New Steganographic scheme based of Reed-Solomon codes

I. DIOP; S .M FARSSI ;O. KHOUMA ; H. B DIOUF ; K .TALL ;
K .SYLLA

Ecole Supérieure Polytechnique de l'Université Dakar Sénégal

Email: idydiop@yahoo.fr; farsism@yahoo.com;

Abstract

Modern steganography [1] is a new science that makes a secret communication. Using the technique of Matrix Embedding in steganography schemes tends to reduce distortion during insertion. Recently, Fontaine and Galand [2] showed that the Reed-Solomon codes can be good tools for the design of a Steganographic scheme. In this paper, we present an implementation of the technique Matrix Embedding using the Reed-Solomon codes. The advantage of these codes is that they allow easy way to solve the problem of bounded syndrome decoding, a problem which is the basis of the technique of embedding matrix.

KEYWORDS: *Matrix Embedding, Steganography, Reed-Solomon code.*

1 Introduction

The steganography or the art of the secret communication consists to dissimulate a message in any medium (picture, video, audio….) so that the hidden message is furtive to the eyes of anyone other than the sender and the recipient.

The main works in this domain are characterized by the minimization of the number of embedding changes. Many Steganographic scheme based on the error correcting codes (BCH, HAMMING…) were proposed [3] [4] and the works of Galand and Fontaine showed that the Reed-Solomon codes can be used in this domain.

Some methods present insufficiencies sometimes related to the size of the messages to be inserted, to the complexity of the decoding programs of the error correcting codes, to the problem of locking of the symbols and the measurement of distortion. All these reasons push the researchers today not only to improve the schemes existing but also to propose other approaches.

The work that we propose presents implementation of the Steganographic scheme based on the Reed-Solomon codes by using the method of computation by prolonged syndrome in order to determine the vector of change.

2 Steganography and error correcting code

2.1 Steganography

A modern stéganographie scheme is based on two algorithms:

DOI : 10.5121/ijdps.2012.3207

- Embedding function $Emb(x, m)$: embedding algorithm of the message m in a cover-medium X. It turns over a stego-medium Y.
- Extraction Function Ext (Y): extraction algorithm of the message dissimulated in stego-medium Y. It turns over the message m.

The word detection is also used when it is about of checking the presence of information (represented thanks to a signal, a particular characteristic of the medium…) in the stego-medium, without to want to extract it Several techniques of Steganographic scheme have been proposed among which one can retain the technique “matrix embedding”.

Matrix embedding is a general principle that can be applied to most Steganographic schemes to improve their embedding efficiency, which is defined as the expected number of random message bits embedded per one embedding change. [5]

All the trick consists in modifying the picture (cover-medium) so that the syndromes calculated at the reception represent the message and also so that the picture is the less modified. Let $x \in F_q^n$ be a vector extracted of the cover-medium and $m \in F_q^{n-k}$ the message. It is about inserting the message m in x by modifying it least possible.

For that the technique is to modify x to $y \in F_q^n$ such as $Hy = m$

Where H represents the matrix of parity control of the code.

To construct y, we need a word e such that its syndrome is $m - Hx$. Thus, we can set $y = e + x$, which leads to $Hy = Hx + He = Hx + m - Hx = m$

Moreover, the Hamming weight of y $(w(y))$ is precisely the number of changes we apply to go from x to y; so, we need $w(y) \leq R$, where R is equal to the covering radius of the code. The covering radius R is the maximum distance between any word of F_q^n and the set of all codewords.

The classical Hamming weight $w(x)$ of a word x is the number of coordinates that is different from zero, and the Hamming distance $d(x, y)$ between two words x, y denotes the weight of their difference, that is, the number of coordinates in which they differ.

2.2 Reed-Solomon codes

2.2.1 Definition :

The Reed-Solomon codes are linear blocks codes with non binary symbols [6].

Let $x_1, \dots, x_n \in F_q$ be a Galois Field with $q = 2^m$ elements where m is the number of bits per symbol, two to two distinct, with $n < q$,

The evaluation function ev associated to $x_1, \dots, x_n$ is

$$ev: \quad F_q[x] \rightarrow F_q^n$$
$$f \quad \rightarrow (f(x_1), \dots, f(x_n))$$

Let $L_k = \{f \in F_q[x];\ deg\, f < k\}$, be the set of the polynomials of degree lower than k with $0 \leq k \leq n$. The Reed-Solomon code of dimension k is

$$C = ev(L_k).$$

2.2.2 Parameters:

The parameters of a Reed-Solomon code correcting errors t in block of n symbols q-ary are thus:

$n = q - 1 = 2^m - 1$; $k = n - 2t$; $d_{min} = n - k + 1$.

where n corresponds to the number of symbol of a codeword, k corresponds to the number of symbol of the parity segment and d_{min} corresponds to the minimal distance of the code.

