QUANTIZED PROJECTION DATA HIDING

Fernando Pérez-González and Félix Balado[*]

Signal Theory and Communications Department
University of Vigo, E-36200 Vigo, Spain
{fperez,fiz}@tsc.uvigo.es

ABSTRACT

In this paper we propose a novel data hiding procedure called Quantized Projection (QP), that combines elements from quantization (i.e. Quantization Index Modulation, QIM) and spread-spectrum methods. The method is based in quantizing a diversity projection of the host signal, inspired in the statistic used for detection in spread-spectrum algorithms. We carry on a theoretical analysis of QP together with its empirical validation to rigorously show that it offers an excellent performance: QP features probabilities of decoding error several orders of magnitude lower than the aforementioned families of methods for the same dimensionality (diversity) and attacking distortion level. In addition we introduce a Costa-based improvement of the basic QP method named Distortion Compensated QP.

1. INTRODUCTION

Quantized Projection (QP) methods aim at effectively combining the advantages of two big groups of methods: 1) Known-host-state methods, such as QIM [1], that use the host signal state for the embedding together with minimum distance decoding, without using host signal characterization; 2) Known-host-statistics methods, such as spread-spectrum ones, that do not use host signal state information, but perform statistical detection using host signal statistics. The idea is to employ a projection function that produces a scalar decision variable playing the same role as the decision statistic in known-host-statistics methods that is afterwards uniformly quantized in a similar way as QIM methods.

The description of this procedure resembles the so-called Spread Transform Dither Modulation (STDM) [2]. However, a crucial observation not made there is that, for a given quantization step, the larger cells in STDM allow for smaller embedding distortions. Alternatively, for a comparable level of watermarking distortions, the scalar quantization step can be made much larger. This adds a high degree of robustness to the method as it becomes more and more difficult for the attacker to change the projection to a wrong quantization centroid when the projection function is unknown. Now the hypothesis of host data uniform inside the Voronoi cells is no longer valid, what demands a new theoretical analysis. Also, an interesting side effect of QP is that the projection of the attacking distortion has a probability distribution function (pdf) that can be assumed to be Gaussian for a large variety of distortions.

For analyzing QP we will hide one information symbol $b = \pm 1$ into a host signal $\boldsymbol{x}$ using L pseudorandomly chosen samples indexed by the set $\mathcal{S}$. Without loss of generality we write the watermarked signal as the addition $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}$. We will denote by D_w the *embedding distortion*, that is, the power of the embedded watermark. Also $\lambda \triangleq \sqrt{\sigma_x^2 / D_w}$, with $\sigma_x^2 = E\{x^2[k]\}$, allows us to define the *document-to-watermark ratio*, $\mathrm{DWR} = 20 \log_{10} \lambda$. Before decoding, $\boldsymbol{y}$ undergoes a channel represented by additive noise $\boldsymbol{n}$ independent of $\boldsymbol{x}$ and following an arbitrary zero-mean pdf, yielding a received signal $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{n}$. Let $D_c = \sigma_n^2$ denote the *channel distortion*, i.e. the power of the distortion produced by the channel. Last we define $\xi \triangleq \sqrt{D_w / D_c}$, calling *watermark-to-noise ratio* to $\mathrm{WNR} = 20 \log_{10} \xi$.

2. QUANTIZED PROJECTION

The projection function consists in computing the cross-correlation between the watermarked image and the watermark, so for a single transmitted bit the projection r is such that

$$r = \sum_{k \in \mathcal{S}} y[k] \alpha[k] s[k] \tag{1}$$

with $s[k]$ a pseudorandom sequence such that $E\{s^2\} = 1$ and $\alpha[k]$ a perceptual restriction limit. The use of this projection function is based in its linearity and in the fact that it would be the optimal ML decoding function if the statistics of the host image were Gaussian [3]. Alternatively, this projection would result in the case of i.i.d. additive Gaussian noise in the channel. Note that an i.i.d. channel noise is not optimal from the perspective of its degree of invisibility; nevertheless, we make here this assumption in order to

[*]This work has been partly supported by the *Xunta de Galicia* under project PGIDT99 PX132203B, the European project Certimark (Certification of Watermarking Technologies), IST-1999-10987, and the CYCIT project AMULET, reference TIC2001-3697-C03-01.

