# Minimisation de l'impact d'insertion en stéganographie avec les codes polaires

Birahime Diouf[1], Idy Diop[1], Sidi Mohamed Farssi[1] et Marc Chaumont[2,3,4]

[1]Ecole Supérieure Polytechnique (ESP), Université Cheikh Anta Diop (UCAD), Dakar, Sénégal
[2]Université de Nîmes, F-30021 Nîmes Cedex 1, France
[3]Université de Montpellier 2, UMR5506-LIRMM, F-34095 Montpellier Cedex 5, France
[4]CNRS, UMR5506-LIRMM, F-34392 Montpellier Cedex 5, France

dioufbira11@yahoo.fr, idydiop@yahoo.fr, farsism@yahoo.com, marc.chaumont@lirmm.fr

## Résumé

*Ce document propose deux approches permettant de réduire la complexité de la stéganographie basée sur les codes polaires PCS (Polar Coding Steganography). La première est basée sur les tables de correspondance et la seconde exploite la forme du syndrome, calculé à partir du vecteur de couverture et du message, pour évaluer la position des modifications sur le médium de couverture. La méthode proposée dans ce document permet la minimisation de l'impact d'insertion et fournit des résultats similaires à ceux offerts par le schéma PCS avec une complexité temporelle réduite.*

## Mots clefs

Complexité, codes polaires, matrix embedding, stéganographie, tables de correspondance.

# 1 Introduction

La stéganographie est une technique de dissimulation d'informations qui permet de cacher un message dans un médium de couverture qui peut être une image (utilisée dans ce document), un son ou une vidéo. Son principal objectif est l'indétectabilité du message.

L'insertion doit se faire de façon à rendre, le moins perceptible possible, les modifications sur médium de couverture. Dans le domaine spatial, les bits du message secret sont insérés au niveau des LSBs (Least Significant Bits) des pixels de l'image de couverture. Pour améliorer cette méthode dite du LSB, plusieurs schémas basés sur les codes correcteurs d'erreurs ont été proposés. Ces schémas implémentent la technique de matrix embedding ; il s'agit d'une utilisation détournée des codes (rencontrés généralement dans la chaîne de transmission numérique). Parmi ces codes utilisés en stéganographie nous pouvons citer les codes de Hamming [1], BCH (Bose-Chaudhuri-Hocquenghem) [2, 3], STC (Syndrome Trellis Codes) [4] et LDPC (Low Density Parity Check) [5].

Après avoir introduits les codes polaires en stéganographie PCS (Polar Coding Steganography) [6], nous proposons dans ce document deux méthodes de minimisation de l'impact d'insertion permettant de réduire la complexité algorithmique. L'originalité de ce travail réside dans la définition d'un algorithme de calcul du vecteur de modifications en une seule étape comparé à PCS [6]. En effet nous proposons deux approches : la première est basée les tables de correspondance et la seconde exploite la forme du syndrome.

Ce document est organisé comme suit. La Section 2 fait un bref rappel de quelques concepts de base de la stéganographie, des tables de correspondance et des codes polaires. Dans la Section 3, nous présentons le schéma de stéganographie basé sur les codes polaires. Les méthodes de réduction de la complexité sont étudiées dans la Section 4. La Section 5 fait la comparaison en termes de complexité du schéma PCS [6] avec celle du nouvel algorithme proposé. La Section 6 conclut le document.

# 2 Définitions et concepts de base

Nous considérons le vecteur de couverture x constitué des LSBs de l'image de couverture en niveaux de gris $I \in \{0, \dots, 255\}^{n_1 \times n_2}$ représentée dans le domaine spatial, le message binaire m, le vecteur des modifications e, le vecteur stégo y ($y = x + e$) et une matrice de contrôle de parité H du code utilisé.

## 2.1 Concepts de base en stéganographie

Introduit par Crandall [7], la technique de matrix embedding est utilisée en stéganographie pour minimiser le nombre de modifications sur le médium de couverture. Elle est basée sur le décodage par syndrome des codes correcteurs d'erreurs.

### *2.1.1 Décodage par syndrome*

Comme son nom l'indique, ce type de décodage utilise le calcul de syndrome pour corriger les erreurs. Le mot reçu $r = z + e$ et la séquence d'erreurs e ont même syndrome :

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

avec z le mot de code correspondant au mot reçu r.

Le problème de décodage peut donc se résumer à trouver le vecteur e de poids minimal dans la classe latérale de r.