The covering radius R of a Reed-Solomon code of length n and dimension k is known and is equal to $R = n - k$.

2.2.3 Codage

- ✓ Generator polynomial

Reed-Solomon codes are valid if and only if they are divisible by their generator polynomial $g(x)$.

The generator polynomial is :

$$g(x) = (x + \alpha^1)(x + \alpha^2) \ldots (x + \alpha^{2t})$$

$$g(x) = g_{2t}x^{2t} + g_{2t-1}x^{2t-1} + \cdots + g_1x + g_0 \qquad (1)$$

where $g_i \in GF(2^m)$ and α^i is a roots of $g(x)$

- ✓ Theory of Coding

The equation defining the systematic encoding of Reed – Solomon $codes\ (n, k)$ is :

$$c(x) = m(x)x^{n-k} + [m(x)x^{n-k}]mod(gx) \qquad (2)$$

where :

$c(x)$: denotes codeword polynomial of degree $n - 1$

$m(x)$: denotes information polynomial of degree $k - 1$

$[m(x)x^{n-k}]mod(gx)$: denotes parity polynomial of degree $n - k - 1$

$g(x)$: denotes generator polynomial of degree $n - k$

Systematic coding means that information is coded in the high degree of the codeword and that the redundancy symbols are introduced after the information words.

Let $m(x) = m_0 + m_1x + \cdots + m_{k-1}x^{k-1}$ be the message polynomial to be encoded, $m_i \in GF(2^m)$ and $k = n - 2t$. Dividing $x^{n-k}m(x)$ by $g(x)$, we have

$x^{n-k}m(x) = a(x)g(x) + b(x)$ were

$b(x) = b_0 + b_1x + \cdots + b_{2t-1}x^{2t-1}$ is a remainder. Then $b(x) + x^{n-k}m(x) is$ the codeword polynomial for the message $m(x)$.

2.2.4 Decoding

The problem of decoding is a difficult work, for which certain algorithms have very bad complexities.

According to the paper of Madhu Sudan [7], one defines the problem of decoding as follows:

NCP (Nearest Codeword Problem) the problem is to find the codeword nearest to the word received to the sense of the Hamming distance.
LD (List Decoding) a bound e is given. The problem is to find all (possibly none) the code words at distance at most e from the received word.
BDD (Bounded Distance Decoding) a bound e is given. The problem is to find a word among the code words at distance e from the received word.
UD (Unambiguous Decoding) Here one gives itself $e = (d-1)/2$, where d is the minimal distance of the code, and one looks for the codeword at distance e from the received word (if he exists).

In a classic way, the problem studied in theory of the codes is this last problem (UD). It has been solved in the case of the Reed-Solomon codes, efficiently.

Classic decoding

Let $c(x)$, the codeword transmitted then , the received word $r(x)$ can be always put in the form:

$$r(x) = c(x) + e(x)$$

where $e(x)$ is the error polynomial

$$e(x) = e_0 + e_1x + \cdots + e_{n-1}x^{n-1}, \ \ e_i \in F_q \ \ \forall\, i$$

when $e_i \neq 0$, an error is in position i.
the objective of decoding is to find the codeword emitted from the word received , potentially erroneous.
The decoding of the Reed-Solomon codes can be achieved from a vector with 2t component $S = [S_1 \cdots S_{2t}]$, called syndrome.

$$S_i = r(\alpha^i) = e(\alpha^i), \qquad i = 1, \cdots ,2t \qquad\qquad (3)$$

Decoding consists of the following steps :
- Syndrome computation.
- determination of error-locator polynomial and the error-evaluator polynomial
- finding the roots and evaluation of the two polynomials
- Summation of the polynomial made up and the polynomial received to reconstitute the starting information without error.

The error-locator polynomial and the error evaluator polynomial can be calculated thanks to Euclide algorithm, Berlekamp-Massey algorithm… [8][9].

List decoding

The list decoding can be decoding e errors with $e > d/2$, where d is the minimal distance from the code to be decoded.
The list decoding is done in two steps
A step of interpolation we search a polynomial $Q(x, y)$ satisfying certain conditions.
A step of factorization/search for roots of $Q(x, y)$
V. Guruswami and M. Sudan quickly proposed a radical improvement of the Sudan's algorithm. It makes it possible to always decode more than the classic radius of decoding single decoding. They introduce in the step of interpolation a parameter of multiplicity.

