Informed secure watermarking using optimal transport

Patrick Bas

INFORMED SECURE WATERMARKING USING OPTIMAL TRANSPORT

Patrick Bas

CNRS - Lagis , Ecole Centrale de Lille,
Avenue Paul Langevin BP 48, 59651 Villeneuve d'Ascq, France
Patrick.Bas@ec-lille.fr

ABSTRACT

This paper presents several watermarking methods preventing the estimation of the secret key by an adversary. The constraints for secure embedding using distribution matching, where the decoding regions rely implicitly on the distribution of the host signal, are first formulated. In order to perform informed coding, different decoding regions are associated with the same message using an appropriate partitioning function. The minimization of the embedding distortion is afterwards casted into an optimal transport problem. Three new secure embeddings are presented and the performances of the proposed embedding functions regarding the AWGN channel for different WCRs are evaluated. Depending on the embedding and noise distortions, informed secure coding can outperform classical secure coding or classical insecure coding such as ISS or SCS.

Index Terms— Security - Secure embedding - Informed coding - Optimal transport

1. INTRODUCTION

The robustness of a data-hiding scheme denotes its ability to transmit a message whenever the content undergoes various processes such as noise addition, format conversion, resynchronization, compression, ... On the contrary, the security of a data-hiding scheme denotes its ability to face an adversary who can build his attack using different materials such as the knowledge of the data-hiding scheme (Kerckhoffs' principle [1]), a pool of watermarked contents, or the watermark detector (Oracle attacks).

In the case of symmetric data-hiding, one objective of the adversary is to estimate the secret key used by the embedding and detection scheme. Once the decoding regions are estimated, the adversary can tamper or remove the embedded message [2]. Moreover, if the mapping function between the possible messages and the decoding regions is known by the adversary, he can also copy the message into another content or modify it at will.

The sequel of this paper addresses the Watermarked content Only Attack (WOA) setting[3], where the adversary uses a pool of content watermarked with the same key to forge its attack. In this framework, secure embedding schemes can be designed in order to prevent the adversary to estimate the decoding regions [4]. One example of such secure embedding scheme is called Natural Watermarking [4] : the embedding is performed in such a way that the distributions of original and watermarked contents are the same, consequently the adversary is unable to estimate the location of the decoding regions. In this paper we propose new embedding schemes relying on informed coding in order to increase the robustness while keeping a minimum distortion.

1.1. Notations

The vector $\mathbf{x}$ denotes a d-dimensional (random) host vector (all vectors considered in the paper are d-dimensional). We assume that $\mathbf{x}$ is composed of samples x_i of variance σ_x^2, independent and identically distributed (iid) according to a distribution function $p_X(x)$.

The embedding of a message $\mathbf{m}$ is performed by applying an embedding function $f(.)$ on the host vector $\mathbf{x}$ to generate $\mathbf{y} = f(\mathbf{x}, \mathbf{m}, K)$ (K denoting the secret key). The watermark signal $\mathbf{w}$ is given by $\mathbf{w} = \mathbf{y} - \mathbf{x}$ and the variance of its samples is denoted σ_w^2. WCR and WNR denote respectively the Watermark to Content Ratio and the Watermark to Noise Ratio and are expressed in dB.

2. SECURE EMBEDDING BY DISTRIBUTION MATCHING

The goal of this section is to present secure embedding functions $f(\mathbf{x}, \mathbf{m}, K)$, such that for all secret keys K :

$$p_X = p_{Y|K}. \tag{1}$$

These classes of embedding functions enable to obtain *stego-security* [4]. The name *stego-security* comes from the fact that this definition is similar to Cachin's definition of perfect secrecy in steganography [5].
We assume that the data-hiding scheme embeds a binary message with equal probability. In order to achieve *stego-security*, the density functions for each bit have to satisfy :