keep the discussion simple. Improvements are possible using nonlinear projections adapted to the statistics of $\boldsymbol{x}$, but we will not pursue this issue here. It is possible to rewrite (1) as

$$r = r_x + r_w = \sum_{k\in\mathcal{S}} x[k]\alpha[k]s[k] + \sum_{k\in\mathcal{S}} w[k]\alpha[k]s[k] \quad (2)$$

where r_x is the projected host image and r_w is the projected watermark. The embedding of b is made by quantizing r with one of two uniform scalar quantizers whose centroids are given by the unidimensional lattices $\Lambda_{-1} = 2\Delta\mathbb{Z} - \Delta/2$ and $\Lambda_1 = 2\Delta\mathbb{Z} + 3\Delta/2$. Thus, the embedder finds r_w with the smallest magnitude such that $r_x + r_w \in \Lambda_b$, i.e. $r_w = Q_b(r_x) - r_x$, with $Q_b(x)$ the closest centroid to x in the lattice Λ_b.

There are infinitely many ways to select the watermark samples $w[k]$, $k \in \mathcal{S}$ so that this condition is satisfied; here we will content ourselves with choosing $w[k]$ proportional to the perceptual mask $\alpha[k]$. Then,

$$w[k] = \rho\alpha[k]s[k] \quad (3)$$

It is easy to see that $\rho = r_w / \sum_{k\in\mathcal{S}} \alpha^2[k]$. Noticing that r_w and $s[k]$ are statistically independent and that $E\{s^2\} = 1$, it is immediate to write

$$\text{Var}\{w[k]\} = \frac{\text{Var}\{r_w\}\alpha^2[k]}{\left(\sum_{k\in\mathcal{S}} \alpha^2[k]\right)^2} \quad (4)$$

In order to simplify the performance analysis while producing illustrative results let us assume a constant perceptual mask, i.e., $\alpha[k] = \alpha$, for all $k \in \mathcal{S}$. With this assumption (3) becomes $w[k] = r_w s[k]/L\alpha,\ \ k \in \mathcal{S}$, and (4) becomes

$$D_w = \text{Var}\{w[k]\} = \frac{\text{Var}\{r_w\}}{L^2\alpha^2} \quad (5)$$

For evaluating $\text{Var}\{r_w\}$ it is necessary to statistically characterize the random variable r_x. Since the $s[k]$, $k \in \mathcal{S}$, are statistically independent, it is possible to resort to the central limit theorem (CLT) to show that, for large L, r_x can be accurately modeled by a Gaussian pdf with zero mean and variance $\sigma_{r_x}^2 = L\alpha^2\sigma_x^2$. As the pdf of r_x is symmetrical, if the binary values of b are equally likely it is trivial to see that $E\{r_w\} = 0$. Now, assuming an equiprobable information bit b we have

$$\text{Var}\{r_w\} = \frac{E\{r_w^2|b=1\} + E\{r_w^2|b=-1\}}{2} \quad (6)$$

where

$$E\{r_w^2|b=1\} = \sum_{i=-\infty}^{\infty} \int_{\Delta(2i-\frac{1}{2})}^{\Delta(2i+\frac{3}{2})} f_{r_x}(r_x)(2i\Delta + \Delta/2 - r_x)^2 dr_x \quad (7)$$

where $f_{r_x}(r_x)$ is the pdf of r_x. A similar analysis applies for $E\{r_w^2|b=-1\}$, and substituting these two expectations into (6) and operating, we have

$$\text{Var}\{r_w\} = \Delta^2\left(\frac{1}{4} + I(\sigma_{r_x}/\Delta)\right) \quad (8)$$

where

$$I(\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=-\infty}^{\infty} \int_{-1/2}^{1/2} e^{-(r_x+i)^2/2\sigma^2}\, r_x^2\, dr_x \quad (9)$$

It is useful to note that for small σ_{r_x}/Δ, $I(\sigma_{r_x}/\Delta)$ can be approximated by $(\sigma_{r_x}/\Delta)^2$, which is in fact an upper bound. Having obtained the distortion D_w for arbitrary Δ, σ_x and α, we will determine the bit error probability at the channel output. For the proposed channel model the projection of $\boldsymbol{z}$ becomes

$$r = \sum_{k\in\mathcal{S}} z[k]\alpha[k]s[k] = r_x + r_w + r_n \quad (10)$$

where r_x and r_w were defined in (2) and the projected noise is $r_n = \sum_{k\in\mathcal{S}} n[k]\alpha[k]s[k]$. For L large, we can apply again the CLT to state that for a wide class of distributions for $n[k]$, the pdf of r_n can be approximated by a zero-mean Gaussian pdf with variance $\sigma_{r_n}^2 = L\alpha^2\sigma_n^2 = L\alpha^2 D_c$, assuming invariance in $\alpha[k]$.

