

ON OPTIMAL WATERMARKING SCHEMES IN UNCERTAIN GAUSSIAN CHANNELS

Alvaro A. Cárdenas*, George V. Moustakides[†] and John S. Baras[‡]

ABSTRACT

This paper describes the analytical derivation of a new watermarking algorithm satisfying optimality properties when the distortion of the watermarked signal is caused by a Gaussian process. We also extend previous work under the same assumptions and obtain more general solutions.

Index Terms— Watermarking, Gaussian Attacks, minimax optimization

I. INTRODUCTION

One of the biggest challenges the designer of a watermarking algorithm faces, is the fact that the degradations a watermarked signal is going to suffer, before reaching the detection algorithm, are highly uncertain. These degradations can be caused by several reasons, such as noisy channels, benign filtering, compression, or even, adversarial attacks.

In order to understand and characterize the performance guarantees of watermarking algorithms facing uncertain degradation channels, recent research has tried to model the problem with a robust detection approach. The main idea is to parameterize the channel with a given set of uncertain parameters \mathcal{A} , and then find the optimal embedding parameters \mathcal{E}^* against the least favorable uncertain channel model \mathcal{A}^* . A recent survey of the subject is presented in [2].

Although this research aims to design watermarking algorithms with provable performance guarantees against an uncertain channel \mathcal{A} , its major drawback is that to find the optimal \mathcal{E}^* and the least favorable \mathcal{A}^* , one has to solve a min-max optimization problem that is often intractable. Therefore, the vast majority of the literature has focused in understanding the parameters of well known watermarking algorithms, such

as, spread spectrum watermarking, or quantized index modulation. Very little work has been done in developing new watermarking algorithms that are efficient and have also analytically derived performance guarantees. In this paper we introduce a new robust watermarking algorithm, and prove how our algorithm outperforms spread spectrum watermarking against least favorable Gaussian channels.

II. PROBLEM DESCRIPTION

In the watermark verification problem, the watermark embedder \mathcal{E} receives as inputs a signal s and a bit m . The objective of the embedder is to produce a signal x with no major loss of information (or perceptual differences) from s , and carrying information about m . The general formulation of this property is to force the embedder to satisfy an average distortion constraint $\mathbb{E}[d(S, X)] \leq D_W$, where d is a distortion function, and D_w is an upper bound on the amount of distortion allowed for embedding.

The signal x is then transmitted through an uncertain channel \mathcal{A} . The output from \mathcal{A} is another signal y satisfying the constraint $\mathbb{E}[d(S, Y)] \leq D_a$.

When the detection algorithm \mathcal{D} receives the signal y , it has to determine if y was originally embedded with $m = 1$ or not.

To find a solution to this problem we need to make some assumptions about the signal source, the distortion functions, the metric of performance, and the degrading channel. In this paper we follow a model originally proposed by [1]. Contrary to the results in [1], we relax two assumptions. First, we relax the assumption of spread spectrum watermarking and instead search for the optimal watermarking algorithm in this model formulation. Second, we relax the assumption of diagonal processors (an assumption made because the embedding algorithm used spread spectrum watermarking) and therefore, obtain results for a more general case. In the following subsection we describe the model of the problem and obtain our results. Then in section III we discuss our results and compare them to [1].

A. Mathematical Model

Given $s \in \mathbb{R}^N$ and $m \in \{0, 1\}$, we assume an ad-

Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, 94720, USA. cardenas@eecs.berkeley.edu

Department of Electrical and Computer Engineering, University of Patras 26500 Rio, Greece. moustaki@ece.upatras.gr

Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, 20742, USA. baras@isr.umd.edu

Research supported by the U.S. Army Research Office under CIP URI grant No. DAAD19-01-1-0494 and by the Communications and Networks Consortium sponsored by the U.S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011.

ditive embedder \mathcal{E} that outputs $x = \Phi(s + pm)$, where Φ is an $N \times N$ matrix and where $p \in \mathbb{R}^N$ is a pattern sampled from a distribution with probability density function (pdf) $h(p)$.

The channel \mathcal{A} is modeled by $y = \Gamma x + e$, where Γ is an $N \times N$ matrix and e is a zero-mean (because any non-zero mean random vector is suboptimal [1]) Gaussian random vector with correlation matrix R_e .