$$p_{Y|K} = (p_{Y|K,m=0} + p_{Y|K,m=1})/2. \tag{2}$$

One way to fulfill constraint (1) is to choose a partitioning function $g : \mathbb{R}^d \rightarrow \{0;1\}$ such that $\int_{\mathbb{R}} g(\mathbf{x})p_X(\mathbf{x}) = 1/2$. $p_{Y|K,m=0}$ and $p_{Y|K,m=1}$ are consequently given by :

$$p_{Y|K,m=0}(\mathbf{x}) = 2g(\mathbf{x})p_X(\mathbf{x}), \tag{3}$$

and

$$p_{Y|K,m=1}(\mathbf{x}) = 2(1-g(\mathbf{x}))p_X(\mathbf{x}). \tag{4}$$

The embedding function can now be considered as a set of two mapping functions $f_0(\mathbf{x}, m = 0, K)$ and $f_1(\mathbf{x}, m = 1, K)$ satisfying respectively (3) and (4). The message decoding is performed using the partitioning function $g()$:

$$g(\mathbf{z}) = 1 \Rightarrow \hat{m} = 0\,;\; g(\mathbf{z}) = 0 \Rightarrow \hat{m} = 1. \tag{5}$$

One straightforward solution to find mappings respecting constraint (1) is to choose $g()$ and generate random variables of pdf $p_{Y|K,m=0}$ and $p_{Y|K,m=1}$ (similar implementations have been proposed in [6]in the steganography context). However this solution doesn't take into account the distortion constraint $D = \sigma_w^2$. The goal of the next section is to find mappings respecting (1) and (2) while minimizing the average embedding distortion. Next section explains how to find such an optimal mapping.

3. DISTORTION MINIMIZATION USING OPTIMAL TRANSPORT

Optimal transportation, also called optimal coupling, has been defined by Monge in 1781 and consists in finding the transport from one density function p_1 to another p_2 that minimizes a cost function $c(\mathbf{x}_1, \mathbf{x}_2)$ representing the average transport [7]. Literally it is equivalent to the Monge's optimal transportation problem :

$$\text{Minimize} \qquad \int_X c(\mathbf{x}, f(\mathbf{x}))dp_1(\mathbf{x}) \tag{6}$$

over all the mapping functions f such that their transport on p_1 equals p_2.

If we consider the data-hiding framework we have $p_1 = p_X$ and $p_2 = p_{Y|K,m}$. Moreover we want to minimize the embedding distortion usually formulated as a quadratic cost : $c(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||^2/d$. We can afterwards use the results of transportation theory in order to find the embedding functions $f_0(\mathbf{x}, m = 0, K)$ and $f_1(\mathbf{x}, m = 1, K)$ as the optimal mappings.

For $d = 1$, given F_X the cumulative distribution function (cdf) of X, $F_{Y|K,m}$ the cdf of $Y|K, m$ and $F^{-1}_{Y|K,m}$ the quantile function, one result of optimal transport f_{opt} is that for semi-continuous cdf, the optimal mapping for the quadratic cost can be calculated as

$$f_{opt}(x, m) = F^{-1}_{Y|K,m} \circ F_X(x). \tag{7}$$

The minimum distortion is given by

$$D_{\min} = \int_0^1 c(F^{-1}_{Y|K,m}(x), F^{-1}_X(x))dx. \tag{8}$$

Note that for $d > 1$, closed-form solutions provided by optimal transportation exist for mappings such as projections, radial functions or densities of independent variables [8].

4. SECURE EMBEDDING FOR GAUSSIAN HOSTS

We assume in the sequel that the host components are iid Gaussian and that $d = 1$. This hypothesis can be practically verified since x can be the result of a projection of a vector (the wavelet coefficients of an image for example [9]) on a pseudo random carrier generated using the secret key K.

In the sequel, we use (7) to compute the optimal mapping, it is based on $F_X(x) = 0.5(1 + \text{erf}(\text{x}/\sqrt{2\sigma_\text{x}^2}))$. $F_{Y|K,m}(x)$ is computed for different partitioning functions, and message decoding is performed using Eq. 5.Without loss of generality, we assume $\sigma_x^2 = 1$. Three partitioning functions and associated embedding schemes are presented.

