Using of Polar Codes in Steganography

Idy Diop ,Birahime Diouf, Sidi Mohamed Farssi, K.Tall,P.A.Fall, A.K.Diop, K.Sylla
Department of Computer Science, ESP/UCAD, Dakar, Senegal
dioufbira11@yahoo.fr, idydiop@yahoo.fr, farsism@yahoo.com.

***Abstract*—In this paper, we propose a new steganographic scheme based on the polar codes. The scheme works according to two steps. The first offers a stego vector from given cover vector and message. The stego vector provided by the first method can be the optimal; in this case, the insertion is successful with a very low complexity. Otherwise, we formalize our problem in a linear program form with initial solution the stego vector given by the first method, to converge to the optimal solution. Our scheme works with the case of a constant profile as well with any profile; it is then adapted to the case of wet paper. Tests on multiple gray scale images showed its good performance in terms of minimizing the embedding impact.**

Keywords-linear programming, matrix embedding, polar codes, steganography, wet paper codes.

I. INTRODUCTION

Steganography is a technique that allows hiding information in an unsuspected medium (image, sound or video) so that it was undetectable. To reach this objective it is indispensable to use a technique in order to reduce the distortion induced by the hiding of the secret message. The matrix embedding technique introduced by Crandall [1] has allowed the definition of steganographic schemes that minimize the embedding impact. The first implementation was created with the work of Westfeld [2] in which the Hamming codes were used. Afterwards, Bose-Chaudhuri-Hocquenghem (BCH) codes [3], [4], Reed-Solomon codes (RS) [5] and Syndrome-Trellis-Codes (STC) [6] are used in steganography. Combination of LSB, matrix embedding and wet paper techniques allowed building more effective and more reliable steganographic schemes. Our works is a contribution to schemes of minimizing embedding impact. We propose in this paper a new steganographic scheme based on the polar codes. The scheme is applied to the cases of constant profile and of wet paper.

We will consider, thought all the paper, the cover vector **v** made up of the LSBs of the cover image, the stego vector **y**, the changes vector **e** (y=v+e), the secret message **m** and the parity check matrix **H** of the polar code used.

This paper is organized as following. Section II describes matrix embedding and minimizing embedding impact. In Section III, we study the linear programming. The polar codes, used to implement our scheme, are presented in Section IV. In Section V, we propose the scheme. Section VI shows the results obtained when the scheme is applied on images. Section VII concludes the paper.

II. STEGANOGRAPHY AND MATRIX EMBEDDING

A. Steganography

Steganography or the art of secret communication aims to hide a message in an apparently innocuous cover medium. Steganography schemes are characterized by different parameters. The most used to evaluate the performance of a steganography scheme is the embedding efficiency. It is the number of bits of the message by distortion.

B. Peak Signal Noise Ratio (PSNR)

The PSNR is a distortion measure between two images. It is calculate from MSE (Mean Square Error) and is expressed in dB. Let I_0 and I_r be, respectively, the original and reconstructed images of same length $M \times N$.

The PSNR and the MSE are given by [7]:

$$\mathrm{PSNR}(I_0, I_r) = 10\log_{10}(255^2 / \mathrm{MSE}(I_0, I_r)) \quad (1)$$

$$\mathrm{MSE}(I_0, I_r) = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_0(i,j) - I_r(i,j)\right]^2 \quad (2)$$

A PSNR greater than 35 dB means that there is no visible difference between these two images [7].

C. Principle of Matrix Embedding

It consists in finding the stego vector y closest to v such that yH^T=m. Replacing y by v+e, we have eH^T= m-vH^T.

The sender researches the minimum weight vector e (leader) in the coset C(m-vH^T). At the reception, the decoding is done by the matrix product m=yH^T.