### *2.1.2 Matrix embedding*

Basé sur le décodage par syndrome, le matrix embedding consiste à trouver le vecteur stégo y le plus proche du vecteur de couverture x tel que:

$$\mathrm{yH}^{\mathrm{T}} = \mathrm{m} \quad (2)$$

En remplaçant y par $\mathrm{x} + \mathrm{e}$, nous avons :

$$\mathrm{eH}^{\mathrm{T}} = \mathrm{m} - \mathrm{xH}^{\mathrm{T}} \quad (3)$$

L'émetteur utilise (3). Son objectif est de trouver le vecteur e de poids minimal dans le coset $\mathcal{C}(\mathrm{m} - \mathrm{xH}^{\mathrm{T}})$. A la réception, l'extraction du message se fait avec (2).

### *2.1.3 Impact d'insertion et sa minimisation*

En supposant que les modifications n'interagissent pas entre elles, l'impact d'insertion total (ou distorsion) est la somme des modifications des différents pixels [4]:

$$D(\mathrm{x},\mathrm{y}) = \sum_{i=1}^{n} \rho_i |x_i - y_i| \quad (4)$$

avec $0 \leq \rho_i \leq \infty$ le coût du changement du LSB d'un pixel $x_i$ en $y_i$. Si $\rho_i = 1$ pour tout $i$ (profil constant), la minimisation de la distorsion $D$ se réduit à la minimisation du nombre de modifications sur l'image de couverture.
Les fonctions d'insertion et d'extraction sont respectivement définies par :

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

### *2.1.4 Les codes à papier mouillé*

Dans la pratique certains pixels de l'image de couverture peuvent être plus sensibles à la modification que d'autres. Les premiers appelés *pixels mouillés* (avec $\rho_i = \infty$) ne doivent pas être changés et les seconds dits *pixels secs* (avec $\rho_i = 1$) peuvent être modifiés. On parle dans ce cas de canal à papier mouillé [8]. Le codage par syndrome est également appliqué sur ce type de canal en utilisant les codes à papier mouillé [4, 6, 9].

## 2.2 Tables de correspondance

Une table de correspondance est une structure de données utilisée pour remplacer un calcul par une opération de consultation (recherche d'une valeur en mémoire) qui est souvent plus rapide. L'utilisation des tables de correspondance peut permettre de réduire le temps d'exécution d'un calcul complexe.
Schönfeld et Winkler [2] ont introduit les codes BCH en stéganographie en proposant deux manières de calcul de syndrome. Une première approche basée sur la recherche d'un leader de coset en utilisant une matrice de contrôle de parité et une seconde approche, utilisant un polynôme générateur pour la recherche des racines. Après le calcul du syndrome s, l'étape suivante consiste à chercher le vecteur de modifications e, leader du coset de s. Grâce aux tables de correspondance, on peut réduire le temps de calcul du leader du coset. En effet, l'utilisation des tables de correspondance par Zhang et al. [3] a permis de réduire la complexité comparée à l'approche basée sur la recherche exhaustive des racines [2].

## 2.3 Codes polaires

Un code polaire de longueur $n$ et de dimension $k$ sera noté $PC(n,k)$. $u_1^n$ désigne le mot d'information, $x_1^n$ le mot de code, $y_1^n$ le mot reçu, $\mathrm{G}_n$ une matrice génératrice, $W$ le canal de transmission et $u_1^i = (u_1, \ldots, u_i)$, avec $1 \leq i \leq n$. La capacité symétrique [11] de $W$ est notée par $I(W)$ et le paramètre de fiabilité par $Z(W)$. Plus $Z(W)$ est petite, plus $W$ est fiable [10]. $A$ et son complémentaire dans $\{1, \ldots, n\}$ $A^c$ désignent respectivement l'ensemble des bits d'information et celui des bits de redondance.
La construction des codes polaires est basée sur la polarisation de canal.

### *2.3.1 Polarisation et Transformation de canal*

-Polarisation de canal : elle consiste à faire la synthèse de $n$ copies indépendantes d'un B-DMC (Binary-input Discrete Memoryless Channel) $W$ donné pour construire $n$ autres canaux $W_n^{(i)}$, $1 \leq i \leq n$. Elle est composée de deux étapes : la combinaison de canaux et le partage de canal [10].

$$(W, W, \ldots, W) \xrightarrow{combinaison} W_n \xrightarrow{partage} \{W_n^{(i)}\}_{i=1,\ldots,n} \quad (6)$$

La combinaison de canal regroupe $n$ copies d'un B-DMC $W$ en un canal $W_n$. Elle est faite de façon récursive en associant deux copies de $W_{n/2}$ [6]. Lors du partage de canal, on subdivise $W_n$ en $n$ canaux $W_n^{(i)}$, $1 \leq i \leq n$.
-Transformation de canal : la polarisation de canal n'est en réalité qu'un processus de transformation récursive [10] :

$$\left(W_n^{(i)}, W_n^{(i)}\right) \xrightarrow[on\ construit]{} \left(W_{2n}^{(2i-1)}, W_{2n}^{(2i)}\right) \quad (7)$$

