Recent results on rank based cryptography

Philippe Gaborit

University of Limoges, France

Caen 20 juin 2018

Summary

Post-quantum cryptography

General problems

Cryptography needs different difficult problems

- factorization
- discrete log
- SVP for lattices
- syndrome decoding problem

For code-based cryptography, the security of cryptosystems is usually related to the problem of syndrome decoding for a special metric.

PQ Crypto

Consider the simple linear system problem :
H a random $(n-k) \times n$ matrix over $GF(q)$

Knowing $s \in GF(q)^{n-k}$ is it possible to recover a given $x \in GF(q)^n$ such that $H.x^t = s$?

Easy problem :

- fix $n-k$ columns of H , one gets a $(n-k) \times (n-k)$ submatrix A of H
- A invertible with good probability, $x = A^{-1}s$.

How to make this problem difficult ?

(1) add a constraint to x : x of small weight for a particular metric

- metric = Hamming distance $\Rightarrow$ **code-based cryptography**
- metric = Euclidean distance $\Rightarrow$ **lattice-based cryptography**
- metric = Rank distance $\Rightarrow$ **rank-based cryptography**

$\Rightarrow$ only difference : the metric considered, and its associated properties ! !

(2) consider rather a multivariable non linear system : quadratic, cubic etc...

$\Rightarrow$ Mutivariate cryptography

General interest of post-quantum cryptogrphy

- a priori resistant to a quantum computer
- usually faster than number-theory based cryptography
- easier to protect against side-channel attacks
- size of keys may be larger

Lattice-based cryptography

- Knapsack '78, NTRU '96, GGH '97
- Regev '04 LWE
- difficult problem : finding short vectors in lattices
- cryptanalysis : LLL algorithm with heuristics
- FHE, better security reduction ?, reasonable size of keys

Code-based cryptography

- McEliece '78, Stern '93,CFS '01, Aleknovich '03, G. '05, MDPC '13
- difficult problem : syndrome decoding problem
- cryptanalysis : ISD, closed formulae
- faster than lattices ?, reasonable size of keys with cyclicity, security reduction ?

Multivariate cryptography

- Matsumoto-Imai ’88, HFE ’95, SFlash ’96, Rainbow ’05, QUAD ’06....
- difficult problem : solving a multivariable system
- cryptanalysis : Groebner basis
- many instation broken (Crypto ’07), security reduction ?, unreasonable size of keys

Rank Codes : definition and basic properties

Rank metric codes

The rank metric is defined in finite extensions.

- $GF(q)$ a finite field with q a power of a prime.
- $GF(q^m)$ an extension of degree m of $GF(q)$.
- $B = (b_1, ..., b_m)$ a basis of $GF(q^m)$ over $GF(q)$.

$GF(q^m)$ **can be seen as a vector space on** $GF(q)$**.**

- $\mathcal{C}$ a linear code over $GF(q^m)$ of dimension k and length n.
- G a $k \times n$ generator matrix of the code $\mathcal{C}$.
- H a $(n-k) \times n$ parity check matrix of $\mathcal{C}$, $G.H^t = 0$.
- H a dual matrix, $x \in GF(q^m)^n \rightarrow$ syndrome of $x =$ $H.x^t \in GF(q^m)^{n-k}$

Rank metric

Words of the code $\mathcal{C}$ are n-uplets with coordinates in $GF(q^m)$.

$$v = (v_1, \ldots, v_n)$$

with $v_j \in GF(q^m)$.
Any coordinate $v_j = \sum_{i=1}^m v_{ij} b_i$ with $v_{ij} \in GF(q)$.

$$v(v_1, \ldots, v_n) \rightarrow V = \begin{pmatrix} v_{11} & v_{12} & \ldots & v_{1n} \\ v_{21} & v_{22} & \ldots & v_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ v_{m1} & v_{m2} & \ldots & v_{mn} \end{pmatrix}$$

Definition (Rank weight of word)

v has rank $r = Rank(v)$ iff the rank of $V = (v_{ij})_{ij}$ is r.
equivalently $Rank(v) = r <=> v_j \in V_r \subset GF(q^m)^n$ with dim(V_r)=r.

the determinant of V does not depend on the basis

Definition (Rank distance)

Let $x, y \in GF(q^m)^n$, the rank distance between x and y is defined by $d_R(x, y) = Rank(x - y)$.

Definition (Minimum distance)

Let C be a $[n, k]$ rank code over $GF(q^m)$, the minimum rank distance d of C is $d = min\{d_R(x, y) | x, y \in C, x \neq y\}$;

Theorem (Unique decoding)

Let $C[n, k, d]$ be a rank code over $GF(q^m)$. Let e an error vector with $r = Rank(e) \leq \frac{d-1}{2}$, and $c \in C$:
if $y = c + e$ then there exists a unique element $c' \in C$ such that $d(y, c') = r$. Therefore $c' = c$.

proof : same as for Hamming, distance property.

Rank isometry

Notion of **isometry** : weight preservation

- Hamming distance : $n \times n$ permutation matrices
- Rank distance : $n \times n$ invertible matrices over $GF(q)$

proof : multiplying a codeword $x \in GF(q^m)^n$ by an $n \times n$ invertible matrix **over the base field GF(q)** does not change the rank (see x as a $m \times n$ matrix over $GF(q)$).

remark : for any $x \in GF(q^m)^n$: $Rank(x) \leq w_H(x)$: potential linear combinations on the x_i may only decrease the rank weight.

Support analogy

An important insight between Rank and Hamming distances tool : support analogy

- **support of a word of $GF(q)^n$ in Hamming metric**
  $x(x_1, x2, \cdots, x_n)$: set of positions $x_i \neq 0$
- **support of a word of $GF(q)^n$ in rank metric**
  $x(x_1, x2, \cdots, x_n)$: the subspace over $GF(q)$, $E \subset GF(q^m)$ generated by $\{x_1, \cdots, x_n\}$
- in both cases if the order of size of the support is small, knowing the support of x and syndrome $s = H.x^t$ permits to recover the complete coordinates of x.

