Minimizing Embedding Impact in Steganography using Polar Codes

Conference Paper · April 2014

DOI: 10.1109/ICMCS.2014.6911224

4 authors:

Birahime Diouf
Cheikh Anta Diop University, Dakar
15 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Idy Diop
Cheikh Anta Diop University, Dakar
44 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Sidi Mohamed Farssi
Cheikh Anta Diop University, Dakar
36 PUBLICATIONS **57** CITATIONS
SEE PROFILE

Khouma Ousmane
Cheikh Anta Diop University, Dakar
9 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Minimizing Embedding Impact in Steganography using Polar Codes

Birahime Diouf *, Idy Diop, Sidi Mohamed Farssi and Ousmane Khouma
Department of Computer Science
Polytechnic High Institute (ESP) / Cheikh Anta Diop University (UCAD)
Dakar, Senegal
dioufbira11@yahoo.fr, idydiop@yahoo.fr, farsism@yahoo.com, oussoukhouma@gmail.com

***Abstract*—This paper proposes two approaches that reduce the complexity of the Polar Coding Steganography (PCS). The first is based on lookup tables and the second exploits the form of the syndrome, calculated from cover vector and secret message, to evaluate la position of the cover vector changes. The scheme proposed in this paper allows minimizing the embedding impact and gives similar results as those of PCS scheme with a reduced time complexity.**

Keywords-Complexity; lookup tables; matrix embedding; polar codes; steganography

I. INTRODUCTION

Steganography is an information hiding technique that consists in concealing a message in a cover medium which can be an image (used in this paper), a sound or a video. The main objective of steganography is the indetectability of the embedded secret message (visual and statistical). The embedding must be done by making the cover medium changes less noticeable as possible. In spatial domain, secret message bits are inserted at the LSBs (Least Significant Bits) of the cover image pixels. To improve this so called LSB technique, several schemes based on error correcting codes were proposed. These schemes implement matrix embedding technique; it consists in using of error correcting codes (usually met in a digital transmission channel) in steganography. Among the used codes we can quote Hamming [1], BCH (Bose-Chaudhuri-Hocquenghem) [2, 3], STC (Syndrome Trellis Codes) [4] and LDPC (Low Density Parity Check) [5].

After having introduced polar codes in steganography PCS (Polar Coding Steganography) [6], we propose in this paper two methods for minimizing embedding impact allowing to reduce the time complexity. The originality of this work lies in definition of an algorithm which calculates the change vector (vector that, added to cover vector, gives stego vector) in a single step compared to PCS [6]. Indeed, we propose two approaches: the first is based on lookup tables and the second uses the form of the syndrome to calculate the change vector.

This paper is organized as follows. Section II gives a brief review of basic concepts and preliminaries in steganography, lookup tables and polar codes. In Section III, we present PCS. New steganographic scheme is studied in Section IV. Section V provides a time complexity comparison between PCS scheme [6] and the proposed new algorithm. Section VI concludes the paper.

II. BASIC CONCEPTS AND PRALIMINARIES

We consider, in this paper, the cover vector x consisting of the LSBs of the gray scale cover image $I \in \{0, ..., 255\}^{n_1 \times n_2}$ represented in spatial domain, the message m, the change vector e, the stego vector y (y = x + e) and a parity check matrix H of the used polar code.

A. Basic Concepts in Steganography

Introduced by Crandall [7], matrix embedding technique is used in steganography to minimize the number of the cover medium changes. Subsequently, several codes are used to implement this technique in steganography. It is based on syndrome decoding of error correcting codes.

1) Syndrome Decoding: As its name indicates it, this decoding type uses the syndrome calculation to correct the errors occurred during the transmission of a code word c. The received word r = c + e and error sequence e have the same syndrome

$$rH^T = cH^T + eH^T = eH^T, \quad (1)$$

The decoding problem can then be come down to search the minimal weight vector e (coset leader) in the coset of r.

2) Matrix Embedding: It uses syndrome decoding and consists in searching the stego vector y closest to x so that

$$yH^T = m. \quad (2)$$

If we replace y by x + e, we will have

$$eH^T = m - xH^T. \quad (3)$$

The sender uses (3). His objective is to find the coset leader of e in the coset $\mathcal{C}(m - xH^T)$. At the reception, the decoding is done with the matrix product (2).

