Quantization-Based Methods: Additive Attacks Performance Analysis

Vila Forcen, Jose Emilio; Voloshynovskyy, Svyatoslav; Pérez-González, Fernando; Koval, Oleksiy; Pun, Thierry

Quantization-Based Methods: Additive Attacks Performance Analysis

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 $|\mathcal{M}|$-ary quantization-based data-hiding methods using as design criteria the error probability and the maximum achievable rate of reliable communications. Our analysis focuses 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 is 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 obtained results demonstrate 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 additive white Gaussian or uniform noise attacks.**

***Index Terms*— Quantization-based, data-hiding, additive attacks, distortion compensation, dither modulation, probability of error, mutual information**

I. Introduction

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. 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.

In general, the digital data-hiding can be considered as a game between the data-hider and the attacker. This three-parties two-players game were already investigated by O'Sullivan *et al.* [1] where two set-ups are analyzed. In the first one, the host is assumed to be available at both encoder and decoder prior to the transmission, the so-called *private game*. In the second one, the host is only available at the encoder as in Figure 1, i.e., the *public game*. The performance is analyzed with respect to the maximum achievable rate when the decoder is aware of the attacking channel and therefore the *maximum likelihood* (ML) decoding is applied.

The knowledge of the attacking channel at the decoder is not a realistic case for most practical applications. Somekh-Baruch and Merhav considered the data-hiding problem in terms of maximum achievable rates and error exponents. They assumed that the host data is available either at both encoder and decoder [2] or only at the encoder [3] and supposed that neither encoder nor decoder is aware of the attacker strategy. In their consideration, the class of potentially applied attacks is significantly broader than in the previous study case [1] and includes any conditional pdf that satisfies a certain energy constraint. Although the solution of the problem is classically presented in terms of the achievable rate establishing the maximum number of messages $|\mathcal{M}|$ that can be reliably communicated, the error exponents solution is interesting in many practical applications where the objective is to minimize the probability of error at a given communications rate.

Quantization-based data-hiding methods have attracted attention in the watermarking community. They are a practical implementation of a binning technique for channels whose state is non-causally available at the encoder considered by Gel'fand-Pinsker [4]. Recently it has been also demonstrated [5] that quantization-based data-hiding performance coincides with the spread-spectrum (SS) data-hiding at the low-WNR by taking into account the host statistics and by abandoning the assumption of an infinite image to watermark ratio.

The quantization-based methods have been widely tested against a fixed channel and assuming that the channel transition pdf is available at the decoder. A *minimum Euclidean distance* (MD) decoder is implemented as a low-complexity equivalent of the ML decoder under the assumption of a pdf created by the symmetric extension of a monotonically non-increasing function [6].

It is a common practice in the data-hiding community to measure the performance in terms of the error rate for a given decoding rule as well as the maximum achievable rate of reliable communications. In this paper we will analyze the WCAA using both criteria.

In this paper we restrict the encoding to the quantization-based one and the channel to the class of additive attacks only. We assume that the attacker might be informed of the encoding strategy and also of the decoding one for the error exponent analysis, while both encoder and decoder are uninformed of the channel. Furthermore, the encoder is aware of the host image but not of the attacking strategy.

It is important to note that the optimality of the attack critically relies on the input alphabet even under power-limited attacks. McKellips and Verdu showed that the additive white Gaussian noise (AWGN) is not the WCAA for discrete input alphabets such as pulse amplitude modulation [7]. Similar conclusion for data-hiding was obtained by Pérez-González *et al.*

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

[8], who demonstrated that the uniform noise attack performs worse that the AWGN attack for some watermark-to-noise ratios (WNRs). In [9], Pérez-González *et al.* demonstrated that the AWGN cannot indeed be the WCAA because of its infinite support. Vila-Forcén *et al.* [10] and Goteti and Moulin [11] solved independently the min-max problem for distortion-compensated dither modulation (DC-DM) [12] in terms of probability of error for the fixed decoder, binary signaling and the subclass of additive attacks. Simultaneously, Vila-Forcén *et al.* [13] and Tzoschoppe *et al.* [14] derived the WCAA for the DC-DM using the mutual information as objective function for the additive attacks and binary signaling.

This paper aims at establishing the information-theoretic limits of $|\mathcal{M}|$-ary quantization-based data-hiding techniques and developing a benchmark that can be used for the fair comparison of different quantization-based methods.

The selection of the distortion compensation parameter α' (see Section II-B) fixes the encoder structure for the quantization-based methods. Although the optimal α' can easily be determined when the power of the noise is available at the encoder prior to the transmission [15], this is not always feasible for various practical scenarios. Nevertheless, the availability of the attacking power and of the attacking pdf is a very common assumption in most data-hiding schemes. We will demonstrate that for a specific decoder (MD decoder) it is possible to calculate the optimal α' independently of the attack variance and pdf for the block error probability as a cost function.

The paper is organized as follows. Problem formulation is given in Section II. The investigation of the WCAA for a fixed quantization-based data-hiding scenario is performed in Section III, where the cost function is the probability of error. The information-theoretic analysis of Section IV derives the information bounds where the cost function is the mutual information between the input message and the channel output.

Notations: We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$ and corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively. An information message and a set of messages with cardinality $|\mathcal{M}|$ is designated as $m \in \mathcal{M}, \mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$, respectively. A host signal distributed according to the pdf $f_{\mathbf{X}}(\mathbf{x})$ is denoted by $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x})$; $\mathbf{Z} \sim f_{\mathbf{Z}}(\mathbf{z})$, $\mathbf{W} \sim f_{\mathbf{W}}(\mathbf{w})$ and $\mathbf{V} \sim f_{\mathbf{V}}(\mathbf{v})$ represents the attack, the watermark and the received signal, respectively. The step of quantization is equal to Δ and the distortion-compensation factor is denoted as α'. The variance of the watermark is σ_W^2 and the variance of the attack is σ_Z^2. The watermark-to-noise ratio (WNR) is given by $\text{WNR} = 10\log_{10}\xi$, where $\xi = \frac{\sigma_W^2}{\sigma_Z^2}$. The set of natural numbers is denoted as $\mathbb{N}$ and $\mathbb{I}_N$ denotes the $N \times N$ identity matrix.

II. Problem formulation

A. Data-hiding formulation of the Gel'fand-Pinsker problem

1) Gel'fand-Pinsker set-up: The Gel'fand-Pinsker problem [4] has been recently revealed as the appropriate theoretical framework of data-hiding communications with side information (Figure 1). The random variable $\mathbf{X}$ stands for the host signal, which is independent and identically distributed (i.i.d.) and available non-causally at the encoder.

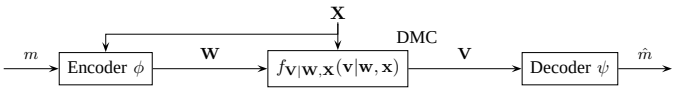


Fig. 1. Gel'fand-Pinsker channel coding with side information available at the encoder.

We define the encoder as a mapping $\phi : \mathcal{M} \times \mathcal{X}^N \to \mathcal{W}^N$, assuming that $m \in \mathcal{M}$, $\mathbf{x} \in \mathcal{X}^N$ and $\mathbf{w} \in \mathcal{W}^N$.

The channel given $\mathbf{X}$ is assumed to be a discrete memoryless channel (DMC) described by the corresponding transition pdf: $f_{\mathbf{V}|\mathbf{W},\mathbf{X}}(\mathbf{v}|\mathbf{w},\mathbf{x}) = \prod_{i=1}^{N} f_{V|W,X}(v_i|w_i,x_i)$.

The decoder estimates the embedded message from the output of the channel $\psi : \mathcal{V}^N \to \mathcal{M}$. Within the Gel'fand-Pinsker set-up, the decoder is aware of the channel pdf and therefore optimal decoding can be performed. A jointly typical decoder is used in [4] as an equivalent to the ML one for the simplicity of the analysis.

For the above channel, the capacity is:

$$C = \max_{p_{U,W|X}(\cdot,\cdot|\cdot)} \left[I(U;V) - I(U;X) \right], \tag{1}$$

where U stands for an auxiliary random variable with $U \in \mathcal{U}$, $|\mathcal{U}| = \min\{|\mathcal{W}|, |\mathcal{Y}|\} + |\mathcal{X}| - 1$.