Sphere packing bound

Counting the number of possible supports for length n and dimension t

- Hamming : number of sets with t elements in sets of n elements : Newton binomial $\binom{n}{t}$ $(\leq 2^n)$
- Rank : number of subspaces of dimension t over $GF(q)$ in the space of dimension n $GF(q^m)$: Gaussian binomial $\begin{bmatrix} n \\ t \end{bmatrix}_q (\sim q^{t(n-t)})$

Sphere packing bound

Theorem (Sphere packing bound)

Let $C[n,k,d]$ be a rank code over $GF(q^m)^n$, the parameters n,k,d and d satisfy : $q^{mk}B(n,m,q,\lfloor\frac{d-1}{2}\rfloor) \leq q^{nm}$

Theorem (Singleton bound)

Let $C[n,k,d]$ be a rank code over $GF(q^m)^n$, the parameters n,k and d satisfy : $d \leq 1 + \lfloor\frac{(n-k)m}{n}\rfloor$

The **rank Gilbert-Varshamov (GVR) bound** for a $C[n,k]$ rank code over $GF(q^m)^n$ with dual matrix H corresponds to the average value of the minimum distance of a random $[n,k]$ rank code.

asymptotically : in the case $m=n$: $\frac{GVR(n,k,m,q)}{n} \sim 1-\sqrt{\frac{k}{n}}$

Decoding in rank metric

Families of decodable codes in rank metric

There exists 3 main families of decodable codes in rank metric

- Gabidulin codes (1985) (analog of Reed-Solomon codes with rank metric and q-polynomials)
- Simple codes (2008,2017)
- LRPC codes (2013)

These codes have different properties, a lot of attention was given to rank metric and especially to subspace metric with the development of Network coding in the years 2000's.

LRPC codes

LDPC : dual with low weight (ie : small support)
$\rightarrow$ equivalent for rank metric : dual with small rank support

Definition (GMRZ13)

A Low Rank Parity Check (LRPC) code of rank d, length n and dimension k over F_{q^m} is a code such that the code has for parity check matrix, a $(n-k) \times n$ matrix $H(h_{ij})$ such that the vector space F of F_{q^m} generated by its coefficients h_{ij} has dimension at most d. We call this dimension the weight of H.

In other terms : all coefficients h_{ij} of H belong to the same 'low' vector space $F < F_1, F_2, \cdots, F_d >$ of F_{q^m} of dimension d.

Decoding LRPC codes

Idea : as usual recover the support and then deduce the coordinates values.

Let $e(e_1, ..., e_n)$ be an error vector of weight r, ie : $\forall e_i : e_i \in E$, and dim(E)=r. Suppose $H.e^t = s = (s_1, ..., s_{n-k})^t$.

$$e_i \in E < E_1, ..., E_r >, h_{ij} \in F < F_1, F_2, \cdots, F_d >$$

$$\Rightarrow s_k \in < E_1F_1, .., E_rF_d >$$

$\Rightarrow$ if $n-k$ is large enough, it is possible to recover the product space $< E_1F_1, .., E_rF_d >$

Decoding LRPC codes

Syndrome $s(s_1, .., s_{n-k})$ **:** $S =< s_1, .., s_{n-k} >\subset< E_1F_1, .., E_rF_d >$

Suppose $S =< E.F > \Rightarrow$ possible to recover E.

Let $S_i = F_i^{-1}.S$, since

$$S =< E.F >=< F_iE_1, F_iE_2, .., F_iE_r, ... >\Rightarrow E \subset S_i$$

$$\mathbf{E} = \mathbf{S_1} \cap \mathbf{S_2} \cap \cdots \cap \mathbf{S_d}$$

General decoding of LRPC codes

Let $y = xG + e$

1. **Syndrome space computation**
   Compute the syndrome vector $H.y^t = s(s_1, \cdots, s_{n-k})$ and the syndrome space $S =< s_1, \cdots, s_{n-k} >$.
2. **Recovering the support E of the error**
   $S_i = F_i^{-1}S$, $E = S_1 \cap S_2 \cap \cdots \cap S_d$,
3. **Recovering the error vector** e Write $e_i (1 \leq i \leq n)$ in the error support as $e_i = \sum_{i=1}^{n} e_{ij} E_j$, solve the system $H.e^t = s$.
4. **Recovering the message** x
   Recover x from the system $xG = y - e$.

Decoding of LRPC

- **Conditions of success**
  - $S =< F.E > \Rightarrow$ rd $\leq$ n-k.
  - possibility that $dim(S) \neq n-k \Rightarrow$ probabilistic decoding with error failure in $q^{-(n-k-rd)}$
  - if $d = 2$ can decode up to $(n-k)/2$ errors.
- **Complexity of decoding** : very fast symbolic matrix inversion $O(m(n-k)^2)$ write the system with unknowns : $e_E = (e_{11}, ..., e_{nr})$: rn unknowns in $GF(q)$, the syndrome s is written in the symbolic basis $\{E_1F_1, ..., E_rF_d\}$, H is written in $h_{ij} = \sum h_{ijk}F_k, \rightarrow nr \times m(n-k)$ matrix in $GF(q)$, can do precomputation.
- Decoding Complexity $O(m(n-k)^2)$ op. in $GF(q)$
- **Comparison with Gabidulin codes :** probabilistic, decoding failure, but as fast

Recent improvement for decoding LRPC codes

Aragon,G., Hauteville, Ruatta , Zémor '18
Remark that if dimension(S)=rd-c then
for $c \leq r$

$$dimension(S_i \cap E) \geq r - c$$

$\rightarrow$ possibility to recover elements of Support(E) even if $dim(S) < rd$
$\rightarrow$ permits a better decoding $\frac{(n-k)}{2} \rightarrow \frac{2(n-k)}{3}$
or smaller failure decoding probability
$q^{-(n-k-rd+1)} \rightarrow q^{-(n-k-2(r+d)+5)}$

Complexity issues : decoding random rankcodes

Rank syndrome decoding

For cryptography we are interested in difficult problems, in the case of rank metric the problem is :

Definition (Rank Syndrome Decoding problem (RSD))

Instance : a $(n-k)\times n$ matrix H over $GF(q^m)$, a syndrome s in $GF(q^m)^{n-k}$ and an integer w
Question : does there exist $x \in GF(q^m)^n$ such that $H.x^t = s$ and $w_R(x) \leq w$?