4.1. p-NW embedding

In this case the real axis is splitted into $2M$ decoding areas of same probability $p = 1/2M$. This can be achieved using as partitioning function :

$$\begin{aligned} g(x) &= 1 \quad \text{if} \quad u(k) < x < u(k+0.5), \\ g(x) &= 0 \quad \text{if} \quad u(k+0.5) < x < u(k+1). \end{aligned} \tag{9}$$

where $u(k) = F_X^{-1}\left(\frac{k}{M}\right)$. The optimal mapping is then :

$$f(x,0) = F_X^{-1}\left(\frac{F_X(x)}{2} + \frac{\lfloor Mx \rfloor}{2M}\right), \tag{10}$$

$$f(x,1) = F_X^{-1}\left(\frac{F_X(x)}{2} + \frac{\lfloor Mx \rfloor + 1}{2M}\right). \tag{11}$$

where $\lfloor x \rfloor$ denotes the floor function.

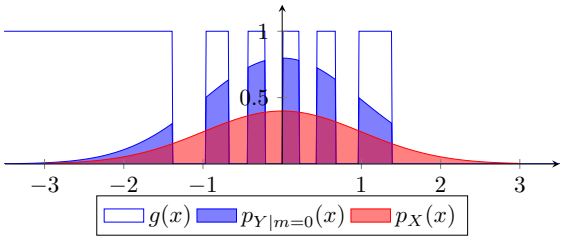
Classical (non-informed) coding is equivalent to 2 decoding regions (one for 1, the other for 0) and $p = 0.5$. Transportation Natural Watermarking or TNW[10] becomes consequently a special case of p-NW.

The partitioning functions $f()$, p_X, $p_{Y|m=0}$ and the optimal mapping are depicted on Fig. 1.

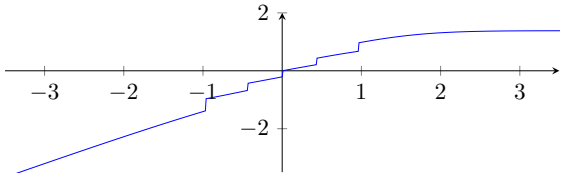
4.2. $\bar{p}$-NW embedding

This mapping can be seen has the symmetrised version of the previous one : the decoding regions are symmetrised according to the x-axis. This enables to have a decoding area around 0 which is twice the size of the decoding area for p-NW embedding. Now the partitioning function is :

$$\begin{aligned} g(x) &= 1 \quad \text{if} \quad u(k) < |x| < u(k+0.5), \\ g(x) &= 0 \quad \text{if} \quad u(k+0.5) < |x| < u(k+1). \end{aligned} \tag{12}$$



(a) : partitioning function and associated pdfs



(b) Optimal mapping $f(x,0)$

Fig. 1. p-NW for $M=6$ and $\sigma_x^2=1$.

And the mapping consists here in using (10) if $x>0$ (resp. $x<0$) and $m=0$ (resp. $m=1$), or using (11) if $x>0$ (resp. $x<0$) and $m=1$ (resp. $m=0$).

4.3. Δ-NW embedding

In order to transpose the Scalar Costa Scheme (SCS) [11] in the secure embedding context, we now choose $g()$ such that each decoded region has the same width Δ. Due to the symmetry of the Gaussian distribution, this is possible if

$$g(x)=\sum_k \Pi_\Delta(x-\frac{\Delta}{2}+2k\Delta), \tag{13}$$

where $\Pi_\Delta(x)$ is the centered rectangular window function of width Δ. Using (7), the optimal mapping can then be expressed as :

$$\begin{aligned} f(x,0) &= F_X^{-1}\left(\frac{F_X(x)-v_{2i}}{2}\right), \\ f(x,1) &= F_X^{-1}\left(\frac{F_X(x)-w_{2i}}{2}\right). \end{aligned} \tag{14}$$

