QIM WATERMARKING GAMES

Anil Kumar Goteti and Pierre Moulin

University of Illinois at Urbana-Champaign
Beckman Inst., Coord. Sci. Lab & ECE Dept.
405 N. Mathews Ave., Urbana, IL 61801, USA
email: {goteti, moulin}@ifp.uiuc.edu

ABSTRACT

Quantization Index Modulation (QIM) methods are widely used for blind data embedding and watermarking. Given a QIM watermarking code, we ask what is the attacker's noise distribution that maximizes probability of error of the detector. For memoryless attacks, the problem is reduced to a convex programming problem. Next, we derive QIM code parameters that are minmax optimal.

1. INTRODUCTION

Current research in blind watermarking has focused on the development of QIM schemes [1, 2], which outperform spread-spectrum modulation techniques in various scenarios. It follows from Erez and Zamir's work [3, 4] that the family of lattice QIM schemes is capacity-achieving for watermarking problems with quadratic distortion constraints for the embedder and Gaussian noise attacks.

This paper extends our recent work on spread-spectrum watermarking games [5, 6] to QIM. We optimize QIM code parameters and attack using a game-theoretic framework.

2. SCALAR-QUANTIZER INDEX MODULATION

To simplify the exposition, consider the problem of embedding one bit $m \in \{0,1\}$ into a length-n data block $\mathbf{s} = \{s_i, 1 \le i \le n\}$. The rate of the code is $\frac{1}{n}$. Each component s_i is marked using the distortion-compensated scalar QIM embedding function

$$x_i = f_i(s_i, m) \;:=\; Q(\alpha s_i + (-1)^m z_i - d_i) + (1-\alpha)s_i - (-1)^m z_i + d_i. \quad (1)$$

Here $Q(\cdot)$ denotes the uniform scalar quantizer with step size Δ, and $\alpha \in (0,1]$ is a parameter to be optimized. The sequences $\mathbf{z} = \{z_i, 1 \le i \le n\}$ and $\mathbf{d} = \{d_i, 1 \le i \le n\}$ are secret dither sequences shared by the watermark embedder and the detector. The sequence $\mathbf{z}$ is bipolar with components in $\{\pm\frac{\Delta}{4}\}$. The sequence $\mathbf{d}$ is made of independent and identically distributed (iid) random variables uniformly distributed over $[0, \Delta)$. This has the effect of making the quantization noise white, independent of the host signal, and uniformly distributed over $[-\frac{\Delta}{2}, \frac{\Delta}{2}]^n$ [4]. The per-sample mean-squared distortion due to embedding is $D_1 = \frac{1}{n}\mathbb{E}\|\mathbf{X} - \mathbf{S}\|^2 = \frac{\Delta^2}{12}$.

WORK SUPPORTED BY NSF GRANTS CCR 00-81268, CCR 02-08809, AND CDA 96-24396.

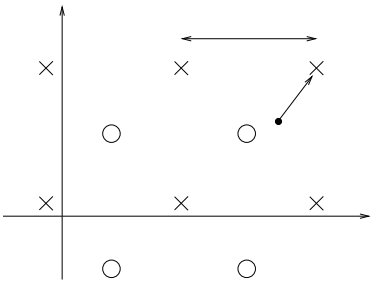


Fig. 1. Using dithered scalar QIM to embed 1 bit in $\mathbf{s} \in \mathbb{R}^2$.

The marked signal $\mathbf{x}$ is corrupted by the attacker, resulting in a noisy signal $\mathbf{y} = \mathbf{x} + \mathbf{w}$. We assume that the noise $\mathbf{w}$ has probability density function (pdf) $p_{\mathbf{w}}$ (to be optimized), is independent of $\mathbf{x}$, and satisfies the distortion constraint $\frac{1}{n}\int \|\mathbf{w}\|^2 p_{\mathbf{W}}(\mathbf{w})\, d\mathbf{w} \le D_2$.

3. DETECTION

3.1. Preprocessor

The QIM detector is a lattice detector. It first reduces the data $\mathbf{y}$ to the statistic

$$\tilde{y}_i := \alpha y_i - d_i \bmod \Delta, \quad 1 \le i \le n. \quad (2)$$

Due to the modulo operation, we have $0 \le \tilde{y}_i \le \Delta$ for all i. The detection problem is of the form

$$\begin{cases} H_0 \;:\; \tilde{\mathbf{Y}} \sim p_0 \\ H_1 \;:\; \tilde{\mathbf{Y}} \sim p_1 \end{cases} \quad (3)$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

where the notation $\tilde{\mathbf{Y}} \sim p$ means that $\tilde{\mathbf{Y}}$ is a random vector with probability distribution $p(\tilde{\mathbf{y}})$.

