Worst case additive attack against quantization-based data-hiding methods

J. E. Vila-Forcen[a], S. Voloshynovskiy[a], O. Koval[a], F. Pérez-González[b] and T. Pun[a]

[a]University of Geneva, Department of Computer Science. 24 rue Général-Dufour, CH 1211, Geneva, Switzerland;

[b]University of Vigo, Signal Theory and Communications Department. E-36200 Vigo, Spain

ABSTRACT

The main goal of this study consists in the development of the worst case additive attack (WCAA) for quantization-based methods using as design criteria the bit error rate probability and the maximum achievable rate of reliable communications. Our analysis is focused on the practical scheme known as a distortion compensation dither modulation (DC-DM). From the mathematical point of view, the problem of the worst case attack (WCA) design using probability of error as a cost function can be formulated as the maximization of the average probability of error subject to the introduced distortion for a given decoding rule. When mutual information is selected as a cost function, a solution of the minimization problem should provide such an attacking noise probability density function (pdf) that will maximally decrease the rate of reliable communications for an arbitrary decoder structure. The results show that within the class of additive noise attacks the developed attack leads to a stronger performance decrease for the considered class of embedding techniques than the AWGN or the uniform noise attacks.

Keywords: quantization-based, data-hiding, watermarking, additive attacks, distortion compensation, DM, probability of error, mutual information.

1. INTRODUCTION

The knowledge of the WCA structure in digital watermarking is especially important because the attacker has an aggressive behaviour. The knowledge of the WCA allows to create a fair benchmark for data-hiding techniques and makes it possible to provide reliable communications with the use of appropriate error correction codes.

Data-hiding techniques aim at reliably communicating the largest possible amount of information under given distortion constraints. Their resistance against different attacks determine the possible application scenarios. It is a common practice in the data-hiding community to measure the performance in terms of two criteria. The first one is based on the estimation of the information-theoretic limits, i.e., the achievable rate. The second one studies the efficiency of the data-hiding embedding techniques in terms of the bit error rate for a given decoding rule.

Quantization-based methods have attracted attention in the watermarking community as a practical implementation of Costa's set-up[1] that enables interference free communications due to the random binning codebook design. At present, quantization-based techniques benchmarking is performed against an additive white Gaussian noise (AWGN) attack even though it has been demonstrated[2] that AWGN is not the WCAA for any WNR.

At the same time, one known problem of quantization-based methods is the selection of the distortion compensation parameter α' (see Section 2). Although the optimal α' can easily be determined when the power of the noise is available at the encoder prior to the transmission,[3] this is not always feasible for various practical scenarios. We will demonstrate that for a specific decoder (minimum distance decoder) it is possible to calculate the optimal α' independently of the attack pdf.

This paper aims at establishing the information-theoretic limits of quantization-based data-hiding techniques and to develop a benchmark that can be used for the fair comparison of different quantization-based methods.

Further information: (Send correspondence to S.Voloshynovskiy): E-mail: svolos@cui.unige.ch

The paper is organized as follows. Fundamentals of quantization-based methods are presented in Section 2. The investigation of the WCAA strategy for a fixed quantization-based watermarking scenario is performed in Section 3, where the decoder is based on minimum distance rule and the cost function is the probability of error. Nevertheless, the decoder might be able to modifying decoding strategy depending on the attack. Hence, the information-theoretic analysis of Section 4 derives the information bounds where the cost function is given as the mutual information between the input message and the output of the channel.

Notations: We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, corresponding small letters x and $\mathbf{x}$ to denote the realizations of respectively scalar and vector random variables. $m \in \mathcal{M} = \{1, 2\}$ represents the information message and a set of messages. $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x})$ denotes the host signal distributed according to the pdf $f_{\mathbf{X}}(\mathbf{x})$, $\mathbf{Z} \sim f_{\mathbf{Z}}(\mathbf{z})$ represents the attack, $\mathbf{W} \sim f_{\mathbf{W}}(\mathbf{w})$ the watermark and $\mathbf{V} \sim f_{\mathbf{V}}(\mathbf{v})$ the received signal. The step of quantization is equal to Δ and the distortion-compensation factor in practical schemes is denoted as α'. The watermark-to-noise ratio (WNR) is represented as $\mathrm{WNR} = 10\log_{10}\xi = 10\log_{10}\frac{\sigma_W^2}{\sigma_Z^2}$ where σ_W^2 and σ_Z^2 stand for the power of watermark and attack, respectively. $\mathbb{N}$ represents the set of natural numbers.

2. PROBLEM FORMULATION

A data-hiding scenario can be represented in a classical communications framework by including an encoder, a channel and a decoder. In our set-up, both encoder and decoder work as a cooperative pair targeting the maximization of the rate of reliable communications while the attacker aims at the opposite goal.

We restrict our analysis to additive data-hiding, where the stego-image $\mathbf{Y}$ is obtained by addition of the host image $\mathbf{X}$ and the watermark $\mathbf{W}$. Quantization-based data-hiding methods apply quantization techniques to avoid host interference by exploiting the non-causal presence of the host image realization at the encoder.

