Cryptographie basée sur les codes correcteurs d'erreurs et arithmétique

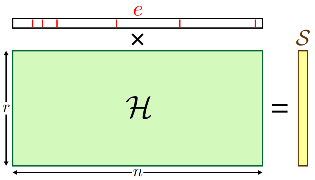
Laboratoire Hubert Curien, UMR CNRS 5516,
Bâtiment F 18 rue du professeur Benoît Lauras
42000 Saint-Etienne
France
pierre.louis.cayrel@univ-st-etienne.fr

16 Novembre 2011

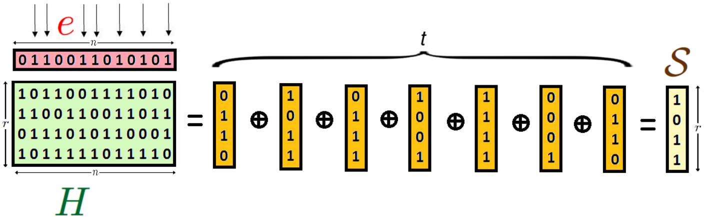
Syndrome decoding problem

1. **Input.**
   H : matrix of size $r \times n$
   $\mathcal{S}$: vector of $\mathbb{F}_2^r$
   t : integer

2. **Problem.** Does there exist a vector e of $\mathbb{F}_2^n$ of weight t such that :



- Problem NP-complete
  E.R. BERLEKAMP, R.J. MCELIECE and H.C. VAN TILBORG 1978


e
n
0 1 1 0 0 1 1 0 1 0 1 0 1
1 0 1 1 0 0 1 1 1 1 0 1 0
1 1 0 0 1 1 0 0 1 1 0 1 1
0 1 1 1 0 1 0 1 1 0 0 0 1
1 0 1 1 1 1 1 0 1 1 1 1 0
r
n
H
t
S
r

Code-based crypto

Error-correcting codes
Encryption with codes
Signature with codes
Identification with codes
Secret-key crypto with codes
Open problems

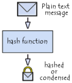
What can we do with this problem ?

- encryption

- signature

- identification

- hash function



- stream cipher

Menu

Error-correcting codes

- make possible the correction of errors when the communication is done on a noisy channel.
  - we add redundancy to the information transmitted.

$$\begin{array}{ccccc} & & \text{Noise} & & \\ & & \downarrow e & & \\ c = \boxed{m}\boxed{r} & \longrightarrow & \boxed{\text{Channel}} & \longrightarrow & y = c + e \end{array}$$

  - by correcting the errors when the message is corrupted.
- stronger than a control of parity, they can detect and correct errors.

We use them :

- DVD,CD : reduce the effects of dust ...
- Phone : improve the quality of the communication.
- cryptography ?

Linear codes

- most used in error correction
- error correcting codes for which redundancy depends linearly on the information
- can be defined by a generator matrix :
  - c is a word of the code $\mathcal{C}$ if and only if :

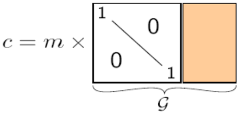


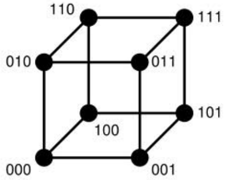
Figure: $\mathcal{G}$: generator matrix in systematic form

The generator matrix $\mathcal{G}$:

- is a $r \times n$ matrix;
- rows of $\mathcal{G}$ form a basis for the code $\mathcal{C}$.

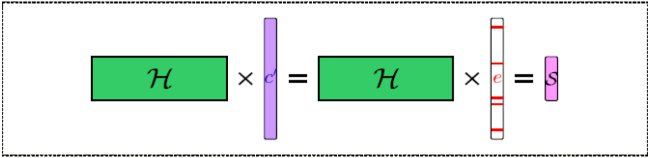
Minimum distance

- The Hamming weight of a word c is the number of non-zero coordinates.
- The minimum distance d of a code is the minimum of the Hamming weight between two words of the code.
- It is also the smallest weight of a non-zero vector.



The parity check matrix $\mathcal{H}$ is orthogonal to $\mathcal{G}$:

- it's a $r \times n$ matrix;
- it's the generator matrix of the dual;
- the code $\mathcal{C}$ is the kernel of $\mathcal{H}$.
  - $c \in \mathcal{C}$ if and only if $\mathcal{H}c = 0$.
- $s = \mathcal{H} \cdot c' = \mathcal{H} \cdot c + \mathcal{H} \cdot e$ is the syndrome of the error.