Many researchers attempt to find the best algorithms to achieve the two steps of the algorithms of list decoding, in order to make them completely effective so that one can program them in electronic circuits.

Classic decoding, containing syndrome, is much faster than decoding in list, which is even heavier, although polynomial.

3 New Steganographic scheme based on Reed-Solomon codes

Galand and Fontaine showed in their paper that the Reed-Solomon codes are good candidates for a design of effective Steganographic scheme. We also consider a new and more general problem, mixing wet papers (locked positions) [10] and simple syndrome coding (low number of changes) [11] in order to face not only passive but also active wardens. However, the implementation of Steganographic scheme is delayed by the complexity of Guruswami-Sudan list decoding, although offering an adaptive compromise between the number of locked positions and the number of changes. Reason for which, to get round this difficulty, we used in this work the techniques of decoding to basis of syndrome.

We employed the calculation of the prolonged syndrome in order to find our vector e inserting our message m.
Let $c(x) \in F_q^n$, be the polynomial representation of the vector extracted from a cover-medium, $r(x) \in F_q^n$ the polynomial representation of the vector stego-medium, $m(x) \in F_q^{n-k}$ the polynomial of the vector of information. We want to modify $c(x)$ into $r(x)$ such that $m(x)$ is embedded in $r(x)$, changing at most R coordinates in $c(x)$. Let us notice that we will regard $c(x)$ as being our codeword.

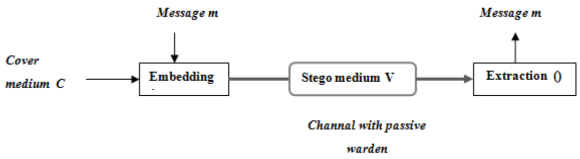


Figure 1-System of Steganography

3.1 Embedding

We can set $r(x) = c(x) + e(x)$ where $e(x)$ the polynomial representation of the vector of information is.
At the reception

$$Hr = m, \Longrightarrow H(c+e) = m \Rightarrow Hc + He = m \Rightarrow He = m - Hc$$

The vector e searched is a word such that its syndrome is $m - Hx$

Algorithm 1: matrix embedding using Reed-Solomon codes

1. Read the next n symboles c from the cover medium and read the next message segment of length $n - k$.
2. Calculate the syndrome Hc.
3. Find vector e such $as\ He = m - Hc$
4. Modify the cover medium c so that $= c + e$; we obtain the stego-medium r.
   If there are no more message symbols to be embedded, stop, otherwise go to 1.

✓ Description of step 3

We can set $S = He = m - Hc$

$$S_i = m_i - c(\alpha^i) \qquad where\ \ i = 1, \ldots, (n-k)$$

What gives the equations following:

$$S_1 = e_{j_1}\alpha^{j_1} + \cdots + e_{j_n}\alpha^{j_n}$$
$$S_2 = e_{j_1}\alpha^{2j_1} + \cdots + e_{j_n}\alpha^{2j_n}$$
$$\vdots$$
$$S_{n-k} = e_{j_1}\alpha^{2tj_1} + \cdots + e_{j_n}\alpha^{(n-k)j_n}$$

The syndrome in polynomial form is:

$$S(x) = S_1 + S_2 x + \cdots + S_{n-k}x^{n-k} + \cdots \quad (4)$$

Let $e = (e_0, \ldots, e_{n-1})$, be the vector of modification of weight w. we define the support of the vector e

$$I = \{i \in \{0, \ldots, n-1\} | e_i \neq 0\}$$

and

$\{X_1, \ldots, X_w\} = \{\alpha^i | i\ \in I\}$, the locators of the modifications of e. We have,

The polynomial syndrome is

$$S(x) = S_0 + S_1 x + S_{n-k-1}x^{n-k-1}$$

The polynomial locator of modification of e is

$$\Lambda(x) = \prod_{i=1}^{w}(1 - X_i\, x) = 1 + \Lambda_1 x + \cdots + \Lambda_w x^w, \qquad (5)$$

The polynomial evaluator of modification of e is

$$W(x) = \sum_{i=1}^{w} e_i X_i \prod_{j=1, j\neq i}^{w}(1 - X_j x) \qquad (6)$$

These three polynomials satisfy the following key equation

$$\frac{w(x)}{\Lambda(x)} = S(x) \quad mod\ x^{n-k+1} \qquad (7)$$

If the two polynomials are known, then we find the X_i by factorization or search for roots (« search for Chien [12] ») and then, we find the e_i by the formula:

$$w(\alpha^{-i}) = e_i \alpha^i \Lambda'(\alpha^{-i}), \qquad (8)$$

where $\Lambda'(x)$ is the derivative of $\Lambda(x)$ (formulas of Forney [13]).