The detection algorithm has to perform the following hypothesis test:

$$\begin{aligned} H_0 : y &= \Gamma\Phi s + e \\ H_1 : y &= \Gamma\Phi s + e + \Gamma\Phi p. \end{aligned}$$

We use the probability of error as objective function $\Psi(\mathcal{E}, \mathcal{A})$. We know that for this objective function, an optimal detection algorithm is the likelihood ratio test. Assuming s is a Gaussian random vector with zero mean (zero mean is assumed without loss of generality) and correlation matrix R_s , and that the priors for m are equally likely, the likelihood ratio test is:

$$p^t \Upsilon^t R_y^{-1} y - \frac{1}{2} p^t \Upsilon^t R_y^{-1} \Upsilon p \stackrel{H_1}{\underset{H_0}{\geq}} 0,$$

where $R_y = \Gamma\Phi R_s \Phi^t \Gamma^t + R_e$, and $\Upsilon = \Gamma\Phi$.

We assume that the distortion constraint that (both) \mathcal{E} and \mathcal{A} need to satisfy is the squared error distortion. The feasible design space is therefore composed of the set

$$\{\mathcal{E} : \mathbb{E}\|X - S\|^2 \leq ND_w\}$$

(where $\mathcal{E} = (\Phi, R_p, h)$) and the set

$$\{\mathcal{A} : \mathbb{E}\|Y - S\|^2 \leq ND_a\}$$

(where $\mathcal{A} = (\Gamma, R_e)$).

B. Optimal Embedding Distribution

Under the assumptions stated in the previous section, we find that the probability of error is

$$\mathbb{E}_p \left[\mathcal{Q} \left(\sqrt{p^t \Omega p} \right) \right] = \int \mathcal{Q} \left(\sqrt{p^t \Omega p} \right) h(p) dp,$$

where $\mathcal{Q}(x)$ is the tail probability of the normal distribution $\mathcal{N}(0, 1)$,

$$\Omega = \frac{1}{2} \Phi^t \Gamma^t (\Gamma\Phi R_s \Phi^t \Gamma^t + R_e)^{-1} \Gamma\Phi$$

and p is the random *embedded* pattern with pdf $h(p)$. Fixing Ω , we would like to determine the form of $h(p)$ that will minimize the error probability.

To solve this problem we rely on the following property of the $\mathcal{Q}(\cdot)$ function,

Lemma 1: The function $\mathcal{Q}(\sqrt{x})$ is a convex function of x . This lemma can be verified by direct differentiation.

Using lemma 1 and applying Jensen's inequality we obtain:

$$\mathbb{E}_x [\mathcal{Q}(\sqrt{x})] \geq \mathcal{Q} \left(\sqrt{\mathbb{E}_x[x]} \right)$$

we have equality iff x is a constant with probability 1 (wp1). Applying this Lemma to our problem we obtain:

$$\begin{aligned} \mathbb{E}_p \left[\mathcal{Q} \left(\sqrt{p^t \Omega p} \right) \right] &\geq \\ \mathcal{Q} \left(\sqrt{\mathbb{E}_p[p^t \Omega p]} \right) &= \mathcal{Q} \left(\sqrt{\text{tr}\{\Omega R_p\}} \right). \end{aligned} \quad (1)$$

Eq. (1) provides a *lower bound* on the error probability for *any* pdf satisfying the covariance constraint $\mathbb{E}[pp^t] = R_p$. We have equality in Eq. (1) iff wp1

$$p^t \Omega p = \text{tr}\{\Omega R_p\}. \quad (2)$$

In other words *every realization* of p must satisfy this equality. Notice that if we find a pdf for p satisfying Eq. (2) *under the constraint* $\mathbb{E}[pp^t] = R_p$, then we attain the lower bound in Eq. (1).

To find a random vector p to achieves these properties, we must do the following. Consider the SVD of the matrix

$$R_p^{1/2} \Omega R_p^{1/2} = U \Sigma U^t \quad (3)$$

where U is orthonormal and $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_K\}$ is diagonal with nonnegative elements. The nonnegativity of σ_i is assured because the matrix is nonnegative definite. Let A be a random vector with i.i.d. elements that take the values ± 1 with probability 0.5. For every vector A we can then define an embedding vector p as

$$p = R_p^{1/2} U A. \quad (4)$$

We now show that this definition satisfies our requirements. First, consider the covariance matrix (which must be equal to R_p). Indeed we have

$$\begin{aligned} \mathbb{E}[pp^t] &= R_p^{1/2} U \mathbb{E}[AA^t] U^t R_p^{1/2} = \\ &R_p^{1/2} U \mathbf{I} U^t R_p^{1/2} = R_p^{1/2} \mathbf{I} R_p^{1/2} = R_p \end{aligned} \quad (5)$$