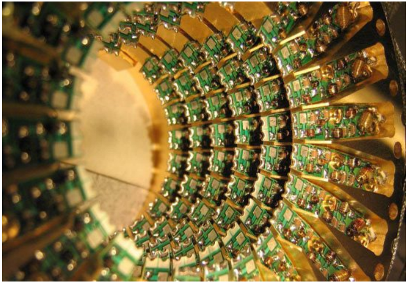


Code based cryptosystems

- introduced at the same time than RSA by McEliece

+ advantages :
  - faster than RSA ;
  - not based on number theory problem (PQ secure) ;
  - does not need cryptoprocessors ;
  - based on hard problem (syndrome decoding problem ...)

– disadvantages :
  - size of public keys (few hundred bits...)



CISCO

Point of View

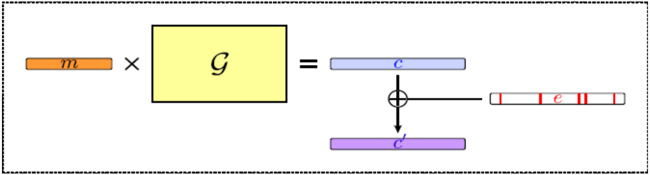
Top 25 Technology Predictions

By Dave Evans, Chief Futurist, Cisco IBSG Innovations Practice

1. By 2029, 11 petabytes of storage will be available for $100—equivalent to 600+ years of continuous, 24-hour-per-day, DVD-quality video. (Source: Cisco IBSG, 2009)
2. In the next 10 years, we will see a 20-time increase in home networking speeds. (Source: Cisco IBSG, 2009)
3. By 2013, wireless network traffic will reach 400 petabytes a month. Today, the entire global network transfers 9 exabytes per month. (Source: FCC Head Julius Genachowski)
4. By the end of 2010, there will be a billion transistors per human—each costing one ten-millionth of a cent. (Sources: Intel Corporation; Cisco IBSG, 2006-2009; IBM)
5. The Internet will evolve to perform instantaneous communication, regardless of distance. (Source: Cisco IBSG, 2009)
6. The first commercial quantum computer will be available by mid-2020. (Source: Cisco IBSG, 2009)
7. By 2020, a $1,000 personal computer will have the raw processing power of a human brain. (Sources: Hans Moravec, Robotics Institute, Carnegie Mellon University, 1998; Cisco IBSG, 2006-2009)

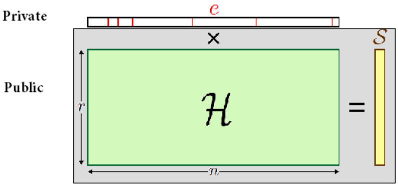
How does the McEliece PKC work ?

- generate a code for which we have a decoding algorithm and $\mathcal{G}'$ the generator matrix.
  - this is the private key.
- transform $\mathcal{G}'$ to obtain $\mathcal{G}$ which seems random.
  - this is the public key.
- encrypt a message m by computing :
  - $c' = m \times \mathcal{G} \oplus e$ with e a random vector of weight t.



A dual construction using $\mathcal{H}$ instead of $\mathcal{G}$?

- Security equivalent to McEliece scheme.
- Private key :
  - $\mathcal{C}$ a $[n, r, d]$ code which corrects t errors,
  - $\mathcal{H}'$ a parity check matrix of $\mathcal{C}$,
  - a $r \times r$ invertible matrix Q,
  - a $n \times n$ permutation matrix P.
- Public key : $\mathcal{H} = Q\mathcal{H}'P$.
- Encryption :
  - $\phi_{n,t} : m \mapsto e$, with e of weight t.
  - $e \mapsto y = \mathcal{H}e$



- Decryption : decode $Q^{-1}y = (Q^{-1}Q)\mathcal{H}'Pe$ in Pe, then $P^{-1}Pe$ gives e.

Arithmetic ?

- Encryption : $\mathcal{O}(n^2)$ binary operations : linear algebra, matrix-vector product
- Decryption : $\mathcal{O}(n^2)$ binary operations : linear algebra, matrix-vector product and a bit more (root finding)
- Size of key : $r \times n$

\+ very fast ;

– public key very big : about 500 000 bits for the original system!

Hardware?

