

# Identity-based identification and signature schemes using correcting codes

Pierre-Louis Cayrel<sup>1</sup>, Philippe Gaborit<sup>1</sup> and Marc Girault<sup>2</sup>

1 Université de Limoges, XLIM-DMI,  
123, Av. Albert Thomas  
87000 Limoges, France

`{pierre-louis.cayrel,philippe.gaborit}@xlim.fr`

2 France Télécom Recherche et Développement  
42, rue des Coutures  
14066 Caen, France  
`marc.girault@orange-ftgroup.com`

January 24, 2007

**Abstract.** In this paper, we propose a new identity-based authentication (and signature) scheme based on error-correcting codes. This scheme is up to date the first identity-based scheme not based on number theory. The scheme combines two well known code-based schemes: the signature scheme of Courtois, Finiasz and Sendrier and the zero-knowledge authentication scheme of Stern (which may also be used for signature). The scheme inherits from the characteristics of the previous schemes: it has a large public key of order 1Mo and necessitates a certain number of exchange rounds. The scheme can also work in signature but leads to a very large signature of size 1Mo.

**Keywords :** Signature, Authentication, Identity based scheme, Correcting codes, Stern, Niederreiter.

## 1 Introduction

The most critical point of classical public key cryptography (RSA, El Gamal...) is in the management of the authenticity of the public key. In fact, if Alice manages to take Bob's identity by cheating her own public key as Bob's one, she would be able to decipher all messages sent to Bob and to sign any message using the stolen identity.

In 1984, Shamir introduced the concept of IDentity-based Public Key Cryptography ID-PKC [15] in order to simplify the management and the authentication of the public key, which time passing by, had become more and more complex.

In the ID-PKC scheme of Shamir, the public key of an user is undeniably linked to his identity on the network (user-id): it can be a concatenation of any publicly known information: his name, his e-mail, his phone number, etc ...

Hence it is not necessary to verify a certificate for the public key or to contact a data base to obtain it. At first glance it seems simple but producing private keys becomes more complex. And since a private user can not derivate his own private key by himself, it is necessary to introduce trusted third party which derivate the private key from the public key and sends it to the user (at least it has to be done once for each user).

In [15] Shamir calls this trusted third party the Key Generation Center (KGC). The KGC is the owner of a secret, namely the master key. After a protocol of authentication of the identity of the user, the KGC computes his private key from the master key, the user-id and a trapdoor function.

In his paper Shamir proposed systems based on RSA or Discrete logarithm but which did not fulfilled the previous requirements. The first efficient identity-based cryptosystem was proposed in 2001 by Boneh and Franklin [2]. This system is based on Weil pairing and elliptic curves. The same year, Cocks [7] published a system based on quadratic residuosity but the system has a very large message expansion which makes it unefficient in practice.

Following the paper by Boneh and Franklin, researches on ID-PKI have made great progresses and lots of schemes have been published all of them based on elliptic curves and bilinear pairings, such as identity-based encryption (IBE) schemes [1, 10], identity-based key agreement schemes [16, 5], identity-based signature (IBS) scheme [14, 13, 6, 11, 4, 28, 29, 20].

In 2004 Bellare, Neven and Namprempre proposed in [?] a general framework to derivate IBI or IBS from signature or authentication scheme, and they applied it to known schemes, but they only considered number theory based schemes.

In this paper we consider a code-based scheme, not considered in their work. Code-based cryptography was introduced at the same time than RSA by MacEliece [23], a variation on the scheme was proposed by Niederreiter in 1986 [?].

The idea of using error-correcting codes for identification purposes is due to Harari, followed by Stern (first protocol) and Girault. But Harari and Girault protocols were subsequently broken, while Stern's one was five-pass and unpractical. At Crypto'93, Stern proposed a new scheme [], which is still today the reference in this area.

For a long time no code-based signature scheme was known, eventually the first (non broken) code-based cryptosystem was proposed by Courtois, Finiasz and Sendrier [9] (CFS) in 2001. At the difference of RSA, the MacEliece or Niederreiter schemes do not rely on purely bijective problems like the modular exponentiation. The basic idea of their signature scheme is to choose parameters such that such an inversion for the Niederreiter scheme is practically possible. This is done at the cost of obtaining rather large parameters (except for the length of the signature) when comparing to other signature schemes but at least it exists!