### *2.3.2 Le codage polaire*

La relation de codage polaire est définie par :

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

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

avec $\mathrm{B}_n$ une matrice de permutation et $\mathrm{G}_1 \triangleq [1]$.
Dans le codage polaire, si $u_1^n$ suit une distribution uniforme alors $W_n^{(i)}$ est le canal effectivement vu par $u_i$. Autrement dit, chaque bit $u_i$ emprunte le canal $W_n^{(i)}$ [6] (Figure 1). Le codage polaire utilise les canaux $W_n^{(i)}$ les plus fiables pour transporter les bits d'information et les canaux les moins fiables portent les bits de redondance :

$$Z(W_n^{(i)}) \leq Z(W_n^{(j)}) \quad (10)$$

pour $i \in A$ et $j \in A^c$.

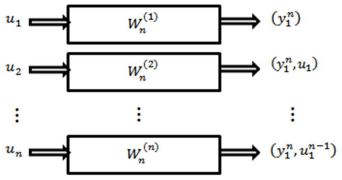


Figure 1 - *Schéma équivalent du codage polaire*

Le type de décodage le plus utilisé pour un code polaire est le Successive Cancellation (SC) [10]. Cependant, en stéganographie c'est le décodage par syndrome (défini dans la sous-Section 2.2.1) qui est utilisé.

# 3 Stéganographie par codage polaire

On notera $\mathcal{S}_{PC}(n, m = n - k)$ pour désigner la stéganographie basé sur un code polaire.

## 3.1 Schéma pour le profil constant

### *3.1.1 Première étape*

Pour déterminer une matrice de contrôle de parité d'un code polaire, on utilise le lemme donné par Goela et al. [12, Lemme 1] qui stipule que si les bits de redondance sont fixés à 0 alors la transposée de la matrice de contrôle de parité H du code polaire est donnée par les colonnes de la matrice génératrice $\mathrm{G}_n$ dont les indices sont dans $A^c$. En exploitant la forme particulière de H et sa transposée obtenue $\mathrm{H}^{\mathrm{T}}$, nous pouvons transformer les équations de la relation (2) en un système permettant de calculer les coefficients du vecteur stégo (voir Réf. [6]):

$$y_i = y_{i+1}H^T_{(i+1),j} + \cdots + y_n H^T_{nj} + m_j,\ j = 1, \ldots, n-k \quad (11)$$

avec $i$ la position du premier 1 sur la colonne $j$ de $\mathrm{H}^{\mathrm{T}}$.
Le vecteur y doit être initialisé au vecteur de couverture x avant les calculs. L'insertion se fait en gardant $k$ bits du vecteur de couverture x inchangés. Ainsi les modifications s'opèrent sur les $n-k$ bits restant de x. Nous obtenons ainsi un vecteur stégo, soit $\mathrm{y_p}$.
L'application de la méthode décrite ci-dessus nous donne une solution vérifiant $\mathrm{yH^T = m}$ mais elle n'est pas toujours la meilleure. On doit donc définit, à partir de $\mathrm{y_p}$, une méthode qui converge vers la solution optimale.

### *3.1.2 Optimisation de la première solution*

L'objectif avec cette étape est de trouver le vecteur stégo y le plus proche de x en utilisant le code polaire $PC(n,k)$. Soit $\mathrm{e_p}$ le vecteur de modifications correspondant au vecteur stégo $\mathrm{y_p}$ trouvé avec la première approche. Nous avons le problème suivant :
***Problème***:

- on a une première solution $\mathrm{e_p} \rightarrow$ solution initiale ;
- on cherche le vecteur e de poids minimal $\rightarrow$ problème de minimisation ;
- vérifiant $\mathrm{eH^T = m - xH^T = s} \rightarrow$ contraintes.

En considérant ces trois points nous avons un problème d'optimisation, notamment un problème de minimisation sous contraintes d'égalité avec comme solution initiale $\mathrm{e_p}$. Ce problème peut être modélisé comme suit :

$$\begin{aligned} &\arg\min_{\mathrm{e}} \quad f_{\mathrm{s.t}}(\mathrm{e}) = \langle \mathrm{c}, \mathrm{e} \rangle = \mathrm{c^T e} \\ &\mathfrak{C} = \begin{cases} \mathrm{e} \in \{0,1\}^n \text{ vecteur binaire} \\ \mathrm{eH^T = m - vH^T = s} \\ \mathrm{e_p} \text{ solution initiale} \iff \mathrm{e_p H^T = s} \end{cases} \end{aligned} \quad (12)$$

avec $f$ la fonction objective, $\mathrm{c} = \rho = \{\rho_i\}_{1 \le i \le n} = \{1\}^n$.
C'est un problème de programmation linéaire [13] écrit sous forme standard avec une contrainte supplémentaire ; le vecteur e est constitué d'éléments binaires.