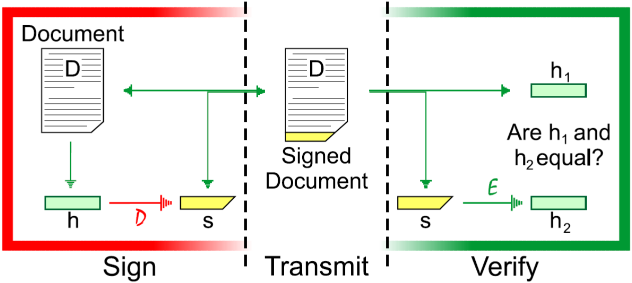
- Eisenbarth *et al.* "MicroEliece: McEliece for Embedded Devices", CHES'09.
- Shoufan *et al.* "A Novel Processor Architecture for McEliece Cryptosystem and FPGA Platforms", ASAP 2009
- Heyse. "Low-Reiter: Niederreiter Encryption Scheme for Embedded Microcontrollers", PQCrypto 2010
- Strenzke. "A Smart Card Implementation of the McEliece PKC", WISTP 2010
- Heyse. "CCA2 secure McEliece based on Quasi Dyadic Goppa Codes for Embedded Devices", PQCrypto 2011

	Method	Platform	Throughput bits/sec
8-bit μC	Niederreiter encryption	ATxMega256@32MHz	119,889
	Niederreiter decryption	ATxMega256@32MHz	1.066
	McEliece encryption	ATxMega192@32MHz	3.889
	McEliece decryption	ATxMega192@32MHz	2.835
	QD-McEliece encryption	ATxMega256@32MHz	6.481
	QD-McEliece decryption	ATxMega256@32MHz	1.229
	ECC-P160	ATMega128@8MHz	197/788[1]
	RSA-1024 $2^{16}+1$	ATMega128@8MHz	2,381/9,524[1]
	RSA-1024 random	ATMega128@8MHz	93/373[1]
FPGA	Niederreiter encryption	Spartan-3 2000-5	14,814,815
	Niederreiter decryption	Spartan-3 2000-5	723,545
	McEliece encryption	Spartan-3AN 1400-5	1,626,517
	McEliece decryption	Spartan-3AN 1400-5	161,829
	ECC-P160	Spartan-3 1000-4	31,200
	RSA-1024 random	Spartan-3E 1500-5	20,275

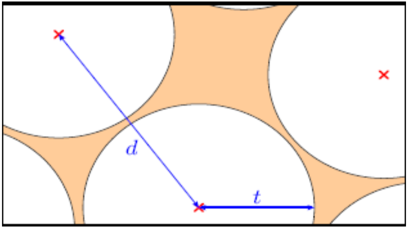
[1] For a fair comparison with our implementations running at 32MHz, timings at lower frequencies were scaled accordingly.

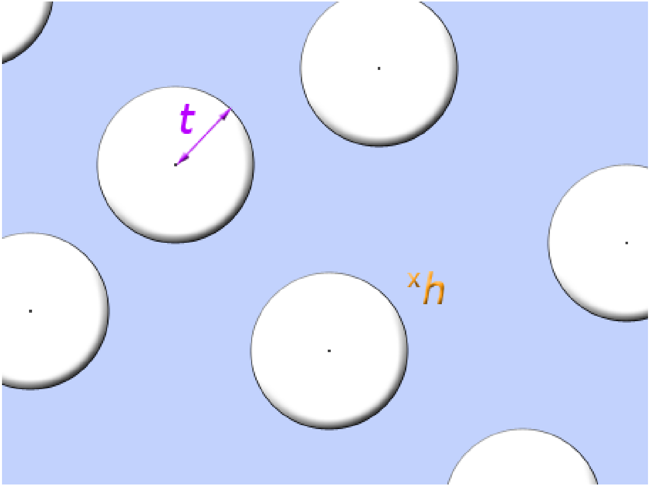
Figure: from Heyse's slides

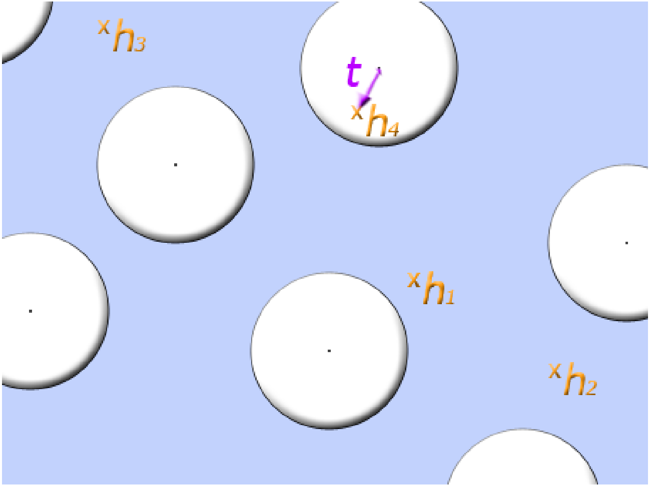
- PKC $\rightarrow$ signature.
    - RSA yes
    - McEliece and Niederreiter no directly



- **Problem:** McEliece and Niederreiter not invertible.
  - if we take $y \in \mathbb{F}_2^n$ random and a code $\mathcal{C}[n, k, d]$ for which we are able to decode $d/2$ errors, it is almost impossible to decode y in a word of $\mathcal{C}$.
- **Solution:**
  - the hash value has to be decodable !