The generalized data-hiding scenario as communications with side information available at the encoder is presented in Fig. 1, where $m \in \mathcal{M} = \{1, 2\}$ represents the message to be embedded. The attacking channel is given by the corresponding transition pdf $p(\mathbf{v}|\mathbf{y}) = \prod_{i=1}^{N} p(v_i|y_i)$ assuming to be an additive independent identically distributed (i.i.d.) noise (discrete memoryless channel, (DMC)). The estimate of the message $\hat{m}$ is obtained at the decoder given the output of the channel $\mathbf{V}$. In general, embedding and decoding is performed key-dependently. However, key-management lies outside of the scope of this paper and the key will not be taken into account in our analysis.

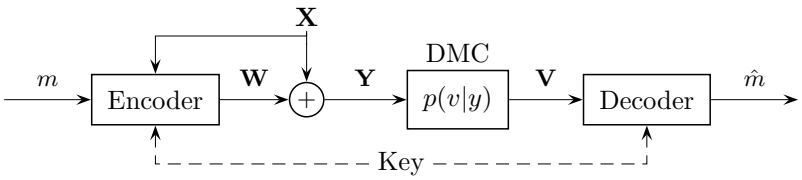


Figure 1. Data-hiding as communications with side information available at the encoder.

Assuming that the channel transition pdf is given by some additive noise pdf, it is possible to write its output as: $\mathbf{V} = \mathbf{X} + \mathbf{W} + \mathbf{Z}$.

Within the class of quantization-based methods, we focus our analysis on DC-DM[3] and dither modulation[4] (DM) as particular cases. Both are approximations of Costa's[1] set-up using a structured codebook aiming at interference free communications.

In the case of DM, the watermarked data is obtained by the application of a message dependent quantizer to the host data, i.e.:

$$y = Q_m(x), m \in \mathcal{M}, \tag{1}$$

as it is presented in Fig. 2.

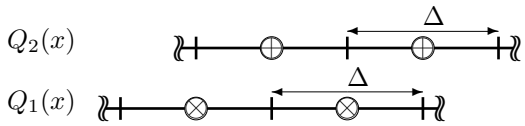


Figure 2. DM embedding quantizers for binary signaling case.

Here, the quantizers are designed using subtractive dithering; each quantizer is a shifted version of the others:

$$Q_m(x) = Q(x + d_m) - d_m, \tag{2}$$

where d_m represents the subtractive dither of the m-th quantizer and $Q(\cdot)$ stands for the fixed quantizer with quantization step Δ.

The variance of the watermark is equal to the variance $D_W = \frac{\Delta^2}{12}$ of a uniform pdf $\mathcal{U}(-\Delta/2, \Delta/2)$. The pdf of the stego data is in this case assumed to be a train of δ functions, as the result of quantization.

For the DC-DM case, the watermarked data is obtained as follows:

$$y = x + \alpha'(Q_m(x) - x), \tag{3}$$

where $0 < \alpha' \leq 1$ is the distortion compensation parameter*. If $\alpha' = 1$, the DC-DM technique simplifies to the DM. The watermark power of the DC-DM embedding is $D_W = \alpha'^2 \frac{\Delta^2}{12}$, and the stego-image pdf is a train of uniform pulses of width $B = 2(1 - \alpha')\Delta$ centered on the quantizer reconstruction levels as a result of the distortion compensation†.

3. PROBABILITY OF ERROR AS A COST FUNCTION

Having fixed the attacking structure and the host pdf, it is possible to implement the optimum decoder for such an attack. However, it is not always possible to know in advance the attacking pdf. The maximum a posteriori optimal decoder under the assumptions of equiprobable messages and asymptotically constant host pdf can consequently be derived as the maximum likelihood (ML) decoder. This decoder is usually implemented assuming the AWGN attack to be the WCAA which corresponds to minimum distance decoding:

$$\hat{m} = \arg \min_{m \in \mathcal{M}} ||v - Q_m(v)||^2. \tag{4}$$

In fact, minimum distance decoding is the optimal ML decoder for all additive attacks whose pdf is symmetric and monotonically non-increasing on each side. The optimum ML decoder for the uniform noise attack is also the minimum distance decoder.[7]

Knowing the channel transition pdf it is possible to implement ML decoding in order to obtain the best possible performance. However, this channel transition pdf is not known in robust watermarking and the actual channel pdf is replaced by the worst case one for the given decoder design.

In the analysis of the WCAA using the probability of error as a cost function, we assume that the minimum distance decoder is used (Fig. 1). Considering the previously explained quantization-based techniques and minimum distance decoder, let $\mathcal{R}_m$ denote the decision region associated to the message m. The probability of correct decoding is determined as:

$$P_c = P\{||v - Q_m(v)||^2 < ||v - Q_{m'}(v)||^2; \forall m' \in \mathcal{M}, m' \neq m\}, \tag{5}$$

that is equal to:

$$P_c = Pr[v \in \mathcal{R}_m | M = m]. \tag{6}$$

*α' does not correspond to Costa α parameter.

