Quantization-based watermarking performance improvement using host statistics: AWGN attack case *

Oleksiy Koval
CVML,CUI - University of Geneva
24, rue General Dufour
1211, Geneva, Switzerland
koval@cui.unige.ch

Sviatoslav Voloshynovskiy
CVML,CUI - University of Geneva
24, rue General Dufour
1211, Geneva, Switzerland
svolos@cui.unige.ch

Fernando Pérez-González
Signal Theory and Communications Department
University of Vigo
E-36200 Vigo, Spain
fperez@tsc.uvigo.es

ABSTRACT

In this paper we consider the problem of performance improvement of known-host-state (quantization-based) watermarking methods undergo Additive White Gaussian noise (AWGN) attack. The motivation of our research is twofold. The first reason concerns the common belief that any knowledge about the host image taken into account designing quantization-based watermarking algorithms can not improve their performance. The second reason refers to the poor practical performance of this class of methods at low Watermark-to-Noise Ratio (WNR) regime in comparison to the known-host-statistics techniques when AWGN attack is applied. We demonstrate in this paper that bit error probability of Dither Modulation (DM) and Distortion-Compensated Dither Modulation (DC-DM) against AWGN attack can be significantly reduced when the quantizers are designed using the statistics of the host data. For the case when the statistics of the data correspond to i.i.d. Laplacian distribution and using Uniform Deadzone Quantizer (UDQ) we develop close-form analytical models for the analysis of bit error probability of DM and DC-DM. Results of performed experiments demonstrate that significant performance improvement of classical DM and DC-DM with respect to bit error probability can be achieved with the minor increase of design complexity.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Models And Principles.

General Terms

Data hiding, known-host-state watermarking.

*(Further information: contact S. Voloshynovskiy, e-mail: svolos@cui.unige.ch).

1. INTRODUCTION

Digital watermarking is targeting reliable non-perceptible communications of secure message via some media channel. There are several requirements to watermarking methods that should be satisfied during their design. At the same time some of them, these requirements are used for benchmarking of different data hiding algorithms. The rate of reliable communications under AWGN [3, 7] and analysis of their practical performance based on bit error rate probability when some additive attack was applied [3, 7] are among the most popular and widely used in practice benchmarking criteria.

The results of comparison with respect to the both criteria demonstrate that embedding rates and robustness against AWGN attack in case of the known-host-state embedding are superior in comparison those based on spread-spectrum watermarking especially at high WNR. However, at low WNR performance of the quantization-based watermarking is significantly decreased with respect to both classification criteria.

Leaving the aspects of information-theoretic properties enhancement of known-host-state methods outside of the scope of this paper, we will concentrate our attention at the improvement of the practical performance of these class of methods undergo AWGN attack.

The paper is organized as follows. In Section 2 we present a review of quantization-based watermarking and propose the solution to the problem of their performance robustness improvement against AWGN attack and provide close-form analytical models for the analysis of bit error probability of DM and DC-DM against this attack. Section 3 contains the proposed set-up and the experimental results of our solution validation versus classical DM and DC-DM. Finally, Section 4 concludes the paper.

Notations. Capital letters are used to denote scalar random variables X and regular letters x designate their realizations. We use $X \sim f_X(x)$ or simply $X \sim f(x)$ ($X \sim p_X(x)$ or $X \sim p(x)$) to indicate that a continuous (discrete) random variable X is distributed according to $f_X(x)(p_X(x))$. The variance of X is denoted σ_X^2. The set of integers is denoted by $\mathbb{Z}$.