t
x h

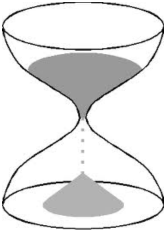

h3
t
h4
h1
h2

- d the message to sign, we compute $M = h(d)$
- h a hash function with values in $\mathbb{F}_2^r$
  - we search $e \in \mathbb{F}_2^n$ of given weight t with $h(M) = \mathcal{H}e$
- let γ be a decoding algorithm

1. $i \leftarrow 0$
2. `while` $h(M|i)$ `is not decodable do` $i \leftarrow i + 1$
3. `compute` $e = \gamma(h(M|i))$

Figure: CFS signature scheme

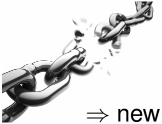
- signer sends $\{e, j\}$ such that $h(M|j) = \mathcal{H}e$



- we need a dense family of codes : Goppa codes
- binary Goppa codes
  - t small
  - the probability for a random element to be decodable (in a ball of radius t centered on the codewords) is $\approx \frac{1}{t!}$
- we take $n = 2^m$, $m = 16$, $t = 9$.
- we have 1 chance over $9! = 362880$ to have a decodable word.

signature cost	$t!t^2m^3$	12×10^{11} op. $\approx$ 1 min on FPGA
signature length	$(t-1) \times m + log_2 t$	131 bits
verification cost	$t^2 m$	1 296 op.
PK size	$tm2^m$	1 MB

- cons :
  - decode several words ($t!$) before to find a good one
    - 70 times slower than RSA
  - t small leads to very big parameters
    - public key of 1 MB



$\Rightarrow$ new PK size : several MB, time to sign : several weeks ...

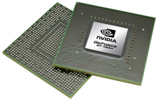
- solution : use structured codes (smaller public key size around 720 KB) and a GPU to have a signature in less than 2 minutes ...

1	2	3	4	5	6	7	8
2	1	4	3	6	5	8	7
3	4	1	2	7	8	5	6
4	3	2	1	8	7	6	5
5	6	7	8	1	2	3	4
6	5	8	7	2	1	4	3
7	8	5	6	3	4	1	2
8	7	6	5	4	3	2	1

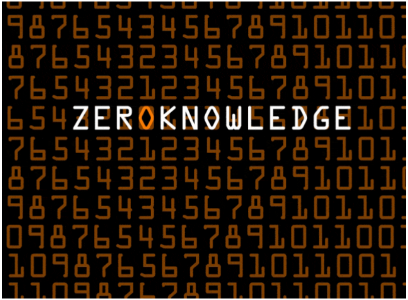
Arithmetic ?

- Signature : matrix-vector product, hash-function (matrix-vector product we will see it later), decoding algorithm (root finding of polynomial over $\mathbb{F}_q$)
- Verification : a hash-function and a matrix vector-product
- Size of key : $r \times n$ (big)

+ very fast verification : a hash value and a matrix vector product ;
+ one of the smallest signature size : around 150 bits ;

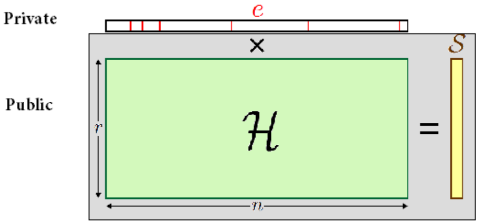


– public key big : about 1MB for the original system!
– signing process very long : around 2 minutes with a GPU !

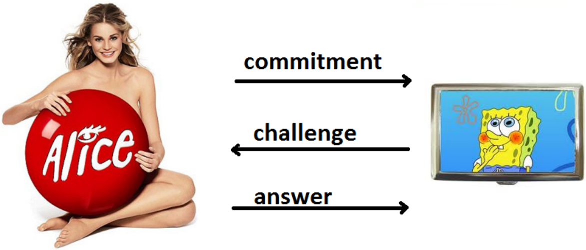


- zero-knowledge,
- the security is based on the syndrome decoding problem.

- generate a random matrix $\mathcal{H}$ of size $r \times n$
- we choose an integer t which is the weight
  - this is the public key $(\mathcal{H}, t)$
- each user receive e of n bits and weight t.
  - this is the private key
- each user compute : $\mathcal{S} = \mathcal{H}e$.
  - just once for $\mathcal{H}$ fixed
  - $\mathcal{S}$ is public



- *A* wants to prove to *B* that she knows the secret but she doesn't want to divulgate it.



- The protocol is on λ rounds and each of them is defined as follows.

