

IMPROVING DATA HIDING PERFORMANCE BY USING QUANTIZATION IN A PROJECTED DOMAIN

Fernando Pérez-González and Félix Balado

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

ABSTRACT

The quantization of a linear projective transformation first proposed by Chen and Wornell is shown to allow for much better performance figures than those yielded by previous approaches. The procedure to achieve this improvement is explained through the proposal and analysis of an improved data hiding method called Quantized Projection (QP), based in the quantization of a statistic similar to those used at detection in spread-spectrum algorithms. Both the theoretical analysis and the empirical validation show that projection-based methods exhibit huge performance improvements over existing ones under the same conditions —i.e. same degree of diversity and level of random additive attacking distortion.

1. INTRODUCTION

Data hiding methods that have received major attention up to now can be roughly subdivided into two families: 1) Known-host-state methods, such as Quantization Index Modulation (QIM) [1], that use host signal state or side information at the encoder and that perform deterministic decoding without resorting to host signal statistical characterization (e.g. minimum Euclidean distance decoding); 2) Known-host-statistics methods, such as spread-spectrum ones [2], that do not use host signal state information, but perform statistical detection using the the host signal characterization (e.g. maximum likelihood decoding). These two families of methods offer a number of advantages and disadvantages that were analyzed in [3]; for instance, known-host-state methods can yield very low or null probabilities of decoding error (P_e) for small distortions, while known-host-statistics methods are less sensitive to noise and can take advantage of the host signal knowledge to improve performance.

The target of the Quantized Projection (QP) method, which we propose in this paper, is to gather the best prop-

This work has been partially supported by the *Xunta de Galicia* under project PGIDT01 PX13204PM, the European project Certimark (Certification of Watermarking Technologies), IST-1999-10987, and the CYCIT project AMULET, reference TIC2001-3697-C03-01.

erties of these two philosophies to get an overall improvement. To this end, QP uses a projection function producing a scalar decision variable —similar to the decision statistic in known-host-statistics methods— that is then uniformly quantized as it occurs in most known-host-state implementations. We note that the idea of combining these two approaches was first proposed by Chen and Wornell in their so-called Spread Transform Dither Modulation (STDM) [4] which the QP method closely resembles; however, we will show some important differences in the way the quantization step is chosen which strongly affect the theoretical analysis and indirectly have an impact on performance.

2. QUANTIZED PROJECTION (QP) DATA HIDING

The scenario in which will carry out the analysis is the following: one information symbol $b = \pm 1$ will be hidden into a zero-mean host signal \mathbf{x} using L pseudorandomly chosen samples from \mathbf{x} indexed by the set \mathcal{S} . Without loss of generality we will write the watermarked signal as the addition $\mathbf{y} = \mathbf{x} + \mathbf{w}$. We will denote by D_w the *embedding distortion*, that is, the power of the embedded watermark. By convenience, the parameter $\lambda \triangleq \sqrt{\sigma_x^2/D_w}$, with σ_x^2 the host signal variance, allows us to define the *document-to-watermark ratio*, $DWR = 20 \log_{10} \lambda$. Before decoding, \mathbf{y} undergoes a channel represented by additive noise \mathbf{n} independent of \mathbf{x} and following an arbitrary zero-mean pdf, yielding a received signal $\mathbf{z} = \mathbf{y} + \mathbf{n}$. By virtue of the pseudorandom choice of the indices in \mathcal{S} we may assume that the samples in \mathbf{n} are also mutually independent. 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 $WNR = 20 \log_{10} \xi$.

At each sample k , the watermark has the form

$$w[k] = \rho \alpha[k] s[k] \quad (1)$$

where $s[k]$ is a key-dependent pseudorandom antipodal sequence, i.e. $s[k] \in \{\pm 1\}$, so that the variance of the watermark becomes proportional to $\alpha^2[k]$. Therefore, $\alpha[k]$ can be regarded as a *perceptual mask* that should take into account

the properties of the Human Visual System, as is customary in modern data-hiding schemes.