978-1-4799-3824-7/14/$31.00 ©2014 IEEE

3) Embedding Impact and its Minimization: Assuming that changes don't interact with each other, the total embedding impact (or distortion) is the sum of the changes of the different pixel [4]:

$$D(\mathrm{x},\mathrm{y}) = \sum_{i=1}^{n} \rho_i \left|\mathrm{x}_i - \mathrm{y}_i\right|, \qquad (4)$$

where $0 \le \rho_i \le \infty$ is change cost of the pixel LSB x_i into y_i. If $\rho_i = 1$ for all i (constant profile), minimizing the distortion D is reduced to minimize the number of changes on the cover image. The insertion and extraction functions are:

$$\begin{aligned} &\mathrm{Emb}(\mathrm{x},\mathrm{m}) = \arg \min_{\mathrm{y} \in \mathcal{C}(\mathrm{m})} D(\mathrm{x},\mathrm{y}) \\ &\mathrm{Ext}(\mathrm{y}) = \mathrm{y}\mathrm{H}^{\mathrm{T}} = \mathrm{m} \end{aligned} . \qquad (5)$$

4) Wet Paper Codes: In practice some pixels of the cover image can be more sensitive to change than others. The first called wet pixels (with $\rho_i = \infty$) must not be changed and the second called dry pixels (with $\rho_i = 1$) can be changed. In this case we say that we have a wet paper channel [8]. The syndrome coding is also applied to this type of channel using wet paper codes [4, 6, 9].

B. Lookup Tables in Steganography

Lookup table is a data structure used to replace a calculation by a lookup operation (search for a value in memory) that is often easier. The use of lookup tables can reduce execution time of a complex calculation because search a value in memory is often faster than doing such calculations.

Schönfeld and Winkler [2] introduced BCH codes in steganography to improve embedding efficiency. They proposed two ways for syndrome calculation. A first approach based on search for a coset leader using a parity check matrix H and a second approach using a generator polynomial to search for roots. After calculating the syndrome s, the next step is to find the coset leader e. With lookup tables, we can reduce the calculation time of the coset leader. Indeed, the use of lookup tables by Kim and al. [3] reduced the complexity compared to the method based on exhaustive research of roots.

C. Polar Codes

Polar codes, based on a new paradigm of coding, are defined as the first codes sequences that achieve the channel capacity, limit established by Shannon [10]. A polar code of length $n = 2^p$ and dimension k will be denoted by $PC(n,k)$, u_1^n denote an information set, x_1^n a code word, y_1^n the received word, G_n a generator matrix, W is the transmission channel and $u_1^i = (u_1, \ldots, u_i)$, with $1 \le i \le n$. The symmetric capacity [11] of W is denoted by I(W) and the reliability parameter is Z(W). Let A and its complementary in $\{1,\ldots, n\}$ A^{c} respectively denote information and frozen bits sets, u_{A} information vector and $u_{\mathrm{A}^{\mathrm{c}}}$ frozen vector. The construction of polar codes is based on channel polarization.

1) Channel Polarisation: It consists in synthesizing of n independent copies of a given B-DMC W to create n others $\{\mathrm{W}_n^{(i)} : 1 \le i \le n\}$. The polarization appears in the sense that $\mathrm{I}(\mathrm{W}_n^{(i)})$ tends to I(W) or 1-I(W), depending on if $\mathrm{I}(\mathrm{W}_n^{(i)})$ is closer to 0 or 1. The operation of channel polarization is made up two steps: channel combining and channel splitting [11]

$$(\mathrm{W},\mathrm{W},\ldots,\mathrm{W}) \xrightarrow{\text{combining}} \mathrm{W}_n \xrightarrow{\text{splitting}} \left\{\mathrm{W}_n^{(i)}\right\}_{i=1,\ldots,n}. \qquad (6)$$

The channel combining combines n copies of a given B-DMC W in a vector channel W_n. It is done recursively by combining two copies of $\mathrm{W}_{n/2}$ [6, 11]. During channel splitting we subdivide W_n into n channels $\mathrm{W}_n^{(i)}$, $1 \le i \le n$.

Channel polarization can be seen as a recursive channel transformation process which can be represented as follows:

$$\left(\mathrm{W}_n^{(i)}, \mathrm{W}_n^{(i)}\right) \xrightarrow[\text{we construct}]{} \left(\mathrm{W}_{2n}^{(2i-1)}, \mathrm{W}_{2n}^{(2i)}\right). \qquad (7)$$

2) Polar Coding: The relations of polar coding are

$$x_1^n = u_1^n \mathrm{G}_n, \qquad (8)$$

$$\mathrm{G}_n = \mathrm{B}_n \begin{bmatrix} \mathrm{G}_{n/2} & 0 \\ \mathrm{G}_{n/2} & \mathrm{G}_{n/2} \end{bmatrix}. \qquad (9)$$

with B_n a permutation matrix and G_1=[1].