2) Gel'fand-Pinsker data-hiding problem: The above Gel'fand-Pinsker set-up describes the general framework of communications with side-information. However, it is needed to introduce the distortion constraints and the key management to convert it to the hidden communications scenario.

The Gel'fand-Pinsker data-hiding set-up is presented in Figure 2. The encoder is now a mapping $\phi : \mathcal{M} \times \mathcal{X}^N \times \mathcal{K} \to \mathcal{W}^N$, where the key $K \in \mathcal{K}, \mathcal{K} = \{1, 2, \ldots, |\mathcal{K}|\}$. The stego data $\mathbf{Y}$ is obtained using the embedding mapping: $\varphi : \mathcal{W}^N \times \mathcal{X}^N \to \mathcal{Y}^N$. The decoder estimates the embedded message as $\psi : \mathcal{V}^N \times \mathcal{K} \to \mathcal{M}$. According to this scheme, a key is available at both encoder and decoder. Nevertheless, key management is outside of the scope of this paper and we will not consider it further.

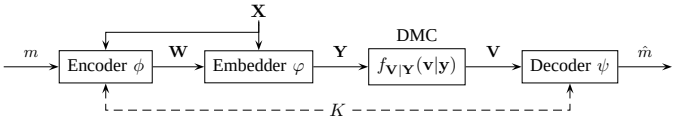


Fig. 2. Gel'fand-Pinsker data-hiding set-up.

Two constraints apply to the Gel'fand-Pinsker framework in the data-hiding scenario: the embedding and the channel constraints [1]. Let $d(\cdot,\cdot)$ be a nonnegative function and σ_W^2, σ_Z^2 be two positive numbers, the embedder is said to satisfy the embedding constraint if:

$$\sum_{\mathbf{x}\in\mathcal{X}^N} \sum_{\mathbf{y}\in\mathcal{Y}^N} d(\mathbf{x},\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) \leq \sigma_W^2, \tag{2}$$

where $d(\mathbf{x},\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} d(x_i, y_i)$.

Analogously, the channel is said to satisfy the channel constraint if:

$$\sum_{\mathbf{y}\in\mathcal{Y}^N}\sum_{\mathbf{v}\in\mathcal{V}^N} d(\mathbf{y},\mathbf{v}) f_{\mathbf{Y},\mathbf{V}}(\mathbf{y},\mathbf{v}) \leq \sigma_Z^2. \quad (3)$$

We define a code $(|\mathcal{M}|, N)$ subject to the distortion constraint σ_W^2 as the message set $m \in \mathcal{M}$, $|\mathcal{M}| = 2^{NR}$, encoding function ϕ and embedding function φ such that the embedding constraint (2) is satisfied.

The average block error probability of a code $(|\mathcal{M}|, N)$ subject to the embedding constraint σ_W^2, a channel transition pdf $f_{\mathbf{V}|\mathbf{Y}}(\mathbf{v}|\mathbf{y})$ subject to the channel constraint σ_Z^2, for a given decoding rule ψ and assuming equiprobable input distribution is:

$$P_B^{(N)} = \frac{1}{|\mathcal{M}|}\sum_{m\in\mathcal{M}} \Pr\left[\psi(\mathbf{V}, K) \neq m | M = m\right]. \quad (4)$$

Given a block error probability, a rate $R = \frac{1}{N}\log_2|\mathcal{M}|$ is said to be achievable for the given distortions pair (σ_W^2, σ_Z^2) if there exists a code $(|\mathcal{M}|, N)$ such that $P_B^{(N)} \to 0$ as $N \to \infty$. In the following we will refer to the block error probability simply as error probability.

Costa set-up: Costa considered the Gel'fand-Pinsker problem for the i.i.d. Gaussian case and mean square error distance [16]. Costa set-up is presented in Figure 3. The embedder φ performs $\mathbf{Y} = \mathbf{W} + \mathbf{X}$, $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2\mathbb{I}_N)$. It is possible to write the channel output as: $\mathbf{V} = \mathbf{X} + \mathbf{W} + \mathbf{Z}$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2\mathbb{I}_N)$, and the estimate of the message $\hat{m}$ is obtained at the decoder given $\mathbf{V}$.

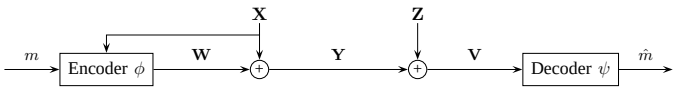


Fig. 3. Costa data-hiding set-up.

The auxiliary random variable was chosen as $\mathbf{U} = \mathbf{W} + \alpha\mathbf{X}$ with optimization parameter α. Costa has shown that an optimal value of this parameter can be chosen as $\alpha_{\text{opt}} = \frac{\sigma_W^2}{\sigma_W^2+\sigma_Z^2}$ assuming that encoder knows in advance the noise variance. In this case, the proposed set-up achieves host interference cancellation and:

$$R(\alpha_{\text{opt}}) = C^{\text{AWGN}} = \frac{1}{2}\log_2\left(1 + \frac{\sigma_W^2}{\sigma_Z^2}\right) \quad (5)$$

that corresponds to the AWGN channel capacity without host interference.

B. Quantization-based data-hiding techniques

Aiming at reducing the Costa codebook exponential complexity, a number of practical data-hiding algorithms exploit *structured codebooks* instead of random ones. The most famous discrete approximations to Costa problem are known as DC-DM [12] and scalar Costa scheme (SCS) [15]. The structured codebooks are designed using quantizers (or lattices

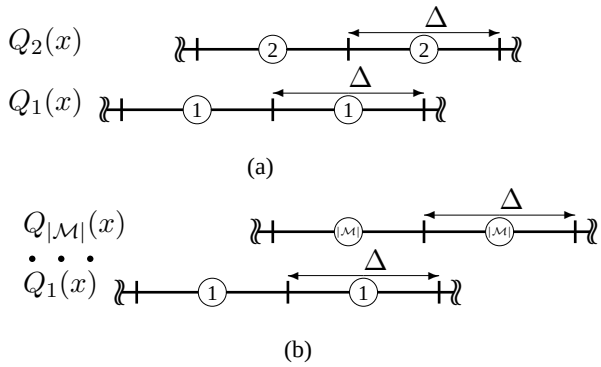


Fig. 4. DM embedding quantizers: (a) binary signaling and (b) M-ary signaling.

[17]) which should achieve host interference cancellation. In this case, the auxiliary random variable is given by:

$$\mathbf{U} = \mathbf{W} + \alpha'\mathbf{X} = \alpha' Q_m(\mathbf{X}), \quad (6)$$

where $Q_m(\cdot)$ denotes a vector or scalar quantizer for the message m.

Assuming that the channel transition pdf is given by some additive noise pdf, within the class of quantization-based methods, we focus our analysis on DC-DM and dither modulation (DM) [12].

In the case of scalar DM, the stego data is obtained by the encoder $\phi : \mathcal{M}\times\mathcal{X} \to \mathcal{Y}$ applying a message dependent quantizer (or lattice) to the host data, i.e.:

$$\phi_{\text{DM}}(m, x) = y = Q_m(x), m \in \mathcal{M}, \quad (7)$$

as it is shown in Figure 4.

Here, we use quantizers designed using subtractive dithering; i.e., each quantizer is a shifted version of the others [18]:

$$Q_m(x) = Q(x + d_m) - d_m, \quad (8)$$

where d_m represents the subtractive dither of the m-th message and $Q(\cdot)$ stands for a fixed quantizer with quantization step Δ assuming high rate quantization regime. The variance of the stego data is equal to the variance of a uniform pdf $\mathcal{U}(-\Delta/2, \Delta/2)$ resulting from the quantization noise, $\sigma_W^2 = \frac{\Delta^2}{12}$. In this case the pdf of the stego data is assumed to be a train of δ-functions as the result of quantization.

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

$$\phi_{\text{DC-DM}}(m, x, \alpha') = y = x + \alpha'(Q_m(x) - x), \quad (9)$$

where $0 < \alpha' \leq 1$ is the analogue of the Costa optimization parameter α. If $\alpha' = 1$, the DC-DM (9) simplifies to the DM (7). The embedding distortion for the DC-DM is $\sigma_W^2 = \alpha'^2\frac{\Delta^2}{12}$. In this case, the pdf of the stego image is represented by a train of uniform pulses of width $2B = (1-\alpha')\Delta$ centered at the quantizer reconstruction level as a result of the distortion compensation[1]. An example of such a pdf corresponding to the communications of the message $m = 1$ is given in Figure 5 where $T_h = \frac{\Delta}{2|\mathcal{M}|}$ denotes the distance between two neighbor quantizer decision and reconstruction levels.