The aforementioned projection function consists in computing a weighted cross-correlation between the watermarked image and the watermark, so for a given signal \mathbf{y} the projection r_y is such that

$$r_y = \sum_{k \in \mathcal{S}} y[k]s[k]/\alpha[k] \quad (2)$$

Although not pursued here, it can be shown that this type of projection is optimal among the class of linear projections under perceptually-shaped additive channel noise. This type of channel distortion amounts to making the variance of each noise sample proportional to $\alpha^2[k]$ as well. It is also possible to show that the latter strategy is optimal from the attacker's point of view under the restriction of imperceptible additive distortions.

The choice of the projection function in (2) makes it computationally appealing due to its linearity. It is also interesting to remark that this projection would be the optimal ML decoding function were the host image statistics Gaussian [2]. Improvements could be possible by using nonlinear projections adapted to the statistics of \mathbf{x} , but they will not be pursued here.

It is useful to rewrite (2) as the addition of two terms

$$r_y = r_x + r_w = \sum_{k \in \mathcal{S}} x[k]s[k]/\alpha[k] + \sum_{k \in \mathcal{S}} w[k]s[k]/\alpha[k] \quad (3)$$

so that r_x is the projected host image and r_w is the projected watermark.

The embedding of b is made by quantizing r_x with one of two uniform scalar quantizers whose centroids are given by the unidimensional lattices $\Lambda_{-1} = 2\Delta\mathbb{Z} - \Delta/2$ and $\Lambda_1 = 2\Delta\mathbb{Z} + \Delta/2$. 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 .

The detector will decide the symbol \hat{b} from the received signal \mathbf{z} after using a minimum Euclidean distance decoder on the projection r_z to determine which of the two lattices is closest:

$$\hat{b} = \arg \min_{-1,1} \|r_z - Q_b(r_z)\|^2 \quad (4)$$

2.1. Election of the quantization step

Before showing how to compute the watermark \mathbf{w} from r_w for getting the desired projection, notice that a crucial observation not explicitly made in the original proposal of STDM is that, as L grows, for a given level of watermarking distortions the scalar quantization step Δ can be made each time larger; this can add a high degree of robustness to the method.

It has to be remarked that STDM was derived under a Mean Square Error (MSE) embedding distortion constraint

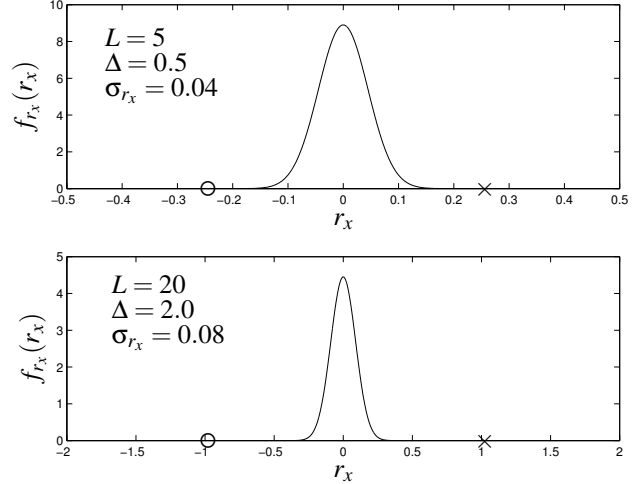


Figure 1: Enlargement of the quantization step Δ with respect to σ_{r_x} when L grows (DWR = 6 dB, $\circ \equiv \Lambda_1$, $\times \equiv \Lambda_{-1}$)

which allows to redistribute distortion across host-signal samples. Depending on the unitary transform that is used, this could be unsuitable for data hiding purposes, as perceptually unacceptable local distortions could be globally compensated. On the other hand, QP has been derived by imposing a stronger structure on the watermark (cf. Eq. 1).

In order to gain some insight on the step growing phenomenon exploited by QP, let us assume for a moment that r_x can be modeled as a zero-mean Gaussian random variable, and that $\sigma_x[k] = \sigma_x$ and $\alpha[k] = \alpha$ for all $k \in \mathcal{S}$. In these conditions, and using the projection (2) with L samples on \mathbf{x} to compute r_x , the maximum allowed energy for r_w grows as L^2 (as the peak amplitude attainable is L) while the variance of the normal variable r_x only grows as $L\sigma_x^2/\alpha^2$. Therefore, if the quantization step Δ is fixed, when L is increased we can use less watermark energy per sample for the encoder to move the projection r_x to the desired quantization centroid.