The minimum-distance decoding rule is given by

$$\hat{m} = \text{argmin}_{m\in\{0,1\}} \|\tilde{\mathbf{y}} + (-1)^m \mathbf{z}\|. \tag{4}$$

This detector determines which of the two lattice cosets $\Lambda_0 = -\mathbf{z} + \Delta\mathbb{Z}$ and $\Lambda_1 = \mathbf{z} + \Delta\mathbb{Z}$ the scaled vector $\alpha\mathbf{y}$ is closest to. The distance between Λ_0 and Λ_1 is

$$d_{\min} = \|2\mathbf{z}\| = \frac{\Delta}{2}\sqrt{n} = \sqrt{3nD_1}.$$

3.2. Modulo Additive Noise Channel

When bit m is embedded, the data $\tilde{\mathbf{y}}$ is the sum (modulo Δ) of the signal $(-1)^{1+m}\mathbf{z}$ and a noise vector $\mathbf{v}$:

$$\tilde{\mathbf{y}} = (-1)^{1+m}\mathbf{z} + \mathbf{v} \bmod \Delta \tag{5}$$

The vector $\mathbf{v}$ is the sum (modulo Δ) of two components:

- Self-noise $\tilde{\mathbf{e}}$, due to quantization (1). Its components are iid with uniform distribution

$$p_{\tilde{E}}(\tilde{e}) = \frac{1}{\Delta(1-\alpha)} 1_{\{|\tilde{e}| \le \frac{\Delta}{2}(1-\alpha)\}}.$$

- Scaled attacker's noise: $\tilde{\mathbf{W}} := \alpha\mathbf{W}$.

The parameter α trades off self-noise against attacker's noise at the receiver. For small α, the self-noise $\tilde{\mathbf{E}}$ dominates $\mathbf{V}$, and the attacker's noise becomes inconsequential. For $\alpha = 1$, the self-noise is zero, and thus $\mathbf{V} = \mathbf{W}$.

Equation (5) describes a Modulo Additive Noise (MAN) channel (see Fig. 2), analogous to the mod AWGN channel in [3]. The detector must decide which of the two signals $\mathbf{z}$ and $-\mathbf{z}$ was transmitted. Since $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{W}}$ are statistically independent, the pdf of $\mathbf{V}$ is the circular convolution of the pdf's of $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{W}}$:

$$p_{\mathbf{V}}(\mathbf{v}) = (p_{\tilde{\mathbf{E}}} \circledast p_{\tilde{\mathbf{W}}})(\mathbf{v}), \quad \mathbf{v} \in [0,\Delta]^n. \tag{6}$$

Therefore the pdf of $\tilde{\mathbf{Y}}$ under H_m takes the form

$$p_m(\tilde{\mathbf{y}}) = p_{\mathbf{V}}(\tilde{\mathbf{y}} + (-1)^m\mathbf{z}), \quad m = 0, 1, \tag{7}$$

which is simply a translate of $p_{\mathbf{V}}$.

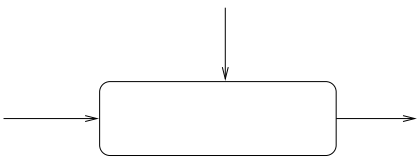


Fig. 2. Modulo Additive Noise Channel.

3.3. Maximum-Likelihood (ML) Detector

Bits $m = 0$ and $m = 1$ are assumed to be equally likely. Assume the detector knows $p_{\mathbf{W}}$ and is thus able to implement the optimal ML detection rule:

$$\frac{p_{\mathbf{V}}(\tilde{\mathbf{y}} - \mathbf{z})}{p_{\mathbf{V}}(\tilde{\mathbf{y}} + \mathbf{z})} \underset{H_0}{\overset{H_1}{\gtrless}} 1. \tag{8}$$

This rule coincides with the nearest-neighbor detection rule (4) when the attacker's noise is iid and 0-mean with a unimodal and symmetric pdf.

The probability of error for the test (8) is

$$P_e = \frac{1}{2}\int_{[0,\Delta]^n} \min(p_0(\tilde{\mathbf{y}}), p_1(\tilde{\mathbf{y}}))\, d\tilde{\mathbf{y}}. \tag{9}$$

4. WORST MEMORYLESS ATTACKS

Here we consider the case of memoryless attacks, where $p_{\mathbf{W}}(\mathbf{w}) = \prod_{i=1}^n p_W(w_i)$. The distortion constraint takes the form

$$\int w^2 p_W(w)\, dw \le D_2. \tag{10}$$

The noise $\mathbf{V}$ at the receiver is also iid with pdf $p_V = p_{\tilde{E}} \circledast p_W$. Define the two shifted pdf's $q_0(\tilde{y}) = p_V(\tilde{y} + \frac{\Delta}{4})$ and $q_1(\tilde{y}) = p_V(\tilde{y} - \frac{\Delta}{4})$. Fig. 3 shows q_0 and q_1 when $D_1/D_2 = 0.1$, $\alpha = 0.09$, and $\mathbf{w} \sim \mathcal{N}(0, D_2 I_n)$.