✓ Computation of the coefficients e_i by transform

It is possible to calculate the coefficients $e_i,\ i = 0, 1, \ldots, (n-1)$ of the error polynomial $e(x)$ without determining the roots of the error locator polynomial $\Lambda(x)$. For that we introduce the syndrome prolonged $S^*(x)$ defined by:

$$S^*(x) = w(x)\frac{1 + x^n}{\Lambda(x} = \sum_{i=1}^{n} S_i x^i \qquad (9)$$

The coefficient e_i is equal to zero (not error) if α^{-i} not root of the error locator polynomial $\Lambda(x)$. in this case, we have $S^*(\alpha^{-i}) = 0$ since $\alpha^{-in} = 1$ (to remember that $n = q - 1$ and $\alpha^{q-1} = 1$).

If α^{-i} is root of the locator polynomial, the e_i coefficient will be not equal to zero (presence of an error) and $S^*(\alpha^{-i})$ is of the form $0/0$. This indetermination can be raised while calculating the derivative of the numerator and the denominator of the expression (9).

$$S^*(\alpha^{-i}) = w(\alpha^{-i})\frac{n\alpha^{-i(n-1)}}{\Lambda'(\alpha^{-i})}$$

By using the equation (8) and by taking account of the fact that $\alpha^{-i(n-1)} = \alpha^i$ and that $na = a$ for n odd in a Galois Field, the coefficient e_i is equal to:

$$\boldsymbol{e_i = S^*(\alpha^{-i})} \qquad (10)$$

Computation of the prolonged syndrome makes itself from the polynomials $\Lambda(x)$ and $w(x)$ by using the following relation deducted of the expression (9).

$$\Lambda(x)S^*(x) = w(x)(1 + x^n) \qquad (11)$$

Another way to get the prolonged polynomial consists in dividing $w(x)(1 + x^n)$ by $\Lambda(x)$ according to the increasing powers

The S_i coefficients of the prolonged syndrome are identical to those of the syndrome $S(x)$ for i going of 1 to $2t$ and are determined while annulling the coefficients of the terms in the product $\Lambda(x)S^*(x)$, for j going of $2t + 1\ to\ n$.

3.2 Extraction

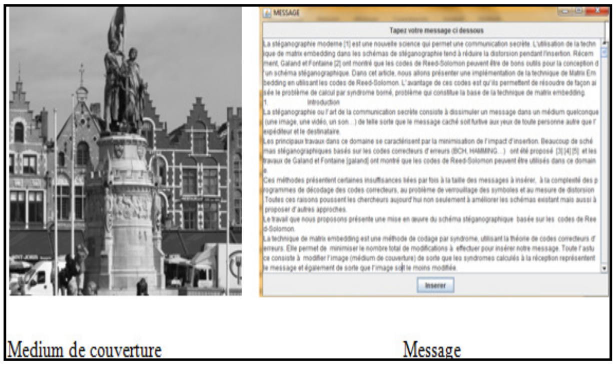
With the reception, we recover the medium stégo (image steganography), we need also to know the size of our message for the extraction. To extract the message we will proceed as follows:

1 Read the n symbols r extracted of the stégo medium.
2 calculate syndrome Hr
3 This found syndrome represents our embedded information block.

4 Presentation and analyses of the Results

In this work, the criteria of analysis considered are the imperceptibility.

The figure 2 represents cover medium before embedding of our message. La figure 4 represents the stego-medium which contains our message



La stéganographie moderne [1] est une nouvelle science qui permet une communication secrète. L'utilisation de la technique de matrix embedding dans les schémas de stéganographie tend à réduire la distorsion pendant l'insertion. Récemment, Galand et Fontaine [2] ont montré que les codes de Reed-Solomon peuvent être de bons outils pour la conception d'un schéma stéganographique. Dans cet article, nous allons présenter une implémentation de la technique de Matrix Embedding en utilisant les codes de Reed-Solomon. L'avantage de ces codes est qu'ils permettent de résoudre de façon aisée le problème de calcul par syndrome borné, problème qui constitue la base de la technique de matrix embedding.