Alternatively, we can use a constant watermark energy value per pixel —ideally the maximum allowed— and increase Δ to get the same effect, thus becoming more and more difficult for the attacker to trick the decoder (i.e. change the projection to a wrong quantization centroid) when the projection function is unknown. In Fig. 1 we depict this effect assuming for illustrative purposes that we take $|s[k]| = 1$ for all $k \in \mathcal{S}$, and that we are always able to choose Δ as large as L without exceeding the allowed embedding distortion. To guarantee this in an average sense becomes a major difficulty for an accurate analysis; actually, the restriction would be always satisfied —not only in average— if a very large value of L were used, since we could consider that r_x is with very high probability around zero compared to the quantization step size. Interestingly, in this case the quantization procedure resembles a simple spread-spectrum scheme with antipodal symbols.

2.2. Computing the watermark

We formalize next the previous intuitive explanation. First, note that the use of larger Voronoi (i.e. quantization) cells makes the hypothesis of host data uniform inside them not necessarily valid, thus demanding for a new theoretical analysis.

Recalling the structure imposed on the watermark in (1) and substituting it into (2) we have that $\rho = r_w/L$. Noticing that r_w and $s[k]$ are statistically independent and that $E\{s^2\} = 1$, it is immediate to write

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

In order to simplify the performance analysis while producing results illustrating the benefits of our scheme, let us assume a constant perceptual mask, i.e., $\alpha[k] = \alpha$, for all $k \in \mathcal{S}$. With this assumption (1) becomes $w[k] = r_w s[k] \alpha / L$, $k \in \mathcal{S}$.

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

$$\sigma_{r_x}^2 = \frac{L\sigma_x^2}{\alpha^2} \quad (6)$$

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

$$\text{Var}\{r_w\} = \frac{E\{r_w^2|b=1\} + E\{r_w^2|b=-1\}}{2} \quad (7)$$

where

$$E\{r_w^2|b=1\} = \sum_{i=-\infty}^{\infty} \int_{\Delta(2i-1/2)}^{\Delta(2i+3/2)} f_{r_x}(r_x) (2i\Delta + \Delta/2 - r_x)^2 dr_x \quad (8)$$

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 (7) and operating, we have

$$\text{Var}\{r_w\} = \Delta^2 \left(\frac{1}{4} + I(\sigma_{r_x}/\Delta) \right) \quad (9)$$

where

$$I(\sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=-\infty}^{\infty} \int_{-1/2}^{1/2} e^{-(r_x+i)^2/2\sigma^2} r_x^2 dr_x \quad (10)$$

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.

2.3. Performance analysis under additive random noise

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 z becomes

$$r_z = \sum_{k \in \mathcal{S}} \frac{z[k]s[k]}{\alpha[k]} = r_x + r_w + r_n \quad (11)$$

where r_x and r_w were defined in (3) and the projected noise is $r_n = \sum_{k \in \mathcal{S}} n[k]s[k]/\alpha[k]$. An interesting side effect of QP is that we can invoke again the CLT to state that, for a wide class of distributions for $n[k]$, the pdf of the projection r_n of the attacking distortion can be approximated by a zero-mean Gaussian pdf with variance $\sigma_{r_n}^2 = L\sigma_n^2/\alpha^2 = LD_c/\alpha^2$, assuming invariance in $\alpha[k]$.

The bit error probability P_e can be determined by considering the decoder (4) and taking into account the symmetry in the problem. Assuming without loss of generality a transmitted $b = 1$ we can write¹

$$P_e = 2 \sum_{k=0}^{\infty} \left\{ Q\left(\frac{(4k+1)\Delta}{2\sigma_{r_n}}\right) - Q\left(\frac{(4k+3)\Delta}{2\sigma_{r_n}}\right) \right\} \quad (12)$$