2. KNOWN-HOST-STATE WATERMARKING

Dither Modulation. The idea of DM was initially presented in [2]. In the scope of this paper we will understand by DM the embedding of the binary value b, $b \in \{-1; 1\}$, using a set of two quantizers. The decision points of the quantizers (Fig. 1, a) belong to the unidimensional lattice [7]:

$$\Lambda_{-1} = 2\Delta\mathbb{Z}, \tag{1}$$

$$\Lambda_1 = 2\Delta\mathbb{Z} + \Delta. \tag{2}$$

The watermarked image is obtained applying one of these quantizers to the host depending on the sign of the bit to be embedded into the host image:

$$y = Q_b(x) = x + w, \tag{3}$$

where y is the stego data, x is the original host data and w is the watermark. $Q_b(x)$ denotes the quantizer corresponding to b-bit embedding. It is evident that the watermark is equivalent to the quantization error:

$$w = Q_b(x) - x. \tag{4}$$

Assuming that high rate quantization conditions are satisfied for embedding [5], the following conjectures are valid:

a) the watermark and the host signal are independent;

b) the watermark is uniformly distributed on the interval $[-\Delta; \Delta]$;

c) embedding distortions are equal to the variance of the quantization noise $D_W = \Delta^2/3$ [7].

The DM decoding is performed using the Minimum Distance Rule based on the attacked stego image z:

$$\hat{b} = argmin_{b\in\{-1,+1\}} \|z - Q_b(z)\|^2, \tag{5}$$

where $z = y + n$, n is the zero-mean AWGN in our case with variance σ_N^2.

Distortion Compensated Dither Modulation. The DC-DM was proposed as a generalization of the DM embedding methods when the watermark (4) is multiplied by some constant $\nu \in [0; 1]$ [2]:

$$w = \nu e = \nu(Q_b(x) - x), \tag{6}$$

$$y = x + w = x + \nu(Q_b(x) - x). \tag{7}$$

It is evident that for ν=1 the DC-DM reduces to the DM.

Scaling by a constant ν changes the dynamic range of the above uniformly distributed watermark to the interval $[-\nu\Delta; \nu\Delta]$ and the embedding distortions are given now by the following expression $D_w = \nu^2\Delta^2/3$ and decoding is performed as in (5).

Deadzone – Based DM and DC – DM. Assumption about the high rate approximation that is used for the design of DM and DC-DM quantizers leads to the uniform quantizer design that was shown to be optimal in source coding [4]. However, practical wavelet image transform coders that are the state-of-the-art in lossy image compression show that it is possible to improve system performance taking into account data statistics in the design of the quantizer even at high rates [6].

Motivated by this observation we propose to take into account host data statistics in design of quantizers for known-host-state watermarking algorithms to provide more favorable conditions for more likely appearing host elements.

Assuming that the data samples can be modeled using i.i.d. Laplacian distribution [7] that is a particular case of the Generalized Gaussian distribution [6] and taking into account the requirement to provide identical robustness of the watermarking method to the positive and negative noise samples we select Uniform Deadzone Quantizer (UDQ) for DC and DC-DM embedding. This type of quantizer satisfies our requirement and outperforms uniform quantizer with respect to rate distortion performance [9].

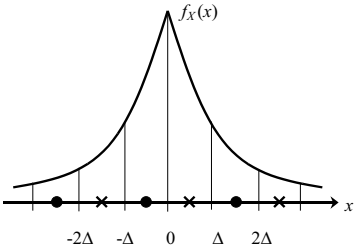
The only difference between the uniform quantizer and the UDQ consists in the wider quantization bin near zero (deadzone) in the latter case. To satisfy the request of the equal embedding conditions of both -1 and $+1$ and taking into account symmetrical character of the host pdf we will use two deadzones symmetrically located near the origin.

Several investigations in lossy source coding were dedicated to the definition of the optimal ratio of the deadzone width to the regular bin width of the UDQ [1, 8]. It was shown that the optimal value is between 1.5 and 2. To our knowledge, no experiments have been performed so far in information hiding to justify the optimal ratio. Therefore, in case of this paper the ratio was selected to be equal 2.

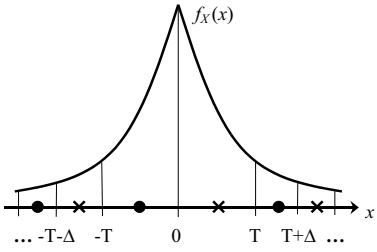
In the following we will refer to the DM and DC-DM designed based on the UDQ as deadzone DM (DDM) and deadzone DC-DM (DDC-DM), respectively.