The bit error probability P_e can be determined by taking into account the symmetry in the problem, assuming without loss of generality a transmitted $b = 1$ we can write[1]

$$P_e = 2\sum_{k=0}^{\infty}\left\{Q\left(\frac{(4k+1)\Delta}{2\sigma_{r_n}}\right) - Q\left(\frac{(4k+3)\Delta}{2\sigma_{r_n}}\right)\right\} \quad (11)$$

When Δ/σ_{r_n} is large, the formula above can be simplified to $P_e \approx 2Q(\Delta/2\sigma_{r_n})$, actually an upper bound to P_e. In order to rewrite P_e in terms of the desired parameters, let us make $\Delta = \tau L\alpha^2$ for some real τ. Then we can write

$$P_e \approx 2Q\left(\frac{\tau\sqrt{L}\alpha}{2\sqrt{D_c}}\right) \quad (12)$$

We want to find an expression for α in terms of D_w. Substituting (8) into (5) and using the expressions of σ_{r_x} and Δ in terms of α and L, it is possible to write

$$\frac{\sqrt{D_w}}{\sigma_x} = \frac{\tau\alpha}{\sigma_x}\sqrt{1/4 + I\left(\frac{\sigma_x}{\tau\sqrt{L}\alpha}\right)} = \frac{1}{\sqrt{L}}F\left(\frac{\tau\sqrt{L}\alpha}{\sigma_x}\right) \quad (13)$$

[1] $Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{\tau^2}{2}}\,d\tau$

with $F(x) \triangleq x\sqrt{1/4 + I(1/x)}$, which can be shown to be one to one and monotonically increasing for $x > 0$. Then, inversion of (13) yields

$$\frac{\tau\sqrt{L}\alpha}{\sigma_x} = F^{-1}\left(\frac{\sqrt{LD_w}}{\sigma_x}\right) \quad (14)$$

Again, after substituting (14) into the first equalty of (13) we obtain

$$\alpha = \frac{\sqrt{D_w}}{\tau} \cdot \left[1/4 + I\left(\frac{1}{F^{-1}\left(\frac{\sqrt{LD_w}}{\sigma_x}\right)}\right)\right]^{-1/2} \quad (15)$$

Noting that Δ increases linearly with τ and L while σ_{r_x} only increases as $\sqrt{L}$, it is possible to considerably simplify equation (15) for large L or large τ, since, for large x $I\left(1/F^{-1}(x)\right) \approx 1/4x^2$ and then

$$\alpha \approx \frac{2\sqrt{D_w}}{\tau} \cdot \left(1 + \frac{1}{\frac{LD_w}{\sigma_x^2}}\right)^{-1/2} \quad (16)$$

so finally we have

$$P_e \approx 2\,Q\left(\frac{\xi}{\sqrt{1+\frac{\lambda^2}{L}}}\sqrt{L}\right) \quad (17)$$

See that P_e does not depend on the size τ of the quantization step for the projection. Also note that, for L large, the denominator in the argument of $Q(\cdot)$ in (17) tends to one and then the asymptotic performance does not depend on the DWR.

3. DISTORTION COMPENSATED QP

The QP scheme can be improved by applying the idea of leaving a residual error when quantizing the scalar variable r, giving raise to Distortion Compensated QP (DC-QP). The projection function and quantization centroids are the same as in the previous section. Given the projected host image r_x, the projected watermark is selected now as $r_w = \nu(Q_b(r_x) - r_x)$. If the $w[k]$ are also selected in the same way the main difference is only that the quantization error in the projected domain is scaled by ν. This in turn implies that the development in the previous section for determining $\mathrm{Var}\{r_w\}$ can be easily adapted to the present case yielding $\mathrm{Var}_{\mathrm{DC-QP}}\{r_w\} = \nu^2\,\mathrm{Var}_{\mathrm{QP}}\{r_w\}$.