D. Minimizing Embedding Impact

Assuming that changes don't interact with each other, the total embedding impact is the sum of those at each pixel [6]:

$$D(\mathrm{v},\mathrm{y}) = \sum_{i=1}^{n}\rho_i\left|\mathrm{v}_i - \mathrm{y}_i\right|, \quad (3)$$

where $0 \le \rho_i \le \infty$ is change cost of the pixel v_i into y_i. We have to minimize D. The insertion and extraction functions are:

$$\begin{aligned}\mathrm{Emb}(\mathrm{v},\mathrm{m}) &= \arg\min_{y \in C(\mathrm{m})} D(\mathrm{v},\mathrm{y}) \\ \mathrm{Ext}(\mathrm{y}) &= \mathrm{y}\mathrm{H}^{\mathrm{T}} = \mathrm{m}\end{aligned} \quad (4)$$

III. LINEAR PROGRAMMING

Linear programming is a central domain of optimization. It can be formulated as follows [8]:

$$\min \text{ or } \max_{x \in \mathfrak{C}} f_{s.t}(x) \tag{5}$$

where f is objective function, x variable and $\mathfrak{C}$ feasible set.

A linear program can be written either in

Canonical form or Standard form

$$\min f_{s.t}(x) = \mathrm{c}^T x \quad \begin{cases} Ax \geq b \\ x \geq 0 \end{cases} \qquad \min f_{s.t}(x') = \mathrm{c}'^T x' \cdot \quad \begin{cases} A'x' = b \\ x' \geq 0 \end{cases}$$

Before solving a linear programming problem, we must put it in standard form. A linear program can be solved by simplex method or interior point methods.

A. Simplex Method

This method was developed by G. Danzig to solve linear programs. This algorithm is based on the following approach [8]: starting from a vertex (initial solution), we move from one extreme point (vertex) to another one along the polyhedron frontier (feasible set). We determine if the current vertex is optimal and if not the case, we move to adjacent vertex that optimizes the objective function.

B. Methods of Interior Points

These methods invented by of Karmarkar [9] in 1984 allow reducing the complexity observed in the simplex algorithm. The interior point methods start from an interior point to the feasible set and, by using a fixed strategy, determines an approximate value of the optimal solution.

IV. POLAR CODES

A. Usual Notations

Let W be B-DMC (Binary input-Discrete Memoryless Channel), $\mathcal{X}$ and $\mathcal{Y}$ respectively input and output alphabets of W, W($y|x$) transition probability ($x \in \mathcal{X}$ and $y \in \mathcal{Y}$), $\otimes$ is Kronecker product, $a_1^n = (a_1, \ldots, a_n)$, $n=2^p$, p positive integer.

B. Definitions

The symmetric capacity (bits/s) and Bhattacharyya parameter or reliability of W are defined by [10]:

$$\mathrm{I}(\mathrm{W}) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} \mathrm{W}(y|x) \log_2 \frac{\mathrm{W}(y \mid x)}{\frac{1}{2}\mathrm{W}(y \mid 0) + \frac{1}{2}\mathrm{W}(y \mid 1)} \tag{6}$$

$$\mathrm{Z}(\mathrm{W}) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{\mathrm{W}(y|0)\,\mathrm{W}(y|1)}\ . \tag{7}$$

The polar coding is based on these two parameters.

C. Channel Polarization and Transformation

1) Channel polarisation: it consists in synthesizing n independent copies of a given B-DMC W in n others $\{\mathrm{W}_n^{(i)} : 1 \leq i \leq n\}$. This is made up of two steps.

a) Channels combination: it is to group n copies of a given W in a channel W_n. We associate recursively two independent copies of $\mathrm{W}_{n/2}$ to create W_n (Fig.1).

$$x_1^n = u_1^n \mathrm{G}_n\,, \tag{8}$$

$$\mathrm{G}_n = \mathrm{B}_n \mathrm{G}_2^{\otimes p} = \mathrm{R}_n (\mathrm{I}_2 \otimes \mathrm{B}_{n/2}) \mathrm{G}_2^{\otimes p}\ . \tag{9}$$