Definition (Syndrome Decoding problem (SD))

Instance : an $r \times n$ matrix $H = [h_1, h_2, \ldots, h_n]$ over a field $GF(q)$, a column vector $s \in GF(q)^r$, an integer w
Question : does there exist $x = (x_1, \ldots, x_n) \in GF(q)^n$ of Hamming weight at most w such that $H^t x = \sum_{i=1}^{n} x_i h_i = s$?

Problem SD proven NP-complete by Berlekamp et al. in 1978.
Computational complexity of RSD : solved in 2014 (G.,Zemor 2014)

Definition (embedding strategy)

Let $m \geq n$ and $Q = q^m$. Let $\alpha = (\alpha_1, \dots \alpha_n)$ be an n-tuple of elements of $GF(Q)$. Define the embedding of $GF(q)^n$ into $GF(Q)^n$

$$\begin{array}{rccc} \psi_\alpha : & GF(q)^n & \to & GF(Q)^n \\ & x = (x_1, \dots, x_n) & \mapsto & \mathbf{x} = (x_1\alpha_1, \dots x_n\alpha_n) \end{array}$$

and for any $GF(q)$-linear code C in $GF(q)^n$, define $\mathcal{C} = \mathcal{C}(C, \alpha)$ as the $GF(Q)$-linear code generated by $\psi_\alpha(C)$, i.e. the set of $GF(Q)$-linear combinations of elements of $\psi_\alpha(C)$.

A randomized reduction

General idea of the embedding :

$$(1,0,0,1,0,1) \rightarrow (\alpha_1,0,0,\alpha_4,0,\alpha_6)$$

Theorem

Let C be a random code over GF(q) and α random, then for convenient m, with a very strong probability :

$$d_H(C) = d_R(\mathcal{C})$$

.

Theorem

If there exists a polynomial time algorithms which solves RSD then P = RP

Best known attacks

There are two types of attacks on the RSD problem :

- Combinatorial attacks
- Algebraic attacks

Depending on type of parameters, the efficiency varies a lot.

Combinatorial attacks

- first attack Chabaud-Stern '96 : basis enumeration
- improvements A.Ourivski and T.Johannson '02
  - Basis enumeration : $\leq (k+r)^3 q^{(r-1)(m-r)+2}$ (amelioration on polynomial part of Chabaud-Stern '96)
  - Coordinates enumeration : $\leq (k+r)^3 r^3 q^{(r-1)(k+1)}$
- improvement : G. et al. '16
  - Support attack : $\mathcal{O}(q^{(r-1)\frac{\lfloor (k+\mathbf{1})m \rfloor}{n}})$
  - improvement Aragon,G.,Hauteville,Tillich ISIT '18 (GRS+) : $\mathcal{O}((nm)^3 q^{r\lceil \frac{km}{n} - m \rceil})$
  - Quantum Speed Up : Grover's algorithm directly applies to GRS+ $\Longrightarrow$ exponent divided by 2.

Basis enumeration Hamming/Rank attacks

- Attack in rank metric to recover the support - a naive approach would consist in trying ALL possible supports : all set of coordinates of weight w

$\Rightarrow$ Of course one never does that ! ! !

- Attack in rank metric to recover the support

The analog of this attack in rank metric : try all possible supports, ie all vector space of dimension r : $q^{(m-r).r}$ such basis, then solve a system.

$\Rightarrow$ it is the Chabaud-Stern ('96) attack - improved by OJ '02

By analogy with the Hamming : it is clearly not optimal In particular the exponent complexity does not depend on n

Improvement : ISD for rank metric

- Information Set Decoding for Hamming distance (simple original approach) : $H.x^t = s$

- syndrome size : $n-k \rightarrow n-k$ equations
- take $n-k$ random columns, if they contain the error support , one can solve a system

- Analog for rank metric :

- syndrome size : $n-k \rightarrow (n-k)m$ equations in $\mathbb{F}_q$
- consider a random space E' of F_q^m of dimension r' which contain E

$\rightarrow$ one can solve if $nr' \geq (n-k)m$

$\rightarrow$ as for ISD for Hamming metric : improve the complexity since easier to find.

Support attack

Detail :
Increasing of searched support : $r' \geq r$ avec $r'n \leq m(n-k)$.

$$e' = \beta U$$

with β a basis of rank r' and U a $r' \times n$ matrix. Operations :

- More support to test : $q^{(r-1)(m-r)} \rightarrow q^{(r'-1)(m-r')}$
- Better probability to find : $\frac{1}{q^{(r-1)(m-r)}} \rightarrow \frac{q^{(r'-m)}}{q^{(r-1)(m-r)}}$

Complexity :
$min(O((n-k)^3 m^3 q^{r\frac{\lfloor km \rfloor}{n}}), O((n-k)^3 m^3 q^{(r-1)\frac{\lfloor (k+1)m \rfloor}{n}}))$

Support attack

Conclusion on the first attack

- Improvement on previous attacks based on $HU^t\beta^t = Hy^t$.
- exponential omplexity in the general case
- Complexité :
  $min(O((n-k)^3m^3q^{r\lfloor\frac{km}{n}\rfloor}), O((n-k)^3m^3q^{(r-1)\lfloor\frac{(k+1)m}{n}\rfloor}))$

Comparison with previous complexities :

- basis enumeration : $\leq (k+r)^3q^{(r-1)(m-r)+2}$
- coordinates enumeration : $\leq (k+r)^3r^3q^{(r-1)(k+1)}$

Remark : when $n = m$ same expoential complexity that OJ '02

Algebraic attacks for rank metric

General idea : translate the problem in equations then try to resolve with grobner basis

Main difficulty : translate in equations the fact that coordinates belong to a same subspace of dimension r in $GF(q^m)$?