## 3.2 Schéma pour une distorsion quelconque

L'algorithme d'insertion devra tenir compte des positions verrouillées. Nous devons définir le schéma de minimisation du nombre de positions modifiées tout en gardant inchangés les pixels verrouillés. Le problème est toujours de minimiser la distorsion $D$ qui peut être réécrite:

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

où $|x_i - y_i| = e_i$ et $\rho_i \in \{1, \infty\}$ pour le cas du papier mouillé ou $0 \le \rho_i \le \infty$ pour une distorsion arbitraire.
Les fonctions d'insertion et d'extraction seront :

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

La fonction objectif $f(\mathrm{e}) = < \mathrm{c}, \mathrm{e} >$ apparait sous forme de produit scalaire dans (13), avec $\mathrm{c} = \rho = \{\rho_i\}_{1 \le i \le n}$.
Comme pour le profil constant, nous avons un problème de programmation linéaire qui peut être résolu de la même manière [6, 13].
PCS [6] est composé de deux étapes dont chacune nécessite un temps de calcul. Le calcul de complexité n'y a pas était réalisé. Cet aspect sera pris en compte dans la définition du nouvel algorithme de stéganographie proposé dans ce document. Nous allons proposons deux approches permettant de réduire la complexité du schéma PCS.

# 4 Nouvelles méthodes proposées

Nous proposons deux méthodes l'une basée sur les tables de correspondance et l'autre exploite la forme du syndrome pour évaluer la position des modifications.

## 4.1 Méthode basée sur les tables de correspondance

Cette méthode utilise les tables de correspondance pour donner le vecteur de modifications à partir du syndrome. Celui-ci est donné par $\mathrm{s = eH^T = m - xH^T}$. Les cas de profil constant et du papier mouillé sont considérés.

### *4.1.1 Cas du profil constant*

La table de correspondance est un tableau composé de 2 colonnes de $2^{n-k}$ lignes chacune. La première colonne contient les différentes configurations de syndrome possibles et la seconde représente les leaders de cosets (ou vecteurs de modifications). Ainsi, sur une ligne $i$, avec $1 \le i \le 2^{n-k}$, nous avons le syndrome $\mathrm{s = S}[i]$ sur la première colonne et le vecteur de modifications $\mathrm{e = E}[i]$ sur la deuxième colonne. Après avoir calculé un syndrome s alors pour trouver e on procède comme suit :

- on parcourt les lignes de la colonne 1;

- si on trouve la position $i$ à laquelle on a $S[i] = s$ alors
- le coset correspondant au syndrome s se trouve à la même position $i$ dans la colonne 2.

L'opération de calcul de leaders de coset est ainsi remplacée par une consultation dans la table stockée. En guise d'exemples, on utilise un code polaire $PC(4,1)$ pour la stéganographie $\mathcal{S}_{PC}(4,3)$ et $PC(8,4)$ pour $\mathcal{S}_{PC}(8,4)$.

Exemple 1: Une matrice de contrôle de parité de $PC(4,1)$

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \text{ et } H^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (15)$$

Les colonnes $H_{.j}$ de H vérifient les égalités suivantes :

$$\begin{aligned} H_{.1} + H_{.2} &= H_{.3} + H_{.4} = (0\,0\,1)^T \\ H_{.1} + H_{.3} &= H_{.2} + H_{.4} = (0\,1\,0)^T \\ H_{.1} + H_{.4} &= H_{.2} + H_{.3} = (0\,1\,1)^T \end{aligned} \quad (16)$$

Le syndrome peut avoir une des trois configurations suivantes :

- **Config 1** : le syndrome est égal à une colonne de H;
- **Config 2** : il est égal au vecteur nul;
- **Config 3** : c'est la somme de 2 colonnes de H avec le système (16).

Ce qui nous permet de dresser le tableau suivant :

**Tableau 1.** Table de correspondance pour $\mathcal{S}_{PC}(4,3)$.

| colonne 1: syndromes | | | colonne 2: modifications | |
|---|---|---|---|---|
| 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 |

Considérons un message $m = (0\,1\,0)$ et un vecteur de couverture $x = (1\,0\,0\,1)$. Le calcul du syndrome donne $s = eH^T = m - xH^T = (0\,0\,1) = S[6]$. Donc le leader de coset correspondant est $E[6] = (1\,1\,0\,0) = e$.