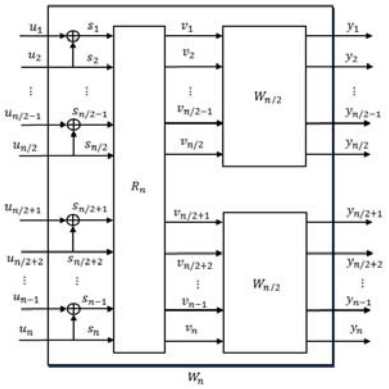

Figure 1. Construction of the channel W_n from two copies of $\mathrm{W}_{n/2}$.

with G_n is generator matrix, B_n and R_n are permutation matrix. $\mathrm{G}_2^{\otimes p}$ is Kronecker product of p copies of $\mathrm{G}_2 = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]$.

a) Channel splitting: one subdivides the channel W_n into n channels $\mathrm{W}_n^{(i)}$, $1 \leq i \leq n$. If u_1^n is uniform on $\mathcal{X}^n$ then $\mathrm{W}_n^{(i)}$ is the channel really seen by u_i (Fig. 2).

2) Recursive channel transformation: the process of channel transformation can be generalized recursively:

$$\left(\mathrm{W}_n^{(i)}, \mathrm{W}_n^{(i)}\right) \xrightarrow[we\ construct]{} \left(\mathrm{W}_{2n}^{(2i-1)}, \mathrm{W}_{2n}^{(2i)}\right). \tag{10}$$

$$\mathrm{Z}(\mathrm{W}_{2n}^{(2i-1)}) \leq 2\mathrm{Z}(\mathrm{W}_n^{(i)}) - \mathrm{Z}(\mathrm{W}_n^{(i)})^2, \tag{11}$$

$$\mathrm{Z}(\mathrm{W}_{2n}^{(2i)}) = \mathrm{Z}(\mathrm{W}_n^{(i)})^2 \tag{12}$$

with equality in (11) if W is a Binary Erasure Channel (BEC).

3) Construction and encoding of polar codes: let A be a subset of dimension k, A^c its complementary in $\{1,\ldots,n\}$, u_{A} information vector and u_{A^c} frozen vector. In general, $u_{\mathrm{A}^c} = 0_1^{n-k}$. Polar codes construction provides A and $\mathrm{W}_n^{(i)}$

such that $Z(W_n^{(i)}) \le Z(W_n^{(j)})$ for $i \in A$ and $j \in A^c$. We use (10) to encode a data-word u_1^n in a code-word x_1^n.

4) Decoding of polar codes: Successive Cancellation (SC) [10], Linear Programming (LP) [11] and Belief Propagation (BP) decoding are used to decode polar codes. The most used is SC but, because of the channel probability involved in its implementation, its application in steganography is not yet possible. Therefore, we use LP.

V. Steganographic Scheme Based on Polar Codes

A. Polar Codes Construction for Steganography

The construction can be summed up in three steps.

Step1: calculation of reliability parameters. Steganographic channel is Binary Symmetric Channel (BSC). Consider that our channel W is chosen such that we have equality in (11):

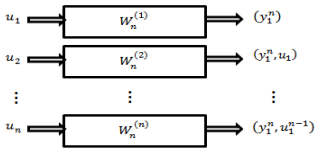

Figure 2. Equivalent scheme of polar coding.

$$Z(W_n^{(j)}) = 2Z(W_{n/2}^{((j+1)/2)}) - Z(W_{n/2}^{((j+1)/2)})^2 \quad \text{if } j \text{ is even,}$$
$$Z(W_n^{(j)}) = Z(W_{n/2}^{(j/2)})^2 \quad \text{if } j \text{ is odd.}$$

with $Z(W_1^{(1)}) = Z(W) = 2\sqrt{W(0|0)W(0|1)} = 2\sqrt{p_e(1-p_e)}$, p_e is the error probability of the channel W.