A chooses y of n bits randomly and a permutation σ of $\{1, 2, \ldots, n\}$.
A sends to B : c_1, c_2, c_3 such that :

$$c_1 = h(\sigma|\mathcal{H}y);\ c_2 = h(\sigma(y));\ c_3 = h(\sigma(y \oplus e))$$

commitment →

A chooses y of n bits randomly and a permutation σ of $\{1, 2, \ldots, n\}$.
A sends to B : c_1, c_2, c_3 such that :

$$c_1 = h(\sigma | \mathcal{H}y); \; c_2 = h(\sigma(y)); \; c_3 = h(\sigma(y \oplus e))$$

commitment →

challenge ←

B sends to A a random $b \in \{0, 1, 2\}$.

A chooses y of n bits randomly and a permutation σ of $\{1, 2, \ldots, n\}$.
A sends to B : c_1, c_2, c_3 such that :

$$c_1 = h(\sigma|\mathcal{H}y);\ c_2 = h(\sigma(y));\ c_3 = h(\sigma(y \oplus e))$$

commitment →

← **challenge**

B sends to A a random $b \in \{0, 1, 2\}$.

Three possibilities:

1. if $b = 0$: A reveals y and σ
2. if $b = 1$: A reveals $(y \oplus e)$ and σ
3. if $b = 2$: A reveals $\sigma(y)$ and $\sigma(e)$

answer →

1. if $b = 0$: B checks that c_1, c_2 are correct
2. if $b = 1$: B checks that c_1, c_3 are correct
3. if $b = 2$: B checks that c_2, c_3 are correct and that $\omega(\sigma(e)) = t$

- for each round : probability to cheat is $\frac{2}{3}$.
  - for a security of $\frac{1}{2^{80}}$, we need 150 rounds.

Idea : Replace the random matrix $\mathcal{H}$ by the parity check matrix of a certain family of codes : *the double-circulant codes.*

- Let ℓ be an integer.
- a random double circulant matrix $\ell \times 2\ell$ $\mathcal{H}$ is defined as :

$$\mathcal{H} = (I|A) \ ,$$

where A is a *cyclic matrix*, of the form :

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_\ell \\ a_\ell & a_1 & a_2 & \cdots & a_{\ell-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{pmatrix} ,$$

where $(a_1, a_2, a_3, \cdots, a_\ell)$ is a random vector of $\mathbb{F}_2^\ell$.

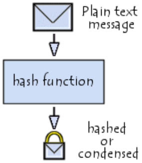
- Store $\mathcal{H}$ needs only ℓ bits.

- the minimum distance is the same as random matrices,
- the syndrom decoding is still hard,
- very interesting for implementation in low ressource devices.

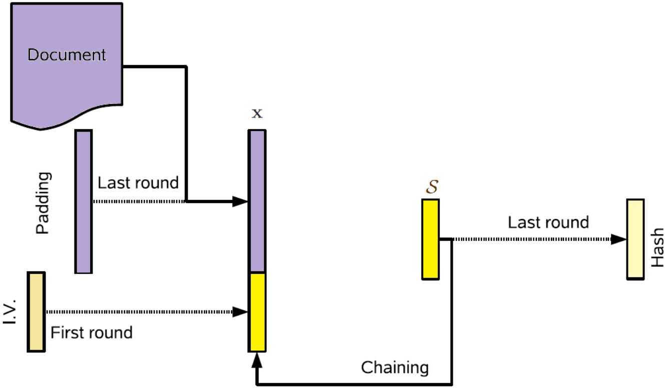
- Let n equal 2ℓ
- **Private data :** the secret e of bit-length n.
- **Public data :** n bits ($\mathcal{S}$ of size ℓ and the first row of H, ℓ bits).

- at least $\ell = 347$ and $t = 74$ for a security of 2^{85}
- public and secret key sizes of $n = 694$ bits

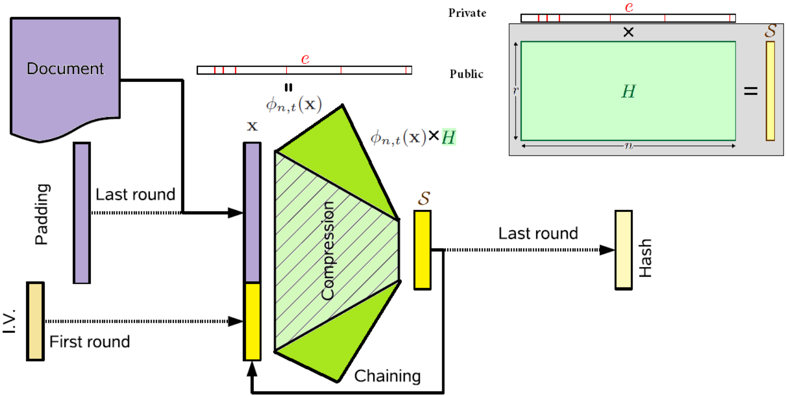
Hash-function and pseudo-random number generator