3. PERFORMANCE ANALYSIS OF DDM AND DDC-DM WATERMARKING

The analysis of the performance of modified known-host-state methods is performed for the case of AWGN attack, assuming i.i.d. Laplacian distribution of the host data, and follows the methodology proposed in [7]. To satisfy fair comparison conditions with [7] we constrained the WNR range, WNR=$10\log_{10}\left(\frac{\sigma_W^2}{\sigma_N^2}\right)$, within the interval [−5, 10] dB. All the obtained results are based on the assumption that high rate quantization conditions are preserved in the deadzones

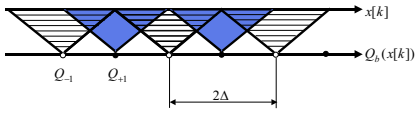


(a)

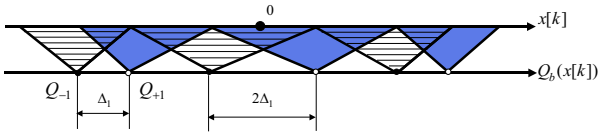


(b)

Figure 1: Quantization of the Laplacian pdf: (a) uniform quantizer case, (b) UDQ-based quantization.



(a)



(b)

Figure 2: DM watermark signaling: (a) classical system, (b) UDQ-based system.

meaning that

(a) watermark and host image are independent;

(b) the watermark pdf is a mixed distribution of the form $f_W(w) = \sum_{i=1}^{2} p_i f_{W_i}(w)$ where $f_{W_i}(w)$ is a uniform distribution over the regular bin (i =1) and over the deadzone (i =2), respectively, and p_i is a Bernoulli mixing density with probability of host sample appearance outside the deadzone equal p_1;

(c) embedding distortions equal to the variance of the quantization noise distributed according to the mixed pdf.

Admit that the stego image y is corrupted by some additive noise n with a pdf $f_N(n)$, $z = y + n = x + w + n$. Let the decision regions associated to $\hat{b}$=-1 and $\hat{b}$=$+1$ are denoted by $\mathcal{R}_{-1}$ and $\mathcal{R}_{+1}$, respectively. In this case bit error probability is determined by the following expression:

$$P_e = P\{\|z - Q_{+1}(z)\|^2 < \|z - Q_{-1}(z)\|^2 \,|b = -1\} = \\ = P\{z \in \mathcal{R}_1 | b = -1\} \quad (8)$$

and

$$P_e = \int_{\mathcal{R}_1} f_Z(z|b=-1) = \int_{\mathcal{R}_1} f_T(t)dt, \quad (9)$$

where $f_T(t)$ is the pdf of equivalent noise. The difference between the classical DM and DC-DM and DDM and DDC-DM consists in dependence of this probability on the bin $Q_{-1}(.)$ where x lies. Therefore, for proper analysis it is necessary to determine the UDQ parameters and compute this probability for all the bins where x can be located.

Determination of the UDQ parameters. According to the selection made in Section 2, the ratio of deadzone width to the regular bin width is equal to 2. Assuming the same embedding distortions as in case of the classical DM (DC-DM) (given by condition (c) in Section 2 and by $p_1 \frac{\Delta_1^2}{4} + (1-p_1)\frac{3\Delta_1^2}{4}$ in case of the DM), one needs to solve the following equation to obtain the regular bin width of the UDQ:

$$\Delta^2 - \Delta_1^2(4 - 3e^{-2\lambda\Delta_1}) = 0, \quad (10)$$

where Δ, Δ_1 are the bin widths of the uniform quantizer and regular bin width of the UDQ, respectively; λ is the parameter of Laplacian distribution. This λ can be determined for the particular Watermark-to-Image Ratio (WIR) regime, WIR=$10\log_{10}\left(\frac{\sigma_W^2}{\sigma_X^2}\right)$ that for the case of this paper is assumed to be $WIR_1 = -6$ dB and $WIR_2 = -16$ dB. Thus, in case of the DDM one has:

$$\lambda = \frac{1}{\Delta}\sqrt{6 \cdot 10^{-0.1WIR}}. \quad (11)$$

Similar expression can be obtained in case of the DDC-DM:

$$\lambda = \frac{1}{\Delta\nu}\sqrt{6 \cdot 10^{-0.1WIR}}. \quad (12)$$

Analysis of the DDM performance against AWGN attack. When the DDM undergo AWGN attack with the variance σ_N^2 we assume that attacking distortions in the MSE sense are equal to this variance $D_n = \sigma_N^2$. The variance range $\sigma_N^2 \in [\sigma_{N_{min}}^2; \sigma_{N_{max}}^2]$ is determined based on the targeting WNR regime, i.e. $\sigma_{N_{min}}^2 = \frac{\Delta^2}{3}10^{-0.1WNR_{max}}$ and $\sigma_{N_{max}}^2 = \frac{\Delta^2}{3}10^{-0.1WNR_{min}}$.

Having these assumptions one can determine the bit error probability as a function of the noise power, developing the integral in (9):

$$\begin{aligned} P_e =& (1 - e^{-2\lambda\Delta_1})\left\{2Q\left(\frac{\Delta_1}{\sigma_N}\right) - Q\left(\frac{2\Delta_1}{\sigma_N}\right)\right\} + \\ &+ 2e^{-2\lambda\Delta_1}\sum_{i=1}^{\infty}\left\{Q\left(\frac{(4i-3)\Delta_1}{2\sigma_N}\right) + Q\left(\frac{(4i-1)\Delta_1}{2\sigma_N}\right)\right\} + \\ &+ 2\sum_{i=1}^{\infty}\left\{P_{2i-1}Q\left(\frac{(4i-1)\Delta_1}{2\sigma_N}\right) - (2P_{2i-1} + P_{2i}) \cdot \right. \\ &\left. \cdot\, Q\left(\frac{(4i+1)\Delta_1}{2\sigma_N}\right) + (P_{2i-1} + 2P_{2i})Q\left(\frac{(4i+3)\Delta_1}{2\sigma_N}\right) - \right. \\ &\left. - P_{2i}Q\left(\frac{(4i+5)\Delta_1}{2\sigma_N}\right)\right\}, \end{aligned} \quad (13)$$

where P_i is a probability that the quantizied sample is within i^{th} bin, $Q(x)$ is a Q-function, $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{t^2}{2}}dt$.

Analysis of the DDC-DM watermarking performance against AWGN attack. In case of the DDC-DM errors are induced by the equivalent noise that is obtained as an additive mixture of the uniform self-noise and AWGN of the attack. Thus, the equivalent noise pdf is given by the convolution of the pdfs' of the mixing components. It should be also taken into account that the bin width is different for the two central bins due to our construction strategy. Therefore, one obtains the following convolution results for the deadzones $f_T^d(t)$ and regular bins $f_T^{\overline{d}}(t)$:

$$\begin{aligned} f_T^d(t) &= \frac{1}{4\Delta_1(1-\nu)}\left\{Q\left(\frac{t - 2\Delta_1(1-\nu)}{\sigma_N}\right) - \right. \\ &- \left. Q\left(\frac{t + 2\Delta_1(1-\nu)}{\sigma_N}\right)\right\}; \end{aligned} \quad (14)$$

$$\begin{aligned} f_T^{\overline{d}}(t) &= \frac{1}{2\Delta_1(1-\nu)}\left\{Q\left(\frac{t - \Delta_1(1-\nu)}{\sigma_N}\right) - \right. \\ &- \left. Q\left(\frac{t + \Delta_1(1-\nu)}{\sigma_N}\right)\right\}. \end{aligned} \quad (15)$$

It is also easy to realize that in the case of DDC-DM the minimal and the maximal target noise variances, $\sigma_{N_{min}}^2$ and $\sigma_{N_{max}}^2$, are determined taking into account distortion compensation factor: $\sigma_{N_{min}}^2 = \frac{\nu^2\Delta^2}{3}10^{-0.1WNR_{max}}$ and $\sigma_{N_{max}}^2 = \frac{\nu^2\Delta^2}{3}10^{-0.1WNR_{min}}$.