Here $v_{2i}<F_X(x)<v_{2i}+2F_X((2i+1)\Delta)$ for $m=0$ and $w_{2i}<F_X(x)<w_{2i}+2F_X(2i\Delta)$ for $m=1$ with

$$v_{2i}=2\sum_{k=-\infty}^{i}\left[F_X((2k-1)\Delta)-F_X(2k\Delta)\right],$$

and

$$w_{2i}=2\sum_{k=-\infty}^{i}\left[F_X((2k-2)\Delta)-F_X((2k-1)\Delta)\right].$$

Note that contrary to p-NW and $\bar{p}$-NW where the embedding distortion cannot be chosen *a-priori* because they rely on the integer M, Δ-NW offers a continuous range of embedding distortions which are function of the scalar Δ.

5. PERFORMANCE COMPARISON

The goal of this section is to compare to different secure embedding schemes for the AWGN channel. In order to provide a fair comparison, each scheme has to be evaluated for the same embedding distortion. For comparison purposes, we also compute the Bit Error Rate (BER) for two robust but insecure embedding schemes : Improved Spread Spectrum (ISS) [12] and the Scalar Costa Scheme (SCS) [11].

Fig. 2 compares the robustness of two implementations of Natural Watermarking : Transportation Natural Watermarking [10] uses optimal mapping but no informed coding, in comparison Δ-NW uses optimal mapping and informed coding. For $WNR\geq -9dB$ the use of informed coding enables to increase the robustness of the scheme and the gap between the two implementations regarding the WNR is above $3dB$. Note that in this case the embedding distortion is fixed for TNW because it depends only of σ_x^2 .

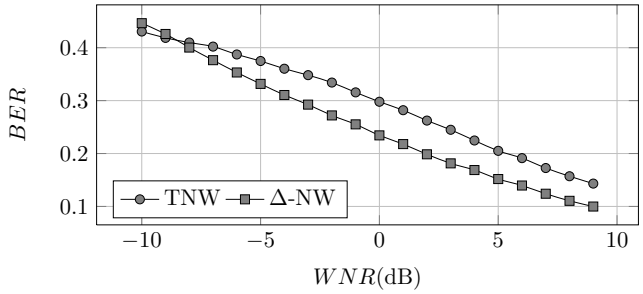


Fig. 2. BER of TNW [10] and Δ-NW, $WCR=-0.7dB$.

Fig. 3 and 4 present the performances of the different embeddings respectively for $WCR=-5dB$ and $WCR=-11dB$. For p-NW and $\bar{p}$-NW these distortions are respectively equivalent to $M=2$ and $M=6$.

The performances of the secure embeddings differ according to the distortion but general remarks can be drawn. For $WCR=-11dB$ and $WNR>-5dB$, the Δ-NW embedding is the secure scheme that provides the best performance and has a BER close to SCS for $WNR=0dB$. Such behavior can be explained by the facts that for $WNR=0dB$ and low $WCRs$, (i) the distributions inside each decoding regions can be approximated as uniform and (ii) the embedding parameter for SCS is $\alpha=0.52$ which is very close to the secure embedding parameter of SCS for uniform hosts ($\alpha_{sec}=0.5$) [13]. For such regimes SCS and Δ-NW embedding are consequently very similar in term of robustness and security.

For $WCR=-5dB$ the decoding regions of Δ-NW are too far for each other and the scheme provides poor performance in comparison with p-NW (for large WNR) and $\bar{p}$-NW (for small WNR). Note also that for small WNR, the proposed secure-embedding schemes can outperform the performance of the ISS (for both $WCRs$) and SCS (for $WCR=-5dB$).