†The analysis is performed here in the framework of Eggers actually disregarding the host pdf impact. For more details, we refer readers to.[5, 6]

The probability of error can be obtained as $P_e = 1 - P_c$. In this analysis we assume independence of the probability of error on the quantization bin where the received signal v lies (because the decision regions $\mathcal{R}_m$ are periodical and the host pdf $f_X(x)$ is assumed to be asymptotically constant within each quantization bin).

In integral form we can represent the probability of error as:

$$P_e = \int\limits_{\bigcup\limits_{m' \neq m} \mathcal{R}_{m'}} f_V(v|M=m)dv. \tag{7}$$

Concerning the DM, the pdf of v is the periodical repetition of the noise pdf. In the DC-DM the pdf is given by the convolution of the attacking pdf with the pdf of the self-noise (periodic uniform pdf).[2]

The analysis of the probability of error has been performed for the AWGN and for the uniform noise. The WCAA has finally been derived for both the DM and the DC-DM.

3.1. Additive white Gaussian noise attack

This section contains the probability of error analysis of the DM and the DC-DM under the AWGN attack.

3.1.1. DM analysis

In the DM case, the probability of error (7) is defined as:

$$P_e = \int\limits_{\bigcup\limits_{m' \neq m} \mathcal{R}_{m'}} \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-v^2/2\sigma_Z^2} dv, \tag{8}$$

where σ_Z^2 denotes the power of the AWGN attack.

In the binary case ($\mathcal{M} = \{1, 2\}$), it is possible to rewrite (8) as[2]:

$$P_e = 2\sum_{k=0}^{\infty} \int_{(4k+1)\Delta/4}^{(4k+3)\Delta/4} \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-v^2/2\sigma_Z^2} dv. \tag{9}$$

Fig. 3(a) depicts the probability of error as a function of the WNR for the DM under the AWGN attack in the binary signaling case.

3.1.2. DC-DM analysis

In the DC-DM case the pdf of the received signal is given by the convolution of the attacking noise pdf (AWGN) with the self-noise pdf (uniform noise pdf). Due to the symmetry of construction, we can write:

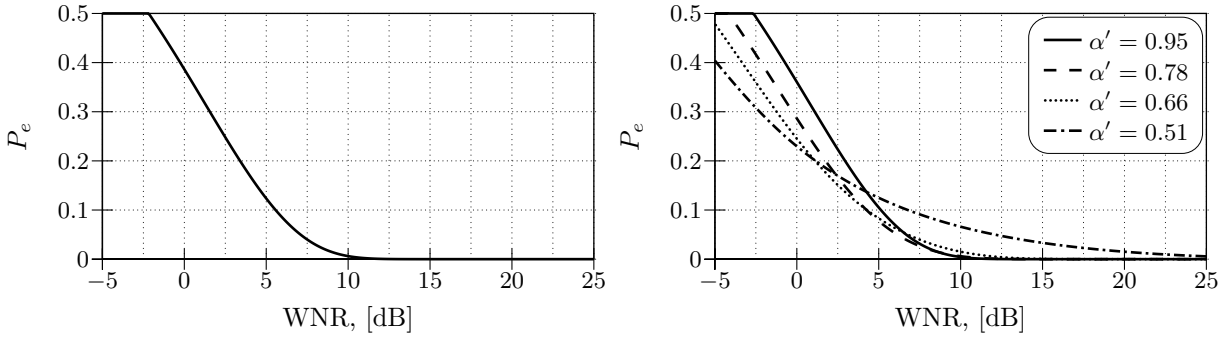
$$f_V(v) = \frac{1}{B}\left(Q\left(\frac{v-B}{\sigma_Z}\right) - Q\left(\frac{v+B}{\sigma_Z}\right)\right), \tag{10}$$

where Q denotes the Q-function, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ and $B = (1-\alpha')\Delta/2$ is the half-width of the self-noise pdf. The analytical expression for the probability of error (7) does not exist, and it is evaluated using numerical integration for the binary signaling case and different distortion compensation parameter values (Fig. 3(b)).