Therefore, it is possible to show using similar approach as in case of the DDM that the probability of error of the DDC-DM undergo AWGN attack is determined by:

$$\begin{aligned} P_e &= (1 - e^{-2\lambda\Delta_1})\left(\int_{\Delta_1}^{2\Delta_1} f_T^d(t)dt + \int_{\Delta_1}^{\infty} f_T^d(t)dt\right) + \\ &+ 2e^{-2\lambda\Delta_1}\sum_{i=1}^{\infty}\int_{\frac{(4i-3)\Delta_1}{2}}^{\frac{(4i-1)\Delta_1}{2}} f_T^{\overline{d}}(t)dt + \\ &+ 2\sum_{i=1}^{\infty} P_{2i-1}\int_{\frac{(4i-1)\Delta_1}{2}}^{\frac{(4i+1)\Delta_1}{2}} f_T^{\overline{d}}(t)dt + \\ &+ (P_{2i-1} + P_{2i})\int_{\frac{(4i+1)\Delta_1}{2}}^{\frac{(4i+3)\Delta_1}{2}} f_T^{\overline{d}}(t)dt + \\ &+ 2P_{2i}\int_{\frac{(4i+3)\Delta_1}{2}}^{\frac{(4i+5)\Delta_1}{2}} f_T^{\overline{d}}(t)dt. \end{aligned} \quad (16)$$

4. EXPERIMENTAL RESULTS

In this section we present the experimental results of comparison of DDM and DDC-DM method robustness with robustness of the classical DM and DC-DM against AWGN attack. As it was already mentioned in the previous sections, the analysis is performed for two different WIR, $WIR_1 = -6$ dB and $WIR_2 = -16$ dB for the $WNR \in [-5; 10]$ dB. The compensation parameter of the DDC-DM is selected equal to ν=0.5 [7]. Obviously, the selection of the optimal compensation parameter for each regions is an important issue that is a subject of our future research.

The results of benchmarking are presented in Fig. 3. These results demonstrate that performance of the classical DM and DC-DM watermarking can be significantly improved taking into account statistics of the host data for the quantizer design even for non-optimal compensation parameter. For quantification of the performance gain we measured the distance in dB between the equal bit error probabilities. For the selected reference point $P_e = 0.3$ the gains of modified methods over classical ones are:

- DM, $WIR = -16$ dB: ≈1.3 dB,

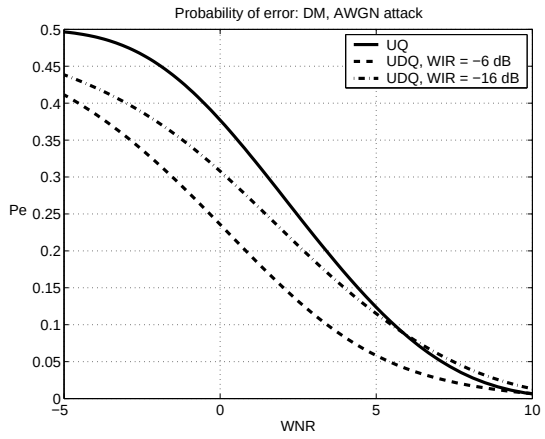
- DM, $WIR = -6$ dB: ≈3 dB;