In this paper we combine the previous signature scheme and the authentication scheme by Stern to obtain an IBI and an IBS scheme.

The basic idea of our scheme is to start from a Niederreiter-like problem which can be inverted like in the CFS scheme. This permits to associate a secret to a random (public) value obtained from the identity of the user. The secret and public values are then used for the Stern zero-knowledge authentication (or signature) scheme.

The paper is organized as follows: in section 2 we recall basic facts on code-based cryptography, in section 3, we recall the cryptosystem of Niederreiter, the signature scheme of Courtois, Finiasz and Sendrier and the protocol of Stern before developping our new protocol in section 4. At last in section 5 we give parameters and security analysis of our scheme and conclude in section 6.

## 2 Code-based cryptography

In this section we recall basic facts about code-based cryptography. We refer to the work of Nicolas Sendrier in [25, 26], for a more general context on these problems and to [24] for a general context on coding theory.

### 2.1 A hard problem

Every public key cryptosystem has to rely on a hard problem. In the case of coding theory, the main problem is:

### 2.2 A hard problem

Every public key cryptosystem has to rely on a hard problem. In the case of coding theory, the main problem used is:

**Problem:** SYNDROME DECODING (SD)

**Instance:** An  $m \times n$  matrix  $H$  over  $F_q$ , a target vector  $s \in F_q^m$  and an integer  $w > 0$ .

**Question:** Is there a vector  $x \in F_q^n$  of weight  $\leq w$ , such that  $Hx^T = s^T$  ? This problem was proven to be NP-complete in [18].

This problem was proven to be NP-complete [18].

### 2.3 Usual attacks: Information Set Decoding

In terl of code-based cryptography there are two kinds of attacks: attacks which try to decode directly a message or structural attacks which try to recover the structure of the code.

The most efficient algorithms in our case are based on the information set decoding. A first analysis was done by MacEliece in [23], then by Lee and in Brickell in [21] and also by Stern in [27] and Leon in [22] and at last by Canteaut and Chabaud in [19].

Consider a  $[n, k, 2t + 1]$  binary code, if one uses information set decoding, one chooses a random set of  $k$  columns, an error is decodable when its support does

not meet the  $k$  random columns. The probability for an error to be decodable (see [26] for more details) is then  $P_{dec} = \frac{\binom{n-k}{t}}{\binom{n}{t}}$ , which leads with the usual binomial approximation to a probability:

$$P_{dec} = O(1).2^{-nH_2(t/n) - (1-k)H_2(t/(n-k))},$$

where  $H_2(x) = -x \log_2(x) - (1-x) \log_2(1-x)$ .

Then the estimated work factor  $WF$  to find a word of weight  $t$  can be estimated as follow:

$$WF = \frac{P(k)}{P_{dec}},$$

where  $P(k)$  corresponds to the cost of a Gaussian elimination,  $P(k)$  can be first thought as a cost in  $O(k^3)$ , in the best improvement of [7] one can consider  $P(k)$  linear or even less. For the parameters we are envisaging it is reasonable to consider them linear to fit the practical results of [7]. This algorithm is currently the best known.

### 3 Signature scheme of Courtois, Finiasz and Sendrier

Before describing the CFS scheme we first recall the Niederreiter scheme:

Let  $C$  be a  $q$ -ary linear code  $t$ -correcting of length  $n$  and of dimension  $k$ . Let  $H$  a matrix of parity of  $C$ . We will use an  $\tilde{H}$  matrix such that :

$$\tilde{H} = VHP \begin{cases} V \text{ is invertible} \\ P \text{ is a permutation matrix} \end{cases}$$

$\tilde{H}$  will be public and its decomposition will be secret, knowing a decoding by syndromes algorithm useful in  $C$ . To be clearer, we recall the various sizes of matrices.

$M$  is  $n \times n - k$ ,  $V$  is  $n - k \times n - k$ ,  $H$  is  $n \times n - k$ ,  $P$  is  $n \times n$ .