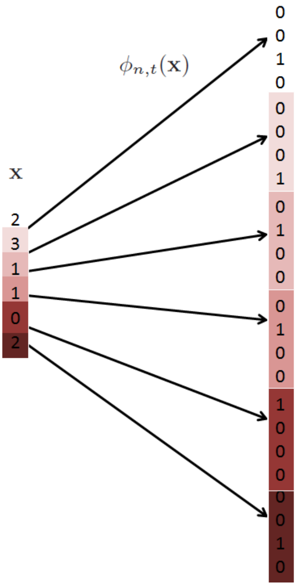
How to hash with codes ?



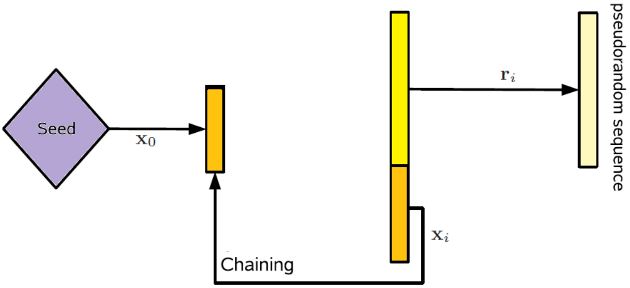
How to hash with codes ?



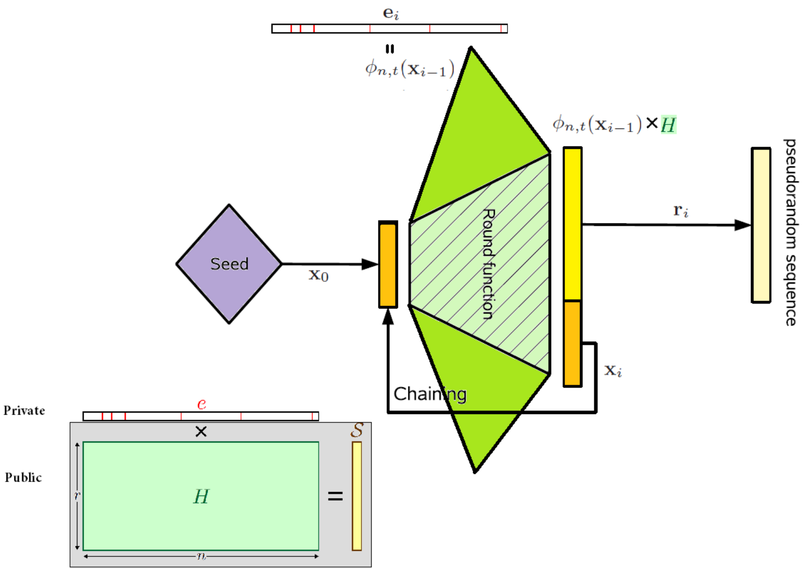
How $\phi_{n,t}$ could work?



How to generate pseudo-random sequences ?



How to generate pseudo-random sequences ?



Encryption :

- Study of the QC/QD constructions ;
- Identity-based encryption.

Signature :

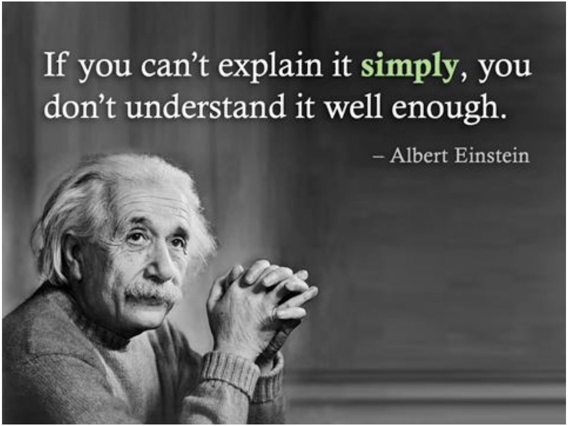
- FPGA implementation ;
- Smaller public keys.

Identification :

- 3-pass and soundness 1/2 ;
- Efficient implementation.

Secret-key :

- Fast schemes ;
- Study of side-channel attacks.