- DC-DM, $WIR = -16$ dB: ≈1.4 dB,

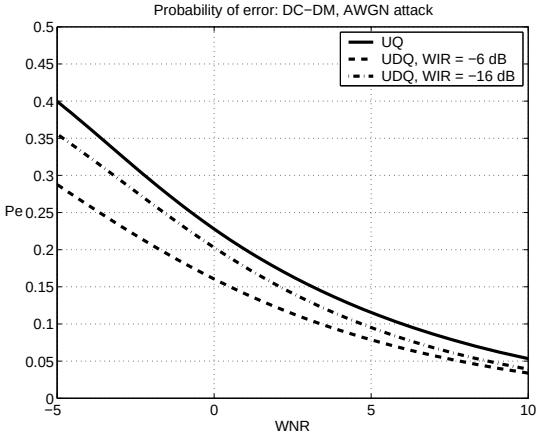
- DC-DM, $WIR = -16$ dB: ≈1 dB.

5. CONCLUSIONS

In this paper we presented a way of performance improvement of known-host-state watermarking methods (DM and



(a)



(b)

Figure 3: Bit error probability of DM and DC-DM schemes for the case of AWGN attack: (a) $WIR = -6$ dB and $WIR = -16$ dB.

DC-DM) designed based on high rate uniform quantization. We showed that, assuming preservation of the conditions of high rate quantization regime, bit error rate probability reduction can be achieved when these methods are developed taking into account host data statistics. We found bit error probability models of DDC and DDC-DM undergo AWGN attack for the host distributed according to the i.i.d. Laplacian distribution and the UDQ with deadzone-to-regular-bin-width ratio equal to 2. Based on these models we demonstrated that, depending on the operational regime for the given range of WNR, the reduction of bit error rate probability of the modified schemes with respect to the classical ones is $1 \div 3$ dB.

As a possible future research direction we are considering the extension of the obtained results to multidimensional case.

6. ACKNOWLEDGMENTS

This paper was partially supported by the SNF Professorship grant No PP002-68653/1, by Interactive Multimodal Information Management (IM2) projects and by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT. The authors are thankful to Jose Vila, Emre Topak, Renato Villan and Yuriy Rytsar for many helpful and interesting discussions.

7. ADDITIONAL AUTHORS

Additional authors: Frederic Deguillaime, CVML, CUI - University of Geneva, email: `Deguillaime@cui.unige.ch`), Thierry Pun, CVML, CUI - University of Geneva, email: `Thierry.Pun@cui.unige.ch`).

8. REFERENCES

[1] I. Balasingham and T. A. Ramstad. On the relevance of the regularity constraint in subband image coding. In *Proc. Asilomar Conference on Signals, Systems, and Computers*, 1997.

[2] B. Chen and G. Wornell. Dither modulation: a new approach to digital watermarking and information embedding. In *IS&T/SPIE's 11th Annual Symposium, Electronic Imaging 1999: Security and Watermarking of Multimedia Content I SPIE: Security and Watermarking of Multimedia Contents*, volume 3657, pages 342–353, San Jose, California, USA, January 1999.

[3] J. Eggers, J. Su, and B. Girod. Performance of a practical blind watermarking scheme. In *Proc. SPIE Security and Watermarking of Multimedia Contents 01*, San Jose, CA, January 2001.

[4] H. Gish and J. P. Pierce. Asymptotically efficient quantizing. *IEEE Trans. on Information Theory*, 14(5):676–683, September 1968.

[5] N. Jayant and P. Noll. *Digital coding of waveforms.* Englewood Cliffs, NJ: Prentice Hall, 1984.

[6] S. LoPresto, K. Ramchandran, and M. Orhard. Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework. In *Data Compression Conference 97*, pages 221–230, Snowbird, Utah, USA, 1997.

[7] F. Pérez-González, F. Balado, and J. R. Hernández. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *IEEE Trans. on Signal Processing, Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery*, 51(4), April 2003.

[8] S. D. Servetto, K. Ramchandran, and M. T. Orchard. Image coding based on morphological representation of wavelet data. *IEEE Trans. on Im. Proc.*, 8(9):1161–1174, 1999.

[9] J. K. Su. *Adaptive Rate-Constrained Transform Video Coding.* PhD thesis, Georgia Institue of Technology, Atlanta, GA, USA, December 1997.