3.2. Uniform noise attack

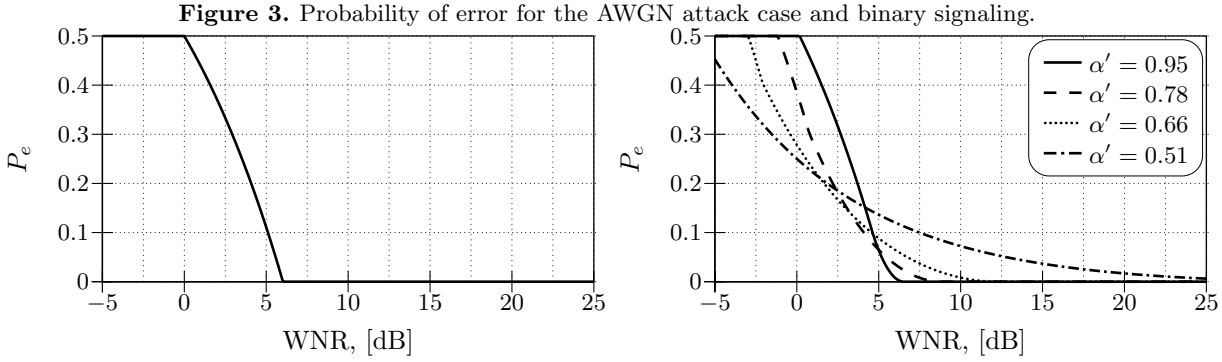
It was shown[2] that the uniform noise attack produces higher probability of error than the AWGN attack for some particular WNR. This fact contradicts the common belief that the AWGN is the WCAA for all data-hiding methods since it has the highest differential entropy for all pdfs with bounded variance.

In our analysis we consider the uniform noise attack $Z \sim \mathcal{U}(-\eta, \eta)$ assuming that the minimum distance decoding rule is used. It can be proven that this decoder is the optimal ML decoder for the uniform noise attack.[7]



(a) DM performance. (b) DC-DM performance.

Figure 3. Probability of error for the AWGN attack case and binary signaling.



(a) DM performance. (b) DC-DM performance.

Figure 4. Probability of error for the uniform noise attack case and binary signaling.

3.2.1. DM analysis

The equivalent noise pdf is given by a train of uniform pulses. In the case when the power of the attack is not strong enough, i.e., all noise samples are within the quantization bin of the sent message, the probability of error is zero. For stronger attacks the probability of error is defined by the integral of the pdf of the received signal over the error region:

$$P_e = \begin{cases} 0, & \eta < \frac{\Delta}{4}, \\ 1 - \frac{\Delta}{4\eta}, & \frac{\Delta}{4} \leq \eta < \frac{1}{2}\frac{\Delta}{2}. \end{cases} \tag{11}$$

The performance of the DM in the uniform noise attack case is presented in Fig. 4(a) for the binary signaling case.

3.2.2. DC-DM analysis

Under the uniform noise attack, the bit error probability is equal to the integral of the equivalent noise pdf (a trapezoidal function) over the error region:

$$P_e = \begin{cases} 0, & B + \eta < \frac{\Delta}{4}, \\ \frac{(4\eta - (2\alpha'-1)\Delta)^2}{32(\alpha'-1)\Delta\eta}, & B + \eta > \frac{\Delta}{4} > |B - \eta|, \\ 1 + \frac{1}{2(\alpha'-1)}, & B - \eta > \frac{\Delta}{4}, \\ 1 - \frac{\Delta}{4\eta}, & \eta - B > \frac{\Delta}{4}, \end{cases} \qquad \text{if } B + \eta < 3\frac{\Delta}{4}. \tag{12}$$

If $B + \eta > \frac{3\Delta}{4}$, the probability of error decreases due to the increase of the attack variance, the equivalent noise pdf enters into the correct decision region $\mathcal{R}_m$. The corresponding performance of the DC-DM under the uniform noise attack is presented in Fig. 4(b) for the binary signaling case.

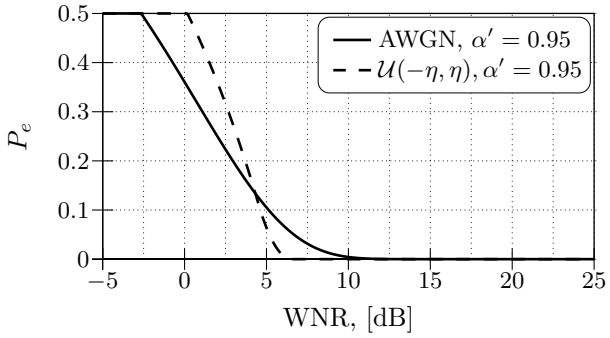


Figure 5. Comparison of the AWGN and the uniform noise attacks using probability of error as a cost function for the minimum distance decoding rule.

The experimental results presented in Fig. 5 allow to conclude that the AWGN attack is not the WCAA strategy for the assumed fixed decoder structure. Thus, the main goal of the following section consists in the development of the WCAA for the considered embedding scenarios.

3.3. The worst case additive attack

The problem of the WCAA design for pulse amplitude communications (PAM) was considered[8] using probability of error as a cost function as well as an energy constraint. In this paper the problem of the WCAA is addressed for quantization-based data-hiding methods.

This problem can be formulated as follows:

$$\max_{f_Z(\cdot)} P_e = \max_{f_Z(\cdot)} \int\limits_{\bigcup\limits_{m' \neq m} \mathcal{R}_{m'}} f_V(v|M=m)dv, \tag{13}$$

subject to the constraints:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1, \tag{14}$$

$$\int_{-\infty}^{\infty} z^2 f_Z(z)dz \leq \sigma_Z^2, \tag{15}$$

where $f_V(\cdot)$ is the pdf of the received signal and σ_Z^2 in (15) constrains the attack power. The constraint (14) follows from the pdf definition.