**Encryption** For a chosen cleartext  $x$  in the  $E_{q,n,t}$  space of  $\mathbb{F}_q^n$  words which Hamming weight  $t$ ,  $y$  is the cryptogram corresponding to  $x$  if and only if :

$$y = \tilde{H}x^T.$$

**Decryption** For  $y = \tilde{H}x^T$ , the knowledge of the secrets allows :

1. to compute  $V^{-1}y (= HPx^T)$ ;
2. to find  $Px^T$  from  $V^{-1}y$  thanks to the decoding by syndromes algorithm used in  $C$ ;
3. to find  $x$  applying  $P^{-1}$  to  $Px^T$ .

The decoding by syndromes algorithm can be, for instance, in the case of Goppa's codes, Patterson's algorithm (see part 6.1).

### 3.1 The CFS signature scheme

As we already mentioned at the difference of the RSA scheme which is naturally invertible, the MacEliece or the Niederreiter schemes are not invertible, ie, if one starts from a random element  $y$  of  $F_2^n$  and a code  $C[n, k, d]$  that we are able to decode up to  $d/2$ , it is almost sure that we won't be able to decode  $y$  into a codeword of  $C$ . This comes from the fact that the density of the whole space which is decodable is very small. The idea of the CFS scheme is to fix parameters  $[n, k, d]$  such that the density of decodable codewords is reasonable and pick up random elements until one is able to decode it.

More precisely, given  $M$  a message to sign and  $h$  a hash function of  $\{0, 1\}^{n-k}$ . We try to find a way to build  $s \in E_{q,n,t}$  such that  $h(M) = \tilde{H}s^T$ . The algorithm works as follows:

1.  $i \leftarrow 0$
2. while  $h(M \oplus i)$  is decodable do  $i \leftarrow i + 1$
3. compute  $s = D(h(M \oplus i))$

We get at the end an  $\{s, j\}$  couple, such that  $h(M \oplus j) = \tilde{H}s^T$ . Let us notice that we can suppose that  $s$  has weight  $t = \lfloor d/2 \rfloor$ .

## 4 Stern's protocol

This scheme was developed in 1993 (see [17]) to aim at providing zero-knowledge authentication scheme, the security of which would not rely on number theory problems. Given  $\tilde{H}$  a matrix of size  $(n - k) \times n$  over  $\mathbb{F}_2$ . This matrix is public. Each user receives a secret key  $s$  of  $n$  bits and of weight  $t$ . A user's public identifier is obtained from :

$$i = \tilde{H}s^T.$$

It is calculated once in the lifetime of  $\tilde{H}$ . It can thus be used by several future identifications. Let us suppose that  $A$  wants to prove to  $B$  that he is indeed the person corresponding to the public identifier  $i_A$ .  $A$  has his own private key  $s_A$  s.t.  $i_A = \tilde{H}s_A^T$ . Our two protagonists can then follow the protocol :

1.  $A$  chooses randomly any word  $y$  of  $n$  bits and a permutation  $\sigma$  of  $\{1, 2, \dots, n\}$ . Then  $A$  sends to  $B$  :  $c_1, c_2, c_3$  such that :

$$c_1 = \langle \sigma, \tilde{H}y^T \rangle; c_2 = \langle y.\sigma \rangle; c_3 = \langle (y \oplus s).\sigma \rangle$$

where  $arg_1, arg_2$  notes the concatenation of  $arg_1$  and  $arg_2$ ,  $\langle arg_1 \rangle$  the action of a hash function on  $arg_1$  and  $arg.\sigma$  is the image of  $arg$  by  $\sigma$ .

2.  $B$  sends to  $A$ ,  $b \in \{0, 1, 2\}$ .
3. Three possibilities:
  - if  $b = 0$  :  $A$  reveals  $y$  and  $\sigma$ .
  - if  $b = 1$  :  $A$  reveals  $(y \oplus s)$  and  $\sigma$ .
  - if  $b = 2$  :  $A$  reveals  $y.\sigma$  and  $s.\sigma$ .
4. Three possibilities:
  - if  $b = 0$  :  $B$  verifies that the  $c_1, c_2$  received at the second round have really been honestly calculated.
  - if  $b = 1$  :  $B$  verifies that the  $c_1, c_3$  received at the second round have really been honestly calculated. For  $c_1$  we can note that  $\tilde{H}y^T$  derives directly from  $\tilde{H}(y \oplus s)^T$  by :

$$\tilde{H}y^T = \tilde{H}(y \oplus s)^T \oplus i = \tilde{H}(y \oplus s)^T \oplus \tilde{H}s^T = \tilde{H}y^T$$

- if  $b = 2$  :  $B$  verifies that the  $c_2, c_3$  received at the second round have really been honestly calculated, and that the weight of  $s.\sigma$  is really equal to  $t$ .
5. Reiterate the steps 1,2,3,4 while the expected security is not reached.

