., the error exponents.

The optimal LRT is hard to evaluate. Instead, we propose the suboptimal watermark detector depicted as in Fig. 4. The degraded signal Y^n is scaled by α and then quantized using the lattice quantizer Q. The quantization noise becomes

$$E^n \;=\; \alpha Y^n \mod \Lambda \;=\; \alpha Y^n - Q(\alpha Y^n). \tag{19}$$

The lattice detector compares the squared norm of E^n with a threshold

$$T(Y^n|K^n) = \|E^n\|^2 \underset{\hat{M}=1}{\overset{\hat{M}=0}{\gtrless}} n\rho. \tag{20}$$

The (possibly slight) loss of optimality is due to the quantization operation, which discards some information contained in Y^n.

4.2. Error-Exponent Analysis

When $M = 0$, it follows from (2), (4) and (19) that

$$E^n = \alpha(S^n + W^n) \mod \Lambda. \tag{21}$$

Clearly, we have $\frac{1}{n}\mathbb{E}\,\|\alpha(S^n + W^n)\|^2 \gg D_1$ due to the low-distortion assumption (3). Therefore

$$E^n \sim \mathbb{U}(\mathcal{V}_0). \tag{22}$$

When $M = 1$, it follows from (2), (17) and (19) that

$$\begin{aligned} E^n &= (U^n + (\alpha - 1)Z^n + \alpha W^n) \mod \Lambda \\ &= ((\alpha - 1)Z^n + \alpha W^n) \mod \Lambda, \end{aligned} \tag{23}$$

where the second equality is because $U^n \in \Lambda$. By (2) and (16), we have

$$\frac{1}{n}\mathbb{E}\,\|(\alpha - 1)Z^n + \alpha W^n)\|^2 \;=\; (\alpha - 1)^2 D_1 + \alpha^2 D_2,$$

which can be made smaller than D_1, the normalized square radius of $\mathcal{V}_0$. Further, the distribution of E^n is complicated by the non-Gaussian term $(\alpha - 1)Z^n$, the mod-Λ operation, and the nonspherical shape of $\mathcal{V}_0$. To derive error exponents, we resort to bounding techniques.

The main result is stated in the following theorem, and illustrated in Fig. 2 (see dashed line parameterized by ρ).

Theorem: *For QIM one-bit watermarking, with $\alpha = \frac{D_1}{D_1+D_2}$, the false-alarm and miss exponents can be simultaneously bounded from below as*

$$E_{FA} \geq \frac{1}{2}\ln\frac{D_1}{\rho} \tag{24}$$

$$E_M \geq \frac{1}{2}\left(\frac{\rho}{D_2}\left(1 + \frac{D_2}{D_1}\right) - \ln\left(\frac{\rho}{D_2}\left(1 + \frac{D_2}{D_1}\right)\right) - 1\right) \tag{25}$$

for any $\frac{D_1 D_2}{D_1+D_2} \leq \rho \leq D_1$.

Outline of Proof: The probability of false alarm can be

bounded from above as follows:

$$\begin{aligned} P_{FA}^{(n)} &= Pr\{\,\|E^n\|^2 < n\rho \,|\, M=0\,\} \\ &= \frac{vol(B^n(\sqrt{n\rho}) \cap \mathcal{V}_0)}{vol(\mathcal{V}_0)} \\ &\leq \frac{vol(B^n(\sqrt{n\rho}))}{vol(\mathcal{V}_0)} \\ &= \left(\frac{n}{n+2}\frac{G_n}{G_n^*}\frac{\rho}{D_1}\right)^{n/2}, \end{aligned} \tag{26}$$

where $vol(\cdot)$ denotes the volume of a set, $B^n(R)$ denotes an n-dimensional ball with a radius of R, and G_n and G_n^* are normalized second moments of the lattice quantizer and the n-dimensional ball, which converge to $1/2\pi e$ as $n \to \infty$ [5]. Hence, (24) follows from (5) and (26).

The probability of miss can be written as

$$P_M^{(n)} = Pr\{\,\|E^n\|^2 > n\rho \,|\, M=1\,\}. \tag{27}$$

We bound this probability in five steps:

Step 1: $P_M^{(n)} \leq Pr\{\,\|(\alpha-1)Z^n + \alpha W^n\|^2 > n\rho\,\}$, following from the use of the Voronoi (nearest neighbor) quantizer.