The result of this optimization problem is a so-called $3-\delta$ attack whose pdf is presented in Fig. 6, where the energy of the attack is concentrated as close as possible to the decision region boundaries in order to introduce the largest error with the minimum energy. The variance of the proposed attack is $\sigma_Z^2 = 2T^2A$.

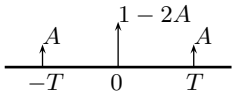
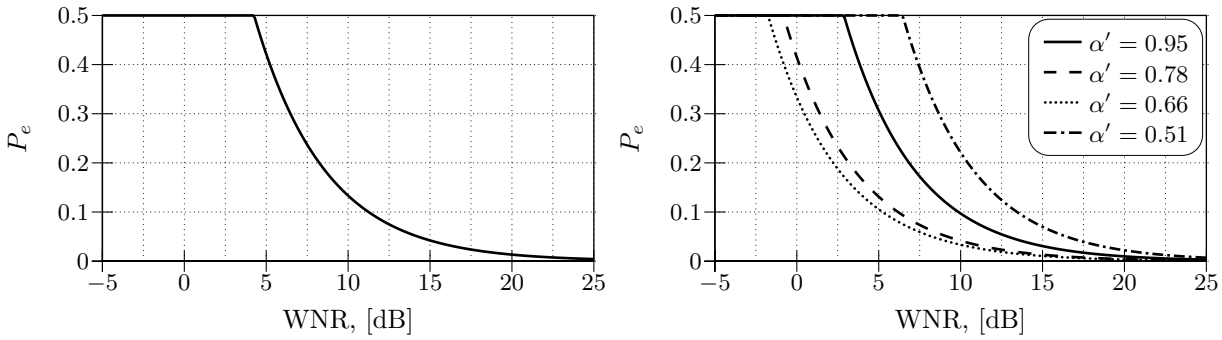


Figure 6. $3-\delta$ attack, $0 \leq A \leq 0.5$.

The optimization of the $3-\delta$ attack has been performed for the DC-DM considering the DM as a particular case for $\alpha' = 1$. When $T - B < Th$ the probability of error is equal to:

$$P_e = \frac{A}{B}(T + B - Th), \tag{16}$$



(a) DM performance.

(b) DC-DM performance.

Figure 7. Probability of error for the $3-\delta$ attack case and binary signaling.

where Th is the threshold of the minimum distance decoder that in the binary case is equal to $Th = \frac{\Delta}{4}$. The probability of error will be maximized for the following selection of T:

$$T = \frac{\Delta}{2}(2\alpha' - 1), \tag{17}$$

and the probability of error for this value of T is given by:

$$P_e = \frac{\sigma_Z^2 \alpha'^2}{12\sigma_W^2(\alpha'(3-2\alpha')-1)}. \tag{18}$$

This result is valid if $T - B < Th$, and this constraint implies $\alpha' < \frac{5}{6}$. For this case, the minimum of the probability of error is achieved at:

$$\alpha' = \frac{2}{3}. \tag{19}$$

In the case when the previous constraint does not hold, the probability of error is defined in the following way:

$$P_e = 2A. \tag{20}$$

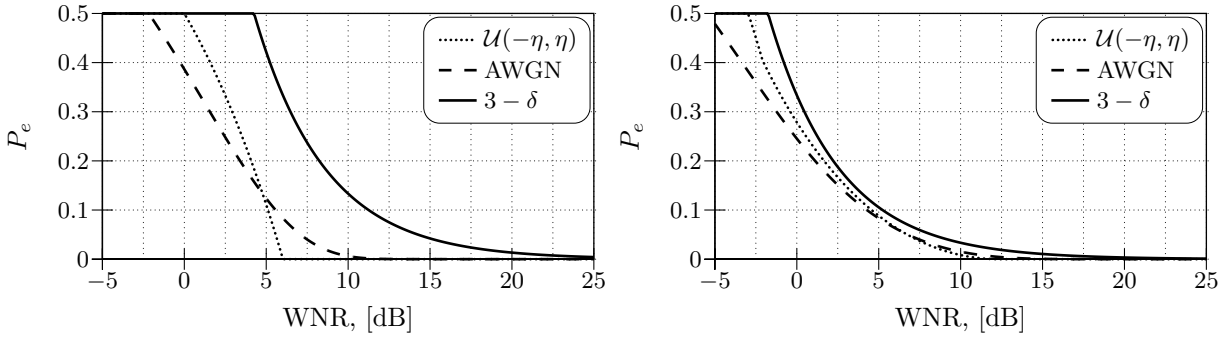
The maximum is found for the minimum possible T, $T = Th + B$, and the probability of error is equal to:

$$Pe = \frac{4\sigma_Z^2 \alpha'^2}{3\sigma_W^2(3-2\alpha')^2}. \tag{21}$$

The corresponding performance for the DM and the DC-DM is presented in Fig. 7(a) and Fig. 7(b), respectively.