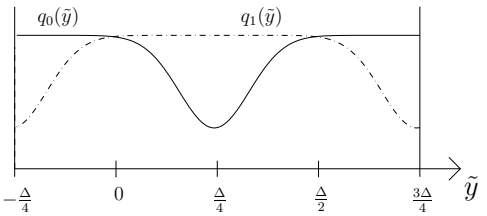


Fig. 3. Example of $q_0(\tilde{y})$ and $q_1(\tilde{y})$.

Given a watermarking code, the attacker wants to design p_W that maximizes P_e. Conversely, the watermark embedder wants to select α that minimizes P_e. Assuming the attacker knows α, the best α and p_W are the ones that achieve

$$\min_{0<\alpha\le 1} \max_{p_W} P_e(\alpha, p_W). \tag{11}$$

Theorem 1. For any α, the noise pdf p_W that minimizes P_e is symmetric around 0. The support set of p_W is $[-\frac{\Delta}{2\alpha}, \frac{\Delta}{2\alpha}]$.

Sketch of the proof. Using noise component w_i or $w_i + k\frac{\Delta}{\alpha}$ for any $k \in \mathbb{Z}$ does not change the value of the test statistic $\tilde{\mathbf{y}}$. Therefore we must have $|w_i| \le \frac{\Delta}{2\alpha}$ to minimize distortion. Symmetry of the worst p_W follows from a convexity argument.

Note. It follows that p_V is also symmetric around 0, and that the rival pdf's $q_0(\tilde{y})$ and $q_1(\tilde{y})$, have means $-\frac{\Delta}{4}$ and $\frac{\Delta}{4}$, respectively.

Property: the minimizing α in (11) is smaller than $\overline{\alpha} = \sqrt{\frac{3D_1}{4D_2}}$. For any $\alpha \geq \overline{\alpha}$, we have $\max_{p_W} P_e(\alpha, p_W) = \frac{1}{2}$.

Proof. An ideal p_W for the attacker would be one that assigns mass $\frac{1}{2}$ to $w = \frac{\Delta}{4\alpha}$ and to $w = -\frac{\Delta}{4\alpha}$, because p_0 and p_1 are identical in this case, and therefore $P_e = \frac{1}{2}$. To be feasible, such p_W must satistfy the distortion constraint $\mathbb{E}W^2 = (\frac{\Delta}{4\alpha})^2 = \frac{3D_1}{4\alpha^2} \leq D_2$. This is possible only if $\alpha \geq \sqrt{\frac{3D_1}{4D_2}}$.

4.1. Bhattacharyya bound

The Bhattacharyya bound on P_e is given by [7]

$$P_e \leq \frac{1}{2} e^{-nB(q_0,q_1)} \tag{12}$$

where

$$B(q_0, q_1) = -\ln \int_0^{\Delta} \sqrt{q_0(\tilde{y}) q_1(\tilde{y})}\, d\tilde{y} \tag{13}$$

is the Bhattacharyya distance between the pdf's q_0 and q_1. Moreover, the bound is tight in the exponent: [1]

$$\lim_{n\to\infty} [-\frac{1}{n} \ln P_e] = B(q_0, q_1).$$

The performance index (13) is much easier to evaluate than the n-dimensional integral (9), and can be used to determine how large n would be to guarantee a prescribed P_e.

The Bhattacharyya coefficient $B(q_0, q_1)$ may be viewed as a function b of α and p_W. Since the problem (11) is numerically hard to solve, we replace P_e with the Bhattacharyya bound (12) and solve the simpler (and asymptotically equivalent) problem

$$\max_{0<\alpha\leq 1} \min_{p_W} b(\alpha, p_W). \tag{14}$$

The function $b(\alpha, p_W)$ is convex in p_W and the distortion constraint (10) is linear in p_W. We solve (14) using the convex programming resources of [8].

Property: For $\frac{D_1}{D_2} > \frac{4}{3}$, we have $\min_{p_W} b(1, p_W) \leq \ln \frac{3D_1}{4D_2}$. Gaussian p_W is severely suboptimal for large $\frac{D_1}{D_2}$: $b(1, p_W) \sim \frac{D_1}{8D_2}$.