The protocol has to be iterated long enough to make the  $k$  numbers of rounds  $(2/3)^k$  close to the level of confidence wanted, where  $(2/3)$  is the probability that a dishonest person cheats during a round. Apart from the number of turns, the security of this scheme relies on the difficulty to invert the function :

$$x \mapsto \tilde{H}x^T.$$

## 5 An identity-based identification protocol : the Stern-Niederreiter's protocol

Given  $C$  a  $q$ -ary linear code of length  $n$  and of dimension  $k$ . Let  $H$  be a matrix of parity of  $C$ . Given  $\tilde{H} = VHP$  with  $V$  invertible and  $P$  a matrix of permutation. Let  $h$  a hash function with values in  $\{0, 1\}^{n-k}$ . Let  $id_A$  Alice's identity,  $id_A$  can be compute by everyone. Similarly,  $\tilde{H}$  is public. The decomposition of  $\tilde{H}$  is, on the contrary, a secret of the authority and not of Alice. We shall describe an identity-based authentication method : Alice the prover is identifying herself to Bob the verifier.

**Preliminary : key deliverance** Alice has to authenticate herself in a classic way, to get the private key which will then allow her to authenticate herself to a third person as Bob. For that purpose, we use variation on identity. Let us admit that we know Bob's identity  $id_B$ . Given  $h$  a hash function with values in  $\{0, 1\}^{n-k}$ . We search a way to find  $s \in E_{q,n,t}$  such that  $h(id_B) = \tilde{H}s^T$ . The main point is to decode  $h(id_B)$ . The main problem is that  $h(id_B)$  is not *in principle* in the arrival space of  $x \rightarrow \tilde{H}x^T$ . That is to say that  $h(id_B)$  is not *in principle* in the space of decodable elements of  $F_2^n$ . That problem can be solved thanks to the following algorithm. Given  $D()$  a decoding algorithm for the hidden code:

1.  $i \leftarrow 0$
2. while  $h(id_B \oplus i)$  is not decodable do  $i \leftarrow i + 1$
3. compute  $s = D(h(id_B \oplus i))$

We get at the end a couple  $\{s, j\}$ , such that  $h(id_B \oplus j) = \tilde{H}s^T$ . We can note that we have necessarily  $s$  of weight  $t$ .

**Authentication by Bob.** We use a slight derivation of Stern's protocol (section 4). We suppose in that protocol that  $A$  obtained a couple  $\{s, j\}$  verifying :  $h(id_A \oplus j) = \tilde{H}s^T$ .  $h(id_A \oplus j)$  is  $A$ 's public key. The new protocol is based on Stern's protocol but with two changes, first  $A$  sends  $j$  to  $B$  at the step one and second, we change the step 4 with :

4bis. Three possibilities:

- if  $b = 0$  :  $Bob$  verifies that the  $c_1, c_2$  received at the second round have really been honestly computed.
- if  $b = 1$  :  $Bob$  verifies that the  $c_1, c_3$  received at the second round have really been honestly computed. For  $c_1$  we can note that  $\tilde{H}y^T$  derives directly from  $\tilde{H}(y \oplus s)^T$  by :

$$\tilde{H}y^T = \tilde{H}(y \oplus s)^T \oplus h(id_A \oplus j) = \tilde{H}(y \oplus s)^T \oplus \tilde{H}s^T$$

- if  $b = 2$  :  $Bob$  verifies that the  $c_2, c_3$  received at the second round have really been honestly computed, and that the weight of  $s \cdot \sigma$  is really equal to  $t$ .

The knowledge of  $j$  doesn't permit to find  $s$  such that  $h(id_A \oplus j) = \tilde{H}s^T$ . The security of this system is the same as the security of Stern's one (see section 2).

## 6 Security Analysis

We shall here deal with the security of *classical* protocol as their applicability and finally end with our protocol.