In polar coding if u_1^n follows a uniform distribution then $\mathrm{W}_n^{(i)}$ is the channel really seen by u_i. In others words, each bit u_i take the channel $\mathrm{W}_n^{(i)}$ [6] (Fig. 1). Polar coding uses the channels $\mathrm{W}_n^{(i)}$ the most reliable to carry the information bits and the least reliable the frozen bits:

$$Z(\mathrm{W}_n^{(i)}) \le Z(\mathrm{W}_n^{(j)}), \qquad (10)$$

with $i \in \mathrm{A}$ and $j \in \mathrm{A}^{\mathrm{c}}$.

Polar codes are several applications in information theory. Among these applications, the one that we are interested in is steganography [8].

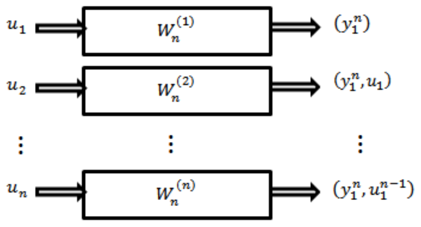

Figure 1. Equivalent scheme of polar coding.

III. First Polar Coding Steganography

Denote by $\mathcal{S}_{\mathrm{PC}}(n,m=n\text{-}k)$ the steganography based on polar code *PC*(*n*,*k*). The constant profile and wet paper case are considered separately.

A. Steganography for Constant Profile

This steganographic scheme (PCS) consists of two steps.

1) First Step: To determinate a parity check matrix of a polar code, we use the lemma given by Goela and al. [12, Lemma 1] which states that if the frozen bits are equal to 0 then the transpose of the parity check matrix H of the polar code is given by the columns of the generator matrix G_n whose indices are in A^c. By making the most of the particular form of H and its transpose obtained H^T, we can transform the equations of (2) a system allowing to calculate the coefficients of the stego vector y (see [6]):

$$y_i = y_{i+1}H^{\mathrm{T}}_{(i+1),j} + \cdots + y_n H^{\mathrm{T}}_{n\,j} + m_j \;\; ; j = n-k \text{ down to } 1, \quad (11)$$

with *i* the position of first 1 on column *j*. For each *j*, we calculate the corresponding y_i.

The vector y must be initialized to the cover vector x before the calculations. The embedding is done by keeping unchanged *k* bits of x. Thus, the changes occur on the *n-k* remaining bits of x. Then, we obtain a stego vector y_p resulting from *d* modifications of the cover vector x.

This method described above gives a solution verifying $yH^T = m$ but it is not necessary the best. We must define, from y_p, a method that converges to the optimal solution.

2) Optimization of The First Solution: The objective of this method is to find the stego vector y closest to x by using the polar code *PC(n,k)*. Let e_p be the change vector corresponding to the stego vector y_p found with the first step. We have the following problem [6]:

Problem:

- we have an initial solution $e_p \mapsto$ initial solution;
- we have to minimize the distortion $D \mapsto$ minimization problem;
- verifying $eH^T = m - xH^T = s \mapsto$ constraints.

Considering these three points, we have an optimization problem, particularly a minimization problem under equalities constraints, with initial solution e_p. The problem can be formalized as follows:

$$\underset{\mathrm{e}}{\operatorname{argmin}} \quad \underset{\mathrm{s.t}}{f(\mathrm{e})} = \langle \rho, \mathrm{e} \rangle = \rho^{\mathrm{T}}\mathrm{e}$$

$$\begin{cases} \mathrm{e} \in \{0,1\}^n \text{ binary vector} \\ \mathrm{eH}^{\mathrm{T}} = \mathrm{m} - \mathrm{xH}^{\mathrm{T}} = \mathrm{s} \\ \mathrm{e_p} \text{ initial solution} \Leftrightarrow \mathrm{e_p H}^{\mathrm{T}} = \mathrm{s} \end{cases} \quad (12)$$

with *f* the objective function and $\rho = \{\rho_i\}_{1\leq i\leq n} = \{1\}^n$ the change cost vector.

This is a problem of linear programming [13] written in standard form with an additional constraint; the vector e is a binary vector.

B. PCS Scheme for Arbitrary Distortion

We keep the first step of constant profile case. The embedding algorithm should consider the locked positions. We need to define a scheme that minimization the number of changed positions while keeping unchanged the locked pixels. The problem is to minimize the distortion *D* which can be rewritten as follows:

$$D(\mathrm{e}) = \sum_{i=1}^{n} \rho_i \mathrm{e}_i \,. \quad (13)$$

with $|x_i - y_i| = e_i$ and $\rho_i = 1$ for constant profile and $\rho_i = \{1,\infty\}$ for wet paper case. The insertion and extraction functions become:

$$\begin{aligned} &Emb(\mathrm{x},\mathrm{m}) = \arg \min_{\mathrm{e}\in \mathcal{C}(\mathrm{s})} D(\mathrm{e}) \\ &Ext(\mathrm{y}) = \mathrm{yH}^{\mathrm{T}} = \mathrm{m} \Leftrightarrow \mathrm{eH}^{\mathrm{T}} = \mathrm{s} = \mathrm{m} - \mathrm{xH}^{\mathrm{T}} \end{aligned} \quad . \quad (14)$$

As constant profile case, we have a linear programming problem which can be solved in the same way [6, 13]. Indeed, the objective function $f(\mathrm{e}) = \langle \rho, \mathrm{e} \rangle$ appears as a scalar product in (13) and the constraints are the same as constant profile.

