WEAKNESS ON CRYPTOGRAPHIC SCHEMES BASED ON REGULAR LDPC CODES

Omessaad Hamdi[1] , Manel abdelhedi[2], Ammar Bouallegue[2], Sami Harari[3]

[1] SYSTEL, SUPCOM, Tunisia

hamdi@univ-tln.fr

[2] SYSCOM, ENIT, Tunisia

abdelhedi_manel@yahoo.fr, ammar.bouallegue@enit.rnu.tn

[3] SIS, USTV, France

harari@univ-tln.fr

ABSTRACT

We propose a method to recover the structure of a randomly permuted chained code and how to cryptanalyse cryptographic schemes based on these kinds of error coding. As application of these methods is a cryptographic schema using regular Low Density Parity Check (LDPC) Codes. This result prohibits the use of chained code and particularly regular LDPC codes on cryptography.

KEYWORDS:

Cryptography, Chained Codes, LDPC Codes, Attack, Complexity.

1. INTRODUCTION

RSA and McEleice are the oldest public key cryptosystems. They are based respectively on intractability of factorization and syndrome decoding problems [1]. However, McEliece [2] was not quite as successful as RSA, partiallydue to its large public key and to the belief that could not be used in signature. In 2001, Courtois, Finiasz and Sendrier [3] show a new method to build practical signature schemes with the McEliece public key cryptosystem. This scheme has the drawback of a high signature cost. One idea to counter this drawback consists in replacing Goppa code by other codes which have faster decoding algorithms like chained codes. In this paper, we show an invariant in the structure of chained codes which makes a weakness in cryptographic schemes based on chained codes. Our approach is based on the fact that any given chained equivalent code can be transformed in a systematic code which has a special generator matrix representation. Regular LDPC code is generally a chained repetition code. We show that these codes are useless in cryptography.

2. CHAINED CODE

A chained code C is defined as a direct sum of γ elementary codes C_i. This code is of length $N = \sum_{i=1}^{\gamma} n_i$ and of dimension $K = \sum_{i=1}^{\gamma} k_i$.

$$C = \bigoplus_{i=1}^{\gamma} C_i = \{(u_1,...,u_\gamma); u_1 \in C_1,...,u_\gamma \in C_\gamma\}$$

To encode an information $m = (m_1,...,m_\gamma)$, where m_i is k_i bits, we simply multiply it by the generator matrix to obtain the codeword $u = m.G = (u_1,...,u_\gamma)$ with u_i is the n_i bits codeword obtained from m_i using the elementary code C_i. So, G is a diagonal matrix in blocs and whose diagonal is formed by elementary generator matrices G_i of the code C_i.

We assume that we have an efficient decoding algorithm for each elementary code C_i. To decode $u = (u_1,...,u_\gamma)$, we apply for each codeword u_i its correspondent decoding algorithm $dec_{C_i}(\ )$. The decoded word is $m = (m_1,...,m_\gamma)$ with $m_i = dec_{C_i}(u_i)$.

We define the support of a non zero word $x = (x_1,...,x_n)$, denoted $\sup(x)$, as the set of its non zero positions. $\sup(x) = \{i \in \{1,..,n\}, x_i \neq 0\}$ and the support of a set $S = \{y_1,...,y_\gamma\}$ as the union of the supports of its words $\sup(S) = \bigcup_{y_i \in S} \sup(y_i)$. So the support of a code $C(N,K)$ is the union of its 2^k codeword supports.

Two words x and y are said to be connected if their supports are not disjoints i.e $\sup(x) \cap \sup(y) = \Theta$ and two sets I and J are said to be disjoints if there is no connection subset between them.

A non zero codeword x of C is said to be minimal support if there is no codeword $y \in C$ such that $\sup(y) \subset \sup(x)$.

Two codes $C(N,K)$ and $C'(N,K)$ are said to be equivalents if there is a permutation σ of $\{1,..,N\}$ such as: $C' = \sigma(C) = \{c_{\sigma(1)},..,c_{\sigma(N)}\}$. In other words, C and C' are equivalents if there is a permutation matrix such as for any generator matrix G of C, the matrix $G' = G.P$ is a generator matrix of C'.