Exemple 2:Une matrice de contrôle de parité de $PC(8,4)$

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \text{ et } H^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (17)$$

Comme à l'exemple précédent, nous avons le système :

$$\begin{aligned} H_{.1} + H_{.2} &= H_{.3} + H_{.4} = H_{.5} + H_{.6} = H_{.7} + H_{.8} = (0\,0\,0\,1)^T \\ H_{.1} + H_{.3} &= H_{.2} + H_{.4} = H_{.5} + H_{.7} = H_{.6} + H_{.8} = (0\,0\,1\,0)^T \\ H_{.1} + H_{.4} &= H_{.2} + H_{.3} = H_{.5} + H_{.8} = H_{.6} + H_{.7} = (0\,0\,1\,1)^T \\ H_{.1} + H_{.5} &= H_{.2} + H_{.6} = H_{.3} + H_{.7} = H_{.4} + H_{.8} = (0\,1\,0\,0)^T \\ H_{.1} + H_{.6} &= H_{.2} + H_{.5} = H_{.3} + H_{.8} = H_{.4} + H_{.7} = (0\,1\,0\,1)^T \\ H_{.1} + H_{.7} &= H_{.2} + H_{.8} = H_{.3} + H_{.5} = H_{.4} + H_{.6} = (0\,1\,1\,0)^T \\ H_{.1} + H_{.8} &= H_{.2} + H_{.7} = H_{.3} + H_{.6} = H_{.4} + H_{.5} = (0\,1\,1\,1)^T \end{aligned} \quad (18)$$

Nous avons **Config 1, 2, 3** de l'exemple précédant et le **Tableau 2**.

**Tableau 2.** Table de correspondance pour $\mathcal{S}_{PC}(8,4)$.

| colonne 1: syndromes | | | colonne 2: modifications | |
|---|---|---|---|---|
| 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 |

Soient le vecteur de couverture $x = (1\,0\,1\,0\,1\,0\,0\,1)$ et le message $m = (1\,1\,0\,1)$. Le syndrome est $s = (1\,1\,0\,0) = S[5]$ et le leader de coset est $E[6] = (0\,0\,0\,0\,1\,0\,0\,0) = e$.

### *4.1.2 Cas du papier mouillé*

Soit $\mathcal{J}$ l'ensemble des positions à verrouiller (éléments mouillés). Si on trouve un vecteur de modifications e avec un bit égal à 1 (un 1-bit) à une position $pos$, devant être verrouillée, alors une des trois configurations se présente.

- Le syndrome calculé s se trouve dans la **Config 1** (c'est-à-dire le vecteur e a un seul 1-bit à une position à verrouiller) et on procède comme suit :

On subdivise la première moitié de la table de correspondance composée de $n$ éléments en $n/4$ sous-ensembles de 4 éléments chacun. Dans chaque sous-ensemble, un syndrome est égal à la somme des 3 autres. Par conséquent, pour verrouiller la position $pos$, on considère le vecteur de modifications (dans le sous-ensemble du syndrome calculé) correspondant à la somme des 3 vecteurs dont la somme de leurs syndromes est égale au syndrome du vecteur de modifications e. Ce qui nous donne un autre vecteur ayant 3 positions $i$, $j$ et $l$ égales à 1 et différentes de $pos$. Si une de ces 3 positions est dans $\mathcal{J}$ alors on cherche un autre vecteur avec des 1-bits à des positions non-verrouillées. Pour ce faire, on remplace le couple du triplet $(i, j, l)$ appartenant à $\mathcal{J}$ par un autre qui ne soit pas dans $\mathcal{J}$ avec les égalités d'un équivalent[1] du système (18).

- Le syndrome est dans la **Config 2**, alors :

Aucun problème ne se pose puisque le vecteur des modifications a tous ses composants nuls.

- Le syndrome est dans la **Config 3** (c'est-à-dire e a deux 1-bits aux positions 1 et $pos$ dont au moins l'une doit être verrouillée), alors :

On cherche le couple $(i, j)$ n'appartenant pas à $\mathcal{J}$ avec les égalités du système. Le vecteur de modifications aura ainsi deux 1-bits aux positions $i$ et $j$.

[1] On parle d'équivalent pour désigner le système obtenu dans le cas d'une stéganographie avec un paramètre différent de 8. Par exemple, pour la stéganographie $\mathcal{S}_{PC}(4,3)$, le système est (16).