Step 2: determination of information and redundancy bits. We select the k channels of the lowest reliabilities parameters for data bits. The indices of these channels form the information set A. The other channels carry the redundancy bits. Their indices constitute A^c.

Step 3: generation of the parity check matrix H. The parity check matrix H is used both for insertion and extraction of the messages. We can use the lemma given by Goela and al. [11, Lemma 1] to calculate H from generator matrix G_n.

$$\text{We have } H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \text{ and } H^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ if } n=8 \text{ and } k=1 \quad (13)$$

B. Steganographic Scheme with a Polar Code

In this part, we consider that the change of any pixel produces the same distortion (constant profile).

1) First method proposed: observing the parity check matrix H of a polar code and its transpose H^T we can make the following remarks (see for exemple (13)):

a) the columns of H *are pairwise independent;*

b) if we scan the columns of H^T*, the position at which we meet the first 1 differs from those of the others;*

c) starting from last column and first line, position at which we meet the first 1 is the first met on this line.

With these remarks, we will define a first steganographic scheme. Consider the relation $yH^T=m$. It can be decomposed by $y_1 H_{1,j}^T + y_2 H_{2,j}^T + \ldots + y_n H_{n,j}^T = m_j$, for j=1,…,*n-k*.

Let i be the position of the first 1 met on the column j, then:

$$y_i = y_{i+1} H_{(i+1),j}^T + \cdots + y_n H_{n,j}^T + m_j \quad (14)$$

To determine y_i in (14), we must find y_{i+l} such that $H_{(i+1),j}^T = 1$. We will assume that these positions are locked ($y_{i+l}=v_{i+l}$). Therefore, before calculating the elements y_i, we first assign it the cover vector. The changes of certain positions y_i will occur as and when we travel over the columns. At the end of this process, we have a stego vector y arising from d modifications of the cover vector v.

The method described above gives a solution satisfying $yH^T=m$ but it is not necessarily the best. This solution is optimal if the number of changes d is less than $(n-k)/2$. To ensure finding the optimal solution, we will define, from the first solution, a method that offers the optimal solution.

2) Second method: let y_p be the stego vector found with the first method ($y_p=v+e_p$). We have to find, from y_p, the optimal vector y_{opt} such as $yH^T=m$. In other words, we have to create an algorithm that, initialized to e_p, converges to e_{opt}.

Recapitulation:

- We have a starting solution e_p → initial solution;
- we search the minimal weight vector e → minimization;
- verifying $eH^T = m\text{-}vH^T$ → constraints.

Considering these three points, we have a minimization problem with initial solution e_p. Find the vector e of minimum weight amounts to search the vector realizing the minimum of the scalar product with the vector $c=\{1\}^n$. From there, we can define our optimization problem as follows:

$$\arg\min_{e} f(e) = \square^T \quad (15)$$
$$\text{s.t } \mathfrak{C} = \begin{cases} e \in \{0,\square\}^n \\ eH^T = m\text{-}vH^T = s \\ e_p \text{ initial} \square \end{cases} \Leftrightarrow {}_p{}^T$$

This problem is a linear optimization, with equalities constraints, written in standard form. It can be solved by simplex [8] or interior points [9] methods (see Section III).

C. Wet Paper Steganographic Scheme

Consider $\rho = \{\rho_i\}_{1 \le i \le n}$, $\rho_i \in [0, \infty]$, our goal is to adapt the scheme proposed previously to this general case of distortion profile [6]. The first method is independent of the profile. Thus, it is applied identically to constant profile case. For the second method, since the problem is the same as constant profile (minimization), we use the same principle of linear programming to find our optimal solution. The initial solution and the constraints have not been changed. But the

objective function will change. Indeed, for an arbitrary profile, the goal is to minimize the distortion function (3). Rewrite it depending on the vector e with $|y_i - x_i| = e_i$:

$$D(e) = \sum_{i=1}^{n} \rho_i e_i \tag{16}$$

We can see that our objective function f(e)=<c,e>, which must be written as a scalar product between the cost vector of the linear program and the variable e, appears well in (16).