[1]The analysis is performed here in the framework of Eggers *et al.* disregarding the host pdf impact. If host pdf is taken into account, we refer readers to [5], [19] for more details.

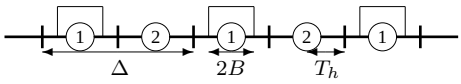


Fig. 5. DC-DM output pdf for the message $m = 1$ and binary signaling.

It is not always possible to know in advance the attacking pdf needed for the optimal ML decoding. Within the class of additive attacks, one can nevertheless assume that if the attack pdf is created by the symmetric extension of a monotonically non-increasing function, the ML decoder reduces to the MD decoder [6]:

$$\hat{m}^{\text{MD}} = \arg \min_{m \in \mathcal{M}} ||v - Q_m(v)||^2. \tag{10}$$

Using the MD decoding rule, the correct decoding region $\mathcal{R}_m$ and the complementary error region $\overline{\mathcal{R}}_m$ associated to a message m, are defined as it is depicted in Figure 6 [8].

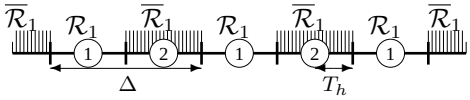


Fig. 6. DM and DC-DM correct decoding region $\mathcal{R}_1$ and error decoding region $\overline{\mathcal{R}}_1$ for the message $m = 1$ and binary signaling when the MD decoder is used.

III. Error probability as a cost function

When the average error probability is selected as a cost function, we formulate the problem of Figure 2 as:

$$P_B^{*(N)} = \min_{\phi,\psi} \max_{f_{V|Y}(\cdot|\cdot)} P_B(\phi, \psi, f_{V|Y}(\cdot|\cdot)). \tag{11}$$

The error probability depends on the particular encoder/decoder pair (ϕ, ψ) and the attacking channel $f_{\mathbf{V}|\mathbf{Y}}(\mathbf{v}|\mathbf{y})$, i.e., $P_B(\phi, \psi, f_{V|Y}(v|y)) = \Pr[\hat{m} \neq m | M = m]$. Here, we assume that the attacker knows both encoder and decoder strategies and selects its attacking strategy accordingly. Both encoder and decoder select their strategy without knowing the attack in advance. Although this is a very conservative set-up, it is also important for various practical scenarios. One can then compute the reliability function for the class of attack channels as:

$$E(R) = \limsup_{N \to \infty} \left[-\frac{1}{N} \log_2 P_B^{*(N)} \right] \tag{12}$$

that can be used for further error exponents analysis.

The more advantageous set-up for the data-hider is based on the assumption that the decoder selects its strategy knowing the attacker choice:

$$\min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} \min_{\psi} P_B(\phi, \psi, f_{V|Y}(\cdot|\cdot)). \tag{13}$$

Here, the attacker knows only the encoding function, which is fixed prior to the attack, and the decoder is assumed to be aware of the attack pdf.

In the general case, Somekh-Baruch and Merhav [2] have shown the following inequalities for the above scenarios:

$$\begin{aligned} \min_{\phi,\psi} \max_{f_{V|Y}(\cdot|\cdot)} & P_B(\phi, \psi, f_{V|Y}(\cdot|\cdot)) \\ & \geq \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} \min_{\psi} P_B(\phi, \psi, f_{V|Y}(\cdot|\cdot)) \quad (14) \\ & = \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} P_B(\phi, \psi^{\text{ML}}, f_{V|Y}(\cdot|\cdot)), \quad (15) \end{aligned}$$

where the equality (15) assumes that the decoder is aware of the attacking pdf and therefore the minimization at the decoder results in the optimal ML decoding strategy ψ^{ML}.

In many practical benchmarking approaches, the performance of various data-hiding methods is measured in front of fixed attacks that are known to be the worst ones in some communication scenarios. Nevertheless, some particular applications might not necessarily use the WCA. Therefore, it is interesting to bound the system performance for any attack: this problem can be formulated as the data-hiding performance for the fixed attack like the AWGN or the uniform noise attacks with a given pdf $\tilde{f}_{V|Y}(\cdot|\cdot)$:

$$\begin{aligned} \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} & P_B(\phi, \psi^{\text{ML}}, f_{V|Y}(\cdot|\cdot)) \\ & \geq \min_{\phi} P_B(\phi, \psi^{\text{ML}}, \tilde{f}_{V|Y}(\cdot|\cdot)), \end{aligned} \tag{16}$$

where the equality holds if, and only if, the fixed attack pdf $\tilde{f}_{V|Y}(\cdot|\cdot)$ coincides with the WCA.

Using (15) one can write:

$$\begin{aligned} \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} & P_B(\phi, \psi^{\text{MD}}, f_{V|Y}(\cdot|\cdot)) \\ & \geq \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} P_B(\phi, \psi^{\text{ML}}, f_{V|Y}(\cdot|\cdot)), \end{aligned} \tag{17}$$

with equality if, and only if, the MD decoder coincides with the optimal ML decoder.

In the analysis of the WCAA using the error probability as a cost function, we will further assume that the MD decoder is applied. Using (16) and assuming the MD decoding rule, one can write:

$$\begin{aligned} \min_{\phi} \max_{f_{V|Y}(\cdot|\cdot)} & P_B(\phi, \psi^{\text{MD}}, f_{V|Y}(\cdot|\cdot)) \\ & \geq \min_{\phi} P_B(\phi, \psi^{\text{MD}}, \tilde{f}_{V|Y}(\cdot|\cdot)), \end{aligned} \tag{18}$$

where the equality holds if, and only if, the fixed attack pdf $\tilde{f}_{V|Y}(v|y)$ coincides with the WCAA.

In the class of additive attacks, the attacking channel transition pdf is only determined by the pdf of the additive noise $f_Z(z)$. Finally, in this analysis we assume independence of the error probability on the quantization bin where the received signal v lies (because the error region $\overline{\mathcal{R}}_m$ (Figure 6) has periodical structure and the host pdf $f_X(x)$ is assumed to be asymptotically constant within each quantization bin).

Applying (18) to the quantization-based data-hiding (Section II-B), assuming an additive attacking scenario (Figure 3), the MD decoding rule (10) and high-rate, one has:

$$\min_{\alpha'} \max_{f_Z(\cdot)} P_B(\alpha', \psi^{\text{MD}}, f_Z(\cdot)) \geq \min_{\alpha'} P_B(\phi, \psi^{\text{MD}}, \tilde{f}_Z(\cdot)), \tag{19}$$

where the encoder optimization is reduced to the selection of an optimal parameter α' and the channel to the selection of the worst additive noise pdf.

The problem (19) implies that the attacker might know both encoding and decoding strategy. Here, we target finding the WCAA pdf and the optimum fixed encoding strategy independently of the particular attacking case which guarantees reliable communications and provides an upper bound on the error probability.

Considering the previously discussed quantization-based techniques and the MD decoder, and assuming that the message m is communicated, the probability of correct decoding P_B^c is determined as [8]:

$$P_B^c = \Pr[||V - Q_m(V)||^2 < ||V - Q_{m'}(V)||^2 : \forall\, m' \in \mathcal{M}, m' \neq m] = \Pr[V \in \mathcal{R}_m | M = m].$$

The error probability can be obtained as $P_B = 1 - P_B^c$. We can represent the error probability as the integral of the equivalent noise pdf $f_{Z_{eq}|M} = f_Z * f_{\text{DC-DM}}$ over the error region $\overline{\mathcal{R}}_m$:

$$P_B = \sum_{m=1}^{|\mathcal{M}|} p_M(m) \int_{\overline{\mathcal{R}}_m} f_{Z_{eq}|M}(z_{eq}|M=m) dz_{eq}. \quad (20)$$

For the $|\mathcal{M}|$-ary case, it is possible to write the probability of error as a sum of integrals as:

$$P_B = 2 \sum_{m=1}^{|\mathcal{M}|} p_M(m) \cdot \sum_{k=0}^{\infty} \int_{k\Delta+\Delta/2|\mathcal{M}|}^{(k+1)\Delta-\Delta/2|\mathcal{M}|} f_{Z_{eq}|M}(z_{eq}|M=m) dz_{eq}. \quad (21)$$

Concerning the DM, the pdf of $f_{Z_{eq}}(\cdot)$ is a periodical repetition of the noise pdf $f_Z(\cdot)$. In the case of DC-DM the pdf is given by the convolution of the attacking pdf with the self-noise pdf (periodic uniform pdf) [8].