Fig. 8 compares the previously analyzed AWGN and the uniform noise attacks with the WCAA, demonstrating that the gap between the AWGN attack and the real worst case attack can be larger than 5 dB in terms of the WNR.

In order to find the optimal compensation parameter value that will allow the data-hider to minimize the probability of error introduced by the WCAA, we analyzed (18) and (21). Surprisingly, it was found that, independently of the operational WNR, $\alpha' = 2/3$ guarantees the lowest probability of error of the analyzed data-hiding techniques under the WCAA (Fig. 9). Having this bound on the probability of error, using error correction codes can guarantee reliable communications.



(a) DM performance. (b) DC-DM performance for $\alpha' = 0.66$.

Figure 8. Analysis of the probability of error for different attacking strategies.

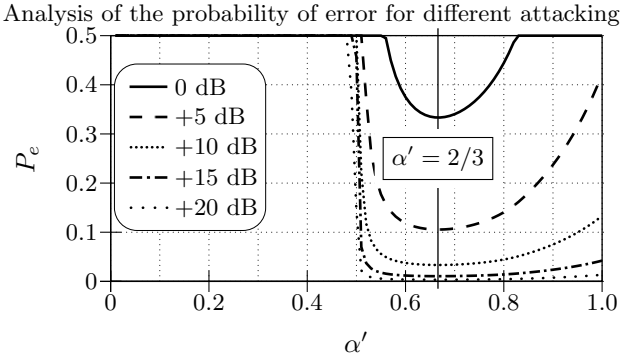


Figure 9. Probability of error as a function of the distortion compensation parameter α'.

4. MUTUAL INFORMATION AS A COST FUNCTION

The analysis of the WCA with mutual information as a cost function is crucial for the fair benchmarking of quantization-based data-hiding techniques. It provides the information-theoretic performance limit (in terms of achievable rate of reliable communications) that can be used for benchmarking of different practical robust data-hiding techniques.

The mutual information is measured between the input message m and the output of the channel $\mathbf{V}$. Thus, the decoder in Fig. 1 is not fixed as in the previous analysis based on the probability of error as a cost function.

It is known[9] that the mutual information can be expressed as a Kullback-Leibler distance (KLD):

$$I(M;V) = D(f_{M,V}(m,v)||f_M(m)f_V(v)), \tag{22}$$

where $f_{M,V}(m,v)$ is the joint pdf distribution of the input message and the output of the channel, $f_M(m)$ denotes the marginal pdf of the input messages and $f_V(v)$ the marginal pdf of the channel output. Thus:

$$\begin{aligned} I(M;V) &= D(\tilde{f}_{M,V}(m,v)||\tilde{f}_V(v)p_M(m)) \\ &= \int \sum_{m=1}^{2} \tilde{f}_{V|M}(v|M=m)p_M(m)\log_2 \frac{\tilde{f}_{V|M}(v|M=m)}{\tilde{f}_V(v)}dv, \end{aligned} \tag{23}$$

where $\tilde{f}(\cdot)$ represents the pdf after modulo operation[10]:

$$\tilde{f}_{V|M}(v|M=m) \triangleq \begin{cases} \sum_i f_{V|M}(v-i\Delta|M=m), & v \in \mathcal{R}_m; \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

In fact, (23) can be written as the KLD between the received pdf when one of the symbols has been sent, and the average of the pdfs of all possible symbols. Assuming equiprobable symbols in the binary signaling case, one

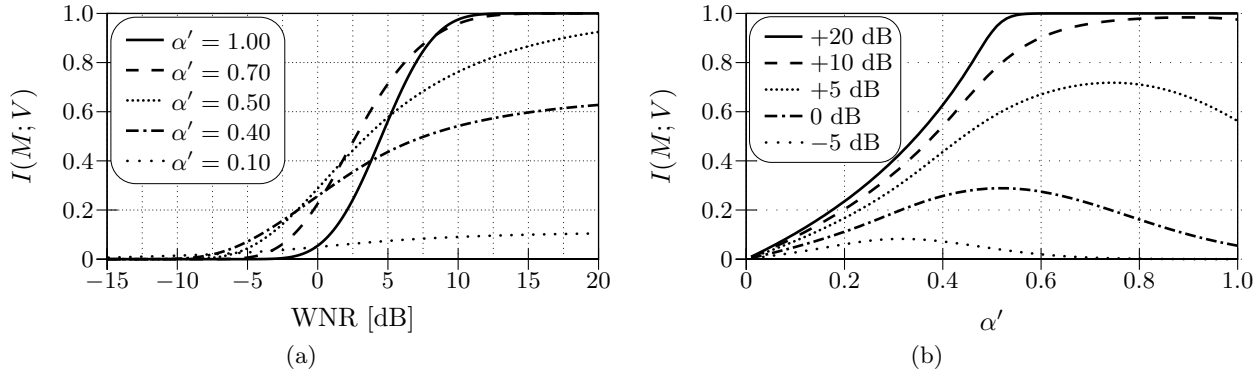


Figure 10. Mutual information for the AWGN attack case and binary signaling.

obtains:

$$I(M;V) = D\left(\tilde{f}_{V|M}(v|M=1)||\ \frac{1}{2}\sum_{m=1}^{2}\tilde{f}_V(v|M=m)\right). \tag{25}$$

As in the case when the probability of error was used as a cost function, the next section is dedicated to the analysis of the DM and the DC-DM under AWGN attack, uniform noise attack and WCAA.