Consequently, the elements of the cost vector are represented by the costs of pixels change. Let c=ρ, we have still a linear program.

VI. Experimental Results

To verify the efficiency of our scheme and the invisibility of the hidden messages using this scheme, we test it on different ***(512×512)*** pixels-images in gray scale *pgm* format. The images are taken from *BOSS* (Break Our Stego System: competition of steganography attacks) database.

To make the message less detectable, we choose to permute the pixels of the cover image before the insertion. Because 512 is a power of 2, we can use the permutation matrix B_n with n=512. Thus the changes will be spread over isolated pixels of the image making less detectable the secret message and allowing a more secure insertion. To find the initial order of cover image pixels, we still use B_n, because it is invertible and equal to its inverse. We insert first, a 3 ko (24576 bits) message in the image 10.pgm of *BOSS* database.

The changes in the stego image are invisible to naked eye, as shown in Fig. 3. Hence, the first and main goal (visual imperceptibility) of steganography is achieved. Comparing cover and stego images, for passive attacker, and their histogram, for a semi-active attacker, we have clearly seen that difference between the two images is almost invisible. The scheme is even more secure that the attacker has only the stego image to see if it contains a message or not.

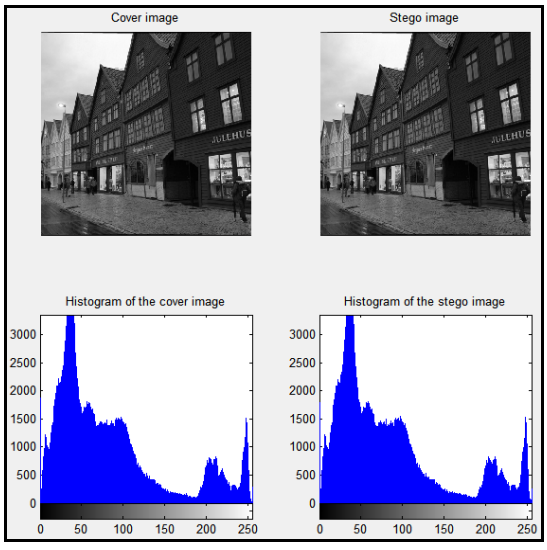

Figure 3. Cover image and stego image and their histograms.

Furthermore, the scheme allows extracting the message, without any alteration. The message inserted in cover image is identical to that extracted from stego image (Fig. 4).

To evaluate the performance of our scheme we calculated the PSNR with (1). We randomly generate 10 messages of different size that we insert in 5 images (1, 10, 100, 1000 and 10000.pgm). We averaged the PNSR (see Fig. 5) which vary between ***65.4 dB*** for α=1/20 and ***55.4 dB*** for α=1/2. These PSNR values are well above ***35 dB*** a value beyond which the difference between two images (cover and stego) is very low. This shows that the proposed scheme has good performance in terms of embedding efficiency.

VII. Conclusion

We have defined a steganographic scheme based on a new type of coding called polar coding with a good embedding efficiency. The proposed scheme consists of two parts: the first gives an initial solution and the second ensures convergence to optimal solution using linear programming. In the case where the first solution corresponds to the optimum, it is not necessary to proceed to the second method. Our scheme is also suitable for the wet paper case.