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

$$\Delta = \frac{\sqrt{D_w}L}{\tau\alpha} \quad (13)$$

with τ such that

$$\tau = \sqrt{1/4 + I(\sigma_{r_x}/\Delta)} \quad (14)$$

Then, by making use of (9) and (5) we can write

$$P_e \approx 2Q\left(\frac{\xi\sqrt{L}}{2\tau}\right) \quad (15)$$

We want to find an expression for τ in terms of the desired parameters, so that Eq. (15) is more easily interpreted. Now, using the previous expressions for σ_{r_x} and Δ , it is possible to write

$$\frac{\sqrt{D_w}}{\sigma_x} = \frac{\Delta}{\sigma_{r_x}} \sqrt{1/4 + I\left(\frac{\sigma_{r_x}}{\Delta}\right)} \quad (16)$$

$$= \frac{1}{\sqrt{L}} F\left(\frac{\Delta}{\sigma_{r_x}}\right) \quad (17)$$

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 (17) yields

$$\frac{\Delta}{\sigma_{r_x}} = F^{-1}\left(\frac{\sqrt{L}}{\lambda}\right) \quad (18)$$

¹ $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt$

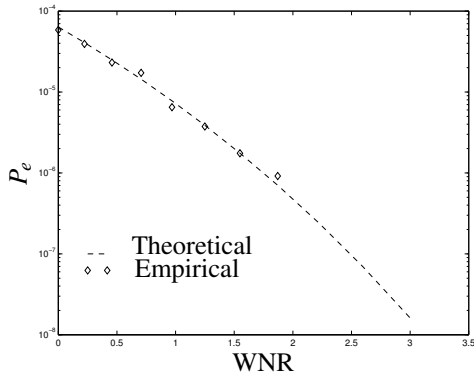


Figure 2: Bit error probability versus WNR for Quantized Projection ($L = 20$), DWR = 7.0 dB

This equation can be readily substituted into (14) to yield an expression for τ that in turn can be plugged into (15) to produce the desired P_e .

Noting that the ratio σ_{r_x}/Δ decreases as $1/\sqrt{L}$, it is possible to considerably simplify the expression for P_e since $I(1/F^{-1}(x)) \approx 1/4x^2$ for large x :

$$P_e \approx 2Q\left(\frac{\xi}{\sqrt{1 + \frac{\lambda^2}{L}}}\sqrt{L}\right) \quad (19)$$

It is interesting to see that for L large, the denominator in the argument of $Q(\cdot)$ in (19) tends to one and then the asymptotic performance does not depend on the DWR.

3. EXPERIMENTAL RESULTS AND CONCLUSIONS

In Fig. 2 the probabilities of decoding error for QP are presented for a decreasing distortion level (increasing WNR). We can see that an impressively low probability of error is attained even with a distortion level as high as the embedding distortion (WNR = 0 dB). Experimental tests show that approximation (19) holds for values of $\lambda^2/L \lesssim 0.25$, as it would be expected from the way it is done. These results outdo by a factor of 10^2 in the P_e those yielded by both spread-spectrum with diversity of L samples and L -dimensional quantization, as it has been analyzed in [3].

Finally, in Fig. 3 the theoretical performance values of QP for different values of L are shown. 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. 3, the QP method is also DWR-dependent for moderate values of L , as it improves its performance for lower DWR values. This is a desirable property exhibited by known-host-statistics methods but not

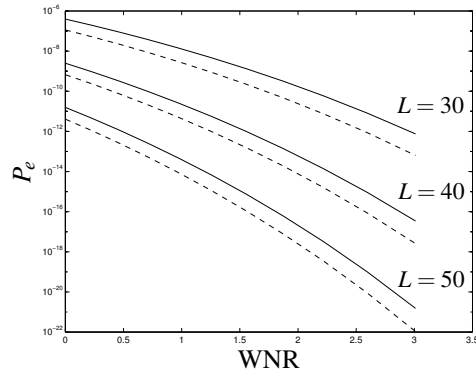


Figure 3: Bit error probability versus WNR compared for QP for increasing values of L (solid line DWR = 7.0 dB, dashed line DWR = 3.0 dB)

by known-host-state methods, this confirming that QP is in fact a hybrid between statistical and side-information methods.

4. 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] 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.
- [3] Fernando Pérez-González, Félix Balado, and Juan R. Hernández, "Performance analysis of existing and new methods for data hiding with known host information in additive channels," *IEEE Trans. on Signal Processing*, 2001, Submitted.
- [4] 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.