PCS [6] consists of two steps, each of which requires a computation time. Thereof was not taken into account. This aspect is taken into account in the definition of the new steganographic algorithm. We propose two approaches that reduce the complexity compared to PCS.

IV. New Polar Coding Steganography

We will propose two methods: the first is based on lookup tables and the second uses the form of the syndrome to evaluate the positions to be changed.

A. Method Based on Lookup Tables

This method uses lookup tables to give the change vector from the syndrome. First, we calculate s. This is given by s = $eH^T = m - xH^T = s$. Constant profile and wet paper case are considered.

1) Constant Profile: The lookup table consists of 2^{n-k} lines and 2 columns. The first column contains the different syndromes and the second column shows the corresponding change vectors. Thus, at line *i*, $1 \leq i \leq n$, we have on the first column the syndrome s = S[*i*] corresponding to the change vector e = E[*i*], on the second column. After caculating a syndrome s, we process as follows :

- we travel the lines of column 1,
- if we find the position *i* with S[*i*] = s then the coset leader corresponding to s is at the same position *i* on column 2.

The coset leaders calculation operation is then replaced by a research in a table stocked first.

As examples, we use a polar code $PC(4,1)$ for the steganography $\mathcal{S}_{PC}(4,3)$ and $PC(8,4)$ for $\mathcal{S}_{PC}(8,4)$.

Example 1: A parity check matrix of $PC(4,1)$ is:

$$\mathrm{H} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \text{ and its transpose } \mathrm{H}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \qquad (15)$$

The columns $H_{.j}$ of H verify the following equalities:

$$\begin{aligned} H_{.1} + H_{.2} &= H_{.3} + H_{.4} = (001)^{\mathrm{T}} \\ H_{.1} + H_{.3} &= H_{.2} + H_{.4} = (010)^{\mathrm{T}} \\ H_{.1} + H_{.4} &= H_{.2} + H_{.3} = (011)^{\mathrm{T}} \end{aligned} \quad . \qquad (16)$$

The syndrome can have one of the following configurations:

- **Config 1:** the syndrome is equal to one column of H
- **Config 2:** it is equal to zero vector ;
- **Config 3:** it is sum of two columns of H with (16).

That allows us to draw up the following table:

TABLE I. LOOKUP TABLE FOR $\mathcal{S}_{PC}(4,3)$

column 1: syndromes			column 2: change vectors	
S [1]	1 0 0		E [1]	1 0 0 0
S [2]	1 0 1		E [2]	0 1 0 0
S [3]	1 1 0		E [3]	0 0 1 0
S [4]	1 1 1	⇔	E [4]	0 0 0 1
S [5]	0 0 0		E [5]	0 0 0 0
S [6]	0 0 1		E [6]	1 1 0 0
S [7]	0 1 0		E [7]	1 0 1 0
S [8]	0 1 1		E [8]	1 0 0 1

Consider a cover vector x = (1 0 0 1) and a message m = (0 1 0). The syndrome is given by s = m - xH$^{\mathrm{T}}$ = (0 0 1). That is equal to syndrome S[6] and then the corresponding change vector (or coset leader) is E[6] = (1 1 0 0) = e.

Example 2: A parity check matrix of $PC(8,4)$ is:

$$\mathrm{H} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \text{ and } \mathrm{H}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \qquad (17)$$

As previous example, we have:

$$\begin{aligned} H_{.1} + H_{.2} &= H_{.3} + H_{.4} = H_{.5} + H_{.6} = H_{.7} + H_{.8} = (0001)^{\mathrm{T}} \\ H_{.1} + H_{.3} &= H_{.2} + H_{.4} = H_{.5} + H_{.7} = H_{.6} + H_{.8} = (0010)^{\mathrm{T}} \\ H_{.1} + H_{.4} &= H_{.2} + H_{.3} = H_{.5} + H_{.8} = H_{.6} + H_{.7} = (0011)^{\mathrm{T}} \\ H_{.1} + H_{.5} &= H_{.2} + H_{.6} = H_{.3} + H_{.7} = H_{.4} + H_{.8} = (0100)^{\mathrm{T}} \\ H_{.1} + H_{.6} &= H_{.2} + H_{.5} = H_{.3} + H_{.8} = H_{.4} + H_{.7} = (0101)^{\mathrm{T}} \end{aligned} \quad . \qquad (18)$$

We have the three configurations **Config 1**, **2** and **3** of previous example and Table II.