We will assume without loss of generality that the symbol $b = 1$ is sent. Then, the probability of bit error can be written as

$$P_e = \sum_i \int_{2i\Delta-\Delta/2}^{2(i+1)\Delta-\Delta/2} f_{r_x}(r_x) P_e(r_x)\,dr_x \quad (18)$$

where $P_e(r_x)$ is the probability of error for a given value of r_x. In order to determine $P_e(r_x)$, note that if

$$r_x \in [2i_*\Delta - \Delta/2,\, 2(i_*+1)\Delta - \Delta/2) \quad (19)$$

for some integer i_*, then the undistorted projected watermarked image $r = r_x + r_w$ becomes

$$r = \nu Q_1(r_x) + (1-\nu)r_x = \nu(2i_*\Delta + \Delta/2) + (1-\nu)r_x \quad (20)$$

When $\Delta/\sigma_{r_x}^2$ is large, it is possible to simplify the analysis by considering that whenever there exists a decoding error it is due to $r_x + r_w + r_n$ lying in the Voronoi cells associated to one of the neighboring centroids to $r_x + r_w$ in Λ_{-1}, namely, $2(i_*+1)\Delta - \Delta/2$ and $2i_*\Delta - \Delta/2$. From here, we can conclude that

$$P_e(r_x) \approx Q\left(\frac{(2i_*+1)\Delta - \nu(2i_*\Delta + \Delta/2) - (1-\nu)r_x}{\sigma_{r_n}}\right) + Q\left(\frac{\nu(2i_*\Delta + \Delta/2) + (1-\nu)r_x - 2i_*\Delta}{\sigma_{r_n}}\right) \quad (21)$$

where i_* is such that (19) holds. Substituting (21) into (18) it is possible to arrive at the following result

$$P_e \approx \sum_i \frac{\Delta}{\sqrt{2\pi}\sigma_{r_x}} \int_{-1/2}^{3/2} e^{-\Delta^2(r_x+2i)^2/2\sigma_{r_x}^2}(R_+ + R_-)\,dr_x \quad (22)$$

with

$$R_\pm = Q\left(\frac{(1 \pm (1-\nu)(1-2r_x))}{\nu\sqrt{1+\frac{\lambda^2}{L}}}\,\xi\sqrt{L}\right) \quad (23)$$

In this result it has been assumed that L is large as to permit the same approximations as in the previous section. The integral in (22) must be evaluated numerically, but it is interesting to see that it depends only on the WNR, the DWR, L and the residual error parameter ν respectively. Bear in mind that the ratio $\Delta^2/\sigma_{r_x}^2$ in (22) may be approximated by $4(L/\lambda^2 + 1)$.

In Fig. 1 we plot P_e as a function of ν for $\xi = 1$, $L = 20$ and for DWR $= 3.5$ and 6.8 dB. Notice that the optimal values of ν, 0.9460 and 0.9420 respectively, are smaller than one; hence, note that the DC-QP method offers a potential improvement over the QP method by choosing an appropriate value of ν.

Since DC-QP resembles Costa's method [4], one might expect that the optimal parameter ν, there derived for maximizing capacity, should become similar to the one here obtained. In Costa's paper, the optimal value is $\nu = 1/(1 + \mathrm{NSR})$ with NSR the noise to signal ratio. In QP the noise to signal ratio after projection becomes $\sigma_{r_n}^2/\mathrm{Var}\{r_w\}$ which results in

$$\frac{\sigma_{r_n}^2}{\mathrm{Var}\{r_w\}} = \frac{L\alpha^2 D_c}{\Delta^2(1/4 + I(\sigma_{r_x}/\Delta))} = \frac{D_c}{LD_w} \quad (24)$$

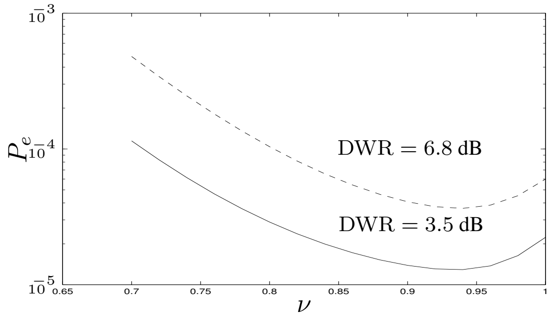

Fig. 1. Bit error probability versus ν for different values of the DWR, $D_c = D_w$ and $L = 20$.

