Manuscript Preparation Guidelines for International Journal of Computer Science and Security

Omessaad Hamdi ohamdi@labri.fr
LABRI Laboratory,
Bordeaux 1, France.

Ammar Bouallegue ammar.bouallegue@enit.rnu.tn
SYSCOM Laboratory,
Ecole Nationale d'ingénieurs
De Tunis, Tunisia

Sami Harari harari@univ-tln.fr
USTV,
Toulon France

Abstract

We discuss the chained randomized linear code and their use in cryptography. We show that the adoption of randomized chained codes in the framework of McEliece cryptosystem expose the cryptosystem to some new attacks.

Key Words: Cryptography, Chained Codes, Attack, Complexity

1. INTRODUCTION

In this paper, a new variant of cryptographic schemes based on error coding is studied. Random based techniques allow to design large families of chained codes. Therefore, in principle, such codes can substitute Goppa codes, originally used by McEliece [2].The McEliece cryptosystem is a public key cryptosystem based on coding theory that has successfully resisted cryptanalysis [1] for thirty years. The original version, based on Goppa codes, is able to guarantee a high level of security, and is faster than computing solutions, like RSA.

Despite this, it has not been considered in practical applications, due to the major drawbacks like the large size of the public key, the low transmission rate. Moreover, there is no efficient signature scheme based on error coding.

Several attempts have been made for overcoming such drawbacks, but the adoption of most families of codes has not been possible without compromising the system security [2], [8], [9]. Chained codes are a particular class, able to join low complexity decoding techniques. One idea consists in adopting this family of codes in some signature schemes.

Recently, however, new attacks have been found that are able to exploit the flaw in the transmission from the private key to the public one [10]. Such attack seems to be effectively countered by changing some constituent matrices like introducing some random vectors.

This works gives an overview of the chained code and weakness related to their structure. A recent randomized version can be considered and its ability to counter the currently known attacks is discussed.

To counter this weakness, we concatenate random rows to the generator matrix. This new structure avoids minimum codewords. However, it does not modify the dual code. Consequently, other attacks can be generated.

The details of chained code design are given in section 2.In sections 3 and 4, a digital signature scheme using chained code and its security are discussed. In section 5, we introduce a digital signature using randomized chained code and before concluding we study its security.

2. CHAINED CODE

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

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

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 $d(\rho, \rho') < \sum_i c_i$

To avoid exhaustive attack, we use at least five different elementary codes and to avoid attack by information set, we use a chained code with length at least equal to 1500 bits.

After developing a digital signature scheme, we discovered a weakness in 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 [6]. 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 [7]. 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.

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]. So, 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.

The complexity of this attack is less than 2^{45} even with so long codes (see FIGURE 1).

4. Randomized chained linear codes

To counter the attack introduced in previous section, one idea consists in concatenating random vectors to the generator matrix. In this section, first, we define randomized chained codes then we introduce a cryptographic scheme based on these codes.

4.1 Random vectors

The randomized chained linear code concatenates random vectors of length N to the chained code. Using Information Theory, a N bit random binary vector is of weight closely to $N/2$ and the distance between two random vectors is of order $N/4$. These approximations are more precise when N is large.

4.2 Construction of randomized chained codes

Lets consider a chained linear code generator matrix G_{CL} as described in section 2. Each elementary linear code is of length n_i and of size k_i. Chained linear code is of length $N = \sum_{i=1}^{\gamma} n_i$ and of dimension $K = \sum_{i=1}^{\gamma} k_i$.

Lets consider a matrix G_r formed by K random rows of length N.

The generator matrix G of the system using randomized linear chained code has the following form: $G = (G_{CL}, G_r)$.

The weight of a row of the systematic generator matrix is about $N/2 + p_i$ where p_i is the weight of i^{th} row of the chained code generator matrix G_{CL}.

4.2.1 Encoding

m is a word of length K to be encoded. The codeword is obtained by multiplying m by the generator matrix G of the randomized chained linear code.

$$c = m.G$$

4.2.2 Decoding

r is the word to be decoded.

$$r = c + e = m.G_{CL} + e_1, m.G_r + e_2$$

Note by $dec_{CL}(\ )$ the chained linear decoding algorithm. Thus, $m = dec_{CL}(m.G_{CL} + e_1)$. The codeword closest to r is $c = m.G$.

5. DIGITAL SIGNATURE USING RANDOMIZED CHAINED LINEAR CODES

5.1 Key generation

- Generate a sequence γ linear codes. Each code is of length n_i and of dimension k_i.
- Build the chained linear code generator matrix G_{CL}. This matrix is of size $N = \sum_{i=1}^{\gamma} n_i \times K = \sum_{i=1}^{\gamma} k_i$
- Generate K random vectors v_i of length N. These vectors will be stored in a matrix G_r of size $K \times N$.