TABLE II. LOOKUP TABLE FOR $\mathcal{S}_{PC}(8,4)$

Column 1: syndromes			Column 2: change vectors	
S [1]	1 0 0 0		E [1]	1 0 0 0 0 0 0 0
S [2]	1 0 0 1		E [2]	0 1 0 0 0 0 0 0
S [3]	1 0 1 0		E [3]	0 0 1 0 0 0 0 0
S [4]	1 0 1 1		E [4]	0 0 0 1 0 0 0 0
S [5]	1 1 0 0		E [5]	0 0 0 0 1 0 0 0
S [6]	1 1 0 1		E [6]	0 0 0 0 0 1 0 0
S [7]	1 1 1 0		E [7]	0 0 0 0 0 0 1 0
S [8]	1 1 1 1	⇔	E [8]	0 0 0 0 0 0 0 1
S [9]	0 0 0 0		E [9]	0 0 0 0 0 0 0 0
S [10]	0 0 0 1		E [10]	1 1 0 0 0 0 0 0
S [11]	0 0 1 0		E [11]	1 0 1 0 0 0 0 0
S [12]	0 0 1 1		E [12]	1 0 0 1 0 0 0 0
S [13]	0 1 0 0		E [13]	1 0 0 0 1 0 0 0
S [14]	0 1 0 1		E [14]	1 0 0 0 0 1 0 0
S [15]	0 1 1 0		E [15]	1 0 0 0 0 0 1 0
S [16]	0 1 1 1		E [16]	1 0 0 0 0 0 0 1

Let m = (1 1 0 1) and x = (1 0 1 0 1 0 0 1) be respectively message and cover vector. Then, the syndrome is s = (1 1 0 0) = S[5] and the corresponding coset leader is E[5] = (0 0 0 0 1 0 0 0) =e.

2) Wet Paper Case: Let $\mathcal{J}$ be the set of positions that must be locked (wet elements). If we find a change vector e with a 1-bit (bit equal to 1) at position *pos*, $1 \le pos \le n$, which must be locked, then we have one of the three configurations.

- Syndrome s is in **Config 1** (i.e. e has a single 1-bit at a wet position), we process as follows:

We subdivide the first half of the lookup table, made up of *n* elements, in *n*/4 subsets of 4 elements each. In each subset, a syndrome is equal to the sum of the three others. Thus, to lock position *pos*, we consider the change vector (in the subset of the calculated syndrome) corresponding to the sum of three change vector whose syndromes sum is equal to that of our change vector e. That give us another vector having three 1-bits at positions *i*, *j* and *l* different to *pos*. If one of these three positions is in $\mathcal{J}$ then we must search another change vector with zero coefficients at locked positions. To do this, we replace the pair in the triplet (*i*, *j*, *l*) belonging to $\mathcal{J}$ by another pair not in $\mathcal{J}$ with the equalities of system (18) or theirs of its equivalent[1].

- The syndrome s is in **Config 2**, then :

No problem arises since the change vector has all the components equal to 0.

- s is in **Config 3** (i.e. e has two 1-bits at positions 1 and *pos* which, at least, one must be locked), then :

We search the couple of indices (*i*, *j*) not belonging to $\mathcal{J}$ with the equalities of system (18) equivalent. Thus change vector will have two 1-bits at positions *i* and *j*.

[1] Equivalent denotes the system obtained for another steganography with a parameter different to 8. If we use $\mathcal{S}_{PC}(4,3)$ then the system is (16).

Example 3: Consider the cover vector x = (1 0 1 0 1 0 0 1) and message m = (1 1 0 1). The syndrome is s = (1 1 0 0) =S[5] and the first change vector found is E[5] = (0 0 0 0 1 0 0 0). If $\mathcal{J}$ = (5, 6, 7) then the only 1-bit is at the 5-th position which must be locked. The syndrome is in **Config. 1**. Thus, we subdivide the first half of Table II in two subsets of four elements each. Our syndrome s is in the second subset. Then s = (1 1 0 0) = S[5] = S[6] + S[7] + S[8] and e = E[5] = E[6] + E[7] + E[8] = (0 0 0 0 0 1 1 1). This vector e has three 1-bits at positions 6, 7 and 8. The pair (6, 7) belong to $\mathcal{J}$; we must replace it by another with the equalities of system (18). According to third line of system (18), we can replace the pair (6, 7) either by (1, 4) or by (2, 3). Thus, we can choose e = (1 0 0 1 0 0 0 1) or e = (0 1 1 0 0 0 0 1).