Consequently, the optimal value of ν approaches 1 asymptotically for $L \to \infty$. Numerical optimization experiments show that this is the case when P_e is minimized in terms of ν and then, for large L the QP and DC-QP methods turn out to be equivalent.

Also, for moderate L, experimentation shows (cf. Fig. 1) that the optimal ν depends on WNR but also on the DWR in contrast with Costa's result in which the host image (there the state) does not show up in the optimal value of ν. The reason for this difference must be found in the fact that Costa's method has unlimited complexity, which is obviously not our case (unless L is made unpractically large). Interestingly, knowledge of the image second-order statistics becomes useful when trying to optimize performance. Moreover, for larger DWR's there is more room for improvement by choosing the proper value of ν.

4. EXPERIMENTAL RESULTS

In Fig. 2 the results for QP and DC-QP are presented. We can see that a impressively low probability of error is attained even when the distortion level is as high as the embedding distortion ($\mathrm{WNR} = 0$ dB). Also note that since the example has a moderate value of DWR, the DC-QP does not offer a salient advantage with respect to QP, even for the optimized value of ν, that in this case is 0.94. Experimental tests show that approximation (16) holds for values of $\lambda^2/L \lesssim 0.25$.

Finally, in Fig. 3 the theoretical performance values of QP and DC-QP for different values of L are shown. The ν parameters optimize DC-QP in each case. The predicted values are so low that its empirical simulation becomes difficult; the soundness of the results is supported by the empirical validation of Fig. 2 and by the fact that the theoretical approach becomes tighter when L grows. As a final observation note that, as depicted in Fig. 1, quantized projection methods are also DWR-dependent, a property exhibited by known-host-statistics methods but not by known-host-state methods.

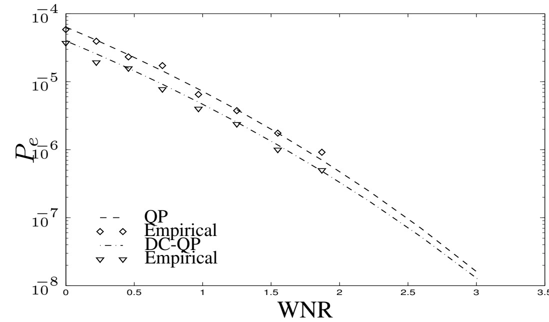

Fig. 2. Bit error probability versus WNR for Quantized Projection ($L = 20$), DWR $= 7.0$ dB

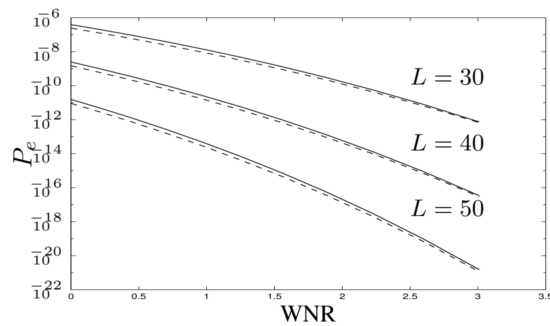

Fig. 3. Bit error probability versus WNR compared for QP (solid line) and DC-QP (dashed line), for increasing values of L (DWR $= 7.0$ dB)

5. REFERENCES

[1] Brian Chen and Gregory W. Wornell, "Provably robust digital watermarking," in *Proc. of SPIE*, San José, USA, 1999, vol. 3845 of *Multimedia Systems and Applications II*, pp. 43–54.

[2] Brian Chen and Gregory W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[3] Juan R. Hernández and Fernando Pérez-González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1142–1166, July 1999, Special Issue on Identification and Protection of Multimedia Information.

[4] Max H.M. Costa, "Writing on dirty paper," *IEEE Trans.*

on Information Theory, vol. 29, no. 3, pp. 439–441, May 1983.