- Levy-Perret '06 : Taking error support as unknown $\rightarrow$ quadratic setting
- Kipnis-Shamir '99 (FLP '08) and others..) : Kernel attack, $(r+1) \times (r+1)$ minors $\rightarrow$ degree $r+1$
- G. et *al.* '16 : annulator polynomial $\rightarrow$ degree q^r

Attack with q-polynomials

Definition (q-polynomials)

A q-polynomial is a polynomal of the form $P(x) = \sum_{i=0}^{r} p_i x^{q^i}$ with $p_r \neq 0$ et $p_i \in \mathbb{F}_{q^m}$.

- Linearity : $P(\alpha x + \beta y) = \alpha P(x) + \beta P(y)$ with $x, y \in \mathbb{F}_{q^m}$ and $\alpha, \beta \in \mathbb{F}_q$.
- $\forall B$ basis of r vectors of $\mathbb{F}_{q^m}$, $\exists! P$ unitary q-polynomial such that $\forall b \in B$, $P(b) = 0$ (Ore '33).

One can then define a subspace of dimension r with a polynomial of q-degree r.

Attack with q-polynomial

Reformulation :

$$c + e = y$$

with c a word of $\mathcal{C}$, e a word of weight r and y known.
There exists a polynomial P of q-degree r such that

$$P(c - y) = 0$$

moreover there exists x such that $c = xG$, which gives :

$$\left(\sum_{i=0}^{r} p_i (xG_1 - y_1)^{q^i}, \ldots, \sum_{i=0}^{r} p_i (xG_n - y_n)^{q^i}\right)$$

with $x \in {\mathbb{F}_{q^m}}^k$, G_j the j-ith column of G and $y \in {\mathbb{F}_{q^m}}^n$ known.

Attack with q-polynomials

Advantages : less unknowns, sparse equations
Disadvantages : higher degree equations $q^r + 1$
Three methods to solve :

- Linearization
- Grobner basis
- Hybrid approach : partial enumeration of unknowns

Conclusion on attacks

- Combinatorial : quadratic in the exponent, usually the best ones but depend on q
- Algebraic : very high when r increases but do not depend on q

$\rightarrow$ best attacks : exponential with quadratic complexity in the exponent. Comparison of this problem with other problems for a 2^n complexity with best known attacks :

problem	size of key	NP-hard problem red.
factorization	$\Omega(n^3)$	no
discrete log (large car.)	$\Omega(n^3)$	no
ECDL	$\Omega(n)$	no
SVP ideal lattices	$\Omega(n)$	no
SD cyclic-codes	$\Omega(n)$	no
SD	$\Omega(n^2)$	yes
SVP	$\Omega(n^2)$	yes
RSD	$\Omega(n^{1.5})$	yes

ENCRYPTION/Key Exchange IN RANK METRIC

- Gabidulin *et al.* '91 : first encryption scheme based on rank metric
- adaptation of McELiece scheme, many variations :

- Parameters
  - B $\{b_1, \cdots, b_m\}$ a basis of $GF(q^m)$ over $GF(q)$
- Private key
  - G generates a Gabidulin code $Gab(m,k)$, $r = \frac{m-k}{2}$
  - S a random $k \times t_1$ matrix in $GF(q^m)$
  - P a random matrix in $GL_{m+t_1}(GF(q))$
  - S a random invertible $k \times k$ matrix in $GF(q^m)$
- Public key

$$G_{pub} = S(G|Z)P$$

- Encryption

$$y = xG_{pub} + e, \quad Rank(e) \leq r$$

- Decryption
  - Compute $yP^{-1} = x(G|Z) + eP^{-1}$
  - Puncture the last t_1 columns and decode

Other variations : G Gabidulin matrix, H : dual matrix

Masking	public matrix	authors
Scrambling matrix	$SG + X$	GPT '91
Right scrambling	$S(G\|Z)P$	Gabi. Ouriv. '01
Subcodes	$\begin{pmatrix} \mathrm{H} \\ \mathrm{A} \end{pmatrix}$	Ber. Loi. '02
Rank Reducible	$\begin{pmatrix} G_1 & 0 \\ \mathrm{A} & G_2 \end{pmatrix}$	[OGHA03],[BL04]
Gabidulin-LRPC	G.H(LRPC)	Loidreau '17

Overbeck's structural attack

Overbeck's attack '06

- general idea : if one consider G=Gab[n,k] and one applies the frobenius : $x \to x^q$ to each coordinate of G then G^q and G have $k-1$ rows in common !
- starting from $G_{pub} = S(G|Z)P$, one can prove there is a rank default in :

$$\begin{pmatrix} G_{pub} \\ \vdots \\ G_{pub}^{q^{n-k-1}} \end{pmatrix}$$

$$thematrixisa$$

k(n-k) $\times(n+t_1)$ matrix, first n columns part : rank $n-1$ and not n !