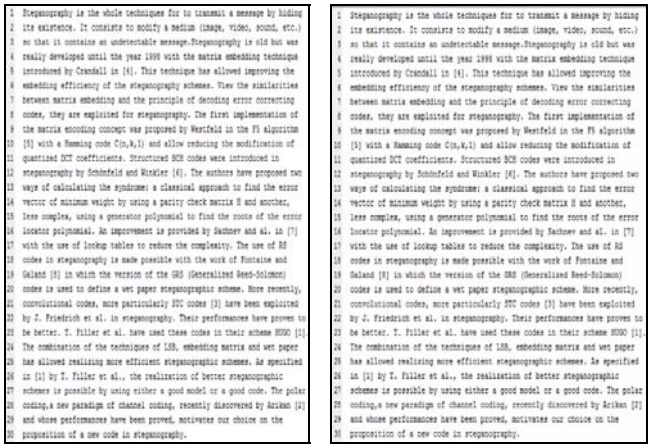

Figure 4. Inserted and extracted message.

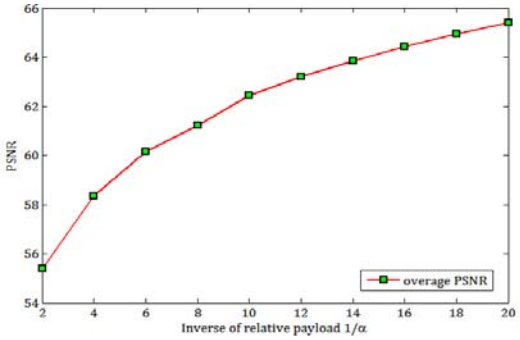

Figure 5. Evolution of PSNR depending on the relative payload.

We showed, by applying it to 5 different images, that the visually undetectable and even statistical, by using histograms, is reached. We also calculated the PSNR with these 5 images and their value varies between 55 dB and 66 dB. That is greater than the limit value 35.

An improvement prospect is to search a scheme in single step by using *LP decoding by changing the polytope* $yH^T=m$.

REFERENCES

[1] R. Crandall, "Some notes on steganography", Posted on Steganography Mailing List (1998).

[2] A. Westfeld, "High capacity despite better steganalysis (F5 – a steganographic algorithm)", In: Moskowitz, I.S. (ed.) IH 2001. LNCS, vol. 2137, pp. 289–302, Springer, Heidelberg (2001).

[3] Schönfeld, D., Winkler, "A Embedding with syndrome coding based on BCH codes", Proceedings of the 8th ACM Workshop on Multimedia and Security, pp. 214 – 223, 2006.

[4] Rongyue Zhang, Vasiliy Sachnev, Hyoung Joong Kim, Fast BCH syndrome coding for steganography; S. Katzenbeisser and A.-R. Sadeghi (Eds.), IH 2009, LNCS 5806, pp. 44-58, Springer-Verlag Berlin Heiderbelg 2009.

[5] F. Galand and C. Fontaine, "How Reed-Solomon Codes Can Improve Steganographic Schemes", In Inform. Hiding, Rennes, France, 2009.

[6] Tomáš Filler, Jan Judas and Jessica Fridrich, "Minimizing Embedding Impact in Steganography using Trellis-Coded Quantization", Department of Electrical and Computer Engineering SUNY Binghamton, Binghamton, NY 13902-6000, USA 2010.

[7] Xuanwen Luo, Qiang Cheng, Joseph Tan, "A Lossless Data Embedding Scheme For Medical in Application of e- Diagnosis," Proceedings of the 25th Annual International Conference of the IEEE EMBS Cancun, Mexico. September 17-21, 2003.

[8] Aaid Djamel, "Étude numérique comparative entre des méthodes de résolution d'un problème de transport à quatre indices avec capacités," Thèse, École Doctorale de Mathématiques, Constantine, 2010. [Online]: http://bu.umc.edu.dz/theses/math/AAI5587.pdf

[9] Narendra Karmarkar (1984). "A New Polynomial Time Algorithm for Linear Programming", Combinatorial, Vol 4, nr. 4, p. 373–395.

[10] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels", IEEE Trans. Inform. Theory, vol. IT-55, pp. 3051–3073, July 2009.

[11] N. Goela, S. B. Korada, and M. Gastpar, "On LP Decoding of Polar Codes," submitted to IEEE Trans. Inform. Theory, 2010.