where we used the independence of the elements of the vector A and the orthonormality of U (\mathbf{I} denotes the identity matrix). Therefore, our random vector has the correct covariance structure. We now show it also satisfies the constraint $p^t \Omega p = \text{tr}\{\Omega R_p\}$ wp1. Indeed for every realization of the random vector A :

$$\begin{aligned} p^t \Omega p &= A^t U^t R_p^{1/2} \Omega R_p^{1/2} U A = A^t \Sigma A \\ &= A_1^2 \sigma_1 + A_2^2 \sigma_2 + \dots + A_K^2 \sigma_K \\ &= \sigma_1 + \sigma_2 + \dots + \sigma_K, \end{aligned}$$

where we use the fact that the elements A_i of A are equal to ± 1 . Notice also that

$$\begin{aligned} \text{tr}\{\Omega R_p\} &= \text{tr}\{R_p^{1/2} \Omega R_p^{1/2}\} = \text{tr}\{U \Sigma U^t\} \\ &= \text{tr}\{\Sigma U^t U\} = \text{tr}\{\Sigma\} = \sigma_1 + \dots + \sigma_K. \end{aligned}$$

This proves the desired equality. We conclude that, although p is a random vector, *all* its realizations satisfy the equality

$$p^t \Omega p = \text{tr}\{\Omega R_p\}.$$

We have thus found the embedding distribution h^* that attains the lower bound in Eq. (1). It is a random mixture of the columns of the matrix $R_p^{1/2} U$ of the form $R_p^{1/2} U A$, where A is a vector with elements ± 1 , R_p is the covariance matrix (which we find in the next section), and U contains the singular vectors of the SVD of $R_p^{1/2} \Omega R_p^{1/2}$. This of course suggests that we can have 2^N different patterns.

C. Least Favorable Channel Parameters

With h^* , the game the embedder and the channel play is:

$$\max_{R_p, \Phi} \min_{R_e, \Gamma} \text{tr}\{(\Gamma \Phi R_s \Phi^t \Gamma^t + R_e)^{-1} \Gamma \Phi R_p \Phi^t \Gamma^t\} \quad (6)$$

Subject to the distortion constraints:

$$\begin{aligned} \text{tr}\{(\Phi - I) R_s (\Phi - I)^t + \Phi R_p \Phi^t\} &\leq N D_w \\ \text{tr}\{(\Gamma \Phi - I) R_s (\Gamma \Phi - I)^t + \Gamma \Phi R_p \Phi^t \Gamma^t + R_e\} &\leq N D_a \end{aligned}$$

Assuming $\Upsilon = \Gamma \Phi$ is fixed, we start by minimizing Eq. (6) with respect to R_e . This minimization problem is addressed with the use of variational techniques. We obtain $R_e^* = \frac{1}{\sqrt{\mu}} (\Upsilon R_p \Upsilon^t)^{1/2} - \Upsilon R_s \Upsilon$, where μ is the solution to:

$$\frac{1}{\sqrt{\mu}} = \frac{2 \text{tr}\{\Upsilon R_s\} - \text{tr}\{R_s\} - \text{tr}\{\Upsilon R_p \Upsilon^t\} + N D_a}{\text{tr}\{(\Upsilon R_p \Upsilon^t)^{1/2}\}}.$$

For the next step, we need to minimize

$$\frac{(\text{tr}\{(\Upsilon R_p \Upsilon^t)^{1/2}\})^2}{2 \text{tr}\{\Upsilon R_s\} - \text{tr}\{R_s\} - \text{tr}\{\Upsilon R_p \Upsilon^t\} + N D_a}$$

over Υ . Proceeding similarly to the previous case we obtain $\Upsilon^* = \Sigma R_p^{-1/2}$, where Σ is the solution to the following nonlinear equation:

$$(A - \Sigma^t) \Sigma (A - \Sigma^t) - c^2 \Sigma^t = 0$$

where $A = R_p^{-1/2} R_s$ and c is a constant determined by the variation.