Remind that in the case of Niederreiter's cryptosystem, its security relies on the *supposed* difficulty of the decoding of a linear code (see section 2).

## 6.1 Parameters and security of the scheme

The protocol has two parts: in the first part one inverts the syndrome decoding problem for a matrix  $\tilde{H}$  in order to construct a private key for the prover and in second part one applies Stern authentication protocol with the same matrix  $\tilde{H}$ .

This shows that the overall parameters of the scheme are equivalent to the security of the CFS scheme.

In particular the scheme has to respect two imperative conditions:

1. make the computation of  $\{s, j\}$ , defined before, difficult without the knowledge of the description of  $H$ ,
2. make the number of trials to determine the correct  $j$  not too important in order to reduce the cost of the computation of  $s$ .

Following [9] the Goppa  $[2^m, 2^m - tm, t]$  codes are a large class of codes which are compatible with condition 2. Indeed, for such a code, the proportion of the decodable syndromes is about  $1/t!$  (which is a relatively good proportion). We also have to choose a relatively small  $t$ .

The  $\{s, j\}$  production process will thus be iterated, about  $t!$  times before finding the correct  $j$ . But each iteration forces to compute  $D(h(id_A \oplus j))$ .

The decoding of the Goppa codes consists of :

- computing a syndrome :  $t^2m^2/2$  binary operations;
- computing a localisator polynomial :  $6t^2m$  binary operations;
- computing its roots :  $2t^2m^2$  binary operations.

We thus get a total cost for the computation of Alice's private key of about :

$$t!t^2m^2(1/2 + 2 + 6/m) \text{ binary operations}$$

The cost of an attack by decoding thanks to the *split syndrome decoding* is estimated to :

$$2^{tm(1/2+o(1))}.$$

The choice of parameters will have to be pertinent enough to conciliate cost and security. Although less crippling, some sizes have also to remain reasonable : the length of  $\{s, j\}$ , the cost of the verification and the size of  $\tilde{H}$ .

The size of  $\tilde{H}$  is  $(n - k) \times n$ , that is for a Goppa code :  $2^m tm$ . The following figure sums up the different parameters :

signature cost	$t!t^2m^2(1/2 + 2 + 6/m)$
signature size	$tm$
verification cost	$t^2m$
attack cost	$2^{tm(1/2+o(1))}$
size of $\tilde{H}$	$2^m tm$

Following [9] we can for example take  $t = 9$  and  $m = 16$ . The cost of the signature stays then relatively reasonable for a security of about  $2^{80}$ . The others sizes remain in that context very acceptable.

## 6.2 Practical values

The big difference when using the parameters associated to the CFS scheme is that the code used is very long,  $2^{16}$  against  $2^9$  for the basic Stern scheme, it dramatically develop the communication cost.

In the next table we sum up for the parameters  $m = 16$ ,  $t = 9$  the general paramaters of the IBI and IBS schemes.

public key	private key	matrix size	communication cost	key generation
$tm$	$tm$	$2^m tm$	$\approx 2^m \times \#rounds$	
144	144	1 Mo	500 Ko (58 rounds)	1 s

Practical values for the IBI scheme for  $m = 16, t = 9$

public key	private key	matrix size	signature length	key generation	Practical
$tm$	$tm$	$2^m tm$	$\approx 2^m \times 150$		
144	144	1 Mo	1.5 Mo	1 s	

values for the IBS scheme for  $m = 16, t = 9$

## 7 Conclusion

In this paper we presented an IBI and INS scheme based on error-correcting code. This scheme is the first non number theory based identity based scheme. The scheme combines two well known schemes and inherits from the worse properties of these schemes: the public data is large, the communication cost for the IBI scheme is large and the signature length for the IBS scheme is also very large but at least the scheme may present an alternative to number theory based schemes.

## References

1. BONEH D., BOYEN X. . Efficient selective-id secure identity based encryption without random oracles. Eurocrypt 2004, LNCS 3027:223–238, 2004.
2. FRANKLIN M. and BONEH D. . Identity-based encryption from the weil pairing. Advances in Cryptology-Crypto'01, 2001.
3. BELLARE M.,NAMPREMPRE C. and NEVEN G. Security proofs sor identity-based authentication and signature schemes. Eurocrypt 2004, volume 3027 of Lecture Notes in Computer Science:268–286, 2004.