The following subsections are dedicated to the analysis of the error probability (21) for the fixed attacks, i.e., AWGN and uniform noise. Finally, the WCAA has been derived for both $|\mathcal{M}|$-ary DM and $|\mathcal{M}|$-ary DC-DM.

A. Additive white Gaussian noise attack

This section contains the error probability analysis of the $|\mathcal{M}|$-ary DM and DC-DM techniques under the AWGN attack.

1) DM analysis: In the DM case, the equivalent noise pdf is given by:

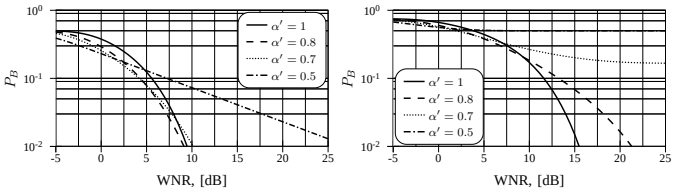
$$f_{Z_{eq}|M}(z_{eq}|M=m) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{{z_{eq}}^2}{2\sigma_Z^2}}, \quad (22)$$

where σ_Z^2 denotes the variance of the attack. The error probability can be therefore calculated using (21).

2) DC-DM analysis: In the DC-DM case the equivalent noise pdf is given by [8]:

$$f_{Z_{eq}|M}(z_{eq}|M=m) = \frac{1}{2B}\left(\mathcal{Q}\left(\frac{z_{eq}-B}{\sigma_Z}\right) - \mathcal{Q}\left(\frac{z_{eq}+B}{\sigma_Z}\right)\right),$$

where $\mathcal{Q}$ is the $\mathcal{Q}$-function, $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2} dt$, and B is the half-width of the self-noise pdf. The analytical expression for the error probability (20) does not exist, and it is evaluated numerically using (21). The error probability for the DM and the DC-DM under the AWGN attack is depicted in Figure 7.



(a) (b)

Fig. 7. Error probability analysis result for the AWGN attack case: (a) binary signaling and (b) quaternary signaling.

B. Uniform noise attack

It was shown [8] that the uniform noise attack produces higher error probability than the AWGN attack for some particular WNR in the binary signaling case. This fact contradicts the common belief that the AWGN is the WCAA for all data-hiding methods since it has the highest differential entropy among all pdfs with bounded variance.

We consider the uniform noise attack $Z \sim \mathcal{U}(-\eta, \eta)$ with variance $\sigma_Z^2 = \frac{\eta^2}{3}$ assuming that the MD decoder is used.

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 error probability is zero. For stronger attacks the error probability is defined by the integral of the equivalent noise pdf (a uniform pdf) over the error region using (21). The analytical solution when $\eta < \frac{2|\mathcal{M}|+1}{|\mathcal{M}|}\frac{\Delta}{2}$ in the $|\mathcal{M}|$-ary case is:

$$P_B(\alpha'=1, \psi^{\text{MD}}, f_Z^{\text{Unif.}}(\cdot)) = \begin{cases} 0, & \eta < \frac{\Delta}{2|\mathcal{M}|}; \\ 1 - \frac{\Delta}{2|\mathcal{M}|\eta}, & \frac{\Delta}{2|\mathcal{M}|} \leq \eta < \frac{2|\mathcal{M}|-1}{|\mathcal{M}|}\frac{\Delta}{2}; \\ \frac{\Delta}{\eta}\frac{|\mathcal{M}|-1}{|\mathcal{M}|}, & \frac{2|\mathcal{M}|-1}{|\mathcal{M}|}\frac{\Delta}{2} \leq \eta < \frac{2|\mathcal{M}|+1}{|\mathcal{M}|}\frac{\Delta}{2}. \end{cases} \quad (23)$$

In the third case, the error probability decreases while the WNR decreases as well. This effect is caused by the entrance of the noise into the nearest correct region and a smaller portion of the attack power is present in the error region. Because of this effect we have a non strictly decreasing probability of error as a function of the WNR. If $\eta > \frac{2|\mathcal{M}|+1}{|\mathcal{M}|}\frac{\Delta}{2}$, the error probability starts increasing again since the received pdf enters again the error region. The performance of the DM in the uniform noise attack is presented in Figure 8.

2) DC-DM analysis: Under the uniform noise attack, the bit error probability is equal to the integral of the equivalent noise pdf $f_{Z_{eq}|M}(z_{eq}|M=m)$ (a trapezoidal function) over the error region (21). The resulting analytical equation for $\eta + B < \Delta - T_h$ in the $|\mathcal{M}|$-ary case is:

$$P_B(\alpha', \psi^{\text{MD}}, f_Z^{\text{Unif.}}(\cdot)) = \begin{cases} 0, & T_h > \eta + B; \\ \frac{k_1}{8|\mathcal{M}|^2}, & |\eta - B| < T_h < \eta + B; \\ \min\{\frac{1}{2B}, \frac{1}{2\eta}\} \cdot k_2, & T_h < |\eta - B|, \end{cases} \quad (24)$$

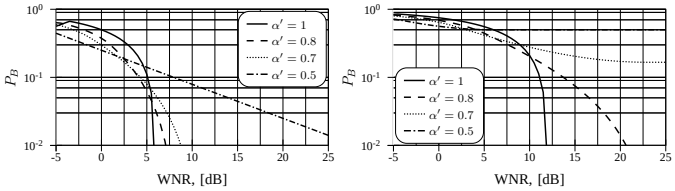


Fig. 8. Error probability for the uniform noise attack case: (a) binary signaling and (b) quaternary signaling.

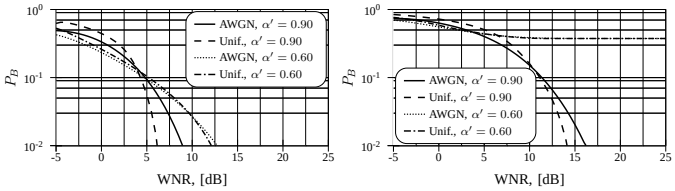


Fig. 9. Comparison of the AWGN and the uniform noise attacks in terms of the error probability for the minimum distance decoding rule: (a) binary signaling and (b) quaternary signaling.

where $k_1 = (2(\eta + B)|\mathcal{M}| - \Delta)(2m|\mathcal{M}|(\eta + B) + 4n|\mathcal{M}| + m\Delta)$, $k_2 = \left(\frac{(\eta+B)-|\eta-B|}{2} + ((\eta - B) - T_h)\right)$, $m = \frac{\min\{1/2B, 1/2\eta\}}{|\eta-B|-(\eta+B)}$ and $n = -m(\eta + B)$. If $\eta + B > \Delta - T_h$, the error probability decreases as in the DM case. The corresponding performance of the DC-DM under the uniform noise attack is presented in Figure 8. Since we are assuming fixed decoder, the error probability for the binary case can be higher than 0.5.

The experimental results presented in Figure 9 allow to conclude that the AWGN attack is not the WCAA for the assumed fixed decoder structure in the $|\mathcal{M}|$-ary case. Thus, the main goal of the following section consists in the development of the WCAA for the considered embedding scenarios.

C. The worst case additive attack

The problem of the WCAA for digital communications based on binary pulse amplitude modulation (PAM) was considered in [7] using the error probability under attack power constraint. In this paper, the problem of the WCAA is addressed for the quantization-based data-hiding methods.