1. Introduction

La stéganographie ou l'art de la communication secrète consiste à dissimuler un message dans un médium quelconque (une image, une vidéo, un son…) de telle sorte que le message caché soit furtive aux yeux de toute personne autre que l'expéditeur et le destinataire.

Les principaux travaux dans ce domaine se caractérisent par la minimisation de l'impact d'insertion. Beaucoup de schémas stéganographiques basés sur les codes correcteurs d'erreurs (BCH, HAMMING…) ont été proposé [3] [4] [5] et les travaux de Galand et Fontaine [galand] ont montré que les codes de Reed-Solomon peuvent être utilisés dans ce domaine.

Ces méthodes présentent certaines insuffisances liées par fois à la taille des messages à insérer, à la complexité des programmes de décodage des codes correcteurs, au problème de verrouillage des symboles et au mesure de distorsion. Toutes ces raisons poussent les chercheurs aujourd'hui non seulement à améliorer les schémas existant mais aussi à proposer d'autres approches.

Le travail que nous proposons présente une mise en œuvre du schéma stéganographique basée sur les codes de Reed-Solomon.

La technique de matrix embedding est une méthode de codage par syndrome, utilisant la théorie de codes correcteurs d'erreurs. Elle permet de minimiser le nombre total de modifications à effectuer pour insérer notre message. Toute l'astuce consiste à modifier l'image (médium de couverture) de sorte que les syndromes calculés à la réception représentent le message et également de sorte que l'image soit le moins modifiée.

Figure 2- ***cover medium and the message to embed***

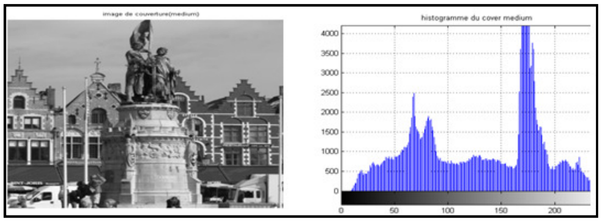


Figure 3 – ***cover medium and its histogram***

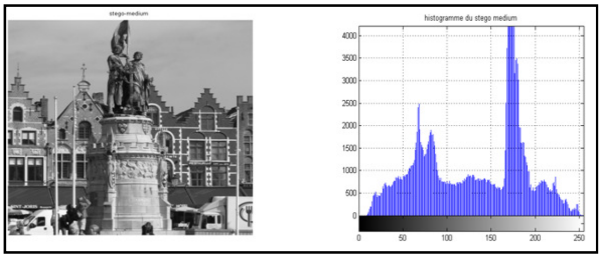
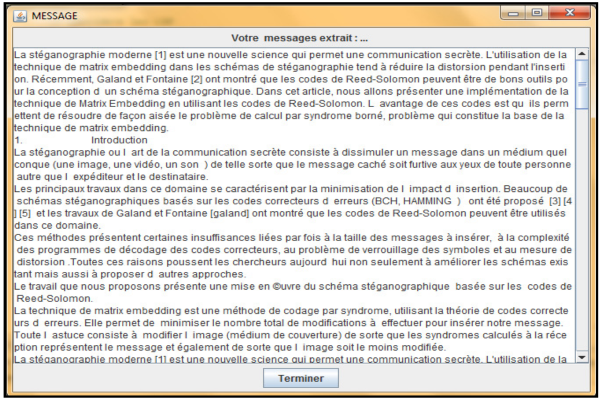


Figure 4 - stego medium ***and its histogram after embedding of the message***

To evaluate the performances of our approach, we made a comparative study in two steps. The first is a preliminary study which consists in comparing the cover medium and the stego-medium. Figures 2 and 4 respectively show the original picture and the picture containing information. We notice that the human visual system does not distinguish the difference generated by the marking.

Analysis of the histograms (figure 3) before and after embedding of the text (figure 4) shows imperceptibility of the message embedded in the picture.