If you can't explain it simply, you don't understand it well enough.
– Albert Einstein

Back-up slides

- My publications in :
  - encryption : page 49
  - signature : page 50
  - identification : page 53
  - secret-key : page 55
  - cryptanalysis : page 56
  - others : page 57
- attack : page 58
- constant weight encoder : page 60
- best weight : page 61

My contributions - Encryption

- ⋆⋆ **Reducing Key Length of the McEliece Cryptosystem**
  *T. P. Berger, **P.-L. Cayrel**, P. Gaborit and A. Otmani*
  AfricaCrypt 2009, LNCS 5580, pages 77-97, Springer-Verlag, 2009
- **McEliece/Niederreiter PKC: sensitivity to fault injection**
  ***P.-L. Cayrel** and P. Dusart*
  FEAS 2010, IEEE
- **Implementation of the McEliece scheme based on compact (flexible) quasi-dyadic public keys**
  ***P.-L. Cayrel** and G. Hoffman*
  eSmart 2010 (not presented)
- **Fault injection's sensitivity of the McEliece PKC**
  ***P.-L. Cayrel** and P. Dusart*
  WEWoRC 2009, pages 84-88

My contributions - Signature - I

- ⋆⋆ **Identity-based Identification and Signature Schemes using Error Correcting Codes**
  P.-L. Cayrel, *P. Gaborit and M. Girault*
  Identity-Based Cryptography, chapter 8, 2009
- ⋆⋆⋆ **A New Efficient Threshold Ring Signature Scheme based on Coding Theory**
  C. Aguilar Melchor, ***P.-L. Cayrel****, P. Gaborit and F. Laguillaumie*
  IEEE Trans. Inf. Theory, number 57(7), pages 4833-4842, 2011
- ⋆ **Quasi Dyadic CFS Signature Scheme**
  P.S.L.M. Barreto, ***P.-L. Cayrel****, R. Misoczki and R. Niebuhr*
  InsCrypt 2010, LNCS 6584, pages 336-349, Springer-Verlag, 2010
- ⋆ **A Lattice-Based Threshold Ring Signature Scheme**
  P.-L. Cayrel*, R. Lindner, M. Rückert and R. Silva*
  LatinCrypt 2010, LNCS 6212, pages 255-272, Springer-Verlag, 2010

My contributions - Signature - II

- ⋆⋆ **A New Efficient Threshold Ring Signature Scheme based on Coding Theory**
  *C. Aguilar Melchor, **P.-L. Cayrel** and P. Gaborit*
  PQCrypto 2008, LNCS 5299, pages 1-16, Springer-Verlag, 2008
- ⋆⋆⋆ **Secure Implementation of the Stern Signature Scheme for Low-Resource Devices**
  ***P.-L. Cayrel**, P. Gaborit and E. Prouff*
  CARDIS 2008, LNCS 5189, pages 191-205, Springer-Verlag, 2008
- **Multi-Signature Scheme based on Coding Theory**
  *M. Meziani and **P.-L. Cayrel***
  ICCCIS 2010, pages 186-192
- **Dual Construction of Stern-based Signature Schemes**
  ***P.-L. Cayrel** and S. M. El Yousfi Alaoui*
  ICCCIS 2010, pages 369-374

My contributions - Signature - III

- **An improved threshold ring signature scheme based on error correcting codes**
  P.-L. Cayrel *and S. M. El Yousfi Alaoui*
  WISSec 2010 (not presented)

⋆⋆ **Identity-based identification and signature schemes using correcting codes**
  P.-L. Cayrel*, P. Gaborit and M. Girault*
  WCC 2007, pages 69-78

My contributions - Identification - I

- **Improved identity-based identification and signature schemes using Quasi-Dyadic Goppa codes**
  *S. M. El Yousfi Alaoui, **P.-L. Cayrel** and M. Meziani*
  ISA 2011, CCIS 200, pages 146-155, Springer-Verlag, 2011

★★★ **A zero-knowledge identification scheme based on the q-ary Syndrome Decoding problem**
__P.-L. Cayrel__, P. Véron and S. M. El Yousfi Alaoui
SAC 2010, LNCS 6544, pages 171-186, Springer-Verlag, 2010

★★ **Improved Zero-knowledge Identification with Lattices**
__P.-L. Cayrel__, R. Lindner, M. Rückert and R. Silva
ProvSec 2010, LNCS 6402, pages 1-16, Springer-Verlag, 2010

★★ **A Lattice-Based Batch Identification Scheme**
R. Silva, __P.-L. Cayrel__ and R. Lindner
ITW 2011, IEEE

My contributions - Identification - II

- **Lattice-based Zero-knowledge Identification with Low Communication Cost**
  *R. Silva, **P.-L. Cayrel** and R. Lindner*
  SBSEG 2011
- **New results on the Stern identification and signature scheme**
  P.-L. Cayrel
  Bulletin of the Transilvania University of Brasov, pages 1-4
- ⋆ **Efficient implementation of code-based identification/signatures schemes**
  ***P.-L. Cayrel**, S. M. El Yousfi Alaoui, Felix Günther, Gerhard Hoffmann and Holger Rother*
  WEWoRC 2011, pages 65-69
- **New results on the Stern identification and signature scheme**
  P.-L. Cayrel
  Colloque Franco Roumain de Mathématiques Appliquées page 53