The problem (18) for the DM with the fixed MD decoder (10) can be reformulated as:

$$\min_{\alpha'} \max_{f_Z(\cdot)} P_B(\alpha', \psi^{\text{MD}}, f_Z(\cdot)), \tag{25}$$

where the encoder is optimized over all α' such that $0 < \alpha' \leq 1$, and the attacker selects the attack pdf $f_Z(\cdot)$ maximizing the error probability P_B. Since the encoder must be fixed in advance in the practical set-ups, we will first solve the above min-max problem as an internal maximization problem for a given encoder/decoder pair:

$$\max_{f_Z(\cdot)} P_B(\alpha', \psi^{\text{MD}}, f_Z(\cdot)) = \max_{f_Z(\cdot)} \int_{\overline{\mathcal{R}}_m} f_V(v|M = m)dv, \tag{26}$$

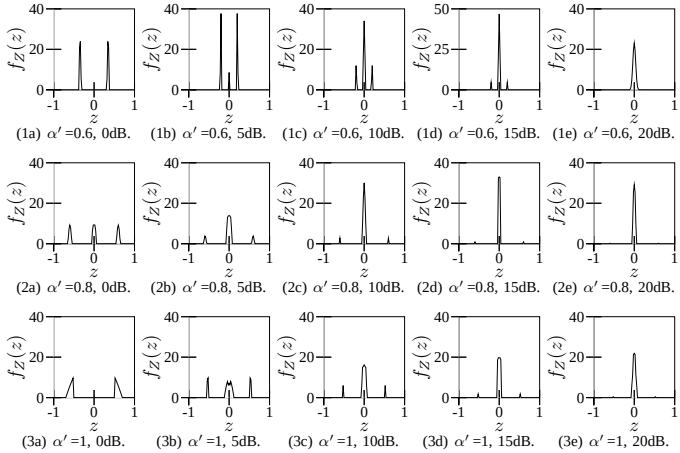


Fig. 10. WCAA optimization resulting pdfs for different α' and WNR, binary signaling.

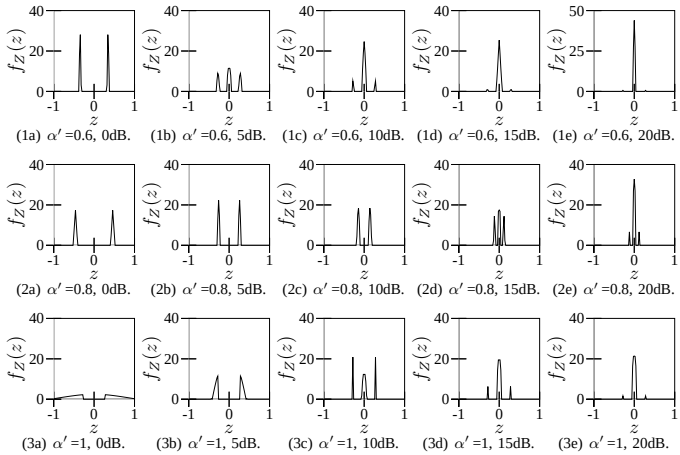


Fig. 11. WCAA optimization resulting pdfs for different α' and WNR, quaternary signaling.

where $0 < \alpha' \leq 1$, subject to the constraints:

$$\int_{-\infty}^{\infty} f_Z(z)dz = 1, \qquad \int_{-\infty}^{\infty} z^2 f_Z(z)dz \leq \sigma_Z^2, \tag{27}$$

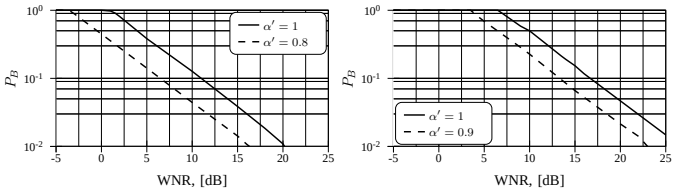
where the first constraint follows from the pdf definition and σ_Z^2 constrains the attack power.

We will derive the WCAA based on (26) for the fixed α' and use it for the solution of (25) accordingly. The distortion compensation parameter α' leading to the minimum error probability will be the solution to (25).

Unfortunately, no close analytical solution has been found. The resulting attacking pdfs obtained using numerical optimization are presented in Figure 10 and Figure 11 for different WNRs and α' values assuming $\Delta = 2$.

The obtained pdfs are non-monotonic functions. Thus, the MD decoder is not equivalent to the ML decoder. The obtained error probabilities are depicted in Figure 12, where the maximum is equal to 1 since we are assuming that the decoder is fixed (MD decoder) and it is completely known to the attacker. In a different decoding case when it is possible to invert the bit values, the maximum error probability will be equal to 0.5.

Motivated by the obtained pdfs and in order to receive



(a) (b)

Fig. 12. WCAA error probability optimization result: (a) binary signaling and (b) quaternary signaling

mathematically tractable results, we approximate the WCAA by a so-called $3-\delta$ attack whose pdf is presented in Figure 13. The $3-\delta$ attack provides a simple and powerful attacking strategy, which approximates the WCAA and might be used for testing different data-hiding algorithms. In order to demonstrate how accurate this approximation is, one needs to compare the average error probability caused by this attack versus the numerically obtained results.

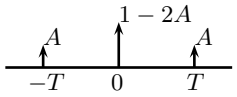


Fig. 13. $3-\delta$ attack, $0 \leq A \leq 0.5$.

For this purpose, the optimization of the $3-\delta$ attack parameters has been performed for the DC-DM considering the DM as a particular case for $\alpha' = 1$. When $T - B < T_h$, the error probability is equal to the integral of the equivalent noise pdf $f_{Z_{eq}|M}(z_{eq}|M=m)$ over the error region $\overline{\mathcal{R}}_m$:

$$P_B = \frac{A}{B}(T + B - T_h), \tag{28}$$

where $2B = (1-\alpha')\Delta$, $T_h = \frac{\Delta}{2|\mathcal{M}|}$ and $A = \frac{\sigma_Z^2}{2T^2}$. It is maximized for the following selection of $T = T_{opt_1}$:

$$T_{opt_1} = \frac{\Delta(1 - |\mathcal{M}|(1-\alpha'))}{|\mathcal{M}|}. \tag{29}$$

The value of T_{opt_1} should be always positive, implying that $\alpha' > \frac{|\mathcal{M}|-1}{|\mathcal{M}|}$. It can be demonstrated that $T_{opt_1} \to 0$ as $\alpha' \to \frac{|\mathcal{M}|-1}{|\mathcal{M}|}$. For a given attack variance $\sigma_Z^2 = 2T^2A > 0$ and $T_{opt_1} \to 0$, one has $A \to 0.5$ (its maximum value to satisfy the technical requirement to pdf in Figure 13). Simplifying (28) for $\alpha' \to \frac{|\mathcal{M}|-1}{|\mathcal{M}|}$ implies that $P_B \to 1$.

If $\alpha' > \frac{|\mathcal{M}|-1}{|\mathcal{M}|}$ and $T = T_{opt_1}$, the error probability is given by:

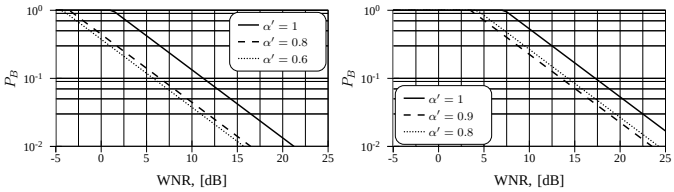
$$P_B = \frac{\sigma_Z^2 |\mathcal{M}| \alpha'^2}{24 \cdot \sigma_W^2 (1-\alpha')(1 - |\mathcal{M}|(1-\alpha'))}. \tag{30}$$

This result is valid if $T_{opt_1} - B < T_h$, and this constraint implies that $\alpha' < 1 - \frac{1}{3|\mathcal{M}|}$. For this case, the minimum of the error probability is achieved at:

$$\alpha'_{\text{opt}} = \frac{2(|\mathcal{M}|-1)}{2|\mathcal{M}|-1}. \tag{31}$$

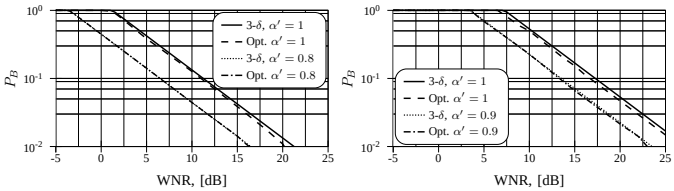
In the case when the previous condition does not hold, the error probability is calculated as: $P_B = 2A$. The maximum is found for the minimum possible $T = T_{opt_2} = T_h + B$, and the error probability is:

$$P_B = \frac{\sigma_Z^2 |\mathcal{M}|^2 \alpha'^2}{3 \cdot \sigma_W^2 (1 + |\mathcal{M}|(1-\alpha'))^2}. \tag{32}$$



(a) (b)

Fig. 14. Error probability analysis result for the $3-\delta$ attack case: (a) binary signaling and (b) quaternary signaling.