Exemple 3 : Considérons toujours le vecteur de couverture $x = (1\,0\,1\,0\,1\,0\,0\,1)$ et le message $m = (1\,1\,0\,1)$. Le syndrome est $s = (1\,1\,0\,0) = S[5]$ et le premier vecteur de modifications trouvé est $E[6] = (0\,0\,0\,0\,1\,0\,0\,0) = e$. Si $\mathcal{J} = (5,6,7)$ alors le seul 1-bit est à la 5-ème position qui doit être verrouillée. Ainsi, on subdivise la première moitié du **Tableau 2** en 2 sous-ensembles de 4 éléments. Le syndrome s se trouve dans le deuxième sous-ensemble. Ainsi, $s = (1\,1\,0\,0) = S[5] = S[6] + S[7] + S[8]$ et $e = E[5] = E[6] + E[7] + E[8] = (0\,0\,0\,0\,0\,1\,1\,1)$. Ce vecteur e a trois 1-bits aux positions 6, 7 et 8. Le couple $(6,7)$ étant dans $\mathcal{J}$; nous devons donc le remplacer par autre couple avec les égalités du système (18). D'après la troisième ligne, nous pouvons remplacer le couple $(6,7)$ soit par $(1,4)$ ou par $(2,3)$. Nous pouvons donc choisir entre $e = (1\,0\,0\,1\,0\,0\,0\,1)$ ou $e = (0\,1\,1\,0\,0\,0\,0\,1)$ pour le vecteur de modifications.
Cet exemple est le pire cas de figure qui puisse se présenter puisque les deux autres configurations sont plus faciles à résoudre.

## 4.2 Méthode de calcul direct du vecteur des modifications

Nous allons considérer le cas de profil constant et du papier mouillé comme nous l'avions fait avec la méthode des tables de correspondance précédente.

### *4.2.1 Profil constant*

La méthode que nous allons proposer exploite une correspondance uniforme entre la valeur du syndrome et la position des coefficients non-nuls du vecteur de modifications correspondant.
Par exemple, sur le **Tableau 2**, les observations suivantes peuvent être faites :

- **Cas 1** : sur la première moitié du tableau (de l'élément 1 à l'élément $8 = n$), les syndromes ont leur premier bit (le bit de poids fort) égal à 1 et leur valeur décimale varie entre $8 = n$ et $15 = 2n - 1$. La position du seul 1-bit du vecteur de modifications va de 1 à $8 = n$ (de la gauche vers la droite);
- **Cas 2** : à la $9 = (n + 1)$-ème ligne, le syndrome et le vecteur de modifications sont tous des vecteurs nuls;
- **Cas 3** : sur la partie restante du tableau (du $10 = n +$ 2-ème au $16 = 2n$-ème élément), les syndromes ont comme premier bit 0 et leur valeur décimale varie entre 1 et $7 = n - 1$. Les vecteurs de modifications ont deux 1-bits le premier à la position 1 et le second à une position qui va de 2 à $8 = n$.

Les trois parties du **Tableau 2** sont séparées par des traits épais. Ces parties et les remarques associées sont aussi valables pour le **Tableau 1 (**avec $n = 4$) et pour les autres tableaux correspondants aux systèmes équivalents. D'après ces remarques, il existe une relation entre la valeur décimale du syndrome s et la position des éléments non-nuls du vecteur de modifications e. Cela est dû au fait que, sur les colonnes de la matrice de contrôle de parité H, figurent les valeurs binaires des nombres de $n$ à $2n - 1$.

Cependant, une condition nécessaire est que le nombre de syndromes doit être égal au double de la taille du vecteur de couverture.

$$2^{n-k} = 2n \tag{19}$$

or $n$ est une puissance de 2, soit $n = 2^p$, alors

$$2^{n-k} = 2 \cdot 2^p = 2^{p+1} \tag{20}$$

Donc

$$n - k = p + 1 \tag{21}$$

Ainsi

$$k = n - 1 - \log_2 n = 2^p - 1 - p \tag{22}$$

Les codes polaires $PC(4,1)$ et $PC(8,4)$, donnés en exemples avec la méthode des tables de correspondance sont tels que (22) est vérifiée. En effet, pour $PC(4,1)$ on a $k = 1 = 4 - 1 - \log_2 4 = 2^2 - 1 - 2$ et $k = 4 = 8 - 1 - \log_2 8 = 2^3 - 1 - 3$ pour $PC(8,4)$.
La validité des observations concernent les valeurs de

$$\begin{aligned} p &\in \{2,3,4,5,6,7\} = \mathcal{P} \\ n &\in \{4,8,\dots,128\} = \mathcal{N} \\ k &\in \{1,4,\dots,120\} = \mathcal{K} \end{aligned} \tag{23}$$

avec $n = 2^p$, $k = 2^p - 1 - p$ et $p \in \mathcal{P}$.
Pour un code polaire $PC(n,k)$ [10], la longueur $n$ est une puissance de 2 et la dimension $k$ est un entier positif dans $\{1, 2, \dots, n-1\}$. Concernant un schéma de stéganographie par codage polaire, la relation d'optimalité [6] est $m = n - k > p = \log_2 n$. Les paramètres de notre code polaire dans l'approche proposée vérifient cette condition d'optimalité car on a $n - k = p + 1 > p$.
L'insertion se fait par couple d'un segment du médium de couverture et d'un segment de message. Le nombre de segments de couverture doit être supérieur ou égal au nombre de segments de message.
Nous proposons l'algorithme suivant:

**Algorithme 1.** Calcul d'un leader de coset d'un syndrome

*Initialisation:*
*p ⟵ un élément de $\mathcal{N}$; $n \leftarrow 2^p$; $k \leftarrow n - 1 - p$;*
e $\leftarrow 0_1^n$; y $\leftarrow$ x *;*
*Déroulement:*
*Si* $xH^T \neq m$ *alors* $s \leftarrow m - xH^T$
*on calcule la valeur décimale du vecteur syndrome binaire*
*($dec \leftarrow conversionDecimale(s)$)*
*si le $1^{er}$ coefficient du syndrome est égal à* 1 *alors*
*on met le $(dec + 1 - n)$-ème coefficient du vecteur* e *à* 1;
*sinon*
*on met le $1^{er}$ et le $(dec + 1)$-ème coefficient de* e *à* 1 ;
*fin (sinon)*
*fin (si)*
*Fin (Si)*

La fonction $conversionDecimale(s)$ permet de faire la conversion d'un vecteur binaire $s$ à sa valeur décimale.

### *4.2.2 Papier mouillé*

Soit un vecteur de modifications e avec au moins un 1-bit à une position à verrouiller, on distingue trois cas :

– le syndrome calculé s se trouve dans le **Cas 1** (e a un seul 1-bit à la position $pos$ à verrouiller) alors

On considère, par quadruplet, les valeurs décimales des syndromes correspondant à des vecteurs de modifications avec une seule position à 1-bit. Les syndromes considérés (au nombre de $n$) représentent la moitié du nombre de configurations de syndromes possibles (qui est de $2n$) et le nombre de quadruplets est de $n/4$. Les valeurs décimales vont de $n$ à $2n-1$ (voir par exemple **Tableau 2** avec $n=8$). Les valeurs décimales des syndromes et leurs quadruplets $Q_i$ se présentent comme suit :

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

Une généralisation de cette représentation donne

$$Q_i\colon\ n+(i-1)\cdot 4\ \rightarrow\ n+(i)\cdot 4-1,\ \text{avec } i=1,\dots,n/4 \quad (25)$$

Pour connaître le quadruplet $Q_i$ auquel appartient un syndrome s, on calcule son indice $i$ par

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

avec $\lceil . \rceil$ l'opération d'arrondi supérieur et $dec(\mathrm{s})$ la valeur décimale du syndrome s.

*Preuve*: D'après (25), la valeur décimale d'un syndrome s varie entre $n+(i-1)\cdot 4$ et $n+(i)\cdot 4-1$. On écrit :

$$dec(\mathrm{s})=de\ \ n+(i-1)\cdot 4 \quad \text{à} \quad n+i\cdot 4-1 \quad (27)$$

Ce qui équivaut à

$$dec(\mathrm{s})-n=de\ 4\cdot i-4 \quad \text{à} \quad i\cdot 4-1 \quad (28)$$

Ainsi

$$dec(\mathrm{s})-n+1=\ de\ \ 4\cdot i-3 \quad \text{à} \quad i\cdot 4 \quad (29)$$

Donc

$$(dec(\mathrm{s})-n+1)/4\ \ =\ de\ \ i-3/4 \quad \text{à} \quad i \quad (30)$$

Le nombre $i$ représente l'arrondi à la borne supérieure des nombres compris entre $i-3/4$ et $i$. D'où (26).

Par exemple pour $\mathcal{S}_{PC}(4,3)$, nous avons un seul quadruplet $(4,5,6,7)$ et pour $\mathcal{S}_{PC}(8,4)$, nous avons deux quadruplets $(8,9,10,11)$ et $(12,13,14,15)$. Ce que nous pouvons vérifier avec les tableaux **1** et **2**. Dans chaque quadruplet, un syndrome est égal à la somme bit-à-bit des 3 autres syndromes. Par conséquent, pour verrouiller la position $pos$, on considère le vecteur de modifications (dans le quadruplet) correspondant à la somme des 3 leaders de cosets dont la somme de leurs syndromes est égale au syndrome de notre vecteur $e$. Ce qui nous donne un autre vecteur ayant 3 positions $i$, $j$ et $l$) non-nuls différentes de $pos$. Si une de ces 3 positions est dans $\mathcal{J}$ alors on procède de la même façon qu'avec la méthode du papier mouillé avec les tables de correspondance. On cherche ainsi un autre vecteur de modifications avec des 1-bits à des positions non-verrouillées. Si un couple du triplet $(i,j,l)$ est dans $\mathcal{J}$, alors on choisit ce couple sinon, si un seul des trois éléments est dans $\mathcal{J}$, on prend un couple contenant cet élément. Le couple choisi est ensuite remplacé par un autre couple qui n'est pas dans $\mathcal{J}$ avec les égalités d'un équivalent du système (18).