3. CHAINED CODES AND CRYPTOGRAPHY

As we mentioned in the introduction, the drawback of the unique digital signature scheme based on error coding is the high signature complexity which is due to Goppa decoding algorithm. One idea to counter this drawback consists in replacing Goppa code by chained code which have faster decoding algorithm.

Generally, the secret key of a cryptographic scheme based on error coding is the code itself, for which an efficient decoding algorithm is known, and the public key is a transformation of the generator or parity check matrices. We consider a digital signature scheme based on chained code, and then we develop an algorithm to discover the private key from public key. This attack is applicable for each cryptographic scheme since it is a structural attack.

Secret key:

- S is a random $(K \times K)$ non singular matrix called the scrambling matrix.
- G is a $(K \times N)$ generator matrix of a chained code
- P is a random $(N \times N)$ permutation matrix

Public key:

- $G' = S.G.P$ is a randomly scrambled et permuted generator matrix. It is a generator matrix of an equivalent non structured code to the chained code $\sum_i c_i$ is the completed correction capacities calculated as [3].
- $h(\ )$ is a hash function.

Signature:

- The signer, first, calculates $y = h(M).P^{-1}$, where $h(M)$ is the N bit message, P^{-1} is the inverse of P. Then he uses the completed decoding algorithm [3] for the original chained code C to obtain $x = S.\sigma$. Finally, the receiver obtains the signature by computing $\sigma = S^{-1}.x$ where S^{-1} is the inverse of S.

Verification:

- The verifier calculates $\rho' = \sigma.G'$ and $\rho = h(M)$
- The signature is valid if $d(\rho, \rho') < \sum_i c_i$

We have introduced a digital signature scheme and then we present the weakness of this scheme. This weakness is due to the fact that chained codes have an invariant. Code equivalence means that one generator matrix is a permutation of the other, because matrix S does not change the code but only performs a modification on the basis of the linear subspace. Canteaut showed that the

matrix S may be important to hide the systematic structure of the Goppa codes, therefore having an important security role [4]. However, Heiman was the first to study this point and states that the random matrix S used in the original McEliece scheme serves no security purpose concerning the protection [5]. We confirm this argument and we show that the random matrix S has no security role for cryptographic schemes based on linear codes. We state also that disjoint elementary code supports is an invariant by permutation.

To avoid exhaustive attack, we used at least five different elementary codes and to avoid attack by information set, we used a chained code with length at least equal to 900 bits.

The attack explores the characteristics of the code transformation in order to identify its building blocks. Its input is a generating matrix G' of a randomly permuted chained code of length N and dimension K. Its output is a structured chained code. The algorithm's steps are:

- Apply a Gauss elimination to the rows of the matrix G' to obtain the systematic form $G_0 = (I_d, Z)$.

  Sendrier shows that rows of any systematic generator matrix of a code C are minimal support codewords of C and that any minimal support codeword of C is a row of a systematic generator matrix of C [4].

  The systematic chained code support is formed by disjoint sets. Each set represents the support of an elementary code. The transformation of any randomly permuted chained code generator matrix into a systematic matrix by linear algebraic algorithms will allow us to find these supports and thus elementary codes.

- Search the disjoint sets of rows of the systematic matrix G_0. Each set forms the elementary code support.

- Use elementary decoding algorithms to decode every message. As application of these codes, regular LDPC codes which represent chained repetition codes. Next sections represent the proprieties of these codes.

4. LOW DENSITY PARITY CHECK CODES

Low-density parity-check (LDPC) codes were first discovered by Gallager [6] in 1962 and have recently been rediscovered by Mackay and Neal [7], [8]. In fact, when LDPC codes have been invented, their decoding was too complicated for the technology, and so they have been forgotten. These codes deliver very good performance when decoded with the belief-propagation (BP) algorithm [7].