Fig. 4 shows the optimal α as a function of D_1/D_2, and Fig. 5 shows the worst-case p_W for three values of D_1/D_2. Note that the optimal α is neither $\frac{D_1}{D_1+D_2}$ [1, 2] nor $\sqrt{\frac{D_1}{D_1+2.7D_2}}$ [9], both of which are the results of optimization for coding problems with iid Gaussian W. The worst-case p_W is strongly non-Gaussian in all examples of Fig. 5.

[1] In general, a Chernoff bound with optimal Chernoff exponent is tight. However, due to the symmetry of p_V and the fact that q_0 and q_1 are translates of p_V, the optimal Chernoff exponent is $\frac{1}{2}$, and thus the optimal bound is the Bhattacharyya bound.

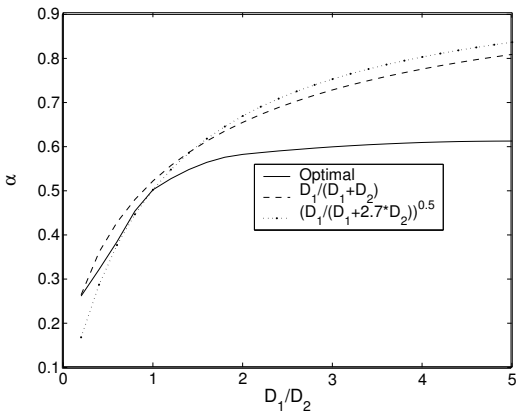
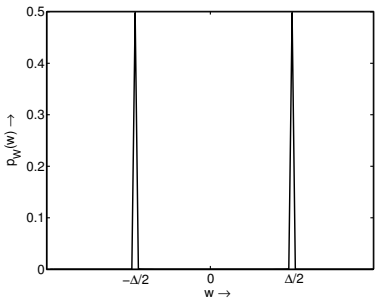
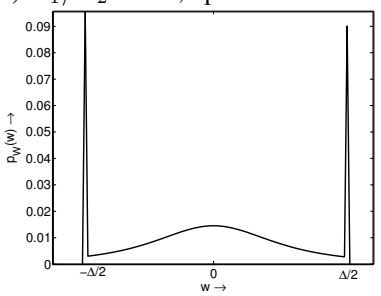


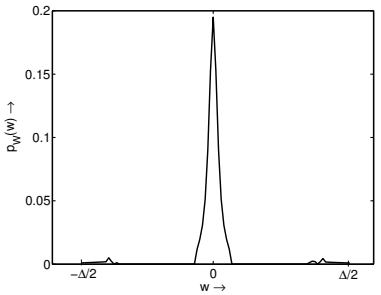
Fig. 4. Optimal α *vs.* D_1/D_2 for worst-case p_W.



(a) $D_1/D_2 = 0.2$, optimal $\alpha \approx 0.25$.



(b) $D_1/D_2 = 1$, optimal $\alpha \approx 0.5$.



(c) $D_1/D_2 = 5$, optimal $\alpha \approx 0.6$.

Fig. 5. Worst-case noise pdf p_W.

5. WORST ATTACKS WITH MEMORY

It is advantageous for the watermark embedder to use a randomization strategy that is more general than the simple dithering ($\mathbf{d}$) used so far. Specifically, we allow a randomized rotation of the lattice. This can be done by selecting a vector $\mathbf{z}$ of length $\frac{\Delta}{4}\sqrt{n}$ according to a pdf $p_{\mathbf{Z}}$, to be optimized. In Secs. 2—4, $p_{\mathbf{Z}}$ allocated uniform mass to the finite set $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}^n$. We wish to solve the problem

$$\min_{0<\alpha\leq 1} \min_{p_{\mathbf{Z}}} \max_{p_{\mathbf{W}}} P_e(\alpha, p_{\mathbf{Z}}, p_{\mathbf{W}}). \tag{15}$$

Theorem 2. The optimal $p_{\mathbf{Z}}$ and $p_{\mathbf{W}}$ in (15) are isotropic.

Proof: analogous to the proof of Theorem 3 in [10].

¿From Theorem 2 we conclude that the best $\mathbf{Z}$ is uniformly distributed on the surface of the n-dimensional sphere with radius $\sqrt{\frac{3}{4}nD_1}$, and the worst noise pdf is characterized by a pdf $p_R(\rho),\ \rho \geq 0$ in the radial direction: $p_{\mathbf{W}}(\mathbf{w}) = p_R(\|\mathbf{w}\|)$. The optimization problem (15) becomes:

$$\min_{0<\alpha\leq 1} \max_{p_R} P_e(\alpha, p_R) \tag{16}$$

where the maximization over p_R is subject to the distortion constraint $\int_0^\infty \rho^2 p_R(\rho)\, d\rho \leq nD_2$.