(a) (b)

Fig. 15. Error probability comparison between the numerical optimization results and the $3-\delta$ attack case: (a) binary signaling and (b) quaternary signaling

The corresponding performance for the DM and the DC-DM under the 3-δ attack is presented in Figure 14. The comparison presented in Figure 15 demonstrates that the 3-δ attack produces asymptotically the same error probability as the optimization results presented in Figure 10 and Figure 11.

The optimization results (Figure 10 and Figure 11) demonstrate that for very low-WNR the WCAA structure does not necessarily corresponds to the 3-δ attack. Nevertheless, the 3-δ attack is one of the possible choices, which achieves the maximal error probability and therefore can be used as the WCAA as it is shown in Figure 15. The previously analyzed AWGN and the uniform noise attacks are compared with the WCAA in Figure 16, demonstrating that the gap between the AWGN attack and the real worst case attack can be larger than 5dB in terms of the WNR.

The error probability as a function of the distortion compensation parameter for a given WNR demonstrates that the $3-\delta$ attacking scheme is worse than either the uniform or the Gaussian ones (Figure 17). If the noise attack is known, it is possible to select such an α' that minimizes the error probability for the given WNR in Figure 17. For example, if WNR $= 0$dB and Gaussian noise is applied, the optimal distortion compensation factor is $\alpha' = 0.53$, resulting in $P_B = 0.23$. Nevertheless, the encoder and the decoder are in general uninformed of the attacking strategy in advance and a mismatch in the attacking scheme may cause a bit error probability[2] of 1, while for $\alpha' = 0.66$ the maximum bit error probability is $P_B = 0.33$.

In order to find the optimal compensation parameter value that will allow the data-hider to upper bound the error

[2]In general the maximum bit error probability is equal to 1 for the fixed MD decoder.

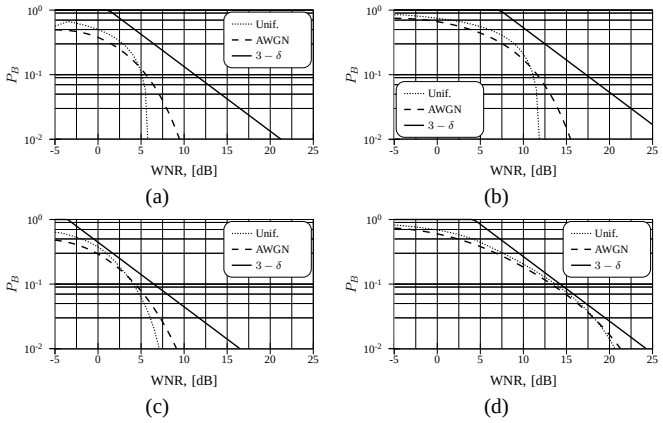


Fig. 16. Error probability analysis result for different attacking strategies: (a) DM performance, binary signaling, (b) DM performance, quaternary signaling, (c) DC-DM for $\alpha' = 0.8$ performance, binary signaling and (d) DC-DM for $\alpha' = 0.8$ performance, quaternary signaling.

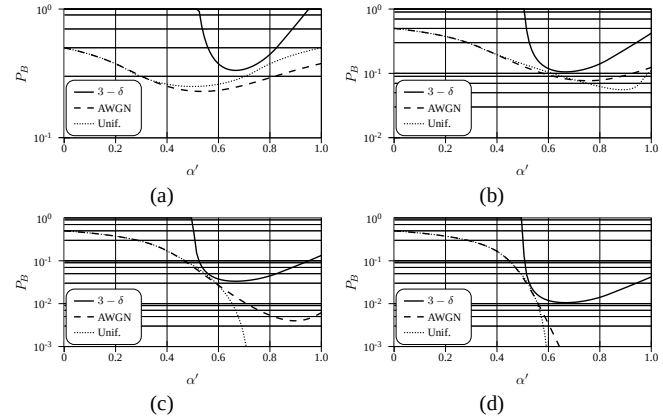


Fig. 17. Error probability comparison as a function of the distortion compensation parameter for the $3-\delta$, Gaussian and uniform attacks and binary signaling: (a) WNR $=$ 0dB, (b) WNR $=$ 5dB, (c) WNR $=$ 10dB and (d) WNR $=$ 15dB.

probability introduced by the WCAA, we analyzed (30) and (32). Surprisingly, it was found that, independently of the operational WNR, $\alpha' = \alpha'_{\text{opt}}$ guarantees the lowest error probability of the analyzed data-hiding techniques under the WCAA (Figure 18). Having this bound on the error probability, it is possible to guarantee reliable communications using proper error correction codes. Therefore, one can select such a fixed distortion compensation parameter $\alpha' = \alpha'_{\text{opt}}$ at the uninformed encoder and the MD decoder, which guarantees a bounded error probability. Substituting (31) into (30), one obtains the upper bound on the error probability:

$$P_B(\alpha'_{\text{opt}}) = \frac{1}{6}|\mathcal{M}|(|\mathcal{M}|-1)\xi^{-1}. \tag{33}$$

IV. Mutual information as a cost function

The analysis of the WCA with mutual information as a cost function is crucial for the fair evaluation 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.

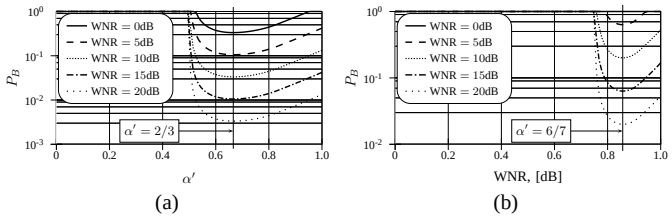


Fig. 18. Error probability analysis result as a function of the distortion compensation parameter α' for the $3-\delta$ attack: (a) binary signaling and (b) quaternary signaling.

P. Moulin *et al.* [1] considered the maximum achievable rate in the Gel'fand-Pinsker set-up (1) as a max-min problem:

$$C = \max_{\phi} \min_{f_{V|Y}(\cdot|\cdot)} \left[I(U;V) - I(U;X) \right], \tag{34}$$

for a blockwise memoryless attack, the embedder distortion constraint σ_W^2 and the attacker distortion constraint σ_Z^2.

In the case of practical quantization-based methods the mutual information is measured between the communicated message M and the channel output V [9]: $I_{\phi, f_{V|Y}(\cdot|\cdot)}(M;V)$, where the subscript means that the mutual information depends on both encoder design and attack pdf.

It was shown in [9] that modulo operation does not reduce the mutual information between V and M if the host is assumed to be flat within the quantization bins. Consequently:

$$I_{\phi, f_{V|Y}(\cdot|\cdot)}(M;V) = I_{\phi, f_{V|Y}(\cdot|\cdot)}(M;V'), \tag{35}$$

where $V' = Q_\Delta(V) - V$, and the above problem can be reformulated as:

$$\max_{\phi} \min_{f_{V|Y}(\cdot|\cdot)} I_{\phi, f_{V|Y}(\cdot|\cdot)}(M;V'). \tag{36}$$

Rewriting the inequalities (14)–(16) for the mutual information as a cost function, we have:

$$\max_{\phi} \min_{f_{V|Y}(\cdot|\cdot)} I_{\phi, f_{V|Y}(\cdot|\cdot)}(M;V') \leq \max_{\phi} I_{\phi, \tilde{f}_{V|Y}(\cdot|\cdot)}(M;V'),$$

with equality if, and only if, the fixed attack $\tilde{f}_{V|Y}(\cdot|\cdot)$ coincides with the WCAA. Thus, the decoder in Figure 3 is not fixed and we assume that the channel attack pdf $f_{V|Y}(\cdot|\cdot)$ is available at the decoder (informed decoder) and, consequently, ML decoding is performed. Under previous assumptions of quantization-based embedding and additive attack, it is possible to rewrite (36) as:

$$\max_{\alpha'} \min_{f_Z(\cdot)} I_{\alpha', f_Z(\cdot)}(M;V'). \tag{37}$$

As for the error probability analysis case, we address the problem of the WCAA and the optimal encoding strategy for the WCAA. It is known [20] that the mutual information can be expressed as a Kullback-Leibler distance (KLD):

$$\begin{aligned} I_{\alpha', f_Z(\cdot)}(M;V') &= D(f_{M,V'}(m,v')||f_{V'}(v')p_M(m)) \\ &= \int f_{M,V'}(m,v') \log_2 \frac{f_{V'|M}(v'|M=m)}{f_{V'}(v')} dv', \end{aligned} \tag{38}$$

where $f_{M,V'}(m,v')$ is the joint pdf of the input message and the modulo of the channel output, $p_M(m)$ denotes the marginal

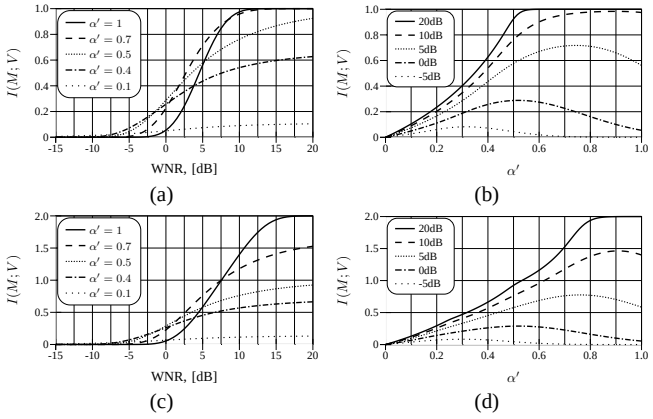


Fig. 19. Mutual information analysis result for the AWGN attack case and different α' and WNR values: (a) and (b) binary signaling and (c) and (d) quaternary signaling.