D. Optimal Embedding Parameters

The objective of the embedding algorithm is:

$$\max_{\Phi, R_p} \frac{(\text{tr}\{(\Phi R_p \Phi^t)^{1/2}\})^2}{2 \text{tr}\{\Phi R_s\} - \text{tr}\{R_s\} - \text{tr}\{\Phi R_p \Phi^t\} + N D_a}$$

Subject to:

$$\text{tr}\{(\Phi - I) R_s (\Phi - I)^t + \Phi R_p \Phi^t\} \leq N D_w. \quad (7)$$

For any Φ and any R_p we have by Schwarz inequality that the optimal value is achieved if and only if $R_p^* = \kappa (\Phi^t \Phi)^{-1}$, where κ is

$$\kappa = \frac{N D_w - \text{tr}\{(\Phi - I) R_s (\Phi - I)^t\}}{N}.$$

Replacing the values of κ and R_p into the original objective function and letting λ be the maximum value the objective function can achieve (as a function of Φ), we conclude (after some algebraic manipulations) that $\Phi^* = (\lambda + 1)^{-1} I$, where λ is the solution to:

$$\frac{\lambda}{\lambda + 1} \text{tr}\{R_s\} + \lambda (D_a - D_w) = N D_w.$$

III. DISCUSSION

So far we have solved

$$\min_{\Phi, R_p} \max_{\Gamma, R_e} \min_h \Psi(\mathcal{E}, \mathcal{A}) \quad (8)$$

where the embedding parameters are $\mathcal{E} = (h, R_p, \Phi)$ and the uncertain parameters are $\mathcal{A} = (\Gamma, R_e)$. Our solution implies that given the optimal Φ^* and R_p^* , we can find the least favorable parameters Γ^* and R_e^* , and given these, we can find the embedding distribution h^* minimizing the probability of error.

The problem with this solution is that we have not shown that Γ^* and R_e^* are the least favorable parameters for h^* . Without loss of generality assume Φ and R_p are fixed, so we can replace \mathcal{E} with h for the following arguments. Furthermore let $h(\mathcal{A})$ denote the embedding distribution as a function of the parameters $\mathcal{A} = (\Gamma, R_e)$ (recall that h depends on \mathcal{A} by the selection of U in Eq. (4)). The problem we have solved is thus:

$$\forall h \forall \mathcal{A} \quad \Psi(h^*(\mathcal{A}), \mathcal{A}) \leq \Psi(h(\mathcal{A}), \mathcal{A})$$

This is true in particular for \mathcal{A}^* , the solution to the full optimization problem from Eq. (8). Moreover, the above is also true for the distribution used in the previous work [1], which assumed a Gaussian embedding distribution h^G :

$$\forall \mathcal{A} \quad \Psi(h^*(\mathcal{A}), \mathcal{A}) \leq \Psi(h^G, \mathcal{A})$$

Notice also that in [1], the solution obtained was

$$\mathcal{A}^G = \arg \max_{\mathcal{A}} \Psi(h^G, \mathcal{A}) \quad (9)$$

Due to some approximations done in [1], $\Psi(h^G, \mathcal{A})$ turns out to be the same objective function given in Eq. (6). Furthermore in [1] there were further approximations in order to obtain linear processors (diagonal matrices). In this work we relaxed this assumption in order to obtain the full solution to Eq. (6). Therefore the general solution (without extra assumptions such as

diagonal matrices) in both cases is the same; i.e., if we denote our solution by:

$$\mathcal{A}^* = \arg \max_{\mathcal{A}} \left\{ \min_h \Psi(h, \mathcal{A}) \right\}$$

then $\mathcal{A}^* = \mathcal{A}^G$. Furthermore,

$$\max_{\mathcal{A}} \Psi(h^*(\mathcal{A}), \mathcal{A}) < \max_{\mathcal{A}} \Psi(h^G, \mathcal{A})$$

However, one of the problems with these solutions is that there might exist \mathcal{A}' such that

$$\Psi(h^*(\mathcal{A}^*), \mathcal{A}^*) < \Psi(h^*(\mathcal{A}'), \mathcal{A}').$$

The reason for this attack, is that the embedding distribution h^* needs to know the channel parameters \mathcal{A} . To prevent these attacks from happening, we can address two problems. With regards to previous work in [1], we want to know if:

$$\forall \mathcal{A} \Psi(h^*(\mathcal{A}^*), \mathcal{A}) \leq \Psi(h^G, \mathcal{A}^*) \quad (10)$$

that is, once we have fixed the operating point \mathcal{A}^* (the least favorable parameters according to Eq. (6)) there are no other channel parameters that will make h^* perform worse than the previous work.