  The obtained code is of length $2N$ and size K. It has the following generator matrix's form $G = (G_{CL}, G_r)$

To hide the code structure, we also generate

- A random invertible matrix S of size $((2.N)-N)\times((2.N)-K)$.
- A permutation matrix P of size $((2.N)\times(2.N))$
- Determine the check parity matrix H as follows $H.(G.P)^t = 0$

Thus, the private key is formed by

- The generator matrix G of size $K \times 2.N$
- The random matrix S of size $((2.N)-N)\times((2.N)-K)$.
- The permutation matrix P of size $((2.N)\times(2.N))$.

The public key is formed by the hidden and permuted parity check Matrix $H' = S.H$ of size $(2.N-K)\times(2.N)$

5.2 Signature algorithm

Let m be a message to be signed. The signer has the private key formed by G, S and P and the hash function $h(\ )$ whose result is of length $2.N$.

- Compute $\rho' = h(m)$ of length $2.N$
- Compute $\rho = \rho'.P^{-1}$.
- Divide ρ in two parts ρ_1 and ρ_2, each one is of length N.

$$\rho = \rho_1 \| \rho_2$$

- Decode ρ_1 using the decoding algorithm of chained linear code to obtain information m of length K .
- Compute $\nu = m.G$ which is a codeword.
- Compute $e' = \rho + \nu$ the error related to the secret code which is closer to $N/2$ .This error has the same syndrome as ρ .
- Compute the error $e = e'.P$ and its weight $p = w(e)$. The error e has the same syndrome as $\rho' = h(m)$ relatively to the public code generated by $G.P$

The signature of m is formed by $\sigma = (e, p)$.

5.3 Verification Algorithm

- The verifier has the matrix H and the hash function $h(\ )$, the message m and the signature σ .
- he checks that $w(e) = p$
- he computes $\rho' = h(m)$.
- he computes $x_1 = H'.e$
- he computes $x_2 = H'.\rho'$

The signature is valid if

$$x_1 = x_2$$

5.4 Soundness

$x_1 = H'.e = H'.(\rho + \nu).P = H'.\rho.P = x_2$ since $\nu.P$ is a codeword of the permuted code having $G.P$ as generator matrix.

5.5 Parameters

Forging a signature consists in determining the signature $\sigma = (e, p)$ message from m or retrieving the secret key. An attacker who has the parity check matrix of size $(2.NK) \times 2.N$, may proceed as follows:

- he transforms H' a systematic matrix $H_0 = \left(R^t, I_{(2.N-K),(2.N-K)}\right)$
- he guess the corresponding matrix G_0 of size $K \times 2.N$:

$$G_0 = (I_K, R)$$

- he computes $\rho = h(M) = (\rho_1, \rho_2)$ with $|\rho_1| = K$ and $|\rho_2| = 2N - K$
- he search the closest codeword $c = (c_1, c_2)$ of length $2.N$ to ρ .

So, he will obtain

- $d(C_1, \rho_1) = 0$
- $d(C_2, \rho_2) = (2.N - K)/2$

To build a secure algorithm, the difference k between p

and $(2.NK)/2$ should be large enough. The table 1 shows parameters for a signature scheme based on randomized chained code. From Table 1, we show that is necessary that used code must have a length $2.N$ greater than 1350.

N	990	1080	1170	1260	1350	1440	1530	1520	1710	1800	1890	1980
K	253	276	299	322	345	368	391	414	437	460	483	506
K	44	48	52	56	60	64	68	72	76	80	84	88

Table 1: Signature parameters

Table 2 shows performances of randomized chained code in terms of execution complexity and public key size.

Signature	**Signature with randomized code**
Public key size (ko)	123
Signature complexity	2^{20}
Verification Complexity	2^{13}

Table 2: Performance of signature based on randomized chained codes

5.6 Solidity

The strength of the scheme depends on the choice of parameters. There are two types of attacks on asymmetric systems.

The starting point was to hide the structure of the chained codes. Possible attack of the new structure consists in enumerating all matrices of size $(2.N-K)\times 2.N$ and test their equivalences with $H^{'}$. The code is formed by γ elementary codes and K random vectors. So, the number of randomized chained code is $\frac{(N!/(N/2!)^2)!}{K!}2^{\nu}$ which is very large considering chosen parameters in section 5. The concatenation of random vectors avoid minimal codewords attack since a codeword is at least of weight $N/2$. Moreover, the new structure avoids support disjunction since the distance between two codewords is in order of N/4.

However, this new structure hides a weakness related to the dual code. In fact, concatenated vectors do not modify the dual code. Consequently, an attacker may proceed as follows:

- Transform $H^{'}$ in a systematic matrix $H_0 = (R^t, I_{2.N-K})$.
- Search minimal codewords of elementary linear codes which have weight smaller than those of random vectors.
- Use the algorithm introduced in section 3 to recover dual code.