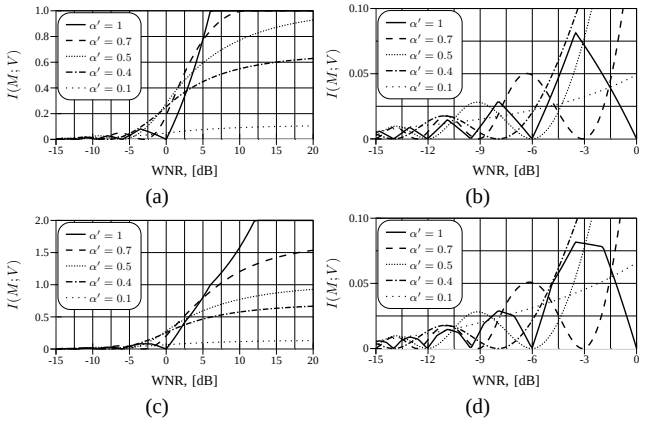


Fig. 20. Mutual information analysis result for the uniform noise attack case: (a) global performance analysis and (b) magnification of the low-WNR regime with binary signaling; (c) global performance analysis and (d) magnification of the low-WNR with quaternary signaling.

pdf of the input messages and $f_{V'}(v')$ the marginal pdf of the modulo of the channel output.

In fact, (38) 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 $|\mathcal{M}|$-ary signaling case, one obtains [9]:

$$I_{\alpha',f_Z(\cdot)}(M;V') = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} D\left(f_{V'|M}(v'|M=m)||f_{V'}(v')\right) = D\left(f_{V'|M}(v'|M=1)||f_{V'}(v')\right), \quad (39)$$

where $D\left(f_{V'|M}(v'|M=m)||f_{V'}(v')\right) = D\left(f_{V'|M}(v'|M=1)||f_{V'}(v')\right)$ since $f_{V'|M}(v'|M=1)$ and $f_{V'|M}(v'|M=m)$ are the same pdf shifted for all $m \in \mathcal{M}$ and $f_{V'}(v') = \frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|} f_{V'|M}(v'|M=m)$.

The next section is dedicated to the analysis of the DM and the DC-DM under the AWGN attack, the uniform noise attack and the WCAA.

A. 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 performed using numerical computations. The results of this analysis for the binary and quaternary cases are shown in Figure 19.

B. Uniform noise attack

It was shown [8] that the uniform noise attack is stronger than the AWGN attack for some WNRs when the error probability is used as a cost function. One of the properties of the KLD measure states that it is equal to zero if, and only if, the 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 (39). It can be demonstrated that $I(M;V') = 0$ when $\xi = \frac{\alpha'^2}{k^2}, k \in \mathbb{N}$ for the $|\mathcal{M}|$-ary signaling. This particular behaviour allows the attacker to achieve zero rate of communication with smaller attacking power than was predicted by the data-hider. the mutual information of quantization-based data-hiding techniques for the uniform noise attacking case with binary and quaternary signaling is depicted in Figure 20. 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 Figure 20(b) and Figure 20(d).

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. As a consequence, $\xi = \alpha'^2$ represents the WNR corresponding to zero rate communication, if the attacking variance satisfies $\sigma_Z^2 \geq \frac{D_w}{\alpha'^2}$.

C. The worst case additive attack

The problem of the WCAA using the mutual information as a cost function can be formulated using (37). Since the encoder must be fixed in advance as for the probability of error analysis case, we solve the max-min problem as a constrained minimization problem:

$$\min_{f_Z(\cdot)} I_{\alpha',f_Z(\cdot)}(M;V') = \min_{f_Z(\cdot)} D\left(f_{V'|M}(v'|M=1||f_{V'}(v')\right), \quad (40)$$

where $0 < \alpha' \leq 1$. The constraints in (40) are the same as with the error probability oriented analysis case (27). Unfortunately, this problem has no closed form solution and it was solved numerically.

The obtained results are presented for different α' values in Figure 21. In comparison with the AWGN and the uniform noise attacks, they demonstrate that the developed attack produces the maximum possible loss in terms of the mutual information for all WNRs (Figure 22).

In the analysis of the WCAA using the error probability 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. In Figure 23 the optimum α' values as a function of the WNR are presented for different input distributions in comparison with the optimum SCS parameter [15]. It demonstrates that SCS optimum distortion compensation parameter designed for the AWGN is also a good approximation for the WCAA case.

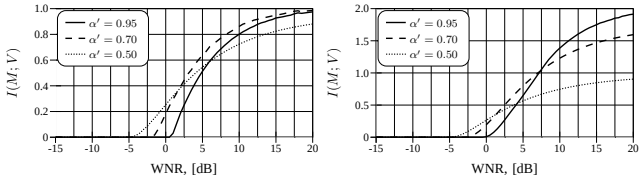


Fig. 21. Mutual information analysis result for the WCAA case: (a) binary signaling and (b) quaternary signaling.

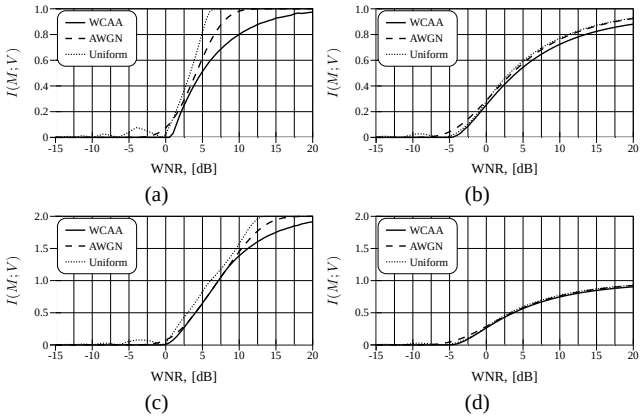


Fig. 22. Comparison of different attacks using mutual information as a cost function: (a) $\alpha' = 0.95$, binary signaling, (b) $\alpha' = 0.5$, binary signaling, (c) $\alpha' = 0.95$, quaternary signaling and (d) $\alpha' = 0.5$, quaternary signaling.

Recalling (37), we can conclude that it is not possible to find a unique optimum α' for the mutual information analysis case, contrarily to the error probability one when the decoder was fixed to the suboptimum MD decoder. The data-hider cannot select a value of the distortion compensation blindly, which guarantees reliable communications at any given rate.

It is possible to observe a saturation of the optimum value of α' in Figure 23 for small dimensionality and large WNR. Therefore, it is possible to select an optimum α' if the WNR range is known, located in the high-WNR regime and requirements of small dimensionality apply. For example, working in the high-WNR with WNR $>$ 5dB and $|\mathcal{M}| = 2$, optimum α' can be chosen as $\alpha' = 0.71$.

Using the optimum α' for each WNR, the resulting mutual information (40) is presented in Figure 24(a) for different cardinality of the input alphabet compared to the performance of the AWGN using the optimized $\alpha = \alpha_{opt}$ parameter [1].The obtained performance demonstrates that the developed WCAA is worser than the AWGN whenever the optimum distortion compensation parameter is selected.