- Si le syndrome est dans le **Cas 2**, alors

Aucun problème ne se pose car le vecteur des modifications a tous ses composants nuls.

- Si le syndrome est dans le **Cas 3** (le vecteur e a deux 1-bits aux positions 1 et $pos$ dont au moins l'une est à verrouiller), alors

On cherche le couple d'indices $(i,j)$ n'appartenant pas à $\mathcal{J}$ avec les égalités du système. Le vecteur de modifications aura ainsi deux 1-bits aux positions $i$ et $j$.

# 5 Comparaison des complexités

Afin de comparer la complexité de l'algorithme de PCS avec celle de la nouvelle méthode, nous mesurons la quantité de ressources (en temps) nécessaire pour la résolution du problème de minimisation de l'impact d'insertion. Pour ce faire, nous avons observé leur temps d'exécution sur un ordinateur. Nous avons fait plusieurs tests à partir d'un ordinateur équipé d'un processeur Intel Pentium Dual CPU 3.46GHz et doté d'une mémoire physique total de 2Go. Nous avons choisi un code polaire de longueur de bloc $n \in \mathcal{N}$ et de dimension $k \in \mathcal{K}$ car notre algorithme s'applique avec ces valeurs (cf. relation (23)). Pour chaque couple $(n,k) \in (\mathcal{N},\mathcal{K})$, on génère de façon aléatoire 20 vecteurs de couverture et 20 messages. On calcule ensuite la valeur moyenne des temps d'exécution (en seconde) de l'insertion des messages dans les vecteurs de couverture. Ce calcul est fait pour les deux algorithmes.

Les résultats obtenus sont représentés par les courbes de la Figure 2. Chaque courbe représente la durée moyenne du temps d'exécution de l'algorithme de recherche du vecteur de modifications correspondant au syndrome calculé à partir du vecteur de couverture et du message générés de façon aléatoire. La courbe de du temps d'exécution de l'algorithme de PCS est en bleu et celle en rouge représente celui du nouvel algorithme proposé.

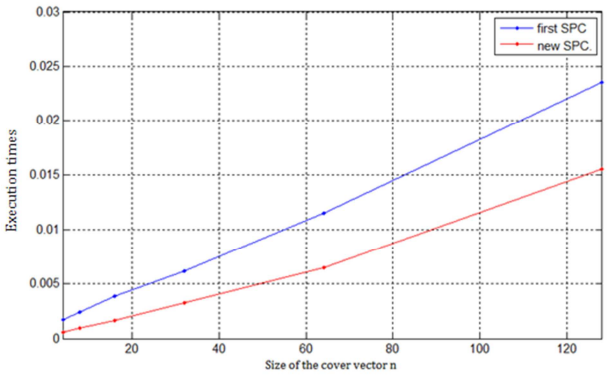


Figure 2 – *Temps d'exécution des deux schémas*

Sur la Figure 2, nous voyons que le temps d'exécution du nouvel algorithme est plus faible que celui du schéma PCS [6]. Le temps d'exécution pour les deux schémas reste inférieur à $0.025s$. L'écart entre les deux courbes augmente avec la taille du vecteur de couverture $n$. Ceci nous permet de nous prononcer sur la réduction de la complexité. Par conséquent, la complexité du nouvel algorithme est plus faible que celle du précédent PCS.

# 6 Conclusion

Nous avons proposé, dans ce document, deux nouvelles approches qui permettent de réduire de façon significative la complexité du schéma de stéganographie basé sur les codes polaires [6]. La première approche utilise les tables de correspondance entre les configurations de syndrome possibles et les séquences de leaders de cosets correspondantes. Pour appliquer cette méthode, nous avons besoin de stocker les tables en mémoire et au besoin les consulter. Afin d'éviter ce stockage, une seconde approche, plus simple, est proposée. Elle exploite la forme des syndromes pour calculer les leaders de cosets. Une relation entre la valeur décimale du syndrome et la position des bits du vecteur de couverture devant être modifiés est établie. Ces deux approches permettent de réduire la complexité du schéma PCS comme le montre la comparaison des courbes de complexité de la méthode utilisée par PCS et celle proposée dans ce document.

Dans le cadre de nos futures recherches, nous prévoyons de proposer un schéma de stéganographie adaptive basé sur les codes polaires comme le cas du STC [4, 9]. Nous prévoyons également de proposer une méthode de stéganalyse pour évaluer sa sécurité.

## Références

[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. 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.

[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.