The given example is the worst case (the most difficult to resolve) that we can met because the two others configurations types are easier to treat.

B. *Direct Calculation Method of Change Vector*

We will consider constant profile and wet paper cases as with lookup tables method.

1) Constant Profile: The method that we will propose exploits the uniform correspondence between the syndrome value and the position of non-zero symbols of the corresponding change vector. For example, on Table II, we have the following cases:

- **Case 1**: on the first half of the table (from line 1 to line 8 = *n*), syndromes have their first bit (most significant bit) equal to 1 and their decimal value varies between 8 = *n* and 15 = 2 *n* -1. The position of the only 1-bit of the change vector goes from 1 to 8 = *n* (from left to right);
- **Case 2**: on the 9 = (*n*+1)-*th* line, syndrome and change vector are both zero vectors;
- **Case 3**: on the remaining part of the table (from line 10=*n*+2 to line 16 = 2·*n*), syndromes have as first bit 0 and there decimal value varies between 1 and 7 = *n* – 1. The change vector have two 1-bits; the first at position 1 and the second at a position which goes from 2 to 8 = *n*.

The three parts of Table II are separated by thick lines and the quadruplets by doubles lines. These parts and associated remarks are also valid for Table I (with *n* = 4) and for the other tables corresponding to equivalent systems. According to these observations, there is a connection between the decimal value of syndrome s and the position of the non-zero elements of the change vector e. This is due to the fact that, on the columns of the parity check matrix H, are represented the binary values of the numbers from *n* to 2*n* -1.

However, a necessary condition is that the number of syndromes must be equal to the double of the cover vector size. It is, thus, equal to the length of the used polar code *PC*(*n,k*):

$$2^{n-k} = 2\cdot n. \qquad (19)$$

n is a power of 2. Let $n = 2^p$, then

$$2^{n-k} = 2\cdot 2^p = 2^{p+1}. \qquad (20)$$

Thus

$$n - k = p + 1. \qquad (21)$$

Consequently

$$k = n - 1 - \log_2 n = 2^p - 1 - p. \qquad (22)$$

The polar codes *PC*(4,1) and *PC*(8,4), given as examples with the lookup tables method are such as (22) is verified. Indeed, for *PC*(4,1), we have k=1=4-1-log$_2$4=2^2-1-2 and k=4=8-1-log$_2$ 8=2^3-1-3 for *PC*(8,4).

The validity of the observations concerns the values

$$\begin{array}{l} p \in \{2,3,4,5,6,7\} = \mathcal{P} \\ n \in \{4,8,\ldots,128\} = \mathcal{N} \;. \\ k \in \{1,4,\ldots,120\} = \mathcal{K} \end{array} \qquad (23)$$

with $n = 2^p$, $k = 2^p - 1 - p$ and $p \in \mathcal{P}$.

For an arbitrary polar code *PC*(*n,k*) [11], the length *n* is a power of 2 and the dimension *k* is a positive integer in {1, 2, …, *n*-1}. For a polar coding steganographic scheme, the optimality condition [6] is $m = n - k > p = \log_2 n$. The parameters of our polar code in the proposed approach satisfy this optimality condition because we have $n - k = p + 1 > p$.

The embedding is done by pair of a cover medium segment and a message segment. The number of cover segments must be equal to or greater than the number of message segments.

We propose the following algorithm:

Algorithm 1. Calculation of a syndrome coset leader

```
Initialization:
$p \leftarrow$ an element of $\mathcal{P}$ ; $n = 2^p$ ; $k \leftarrow n-1-p$ ;
$e \leftarrow 0_1^n$ ; $y \leftarrow x$ ;
Calculation:
If $xH^T \neq m$ then
    $s \leftarrow m - xH^T$ ;
    calculate decimal value of binary syndrome vector
     (dec← decimalConversion (s))
     if 1st  coefficient of syndrome s is equal to 1 then
            affect the (dec+1-n)-th coefficient of e to 1;
    else
            affect 1st  and (dec+1)-th coefficients of e to 1 ;
    end (else)
    end (if)
End (If)
```

The function *decimalConversion*(s) convert a binary vector s into its decimal value.