Step 2: $P_M^{(n)} \leq \left(\frac{R_c}{R_e}\right)^n Pr\{\,\|\widetilde{V}^n\|^2 > n\rho\,\}$, where R_c and R_e are respectively the covering radius and the equivalent radius of $\mathcal{V}_0$, $\widetilde{V}^n = \widetilde{Z}^n + W^n$, and $\widetilde{Z}^n \sim \mathbb{U}(B^n(R_c))$.

Step 3: For any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that for any $n \geq N$ we have $P_M^{(n)} \leq \left(\frac{R_c}{R_e}e^{\varepsilon}\right)^n Pr\{\,\|\hat{V}^n\|^2 > n\rho\,\}$, where $\hat{V}^n = G^n + W^n$ and $G^n \sim \mathcal{N}(0, D_1\mathbf{I}_n)$.

Step 4: (Chernoff bound for χ_n^2 random variables) For any $\varepsilon > 0$ and n large enough, we have

$$P_M^{(n)} \leq \left(\frac{R_c}{R_e}e^{\varepsilon}\right)^n \exp(-nC(\rho,\alpha)),$$

where $C(\rho,\alpha) = \frac{1}{2}\left(\frac{\rho}{\sigma^2} - \ln\frac{\rho}{\sigma^2} - 1\right)$ depends on α via $\sigma^2 = (\alpha-1)^2 D_1 + \alpha^2 D_2$.

Step 5: $C(\rho,\alpha)$ is maximized by $\alpha^* = \frac{D_1}{D_1+D_2}$.

Now choose ε arbitrarily small. We have $\frac{R_c}{R_e} \to 1$ as $n \to \infty$ [5]. Then (25) follows. This completes the proof of the theorem. □

Next consider the total probability of error, which is the sum of P_{FA} and P_M weighted by their priors. The exponent for this probability of error is given by $E_{TE} = \min\{E_{FA}, E_M\}$. The threshold of the test, $n\rho$, can be selected to equalize E_{FA} and E_M. We have the following result, which is illustrated in Fig. 5.

Proposition: *Choosing* $\rho = \frac{D_1 D_2}{D_1+D_2}\left(1 + \ln\left(1 + \frac{D_1}{D_2}\right)\right)$, *we have for (public) QIM embedding*

$$E_{TE} \geq \ln\left(1 + \frac{D_1}{D_2}\right) - \ln\left(1 + \ln\left(1 + \frac{D_1}{D_2}\right)\right). \tag{28}$$

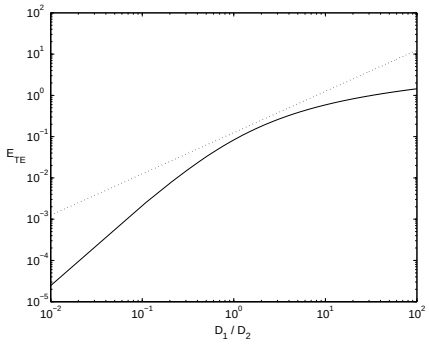

Fig. 5. Lower bound on E_{TE} for (public) QIM embedding (solid line) and exact value of E_{TE} for private spread-spectrum embedding (dotted line).

Choosing $\rho = \frac{1}{2}D_1$, *we have for private spread-spectrum embedding*

$$E_{TE} = \frac{1}{8}\frac{D_1}{D_2}. \tag{29}$$

5. CONCLUDING REMARKS

The lower bounds derived for our lattice detector also provide lower bounds for the optimal QIM error exponents. Our lower bounds are independent of the host-signal distribution. This shows that QIM is a host-interference rejecting method for one-bit watermarking. It is also interesting to note that for QIM embedding there is no symmetry between the false alarm exponent and the miss exponent due to the constraint (4). The error exponents for multi-bit transmission without this constraint can be treated using similar bounding techniques.

6. REFERENCES

[1] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Info. Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001.

[2] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," to appear in *IEEE Trans. Info. Theory*, March 2003.

[3] P. Moulin and A. Ivanović, "The zero-rate spread-spectrum watermarking game," to appear in *IEEE Trans. Signal Processing*, April 2003.

[4] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Info. Theory*, vol. 48, no. 6, pp. 1250-1276, June 2002.

[5] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Info. Theory*, vol. 42, no. 4, pp. 1152-1159, July 1996.