Binary LDPC codes, are linear block codes defined by a sparse parity check matrix $H(M \times N)$, where N denotes the codeword length and M the number of parity-check equations. When the numbers of 1's in each column and row are constant the code is called a regular LDPC code. Otherwise, it's said to be irregular.

4.1. Regular LDPC codes

In this section, we show that the parity check matrix of an LDPC code has a particular structure. The uniqueness of the canonical matrix provides us to recover used codes of any equivalent code.

The support of systematic LDPC code is formed by disjoint sets. Each set represents the support of an elementary repetition code. The transformation of any randomly permuted LDPC code parity check matrix into a systematic matrix by linear algebraic algorithms will allow us to find these supports and thus elementary codes.

The regular LDPC parity check matrix is constructed as follows: it is a concatenation of permuted repetition code.

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

4.2. Parity check matrix properties

We are interested, in this section, on parity check matrix properties which will be used to analyse the regular LDPC code structure. The parity check matrix H of a linear code is not unique, any S.H is also a parity check matrix.

- If the systematic parity check matrix exists then it is unique [4].
- Rows of any systematic parity check matrix of a code C are minimal support codewords of C [4].
- Any minimal support codeword of C is a row of a systematic parity check matrix of C [4].

Consequently, the systematic LDPC code parity check matrix rows are divided in γ disjoint sets. Each set defines the support of a repetition code C_i. This property is invariant by permutation.

Based on this property, we show that, a randomly permuted LDPC parity check matrix $H' = SHP$ has a particular structure. This structure permits to discover easily the hidden matrix H'.

5. RESULTS

5.1. Attack complexity on chained linear codes

The security of cryptographic schemes based on error coding is highly dependent on the class of used codes. Some class of codes reveal their characteristics even when they go through the permutation used to construct the public code. It is the case of chained codes. The starting point was the observation that any systematic matrix is formed by small weight codeword and that chained code contains so many minimal support codewords. These two properties lead to a structural attack of digital signature scheme based on chained code. Figure 1 shows the complexity of the attack of some cryptosystems using chained codes. The complexity is always less 2^{45} even with so long codes (N = 3000). This complexity prohibits using chained code in cryptography.

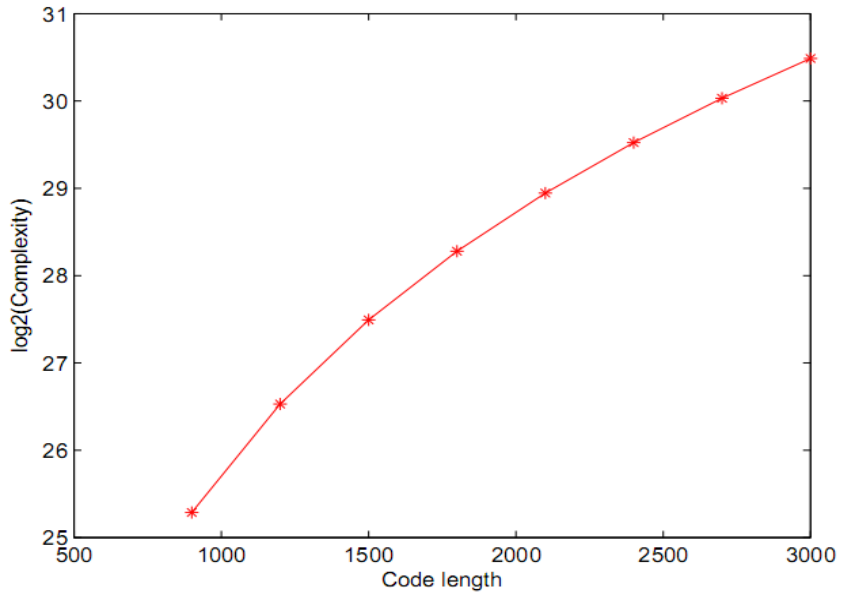

Figure 1: Attack complexity on chained linear codes

5.2. Attack complexity on LDPC Codes

The complexity is the number of binary operations to discover a randomly permuted regular LDPC code structure.