Theorem 3. For any α, the support set of the noise pdf p_R that achieves the maximum in (16) is $[\frac{\Delta\sqrt{n}}{4\alpha}, \frac{\Delta\sqrt{n}}{2\alpha}]$.

6. IMAGE WATERMARKING

To apply the theory of Sec. 4 to image watermarking, we have used a block-DCT decomposition (8×8 blocks), evaluated Just Noticeable Difference (JND) thresholds for each DCT coefficient, and selected a random subset of 1024 significant coefficients for watermarking. ¿From these coefficients, we formed 16 groups of $n = 64$ coefficients with similar JND values. One bit was embedded in each group. A group is treated as a host vector, and a quantization step size Δ equal to twice the JND for that group is evaluated. (This ensures that the maximum error due to embedding remains below the JND). The attacker also computes JND's (and therefore knows each $D_1 = \frac{\Delta^2}{12}$) and chooses $D_2 = 5D_1$. The optimal α and p_W are obtained from (14). Fig. 6 compares Bhattacharyya bounds on P_e for optimal and Gaussian attacks, as a function of D_1/D_2. Similar results were obtained for the actual P_e's. For instance, $P_e \approx 3.2 \times 10^{-3}$ for $D_1/D_2 = 0.2$; the Bhattacharyya bound is about 1.6 times larger. Also note that performance is independent of the image used. Different rate/P_e tradeoffs can be obtained by varying n.

7. DISCUSSION

While iid Gaussian noise is the attack noise that minimizes capacity under squared-error distortion constraints [2], the same noise is suboptimal against specific binning schemes such as scalar QIM. The theory can be extended straightforwardly to more general lattice QIM schemes, including sparse (STDM-like) QIM schemes. To achieve the performance limits given in this paper, we have assumed that the detector knows $p_{\mathbf{W}}$. Probability of error is expected to be worse for GLRT-type detectors [7], which do not know $p_{\mathbf{W}}$.

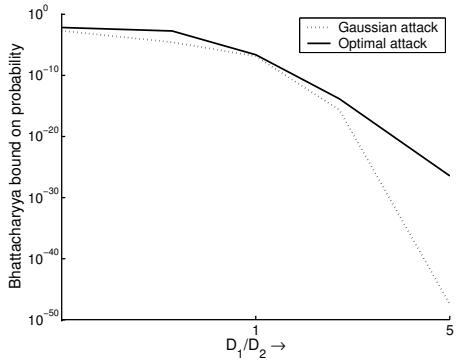


Fig. 6. Bhattacharyya bound (12) on P_e for optimal and Gaussian p_W, as a function of D_1/D_2, when $n = 64$.

8. REFERENCES

[1] B. Chen and G. W. Wornell, "Quantization Index Modulation Methods: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Trans. Info. Thy*, Vol. 47, No. 4, pp. 1423—1443, May 2001.

[2] P. Moulin and J. A. O'Sullivan, "Information–Theoretic Analysis of Information Hiding," *IEEE Trans. Information Theory*, Vol. 49, No. 3, pp. 563-593, March 2003.

[3] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested Linear/Lattice Codes for Structured Multiterminal Binning," *IEEE Tr. Info. Thy*, Vol. 48, pp. 1250—1276, June 2002.

[4] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN Channel with Lattice Encoding and Decoding," *preprint*, May 2001; revised, Sep. 2003.

[5] P. Moulin and A. Ivanović, "The Zero-Rate Spread-Spectrum Watermarking Game," *IEEE Transactions on Signal Processing*, Vol. 51, No. 4, pp. 1098—1117, Apr. 2003.

[6] A. K. Goteti and P. Moulin, "Two Private, Perceptual Data-Hiding Games," *Proc. ICASSP*, Montreal, Canada, May 2004.

[7] H. V. Poor, *An Introduction to Detection and Estimation Theory*, Springer-Verlag, 1994.

[8] `http://www-neos.mcs.anl.gov/neos/`, NEOS.

[9] J. J. Eggers, R. Bäuml, R. Tzschoppe and B. Girod, "Scalar Costa Scheme for Information Embedding," *IEEE Trans. Sig. Proc.*, Vol. 51, No. 4, pp. 1003—1019, Apr. 2003.

[10] T. Liu and P. Moulin, "Error exponents for watermarking game with squared-error constraints," *Proc. Int. Symp. on Info. Theory*, Yokohama, Japan, July 2003.