The pdfs of the WCAA for different cardinality of the input alphabet and WNRs are presented in Figure 25. There is a strong impact of the optimization precision on the pdf shape although the mutual information value remains constant, and therefore high precision has been used to generate the presented results (optimization tolerance up to 10^{-12} was used). Nevertheless, the presented pdfs are not unique and different shapes might achieve the same performance.

Previous results [9] have already proven that the optimal WCAA pdf must be strictly inside the bin (and following the AWGN cannot be the WCAA). However, it is possible to

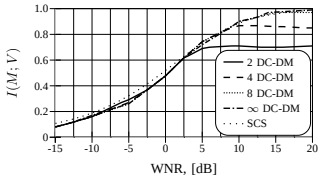


Fig. 23. Optimum distortion compensation parameter α' when the mutual information is selected as a cost function.

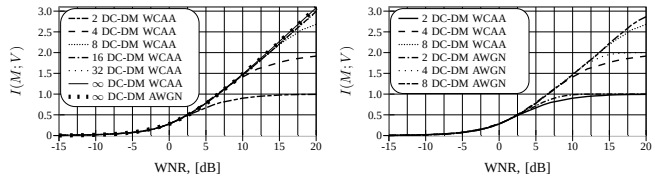


Fig. 24. Maximum achievable rate for different cardinality of the input alphabet under the WCAA compared to the AWGN (a) for $|\mathcal{M}| \to \infty$ and (b) for $|\mathcal{M}| < \infty$.

achieve nearly optimal solutions with larger support of the pdf. The periodical repetition of the constellations yields a similar effective attack whilst using a larger power. In the presented experiments, the bin width was chosen as $\Delta = 2$.

The support of the presented WCAA pdfs does not vary significantly. The optimum distortion compensation parameter α' increments with the WNR and so the power of the embedded signal while the self-noise support decrements. Thus, the support of the attack remains nearly the same for all WNRs. Larger variations can be observed for the high-WNR and high dimensionality, where the optimum α' variation is smaller.

It is possible to observe in Figures 24(a) and 25 that the impact of the WCAA is very similar to a truncated Gaussian and that the difference in terms of the mutual information is negligible. Although the AWGN is not the WCAA, its performance is an accurate and practical approximation to the WCAA in the asymptotic case when $|\mathcal{M}| \to \infty$. For $|\mathcal{M}| < \infty$, the difference might be important for some WNRs and it is needed to consider the real WCAA as it is presented in Figure 24(b).

V. Conclusions

In this paper we addressed the problem of the WCAA for the quantization-based data-hiding techniques from the probability of error and mutual information perspectives. The comparison between the analyzed cost functions demonstrated that in a rigid scenario with a fixed decoder, the attacker can decrease the rate of reliable communication more severely than by using either the AWGN or the uniform noise attacks. We showed that the AWGN attack is not the WCAA in general, and we obtained an accurate and practical analytical approximation to the WCAA, the so-called $3-\delta$ attack, when the cost function is the probability of error for the fixed MD decoder. For the $3-\delta$ attack, $\alpha' = \frac{2(|\mathcal{M}|-1)}{2|\mathcal{M}|-1}$ was found to be the optimal value for the MD decoder that allows to communicate with an upper bounded probability of error for a given WNR. This value could be fixed without prior knowledge of the attacking pdf.

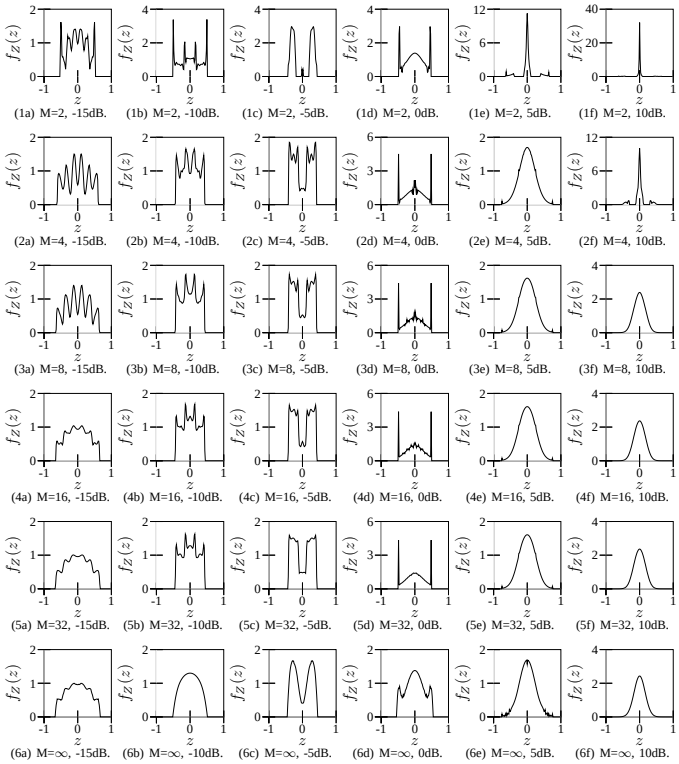


Fig. 25. Pdfs of the WCAA for different input distribution and WNRs.

The analysis results obtained by means of numerical optimization showed that there exists a worse attack than the AWGN when the mutual information was used as a cost function. Contrarily to the error probability analysis case, the optimal distortion compensation parameter (α') depends on the operational WNR for the mutual information analysis case. The particular behaviour of the mutual information under uniform noise attack was considered, achieving zero-rate communication for attacking variances σ_Z^2 such that $\sigma_Z^2 \geq \frac{D_w}{\alpha'^2}$. The presented results should serve as a basis for the development of fair benchmarks for various data-hiding technologies.

ACKNOWLEDGMENT

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] P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. on Information Theory*, vol. 49, no. 3, pp. 563–593, October 2003.

[2] A. Somekh-Baruch and N. Merhav, "On the error exponent and capacity games of private watermarking systems," *IEEE Trans. on Information Theory*, vol. 49, no. 3, pp. 537–562, March 2003.

[3] A. Somekh-Baruch and N. Merhav, "On the capacity game of public watermarking systems," *IEEE Trans. on Information Theory*, vol. 49, no. 3, pp. 511–524, March 2004.

[4] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[5] L. Peréz-Freire, F. Peréz-González, 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.

[6] M. Barni and F. Bartolini, *Watermarking Systems Engineering*. New York: Marcel Dekker, Inc., 2004.

[7] 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*, vol. 43, no. 4, pp. 1256–1264, July 1997.

[8] F. Peréz-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*, vol. 51, no. 4, April 2003.

[9] F. Pérez-González, "The importance of aliasing in structured quantization modulation data hiding," in *International Workshop on Digital Watermarking*, Seoul, Korea, 2003.

[10] J. E. Vila-Forcén, S. Voloshynovskiy, O. Koval, F. Peréz-González, and T. Pun, "Worst case additive attack against quantization-based watermarking techniques," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Siena, Italy, September 29 - October 1 2004.

[11] A. K. Goteti and P. Moulin, "Qim watermarking games," in *Proc. ICIP*, Oct. 2004.

[12] 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*, vol. 47, pp. 1423–1443, 2001.

[13] J. E. Vila-Forcén, S. Voloshynovskiy, O. Koval, F. Peréz-González, and T. Pun, "Worst case additive attack against quantization-based data-hiding methods," in *Proceedings of SPIE Photonics West, Electronic Imaging 2005, Security, Steganography, and Watermarking of Multimedia Contents VII (EI120)*, San Jose, USA, January 16-20 2005.

[14] R. Tzschoppe, R. Bäuml, R. Fischer, A. Kaup, and J. Huber, "Additive Non-Gaussian Attacks on the Scalar Costa Scheme (SCS)," in *Proceedings of SPIE Photonics West, Electronic Imaging 2005, Security, Steganography, and Watermarking of Multimedia Contents VII (EI120)*, vol. 5681, San Jose, USA, January 16-20 2005.

[15] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, April 2003.

[16] M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[17] U. Erez and R. Zamir, "Lattice decoding can achieve 0.5 log(1+snr) over the additive white gaussian noise channel using nested codes," in *Proceedings of IEEE International Symposium on Information Theory*, Washington DC, USA, June 2001, p. 125.

[18] R. A. Wannamaker, "The theory of dithered quantization," Ph.D. dissertation, University of Waterloo, Waterloo, Ontario, Canada, July 1997.

[19] O. Koval, S. Voloshynovskiy, F. Peréz-González, 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.

[20] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley and Sons, New York, 1991.