- Overbeck uses this point to break parameters of all presented GPT-like systems at that time (generalization G.,Otmani,Tale DCC '18)

The NTRU-like family

- NTRU
  - double circulant matrix $(A|B) \to (I|H)$
  - A and B : cyclic with 0 and 1, over Z/qZ (small weight) (q=256), $N \sim 300$
- MDPC
  - double circulant matrix $(A|B) \to (I|H)$
  - A and B : cyclic with 0 and 1, 45 1, (small weight) $N \sim 4500$
- LRPC
  - double circulant matrix $(A|B) \to (I|H)$
  - A and B : cyclic with small weight (small rank)

LRPC codes for cryptography

- We saw that LRPC codes with H [n-k,n] over F of rank d could decode error of rank r with probability $q^{n-k-rd+1}$:

- McEliece setting :
Public key : G LRPC code : $[n, k]$ of weight d which can decode up to errors of weight r
Public key : $G' = MG$
Secret key : M

- **Encryption**
c= mG' +e , e of rank r

- **Decryption**
Decode $H.c^t$ in e, then recover m.

- Smaller size of key : double circulant LRPC codes : H=(I A), A circulant matrix

Application to cryptography

- **Attacks on the system**

- message attack : decode a word of weight r for a $[n, k]$ random code

- structural attack : recover the LRPC structure
$\rightarrow$ a $[n, n-k]$ LRPC matrix of weight d contains a word with $\frac{n}{d}$ first zero positions. Searching for a word of weight d in a $[n - \frac{n}{d}, n - k - \frac{n}{d}]$ code.

- **Attack on the double circulant structure**

Hauteville-Tillich ISIT 2015, same attack than for lattices and codes (Gentry attack), can be avoided by considering an irreducible polynomial for the ideal structure.

Examples of parameters : LAKE

All the times are given in **ms**, performed on an Intel Core i7-4700HQ CPU running at 3.40GHz.

Security	Message/key Size (bits)	KeyGen Time	Encap Time	Decap Time	Probability of failure
128	3,149	0.65	0.13	0.53	$< 2^{-30}$
192	4,717	0.73	0.13	0.88	$< 2^{-32}$
256	6,313	0.77	0.15	1.24	$< 2^{-36}$

Examples of parameters : LOCKER

All the times are given in **ms**, performed on an Intel Core i7-4700HQ CPU running at 3.40GHz.

Security	PK Size (bits)	CT Size (bits)	Encrypt Time	Decrypt Time	Probability of failure
128	5,893	6,405	0.22	1.04	$< 2^{-64}$
192	8,383	8,895	0.23	1.08	$< 2^{-64}$
256	9,523	10,023	0.25	1.58	$< 2^{-64}$
128	12,367	12,879	0.56	1.99	$< 2^{-128}$
192	15,049	15,561	0.56	2.03	$< 2^{-128}$
256	17,113	17,625	0.62	2.76	$< 2^{-128}$

Conclusion for LRPC

- LRPC : new family of rank codes with an efficient probabilistic decoding algorithm
- Application to cryptography in the spirit of NTRU and MDPC (decryption failure, more controlled)
- Very small size of keys, comparable to RSA
- More studies need to be done but very good potentiality
- Security based on recovering small weight random vectors, NOT BASED on decoding random (QC) codes

RQC PKE scheme

RQC scheme Aguilar,Blazy,Deneuville,G., Zemor IEEE IT '18 (first described in 2010) in the spirit of Aleknovich '03

Vectors $\mathbf{x}$ of $\mathbb{F}_{q^m}^n$ seen as elements of $\mathbb{F}_{q^m}[X]/(P)$ for some polynomial P.

$\mathcal{S}_w^n(\mathbb{F}_{q^m}) = \left\{\mathbf{x} \in \mathbb{F}_{q^m}^n \text{ such that } \omega(\mathbf{x}) = w\right\}$

- Public Data : $\mathbf{G}$ is a generator matrix of some public code $\mathcal{C}$
- Secret key $\mathbf{sk} = (\mathbf{x}, \mathbf{y})$, Public key : $\mathbf{pk} = (\mathbf{h}, \mathbf{s} = \mathbf{x} + \mathbf{h}.\mathbf{y})$

Alice		Bob
$\text{seed}_{\mathbf{h}}\{0,1\}^\lambda$, $\mathbf{h} \overset{\text{seed}_{\mathbf{h}}}{\longleftarrow} \mathbb{F}_{q^m}$ $\mathbf{x}, \mathbf{y}\mathcal{S}_\omega^{2n}(\mathbb{F}_{q^m}), \mathbf{s} \leftarrow \mathbf{x} + \mathbf{h}.\mathbf{y}$	$\xrightarrow{\text{seed}_{\mathbf{h}},\mathbf{s}}$	
		$\mathbf{r}_1, \mathbf{r}_2, \mathbf{e} \leftarrow \mathcal{S}_{\omega_r}^{3n}(\mathbb{F}_{q^m})$ $\mathbf{u} \leftarrow \mathbf{r}_1 + \mathbf{h}\mathbf{r}_2$, $\mathbf{v} \leftarrow \mathbf{m}\mathbf{G} + \mathbf{s}\mathbf{r}_2 + \mathbf{e}$
$\mathbf{m} \leftarrow \mathcal{C}.\text{Decode}\,(\mathbf{v} - \mathbf{u}\mathbf{y})$	$\xleftarrow{\mathbf{u},\mathbf{v}}$	

Why does it work ?

$$v - uy = mG + (x + hy)r_2 + e - (r_1 + hr_2)y$$
$$= mG + xr_2 - yr_1 + e.$$

Decrypts whenever the public code $\mathcal{C}$ decodes the small rank weight error $xr_2 - yr_1 + e$ for (x, y) and (r_1, r_2, e) small rank weight vectors.

Choice for $\mathcal{C}$: Gabidulin codes and hence NO decryption failure.

Semantic Security

Theorem

Under the assumption of the hardness of the $[2n, n]$-Decisional-QCRSD and $[3n, n]$-DQCRSD problems, RQC is IND-CPA in the Random Oracle Model.