- $N.M^2/2$ binary operations for Gaussian elimination.
- $M.N$ binary operations to compute all line weights.

Thus, the number of binary operations necessary for this algorithm is equal to $N.M^2/2 + N.M$.

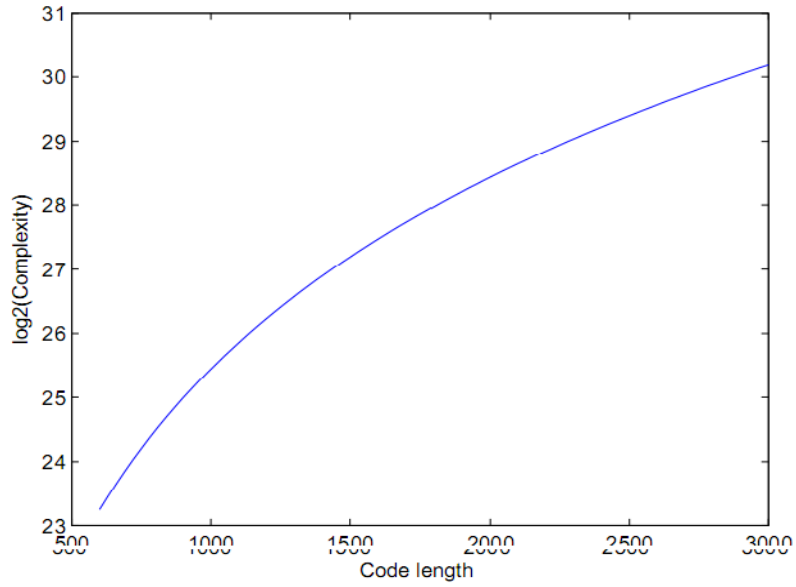

Figure 2: Complexity of the attack on cryptosystem using regular LDPC

Figure 2 shows the complexity of the attack of some cryptosystems using regular LDPC. The complexity is always less 2^{45} even with so long codes (N = 3000). This complexity prohibits using LDPC in cryptography.

6. CONCLUSION

In this paper, we discussed the structure of a randomly permuted chained code. We explored potential threats from systematic generator matrix that has particular structure. Chained code generator matrices have the properties of disconnected elementary code supports. This property is invariant by permutation, which make this kind of code useless in cryptography. Regular LDPC codes have this property.

REFERENCES

[1] E.R. Berlekamp, R.J. McEliece, and H.C.A. van Tilborg, "On the inherent intractability of certain coding problems" IEEE Transactions on Information Theory, Vol.24, No.3,1978, pp.384-386.

[2] R.J. McEliece, "A public-key cryptosystem based on algebraic coding theory" DSN Prog. Rep., Jet Propulsion Laboratory, California Inst. Technol., Pasadena, CA, pp. 114-116,January 1978.

[3] N. Courtois, M. Finiasz, and N. Sendrier, "How to achieve a McEliece-based digital signature scheme" In C. Boyd, editor, Asiacrypt 2001, volume 2248 of LNCS, pages 157-174. Springer-Verlag, 2001.

[4] N.Sendrier, "On the structure of a linear code" AAECC, Vol.9, n3, 1998, pp.221-242.

[5] A. Canteaut, "Attaques de cryptosystmes mots de poids faible et construction de fonctions t-rsilientes" . PhD thesis, Universit Paris 6, October 1996.

[6] R. G. Gallager,"Low-Density Parity-Check codes" PhD thesis, MIT, July 1963.

[7] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices" IEEE Transactions on Information Theory, vol. 45, pp 399–431, March 1999.

[8] D. J. C. Mackay,"Near shannon limit performance of low density parity check codes" Electron. Lett., vol. 33, pp. 457–458, Mars. 1997.

[9] J. Chen and M. P. C. Fossorier, "Near optimum universal belief propagation based decoding of Low-Density Parity Check codes" IEEE Transactions on Communicatons, vol. 50, pp. 406–414, March 2002.