4.1. Additive white Gaussian noise attack

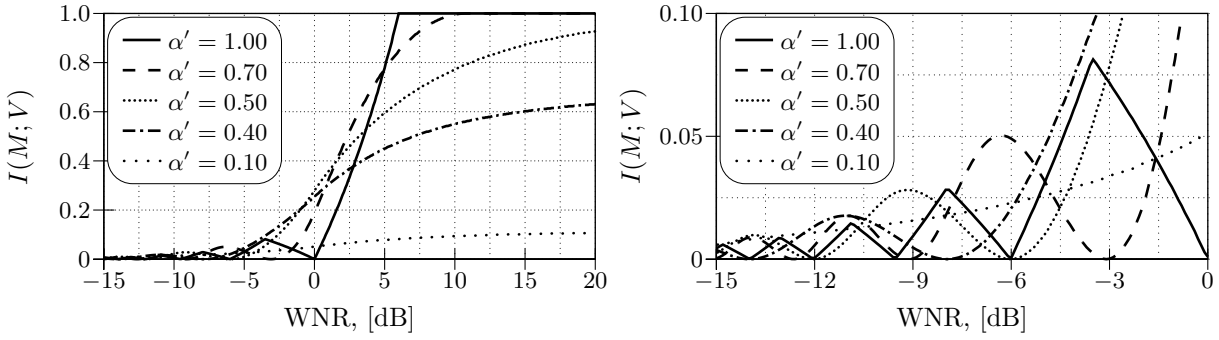
When the DM and the DC-DM undergo the AWGN, no closed analytical solution to the mutual information minimization problem exists; the minimization was therefore solved using numerical computations. The mutual information as a function of the WNR for different distortion compensation parameters is shown in Fig. 10(a) for the binary signaling case. Fig. 10(b) represents the mutual information as a function of the distortion compensation parameter α' for different WNRs represented in Fig. 10(a).

4.2. Uniform noise attack

It was shown[2] that the uniform noise attack is stronger than the AWGN attack for some WNRs when the probability of error is used as a cost function. One of the properties of the KLD measure states that it is equal to zero only if two pdfs are equal. In case the uniform noise attack is applied, this condition holds for some particular values of WNR for the mutual information given by (25). It can be demonstrated that $I(M;V) = 0$ when $\xi = \frac{\alpha'^2}{k^2}, k \in \mathbb{N}$. This particular behaviour allows the attacker to achieve zero rate of communication by applying an attacking power smaller than was predicted by the data-hider. Fig. 11 depicts the mutual information of quantization-based data-hiding techniques for the uniform noise attacking case. It demonstrates that the efficiency of the attack strongly depends on the value of the distortion compensation parameter, and shows the oscillating behaviour at the low-WNR detailed in Fig. 11(b).

The uniform noise attack guarantees that it is not possible to communicate using the DC-DM at $\xi \leq \alpha'^2$, and therefore distortion compensation parameter α' has a strong influence on the performance at the low-WNR when uniform noise attack is applied. As a conclusion, $\xi = \alpha'^2$ represents the WNR corresponding to zero rate of communication if the attacking energy satisfies $\sigma_Z^2 \geq \frac{D_w}{\alpha'^2}$.

For example, if the data-hider anticipates the WNR $= -6$dB, he/she could select $\alpha' = 0.7$ (Fig. 11(b)) to maximize the mutual information. Nevertheless, at the WNR $= -3$dB the mutual information is zero for $\alpha' = 0.7$. Therefore, the attacker can inhibit communications by making less efforts. In this example, to reduce the power of the attack on 3dB from the embedder prediction is favorable for the attacker.



(a) Global performance analysis. (b) Low-WNR.

Figure 11. Mutual information for the uniform noise attack case and binary signaling.

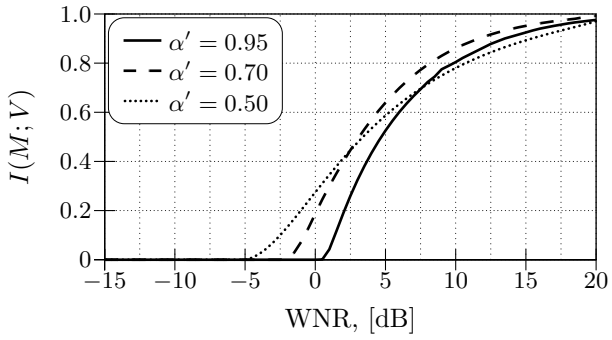


Figure 12. Mutual information for the WCAA case and binary signaling.