- • Applying HHK's transform to RQC PKE $\rightarrow$ IND-CCA2 RQC KEM
- • IND-CCA2 RQC KEM $\rightarrow$ IND-CCA2 RQC Hybrid Encryption.

OUROBOROS-R scheme

Deneuville, G., Zemor PQCrypto '17

Vectors x of $\mathbb{F}_{q^m}^n$ seen as elements of $\mathbb{F}_{q^m}[X]/(P)$ for some polynomial P.

Alice		Bob
$\mathsf{seed}_{\mathbf{h}} \leftarrow \{0,1\}^\lambda$, $\mathbf{h} \overset{\mathsf{seed}_{\mathbf{h}}}{\leftarrow} \mathbb{F}_{q^m}^n$		
$(\mathbf{x},\mathbf{y}) \leftarrow \mathcal{S}_{1,w}^{2n}(\mathbb{F}_{q^m})$, $\mathbf{s} \leftarrow \mathbf{x} + \mathbf{h}\mathbf{y}$	$\xrightarrow{\mathbf{h},\mathbf{s}}$	$(\mathbf{r}_1, \mathbf{r}_2, \mathbf{e}_\mathbf{r}) \leftarrow \mathcal{S}_{w_r}^{3n}(\mathbb{F}_{q^m})$
$\mathbf{F} \leftarrow \mathsf{Supp}(\mathbf{x},\mathbf{y})$		$\mathbf{E} \leftarrow \mathsf{Supp}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{e}_\mathbf{r})$
$\mathbf{e}_\mathbf{c} \leftarrow \mathbf{s}_\mathbf{e} - \mathbf{y}\mathbf{s}_\mathbf{r}$	$\xleftarrow{\mathbf{s}_\mathbf{r}, \mathbf{s}_\mathbf{e}}$	$\mathbf{s}_\mathbf{r} \leftarrow \mathbf{r}_1 + \mathbf{h}\mathbf{r}_2$, $\mathbf{s}_\mathbf{e} \leftarrow \mathbf{s}\mathbf{r}_2 + \mathbf{e}_\mathbf{r}$
$\mathbf{E} \leftarrow \mathsf{QCRS\text{-}Recover}(\mathbf{F}, \mathbf{e}_\mathbf{c}, w_r)$		
$\boxed{\mathsf{Hash}(\mathbf{E})}$	Shared Secret	$\boxed{\mathsf{Hash}(\mathbf{E})}$

Figure 1 – Informal description of OUROBOROS-R. **h** and **s** constitute the public key. **h** can be recovered by publishing only the λ bits of the seed (instead of the n coordinates of **h**).

Why does it work ?

$$\mathbf{e_c} = \mathbf{s_e} - \mathbf{y}\mathbf{s_r} = \mathbf{s}\mathbf{r}_2 + \mathbf{e_r} - \mathbf{y}(\mathbf{r}_1 + \mathbf{h}\mathbf{r}_2)$$
$$= (\mathbf{x} + \mathbf{h}\mathbf{y})\mathbf{r}_2 + \mathbf{e_r} - \mathbf{y}(\mathbf{r}_1 + \mathbf{h}\mathbf{r}_2) = \mathbf{x}\mathbf{r}_2 - \mathbf{y}\mathbf{r}_1 + \mathbf{e_r}$$

$1 \in \mathbf{F}$, coordinates of $\mathbf{e_c}$ generate a subspace of $\mathrm{Supp}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{e_r}) \times \mathrm{Supp}(\mathbf{x}, \mathbf{y})$ on which one can apply the QCRS-Recover algorithm to recover E (LRPC decoder).

In other words : e_c seen as syndrome associated to an LRPC code based on the secret key (x, y)

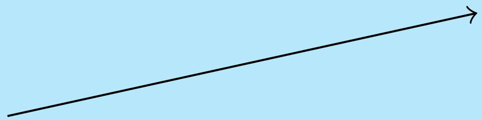
$\rightarrow$ a reasonable decoding algorithm is used to decode a SMALL weight error !

Semantic Security

Theorem

Under the assumption of the hardness of the $[2n, n]$-Decisional-QCRSD and $[3n, n]$-Decisional-QCRSD problems, OUROBOROS-R is IND-CPA in the Random Oracle Model.

	NTRU-like family		Ouroboros family	RLWE-like family
	• McEliece setting / Code generated by small weight vectors • No reconciliation / Polynomial inversion		• Reconciliation • No hidden structure • No polynomial inversion • Small decoded error	• Reconciliation • No hidden structure • No polynomial inversion • Larger decoded error
Security reduction	• Indistinguishability of small weight vectors generated [2n,n] code		• Decisional SD [2n,n] or SD [3n,n] for (ideal/QC) random codes	• Decisional SD [2n,n] or SD [3n,n] for (ideal/QC) random codes
Error form	(e_1, e_2)	(e)	(e_1, e_2, e_3)	(e_1, e_2, e_3)
Decoded word	$x_1\ e_2 + x_2\ e_1$	$x_1 m + pex_2$	$e_3 + x_1\ e_2 + x_2\ e_1$	$mG + e_1 x_2 + e_2 x_1 + e_3$
Decoding algorithm	Bit-flipping like based on (x_1, x_2)	Generic	Noisy bit-flipping like based on (x_1, x_2)	Generic
Euclidean	GuoJohansson '16	NTRU '95 ($N\infty$)	Ouroboros-E '18	RLWE '10 ($N\infty$)
Rank	LRPC '13 (LAKE-LOCKER)		Ouroboros-R '17	RQC '16 (Gabidulin)
Hamming	MDPC '13 (BIKE-2)		Ouroboros '17 (BIKE-3)	HQC '10 - '16 (BCH $\otimes$ repetition code)

Semantic security			Ciphertext size			Keygen computation cost		
NTRUlike	OURlike	RLWElike	NTRUlike	OURlike	RLWElike	NTRUlike	OURlike	RLWElike
			n	$n + recon$	$n + recon$			
								

Authentication

Chen's protocol

In '95 K. Chen proposed a rank metric authentication scheme, in the spirit of the Stern SD protocol for Hamming distance and Shamir's PKP protocol.