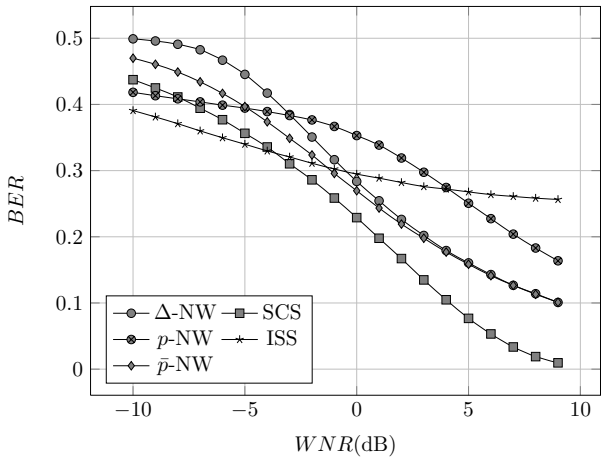


Fig. 3. Comparison between secure and insecure embeddings, $WCR = -5dB$.

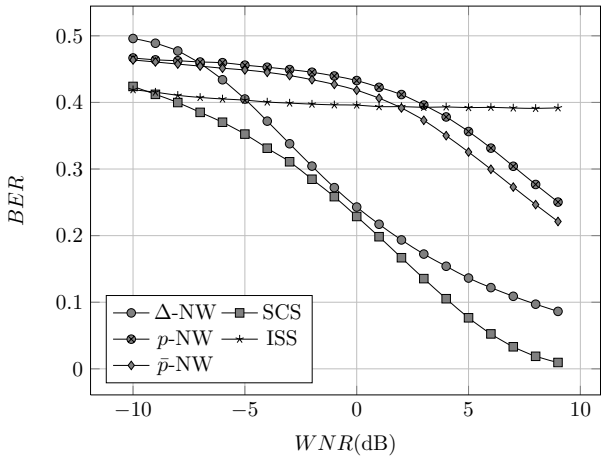


Fig. 4. Comparison between secure and insecure embeddings, $WCR = -11dB$.

6. CONCLUSION

We have proposed different secure embedding functions using informed coding. This is combined with the transportation theory framework which enables to find a mapping minimizing the distortion after the embedding while taking the security constraint into account. We have noticed that the use of infomed coding enables to increase the robustness of the secure scheme and can also compete in some cases with insecure schemes such as ISS or SCS. Our future works will focus on the design on more general partitioning functions and on the computation of the capacity of secure embedding schemes.

7. REFERENCES

[1] A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. IX, 1883.

[2] L. Pérez-Freire and F. Pérez-González, "Spread spectrum watermarking security," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 2–24, Marsh 2009.

[3] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security : theory and practice," *IEEE Trans. Signal Processing*, vol. 53, no. 10, oct 2005.

[4] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for WOA data-hiding," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, March 2008.

[5] C. Cachin, "An information-theoretic model for steganography," in *Information Hiding : Second International Workshop IHW'98*, Portland, Oregon, USA, April 1998.

[6] Y. Wang and P. Moulin, "Steganalysis of block-structured stegotext," *Security, Steganography, and Watermarking Multimedia Contents VI*, vol. 5306, pp. 477–488.

[7] C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, 2003.

[8] J. A. Cuesta-Albertos, L. Ruschendorf, and A. Tuerodiaz, "Optimal coupling of multivariate distributions and stochastic processes," *Journal of Multivariate Analysis*, vol. 46, no. 2, pp. 335–361, August 1993.

[9] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Comparison of secure spread-spectrum modulations applied to still image watermarking," *Annals of Telecommunication*, vol. 64, no. 11-12, pp. 801–813, Dec. 2009.

[10] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Optimization of natural watermarking using transportation theory," in *MM&Sec'09 : Proceedings of the 11th ACM workshop on Multimedia and security*, New York, NY, USA, 2009, pp. 33–38, ACM.

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

[12] H. S. Malvar and D. A. F. Florêncio, "Improved Spread Spectrum : a new modulation technique for robust watermarking," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, April 2003.

[13] L. Pérez-Freire, F. Pérez-González, Teddy Furon, and P. Comesaña, "Security of lattice-based data hiding against the Known Message Attack," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 421–439, December 2006.