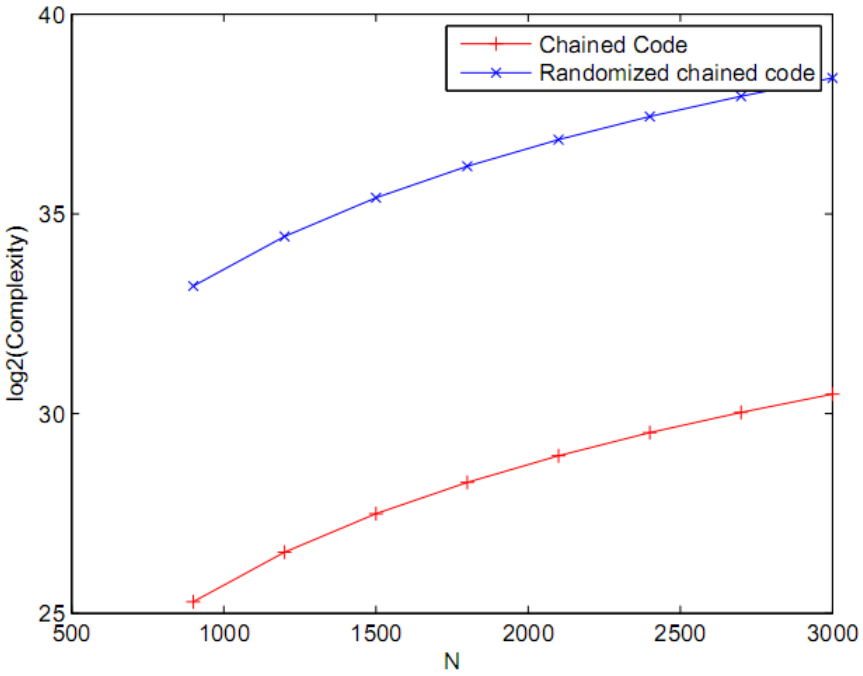

FIGURE 1: Attack Complexity

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 with chained codes and randomized chained codes. The starting point was the observation that any systematic matrix of a chained code is formed by small weight codeword and that the code contains so many minimal support codewords. These two properties lead to a structural attack of digital signature scheme based on chained code.

We have tried to counter this attack by concatenating some random vectors to the generator matrix. However, the added vectors avoid this attack but they do not modify the dual code. Consequently, we discover another structural weakness related to this kind of codes.

Figure 1 shows the complexity of the attacks of some cryptosystems using chained codes and randomized chained code. The complexity is always less than 2^{45} even with so long codes $(N = 3000)$. This complexity prohibits using chained code in cryptography.

6. Conclusion

In this paper, we discussed the structure of a randomly permuted chained code. We explored potential threats from systematic generator matrices that have particular structure. Chained code generator matrices have the properties of disconnected elementary code supports. We have tried to hide this property by concatenating some random vectors to the generator matrix. Unfortunately, these vectors avoid attack by minimum codeword in the code itself. However, they do not modify the dual code which makes weakness on cryptographic scheme based on chained codes. This property is invariant by permutation, which make this kind of code useless in cryptography.

7. 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. *D. J. Bernstein, T. Lange, and C. Peters. Attacking and defending the McEliece cryptosystem. In Post-Quantum Cryptography, volume 5299 of Lecture Notes in Computer Science, pages 31-46. Springer Berlin Heidelberg, 2008.*

4. *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.*

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

6. *A. Canteaut "Attaques de cryptosystemes a mots de poids faible et construction de fonctions t-resilientes" PhD thesis, Universite Paris 6, October 1996.*

7. *R. Heiman "On the security of Cryptosystems Based on Linear Error Correcting codes" MSc. Thesis, Feinberg Graduate School of the Weizmann Institute of Science. August 1987.*

8. *M. Baldi and F. Chiaraluce. Cryptanalysis of a new instance of McEliece cryptosystem based on QC-LDPC codes. In Proc. IEEE International Symposium on Information Theory (ISIT 2007), pages 2591-2595, Nice, France, June 2007.*

9. *A. Otmani, J. P. Tillich, and L. Dallot. Cryptanalysis of two McEliece cryptosystems based on quasi- cyclic codes. In Proc. First International Conference on Symbolic Computation and Cryptography (SCC 2008), Beijing, China, April 2008.*

10. *O. Hamdi, A. Bouallegue, S.Harari, Weakness on Cryptographic Schemes based on Chained Codes, The First International Workshop on Wireless and Mobile Networks Security (WMNS-2009) in conjunction with NSS 2009, October 19~21 2009, Gold Coast, Australia.*