Unfortunately the ZK proof is false.... a good toy example to understand some subtilities of rank metric. [G. *et al.* (2011)]

1. [Commitment step] The prover $\mathcal{P}$ chooses $x \in V_n$, $P \in GL_n(\mathrm{GF}(q$
and $Q \in GL_m(q)$. He sends c_1, c_2, c_3 such that :

$$c_1 = hash(Q|P|Hx^t), c_2 = hash(Q * xP), c_3 = hash(Q * (x + s)P$$

2. [Challenge step] The verifier $\mathcal{V}$ sends $b \in \{0, 1, 2\}$ to P.
3. [Answer step] there are three possibilities :
   - if $b = 0$, $\mathcal{P}$ reveals x and $(Q|P)$
   - if $b = 1$, $\mathcal{P}$ reveals $x + s$ and $(Q|P)$
   - if $b = 2$, $\mathcal{P}$ reveals $Q * xP$ and $Q * sP$
4. [Verification step] there are three possibilities :
   - if $b = 0$, $\mathcal{V}$ checks c_1 and c_2.
   - if $b = 1$, $\mathcal{V}$ checks c_1 and c_3.
   - if $b = 2$, $\mathcal{V}$ checks c_2 and c_3 and that $rank(Q * sP) = r$.

Signature with rank metric

Different approaches for signature

- Signatures by inversion
  - unique inversion : RSA,CFS
  - several inversions : NTRUSign, GGH, GPV
- Signature by proof of knowledge
  - by construction : Schnorr, DSA, Lyubashevski (lattices 2012)
  - generic : Fiat-Shamir paradigm
- one-time signatures : KKS '97, Lyubashevski '07

LRPC with erasure

Input $T =< T_1, \cdots, T_t >$, H a matrix of LRPC, a syndrome $s = H.e^t$, with support E and dim(E) $= t + \frac{n-k}{d}$ and $T \subset E$
Result : the error vector e.

1. **Syndrome computations**

   a) Compute $B = \{F_1 T_1, \cdots, F_d T_t\}$ of the product space $< F.T >$.

   b) Compute the subspace $S =< B \cup \{s_1, \cdots, s_{n-k}\} >$.

2. **Recovering the support E of the error**

   Define $S_i = F_i^{-1} S$, compute $E = S_1 \cap S_2 \cap \cdots \cap S_d$, and compute a basis $\{E_1, E_2, \cdots, E_r\}$ of E.

3. **Recovering the error vector e**

   Write $e_i = \sum_{j=1}^{n} e_{ij} E_j$, and solve a linear system

Corollary (Density of decodable syndromes)

The density of unique support decodable syndromes of rank weight $r = t + r'$ for a fixed random partial support T of dimension t is :

$$\frac{\prod_{i=0}^{r'-1}\left(\frac{q^{m-t-i}-1}{q^{i+1}-1}\right).min(q^{nr}, q^{rd(n-k)})}{q^{(n-k)m}}.$$

$\rightarrow$ very strong constraints on LRPC parameters to obtain a density close to 1.

RankSign$^+$ signature algorithm

1. **Secret key :** H :LRPC, $r' = t + \frac{n-k}{2}$ errors, R random in $GF(q^m)$ invertible in $GF(q^m)$, P invertible in $GF(q)$.
2. **Public key :** the matrix $H' = A(R|H)P$, a small integer value l.
3. **Signature of a message** M **:**
   a) *initialization* : seed $\leftarrow \{0,1\}^l$, pick $(e_1, \cdots, e_t) \in GF(q^m)^t$
   b) *syndrome* : $s \leftarrow hash(M||seed) \in GF(q^m)^{n-k}$
   c) decode by H, $s' = A^{-1}.s^T - R.(e_1, \cdots, e_t)^T$ with $T =< e_1, \cdots, e_t >$ and r' errors
   d) if the decoding works $\rightarrow (e_{t+1}, \cdots, e_{n+t})$ of weight $r = t + r'$, signature=$((e_1, \cdots, e_{n+t}).(P^T)^{-1}, seed)$, else return to a).
4. **Verification** : Verify that $Rank(e) = r = t + r'$ and $H'.e^T = s = hash(M||seed)$.

Structural attacks

- Overbeck attack : irrelevant
- Attack on the dual matrix : $r = t + d$
- Attack on isometry matrix P : recover some positions of P
- Recent attack by Debris-Tillich '18, based on the necessary conditions for inversion $\rightarrow$ breaks the masking $(R \| LRPC)$, possibility to repair

IBE with rank metric

Description of the cryptosystem

G.,Hauteville, Phan, Tillich CRYPTO '17

A PKE consists in three algorithms :

RankPKE.KeyGen(1^λ) :

$\begin{pmatrix} \boldsymbol{s} \end{pmatrix} \xleftarrow{\$} \mathbb{F}_{q^m}^{n-k}$.

$\begin{pmatrix} \\ \boldsymbol{A} \\ \\ \end{pmatrix} \xleftarrow{\$} \mathbb{F}_{q^m}^{(n-k)\times n}$.

$\begin{pmatrix} \boldsymbol{e} \end{pmatrix} \xleftarrow{\$} \mathbb{F}_{q^m}^{n}$ of weight r.

$\boldsymbol{p} = \boldsymbol{s}\boldsymbol{A} + \boldsymbol{e}$.

$\mathcal{C}_{pub}$ a public code of generator matrix $\boldsymbol{G}$ which can decode up to wr errors.

public key $= (\boldsymbol{A}, \boldsymbol{G}, \boldsymbol{p})$. secret key $= \boldsymbol{s}$.