4.3. The worst case additive attack

The problem of the WCAA design using the mutual information as a cost function can be formulated as the following optimization problem:

$$\min_{\tilde{f}_Z(\cdot)} I(M;V) = \min_{\tilde{f}_Z(\cdot)} D\left(\tilde{f}_{V|M}(v|M=1)||\tilde{f}_V(v)\right), \qquad \tilde{f}_V(v) = \frac{1}{2}\sum_{m=1}^{2} \tilde{f}_{V|M}(v|M=m)\,. \tag{26}$$

The constraints in (26) are the same as with the probability of error oriented analysis case (14) and (15). Unfortunately, this problem has no closed form solution and it was solved numerically.

The obtained results are presented for different α' values in Fig. 12. In comparison with the AWGN and the uniform noise attacks, they demonstrate that the developed attack produces higher loss in terms of the mutual information than AWGN and uniform noise attacks for all WNRs (Fig. 13). In the analysis of the WCAA using probability of error as a cost function, the optimal α' parameter was found. Unfortunately, it is not the case in the mutual information oriented analysis, and its value varies with the WNR.

5. CONCLUSIONS

In this paper we have addressed the problem of designing the WCAA against quantization-based techniques from the probability of error and mutual information perspectives. The comparison between the analyzed cost functions demonstrates that in a rigid scenario with fixed decoder the attacker can decrease the rate of reliable communication more severely than with using the AWGN and the uniform noise attacks. We have shown that the AWGN attack is not the WCAA in general, and we have obtained the analytical solution of the WCAA problem design when the cost function is the probability of error. Moreover, $\alpha' = 2/3$ has been found to be the optimal value for the given decoder that allows to communicate with an upper bounded probability of error for the given WNR. This value can be fixed without prior knowledge of the attacking pdf.

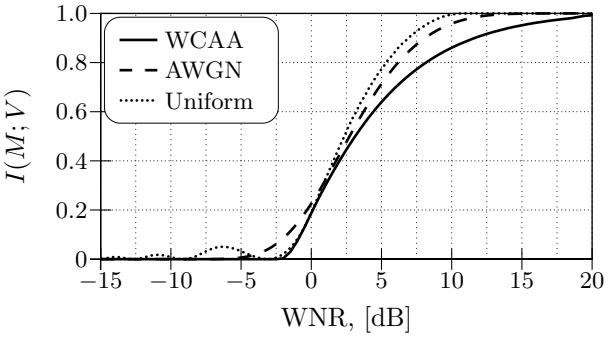


Figure 13. Comparison of different attacks using mutual information as a cost function for $\alpha' = 0.7$.

When the mutual information was used as a cost function, it was also demonstrated by means of numerical optimization that the AWGN is not the WCAA and the optimal distortion compensation parameter (α') depends on the operational WNR. The particular behaviour of the mutual information under uniform noise attack was described, achieving zero-rate performance for attacking energies σ_Z^2 such that $\sigma_Z^2 \geq \frac{D_w}{\alpha'^2}$.

ACKNOWLEDGMENTS

This paper was partially supported by SNF Professorship grant No PP002-68653/1, Interactive Multimodal Information Management (IM2) project and by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The authors are thankful to the members of the Stochastic Image Processing group at University of Geneva and to Pedro Comesaña and Luis Pérez-Freire of the Signal Processing in Communications Group at University of Vigo for many helpful and interesting discussions. The information in this document reflects only the author's views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

REFERENCES

1. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory* **29**, pp. 439–441, May 1983.
2. F. Perez-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**, April 2003.
3. J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Transactions on Signal Processing* **vol. 51**, pp. 1003–1019, April 2003.
4. B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory* **47**, pp. 1423–1443, 2001.
5. O. Koval, S. Voloshynovskiy, F. Perez-Gonzalez, F. Deguillame, and T. Pun, "Quantization-based watermarking performance improvement using host statistics: Awgn attack case," in *ACM Multimedia and Security Workshop 2004*, (Magdeburg, Germany), September 20-21 2004.
6. L. Perez-Freire, F. Perez-Gonzalez, and S. Voloshynovskiy, "Revealing the true achievable rates of scalar costa scheme," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, (Siena, Italy), September 29 - October 1 2004.
7. M. Barni and F. Bartolini, *Watermarking Systems Engineering*, Marcel Dekker, Inc., New York, 2004.
8. A. Mckellips and S. Verdu, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Transactions on Information Theory* **43**, pp. 1256–1264, July 1997.
9. T. Cover and J. Thomas, *Elements of Information Theory*, Wiley and Sons, New York, 1991.
10. F. Pérez-González, "The importance of aliasing in structured quantization modulation data hiding," in *International Workshop on Digital Watermarking*, (Seoul, Korea), 2003.