2) Wet Paper Case: If we have a change vector e with at least a 1-bit at a wet position then we can have one of the three cases:

- Syndrome s is in **Case 1** (i.e. e has a single 1-bit at a position *pos* in $\mathcal{J}$), then:

Consider, by quadruplet, the decimal values of the syndromes corresponding to change vectors with a single 1-bit. These *n* syndromes form the half of the all *2n* syndromes configurations. The number of quadruplets is then *n*/4. The decimal values of the syndromes go from *n* to 2*n* - 1 and the quadruplets Q_i are represented as follows:

$$\begin{array}{lcll} from\ n = n+0\cdot 4 & to & n+3 = n+1\cdot 4-1 & Q_1 \\ n+1\cdot 4 & \rightarrow & n+2\cdot 4-1 & Q_2 \\ n+2\cdot 4 & \rightarrow & n+3\cdot 4-1 & Q_3 \\ & \vdots & & \\ 2\cdot n-4 = n+(n/4-1)\cdot 4 & \rightarrow & 2\cdot n-1 = n+(n/4)\cdot 4-1 & Q_{n/4} \end{array} \quad (24)$$

We generalize this representation by:

$$Q_i : n+(i-1)\cdot 4 \quad \rightarrow \quad n+(i)\cdot 4-1\text{, } with\ i = 1,\ldots,n/4 \quad (25)$$

Thus, to know the quadruplet Q_i which the syndrome s belong to, we calculate its index *i* by:

$$i = \lceil dec(\mathrm{s}) - n + 1\ )/4 \rceil \quad (26)$$

with $\lceil . \rceil$ is the ceil operator and *dec*(s) decimal value of s.

Proof: According to (25), *dec*(s) varies between *n* + (*i*-1)·4 and *n* + (*i*)·4 - 1 for Q_i. Write simply:

$$dec(\mathrm{s}) \quad : from\ \ n+(i-1)\cdot 4\ \ to\ \ n+i\cdot 4-1 \quad (27)$$

That is equivalent to

$$dec(\mathrm{s}) - n \quad : from\ \ i\cdot 4-4\ \ to\ \ i\cdot 4-1 \quad (28)$$

Thus

$$dec(\mathrm{s}) - n+1 \quad : from\ \ i\cdot 4-3\ \ to\ \ i\cdot 4 \quad (29)$$

So

$$(dec(\mathrm{s}) - n+1\ )/4 \quad : from\ \ i-3/4\ \ to\ \ i \quad (30)$$

The rounding to the upper bound of the numbers between *i*-3/4 and *i* give exactly the index *i*.

$$\lceil (dec(\mathrm{s}) - n+1\ )/4 \rceil : from\ \lceil i-3/4 \rceil = i\ \ to\ \ i \quad (31)$$

Whence, we have (26).

For example, if we use $\mathcal{S}_{PC}(4,3)$ then we have a single quadruplet (4, 5, 6, 7). For $\mathcal{S}_{PC}(8,4)$, we have two quadruplets (8, 9, 10, 11) and (12, 13, 14, 15). That can be verified with Table I and Table II. In each quadruplet, a syndrome is equal to the binary sum of three others. Therefore, to lock position *pos*, we consider change vector (in the quadruplet) corresponding to the sum of three others coset leaders whose binary syndromes sum is equal to the change vector e. That gives us another vector with three 1-bits at positions *i*, *j* and *l* different from *pos*. If one of these three positions is in $\mathcal{J}$ then we do the same as wet paper method with lookup tables. We search another change vector with 1-bits at positions not in $\mathcal{J}$. If a pair of triplet (*i*, *j*, *l*) is in $\mathcal{J}$, then we choose this pair else, if only one of them is in $\mathcal{J}$, we choose a pair containing this element. The chosen pair is then replaced by another pair not in $\mathcal{J}$ with the equalities of an equivalent of system (18).

- If syndrome s is in **Case 2**, then:

No problem arises because the change vector has all its components equal to 0.

- If s is in **Case 3** (i.e. e has two 1-bits at positions 1 and *pos* which, at least, one must be locked), then:

We search the pair (*i*, *j*) not in $\mathcal{J}$ with the equalities of system (18) or an equivalent. Thus, the change vector will have two 1-bits at positions *i* and *j*.

V. Complexity Comparison

To compare the complexity of algorithm PCS with the new algorithm, we measure the required time resources amount for solving the problem of minimizing the embedding impact (here, research of the change vector). To do this, we observe their execution time on a computer. We perform several tests on Dual Core CPU running at 3.46 GHz with 2 GB RAM. We chose a polar code of block length $n \in \mathcal{N}$ and dimension $k \in \mathcal{K}$ because our algorithm is applied to these values (see (23)). For each pair $(n,k) \in (\mathcal{N},\mathcal{K})$, 20 cover vectors and 20 messages are randomly generated. Then, we calculate the execution times average (in seconds) of messages embedding in cover vectors. This calculation is done for the two algorithms.

The obtained results are represented by the curves in Fig. 2. Each curve represents specifically the average execution times of the research algorithm of the change vector corresponding to the syndrome calculated from randomly generated cover vector and message. The execution time curve of PCS algorithm is blue and the red one represents the proposed new algorithm.