4. CHA Jae Choon, CHEON Jung Hee. An identity-based signature from gap diffie-hellman groups. Proceedings of the 6th International Workshop on Theory and Practice in Public Key Cryptography: Public Key Cryptography:18–30, January 06-08, 2003.
5. CHEN L. and KUDLA C. Identity based authenticated key agreement from pairings. Cryptology ePrint Archive, Report 2002/184, 2002.
6. CHEN X., ZHANG F., KIM K. . A new id-based group signature scheme from bilinear pairings. WISA 2003, LNCS 2908:585–592, 2003.
7. COCKS C. An identity based encryption scheme based on quadratic residues. Lecture Notes in Computer Science, Vol 2260:360–363, 2001.
8. COHEN Henri . A course in computational algebraic number theory. Springer-Verlag Graduate texts in mathematics, volume 138, 1993.
9. COURTOIS Nicolas, FINIASZ Matthieu , and SENDRIER Nicolas . How to achieve a maceliece-based digital signature scheme. Asiacrypt 2001 volume 248, 2001.
10. GENTRY Craig , SILVERBERG Alice . Hierarchical id-based cryptography. Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology:548–566, December 01-05,2002.
11. HESS Florian. Efficient identity based signature schemes based on pairings. Revised Papers from the 9th Annual International Workshop on Selected Areas in Cryptography:310–324, August 15-16, 2002.
12. NIEDERREITER Harald. Knapsack-type cryptosystems and algebraic coding theory. Prob. Contr. Inform. Theory, 1986.
13. PATERSON K. G. Id-based signatures from pairings on elliptic curves. Cryptology ePrint Archive, Report 2002/003, 2002.
14. SAKAI R., OHGISHI K. and KASAHARA M. Cryptosystems based on pairing. SCIS 2000, 2000.
15. SHAMIR Adi. Identity-based cryptosystems and signature schemes. Advances in Cryptology-Crypto'84, 1984.
16. SMART N. A id-based authenticated key agreement protocol based on the weil pairings. Electron. Lett. 38(13):630–632, 2002.
17. STERN Jacques. A new identification scheme based on syndrome decoding. Lecture Notes in Computer Science 773, 1994.
18. MACELIECE R. BERLEKAMP E. and VAN TILBORG H. On the inherent intractability of certain coding problems. IEEE Transactions on Information Theory, IT-24(3), 1978.
19. CANTEAUT Anne and CHABAUD François. A new algorithm for finding minimum-weight words in a linear code: application to primitive narrow-sense bch codes of length 511. IEEE Transactions on Information Theory, IT-44:367–378, 1988.
20. DUAN Pu CUI Shi and CHAN Choong Wah. An efficient identity-based signature scheme with batch verifications. page 22, 2006.
21. LEE P. and BRICKELL E. An observation on the security of maceliece's public-key cryptosystem. Advances in Cryptology-EUROCRYPT'88, C. Gunter:275–280, 1988.
22. LEON J. A probabilistic algorithm for computing minimum weights of large error correcting codes. IEEE Trans. on Information Theory, IT-34:1354–1359, 1988.
23. MACELIECE R.J. A public-key cryptosystem based on algebraic coding theory. JPL DSN Progress Report 42-44:114–116, 1978.
24. SLOANE N.J.A. MACWILLIAMS F.J. *The Theory of Error Correcting Codes*, north-holland. 1977.

25. SENDRIER N. *Cryptosystèmes à clé publique basés sur les codes correcteurs d'erreurs*. Mémoire d'habilitation, Inria 2002,available at: <http://www-rocq.inria.fr/codes/Nicolas.Sendrier/pub.html>.
26. SENDRIER N. On the security of the maceliece public-key cryptosystem. In: M. Blaum, P.G. Farrell and H. van Tilborg,editors, Information, Coding and Mathematics:141–163, 2002.
27. STERN Jacques. A method for finding codewords of small weight. coding theory and applications. Lecture Notes in Comput. Sci., 388, New York:106–113, 1989.
28. YI X. An identity-based signature scheme from the weil pairing. IEEE Communications Letters 7(2):76–78, 200.
29. YOON H., CHEON J. H. and KIM Y. Batch verifications with id-based signatures. ICISC 2004, LNCS 3506:223–248, 2005.