My contributions - Secret-key

- **S-FSB: An Improved Variant of the FSB Hash Family**
  *M. Meziani, Ö. Dagdelen, **P.-L. Cayrel** and S. M. El Yousfi Alaoui*
  ISA 2011, CCIS 200, pages 132-145, Springer-Verlag, 2011
- **2SC: an Efficient Code-based Stream Cipher**
  *M. Meziani, **P.-L. Cayrel** and S. M. El Yousfi Alaoui*
  ISA 2011, CCIS 200, pages 111-122, Springer-Verlag, 2011
- ⋆ **GPU Implementation of the Keccak Hash Function Family**
  ***P.-L. Cayrel**, G. Hoffmann and M. Schneider*
  ISA 2011, CCIS 200, pages 33-42, Springer-Verlag, 2011
- **Hash Functions Based on Coding Theory**
  *M. Meziani, S. M. El Yousfi Alaoui and **P.-L. Cayrel***
  WCCCS 2011, pages 32-37

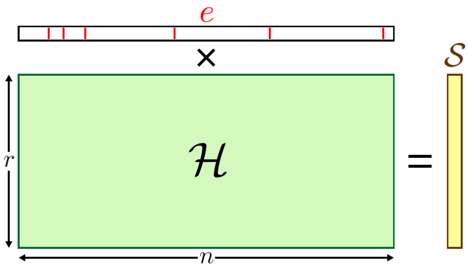
My contributions - Cryptanalysis

- ⋆⋆ **On Kabatianskii-Krouk-Smeets Signatures**
  P.-L. Cayrel, *A. Otmani and D. Vergnaud*
  WAIFI 2007, LNCS 4547, pages 237-251, Springer-Verlag, 2007
- **Improving the efficiency of GBA against certain structured cryptosystems**
  R.Niebuhr, ***P.-L. Cayrel*** *and J. Buchmann*
  WCC 2011, pages 163-172
- **Attacking code/lattice-based cryptosystems using Partial Knowledge**
  R.Niebuhr, ***P.-L. Cayrel****, S. Bulygin and J. Buchmann*
  InsCrypt 2010, Science Press of China
- ⋆ **On lower bounds for Information Set Decoding over Fq**
  R. Niebuhr, ***P.-L. Cayrel****, S. Bulygin and J. Buchmann*
  SCC 2010, pages 143-157

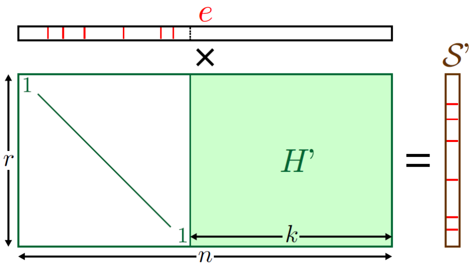
My contributions - Others

- ⋆⋆ **Quasi-cyclic codes as codes over rings of matrices**
  P.-L. Cayrel*, C. Chabot and A. Necer*
  Finite Fields and their Applications, number 16(2), pages 100-115, 2010
- **Recent progress in code-based cryptography**
  P.-L. Cayrel*, S. M. El Yousfi Alaoui, G. Hoffmann, M. Meziani and R. Niebuhr*
  ISA 2011, CCIS 200, pages 21-32, Springer-Verlag, 2011
- **Post-Quantum Cryptography: Code-based Signatures**
  P.-L. Cayrel *and M. Meziani*
  ISA 2010, LNCS 6059, pages 82-99, Springer-Verlag, 2010
- **Side channels attacks in code-based cryptography**
  P.-L. Cayrel *and F. Strenzke*
  COSADE 2010, pages 24-28
- **Improved algorithm to find equations for algebraic attacks for combiners with memory**
  F. Armknecht, ***P.-L. Cayrel****, P. Gaborit and O. Ruatta*
  BFCA 2007, pages 81-98

Information Set Decoding



Information Set Decoding



$\phi : m \mapsto x$ with x of weight t

This application is called a constant weight encoder.

Enumerative coding:

$$\begin{array}{llll} \phi^{-1} : & W_{n,t} & \longrightarrow & \left[0, \binom{n}{t}\right[\\ & (i_0, i_1, \ldots, i_{t-1}) & \longmapsto & \binom{i_0}{1} + \binom{i_1}{2} + \cdots + \binom{i_{t-1}}{t} \end{array}$$

How to choose the weight for an optimal complexity ?