In Fig. 2, we see that the execution time of the new algorithm is lower than the PCS scheme [6]. The execution time for the two schemes is less than 0.025s. The difference between the two curves increases with the size of the cover vector *n*. This allows us to pronounce on complexity reducing. Therefore, the complexity of the new algorithm is lower than the previous PCS. Thus, the scheme proposed in this paper

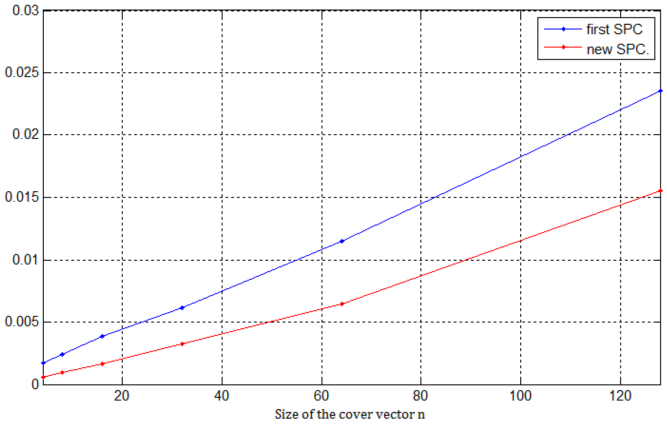

Figure 2. Execution time of the two schemes.

allows minimizing the embedding impact and give similar results as those of PCS scheme with a reduced time complexity.

VI. CONCLUSION

We proposed, in this paper, two new approaches that reduce the complexity of polar coding steganographic scheme [6]. The first approach uses lookup tables between possible syndrome configurations and corresponding coset leaders. To apply this method, we need to store tables in memory. To avoid this storage, a second approach, simpler, is proposed. It avails syndromes form to calculate the change vectors. A connection between the syndrome decimal value and non-zeros bits position of the cover vector is established. As with PCS scheme, this method allows minimizing the embedding impact with a reduced time complexity. The algorithm proposed has a time complexity lower than of the PCS scheme as shown by the comparison of the execution time curves of the two schemes.

As future research, we plan to propose an adaptive steganographic scheme based on polar codes as STC [4, 9]. We also plan to propose a steganalysis method to evaluate its security.

REFERENCES

[1] 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.

[2] Schönfeld, D., Winkler, A, “An Embedding with syndrome coding based on BCH codes,” In: Proceedings of the 8th ACM Workshop on Multimedia and Security, pp. 214 – 223, 2006.

[3] 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.

[4] Tomáš Filler, Jan Judas and Jessica Fridrich, “Minimizing embedding impact in steganography using trellis-coded quantization,” Department of Electrical and Computer Engineering SUNY Binghamton, USA, 2010.

[5] I. Diop, S. M. Farssi, M. Chaumont, O. Khouma, et H. B. Diouf, « Utilisation des codes LDPC en stéganographie, », COmpression et REprésentation des Signaux Audiovisuels (CORESA’2012), pp. 98-104, Lille, France, 24-25 mai, 2012.

[6] Birahime Diouf, Idy Diop, Sidi Mohamed Farssi, K. Tall, P. A. Fall, A. K. Diop, K. Sylla, “Using of polar codes in steganography,” In Proceedings of the 2nd International Conference on Advances in Computer Science and Engineering (CSE 2013), vol. 42, pp. 262-266, Atlantis Press, Los Angeles, July 1-2, 2013.

[7] R. Crandall, “Some notes on steganography,” Posted on Steganography Mailing List, 1998.

[8] J. Fridrich, M. Goljan, P. Lisonek and D. Soukal, “Writing on wet paper,” In IEEE Trans. on Sig. Proc., Third Supplement on Secure Media, vol. 53, pp. 3923–3935, Oct. 2005.

[9] Tomáš Filler, Jan Judas and Jessica Fridrich, “Minimizing additive distortion in steganography using syndrome-trellis codes,” IEEE Transactions on Information Forensics and Security, vol. 6, no. 3, September 2011.

[10] E. Arikan, “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] C. E. Shannon, “A mathematical theory of communication,” Bell System Tech. J., vol. 27, pp. 379–423, 623–656, July-Oct. 1948.

[12] N. Goela, S. B. Korada, and M. Gastpar, “On LP decoding of polar codes,” Submitted to IEEE Trans. Information Theory Workshop-ITW, 2010, Dublin.

[13] 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, pôle de Constantine, 2010.

[14] Tomáš Pevnỳ, Tomáš Filler and Patrick Bas, “Using high-dimensional image models to perform highly undetectable steganography,” Czech Technical University, Prague, Czech Republic; State University, New York in Binghamton, NY, USA; CNRS-LAGIS, Lille, France, 2010.