By applying the algorithm of extraction; it is noticed that we manage to extract the initial text (figure 5).



```
La stéganographie moderne [1] est une nouvelle science qui permet une communication secrète. L'utilisation de la
technique de matrix embedding dans les schémas de stéganographie tend à réduire la distorsion pendant l'inserti
on. Récemment, Galand et Fontaine [2] ont montré que les codes de Reed-Solomon peuvent être de bons outils po
ur la conception d  un schéma stéganographique. Dans cet article, nous allons présenter une implémentation de la
technique de Matrix Embedding en utilisant les codes de Reed-Solomon. L  avantage de ces codes est qu  ils perm
ettent de résoudre de façon aisée le problème de calcul par syndrome borné, problème qui constitue la base de la
technique de matrix embedding.
1.                    Introduction
La stéganographie ou l  art de la communication secrète consiste à dissimuler un message dans un médium quel
conque (une image, une vidéo, un son  ) de telle sorte que le message caché soit furtive aux yeux de toute personne
 autre que l  expéditeur et le destinataire.
Les principaux travaux dans ce domaine se caractérisent par la minimisation de l  impact d  insertion. Beaucoup de
 schémas stéganographiques basés sur les codes correcteurs d  erreurs (BCH, HAMMING  )   ont été proposé  [3] [4
] [5]  et les travaux de Galand et Fontaine [galand] ont montré que les codes de Reed-Solomon peuvent être utilisés
dans ce domaine.
Ces méthodes présentent certaines insuffisances liées par fois à la taille des messages à insérer,  à la complexité
 des programmes de décodage des codes correcteurs, au problème de verrouillage des symboles et au mesure de
 distorsion .Toutes ces raisons poussent les chercheurs aujourd  hui non seulement à améliorer les schémas exis
tant mais aussi à proposer d  autres approches.
Le travail que nous proposons présente une mise en ©uvre du schéma stéganographique  basée sur les  codes de
 Reed-Solomon.
La technique de matrix embedding est une méthode de codage par syndrome, utilisant la théorie de codes correcte
urs d  erreurs. Elle permet de  minimiser le nombre total de modifications à  effectuer pour insérer notre message.
Toute l  astuce consiste à  modifier l  image (médium de couverture) de sorte que les syndromes calculés à la réce
ption représentent le message et également de sorte que l  image soit le moins modifiée.
La stéganographie moderne [1] est une nouvelle science qui permet une communication secrète. L'utilisation de la
```

Figure 5- ***Extraction of our embedded message***

5 Conclusion

In this article, we have just shown an implementation of a Steganographic scheme based on the Reed-Solomon codes. The use of the method of calculating by prolonged syndrome has allowed to us a more effective determination of the vector of modification.

The use of the technique of matrix embedding in the steganography makes it possible to minimize the changes introduced into the image of cover.

A more advanced survey as fixing a measure of distortion would improve the security of our scheme, moreover one improvement of the robustness of our diagram following attacks of types compression, … can also be considered

6 References

[1] J. Fridrich. :Steganography IN DIGITAL MEDIA Principles, Algorithms, and Application. Binghamton University, State University of New York, Cambridge University Press, 2010.

[2] C. Fontaine and F. Galand. :How Reed-Solomon Codes Can Improve Steganographic Schemes. Hindawi Publishing Corporation EURASIP Journal on Information Security Volume 2009, Article ID 274845

[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] Vasiliy Sachnev, Hyoung Joong Kim, Rongyue Zhang. :Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding , MM& Sec '09, Princeton, New Jersey, USA, Septembre 2009.

[5] 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).

[6] Claude BERROU. : Codes et turbocodes. 1e édition, Springer - Verlag, France, 2007.

[7] Madhu Sudan. :Decoding of Reed-Solomon codes beyond the error correction bound. Journal of Complexity, 13(1) :180–193, March 1997.

[8] V. Guruswami and M. Sudan. : Improved decoding of Reed-Solomon codes and algebraic-geometry codes. IEEE Trans. Inform. Theory, vol. 45, no. 6, pp. 1757-1767, Sept. 1999.

[9] Y. Wu. : New list decoding algorithms for Reed-Solomon and BCH codes". Information Theory, IEEE Transactions on, 54(8) : 3611–3630, 2008.

[10] J. Fridrich, M. Goljan, and D. Soukal. :Efficient wet paper codes. In M. Barni, editor, Proceedings, Information Hiding, 7th International Workshop, IH 2005, Barcelona, Spain, June 6–8, 2005, LNCS. Springer, Berlin, 2006.

[11] J. Fridrich and D. Soukal. : Matrix Embedding for Large Payloads. IEEE Transactions on Information Forensics and security, vol .1, no. 3, pp. 390-395

[12] R. T. Chien. :Cyclic decoding procedures for Bose-Chaudhuri-Hocquenghem.IEEE Transactions on Information Theory, 10(4) :357–363, 1964.

[13] G. Forney. :On decoding BCH codes. Information Theory, IEEE Transactions on, 11(4) :549–557, 1965.

[14] R. Crandall. :Some notes on steganography. Posted on Steganography Mailing List (1998).