RankPKE.Enc($\boldsymbol{m}, \boldsymbol{A}, \boldsymbol{G}, \boldsymbol{p}$) :

$$\left(\begin{array}{c} \boldsymbol{A} \\ \\ \hline \boldsymbol{p} \end{array}\right) \boldsymbol{U} + \left(\begin{array}{c} 0 \\ \\ \hline \boldsymbol{mG} \end{array}\right) = \left(\begin{array}{c} \boldsymbol{C} \\ \\ \hline \boldsymbol{x} \end{array}\right)$$

where $\boldsymbol{U}$ is an $(n-k+1) \times n'$ homogeneous matrix of weight w.

Description of the cryptosystem

RankPKE.Dec($\boldsymbol{C}, \boldsymbol{x}, \boldsymbol{s}$) :

compute $(\ \boldsymbol{s}\ |\text{-}1)\left(\begin{array}{c}\boldsymbol{C}\\ \hline \boldsymbol{x}\end{array}\right) = \boldsymbol{sC}\text{-}\ \boldsymbol{x} = \boldsymbol{sAU}\text{-}\ \boldsymbol{pU}\text{-}\ \boldsymbol{mG}$

$= \boldsymbol{sAU}\text{-}\ (\boldsymbol{sA}+\boldsymbol{e})\boldsymbol{U}\text{-}\ \boldsymbol{mG}$

$=$ -$\boldsymbol{eU}$- $\boldsymbol{mG}$ $\boldsymbol{U}$ is homogeneous of weight w and

$|\boldsymbol{e}|_r = r \Rightarrow |\boldsymbol{eU}|_r \leqslant wr$.

$\rightarrow$ compute $\boldsymbol{m}$ with the decoder of $\mathcal{C}_{pub}$.

Security of RankPKE

A new problem, Rank Support Learning :

Let $\boldsymbol{A}$ be a random full-rank matrix of $\mathbb{F}_{q^m}^{(n-k)\times n}$ and V a subspace of $\mathbb{F}_{q^m}$ of dimension w.

$$\begin{pmatrix} \boldsymbol{A} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \xleftarrow{\$} V^{n\times n'} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_{n'} \end{pmatrix}$$

The problem is to recover V given only access to $(\boldsymbol{A}, \boldsymbol{AU})$.

Security of RankPKE

A new problem, Rank Support Learning :
Let $\boldsymbol{A}$ be a random full-rank matrix of $\mathbb{F}_{q^m}^{(n-k)\times n}$ and V a subspace of $\mathbb{F}_{q^m}$ of dimension w.

$$\begin{pmatrix} & \boldsymbol{A} & \end{pmatrix} \begin{pmatrix} & \boldsymbol{U} \xleftarrow{\$} V^{n\times n'} & \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_{n'} \end{pmatrix}$$

The problem is to recover V given only access to $(\boldsymbol{A}, \boldsymbol{AU})$.
The corresponding decisional problem, namely DRSL, is to distinguish $(\boldsymbol{A}, \boldsymbol{AU})$ from $(\boldsymbol{A}, \boldsymbol{Y})$, $\boldsymbol{Y} \xleftarrow{\$} \mathbb{F}_{q^m}^{n\times n'}$.

Security of RankPKE

A new problem, Rank Support Learning :
Let $\boldsymbol{A}$ be a random full-rank matrix of $\mathbb{F}_{q^m}^{(n-k)\times n}$ and V a subspace of $\mathbb{F}_{q^m}$ of dimension w.

$$\begin{pmatrix} \boldsymbol{A} \end{pmatrix} \begin{pmatrix} \boldsymbol{U} \stackrel{\$}{\leftarrow} V^{n\times n'} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_1 & \dots & \boldsymbol{c}_{n'} \end{pmatrix}$$

The problem is to recover V given only access to $(\boldsymbol{A}, \boldsymbol{AU})$.
The corresponding decisional problem, namely DRSL, is to distinguish $(\boldsymbol{A}, \boldsymbol{AU})$ from $(\boldsymbol{A}, \boldsymbol{Y})$, $\boldsymbol{Y} \stackrel{\$}{\leftarrow} \mathbb{F}_{q^m}^{n\times n'}$.

Theorem

Under the assumption that DRSL is hard, the scheme RankPKE is IND-CPA idea of the IBE :

• use the RankSign algorithm to decode a random vector p in a relatively small weight vector $p = sA + e$,

• with RankPKE possibility to decrypt from the knowledge of a small preimage of random vector, the couple (s,e) is used as decryption key from a public key p (random).

• similar to the GPV approach, based on the difficulty of the RSL problem

• recent attack from Debris-Tillich '18 restrains the possible parameters.

Limitations of rank metric

There are two main limitations for rank metric :

• Ceiling limitation : the ratio Singleton/GV is always less than 2 !

$\rightarrow$ limits the possibility to find a collision resistant hash function

• if $x_1, \ldots, x_t$ are small weight vectors then $\sum a_i x_i$ for $a_i \in GF(q)$ does not hide the x_i if their associated syndrome is known

$\rightarrow$ makes Lyubashevski-like signature difficult to obtain at first sight.

GENERAL CONCLUSION

- Rank distance is interesting since small parameters $\rightarrow$ strong resistance
- until recently only one family of decodable codes
- LRPC codes -weak structure-, similar to NTRU or MDPC offer many advantages
- Very efficient solutions for encryption very competitive with tight reduction to decoding RANDOM (QC-ideal) codes.

Open problems

- deterministic reduction to SD rather than only probabilistic ?
- Is it possible to have worst case - average case reduction ?
- Attacks improvements on rank ISD ?
- Better algebraic settings ?
- Optimized implementations ?
- Efficient signature ?
- Security for the ideal case ?
- search to decision reduction (Goldreich-Levin for large field, also for the LRE problem) ?
- advanced encryption (functional encryption, witness encryption, FHE etc...)

THANK YOU