The second problem is in fact more general; it relates to the original intention of minimizing the worst possible error: we want to find h and \mathcal{A} in the following order:

$$\min_h \max_{\mathcal{A}} \Psi(h, \mathcal{A}). \quad (11)$$

A way to show that (h^*, \mathcal{A}^*) satisfies Eq. (11)–and therefore also satisfies Eq. (10)– is to show that the pair (h^*, \mathcal{A}^*) forms a saddle point equilibrium:

$$\forall (h, \mathcal{A}) \quad \Psi(h^*, \mathcal{A}) \leq \Psi(h^*, \mathcal{A}^*) \leq \Psi(h, \mathcal{A}^*). \quad (12)$$

Let \mathcal{E} denote again the triple (h, R_p, Φ) . We are interested in showing that

$$\forall (\mathcal{E}, \mathcal{A}) \quad \Psi(\mathcal{E}^*, \mathcal{A}) \leq \Psi(\mathcal{E}^*, \mathcal{A}^*) \leq \Psi(\mathcal{E}, \mathcal{A}^*) \quad (13)$$

where $(\mathcal{E}^*, \mathcal{A}^*) = (h^*, R_p^*, \Phi^*, R_e^*, \Gamma^*)$ is the solution to Eq. (8).

It is easy to show how the right hand side inequality of Eq. (13) is satisfied, since $\Psi(\mathcal{E}, \mathcal{A}^*)$ equals:

$$\begin{aligned} & \mathbb{E}_p \left[\mathcal{Q} \left(\sqrt{p^t \frac{1}{2} \Phi^t \Gamma^{*t} (\Gamma^* \Phi R_s \Phi^t \Gamma^{*t} + R_e^*)^{-1} \Gamma^* \Phi p} \right) \right] \\ & \geq \mathcal{Q} \left(\sqrt{R_p \frac{1}{2} \Phi^t \Gamma^{*t} (\Gamma^* \Phi R_s \Phi^t \Gamma^{*t} + R_e^*)^{-1} \Gamma^* \Phi} \right) \\ & \geq \mathcal{Q} \left(\sqrt{R_p^* \frac{1}{2} \Phi^{*t} \Gamma^{*t} (\Gamma^* \Phi^* R_s \Phi^{*t} \Gamma^{*t} + R_e^*)^{-1} \Gamma^* \Phi^*} \right) \\ & = \Psi(\mathcal{E}^*, \mathcal{A}^*) \quad \text{by the definition of } h^* \end{aligned}$$

The left hand side of Eq. (13) is more difficult to prove. A particular case where it is satisfied is the *scalar* case, i.e., when $N = 1$. In this case we have the following:

$$\begin{aligned} \Psi(\mathcal{E}^*, \mathcal{A}^*) &= \mathcal{Q} \left(\sqrt{\frac{R_p^* (\Phi^* \Gamma^*)^2}{2((\Gamma^* \Phi^*)^2 R_s + R_e^*)}} \right) \\ &\geq \mathcal{Q} \left(\sqrt{\frac{R_p^* (\Phi^* \Gamma)^2}{2((\Gamma \Phi^*)^2 R_s + R_e)}} \right) \quad \text{by Eq. (6)} \\ &= \mathbb{E}_p \left[\mathcal{Q} \left(\sqrt{\frac{p^2 (\Phi^* \Gamma)^2}{2((\Gamma \Phi^*)^2 R_s + R_e)}} \right) \right] \\ &= \Psi(\mathcal{E}^*, \mathcal{A}^*) \end{aligned}$$

Where we are able to take the expected value outside \mathcal{Q} because p is independent of \mathcal{A} . The independence of p in the scalar case comes from the fact that Eq. (4) yields in this case $p = \sqrt{R_p}$ with probability $\frac{1}{2}$ and $p = -\sqrt{R_p}$ with probability $\frac{1}{2}$. With this distribution Eq. (2) is always satisfied.

This result can be seen as a counterexample against the optimality of spread spectrum watermarking in Gaussian channels: if the channel has Gaussian noise, then the embedding distribution should not be a spread spectrum watermarking, or conversely, if the embedding distribution is spread spectrum, then the least favorable distribution is not Gaussian.

IV. CONCLUSIONS AND FUTURE WORK

We have introduced a new watermarking scheme with provable performance guarantees. Our work improves on previous research done under the same assumptions and gives a counterexample regarding the optimality of spread spectrum for Gaussian channels.

In future work we plan to investigate the practical and theoretical extensions to our work. We plan to investigate the resiliency and efficiency of the algorithm empirically. We also want to investigate under which more general conditions is the left inequality in Eq. (13) satisfied, and also, whether Eq. (10) is true in general. We also plan to extend our work to other evaluation metrics (such as the case when one of the errors is more important than the other).

V. REFERENCES

- [1] Pierre Moulin and Aleksandar Ivanović, “The zero-rate spread-spectrum watermarking game,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1098–1117, April 2003.
- [2] Pierre Moulin and Ralf Koetter, “Data-hiding codes,” *Proceedings of the IEEE*, vol. 93, no. 12, pp. 2083–2126, December 2005.