Rank based cryptography : a credible post-quantum alternative to classical cryptography

P.Gaborit[1], *O.Ruatta*[1], *J.Schrek*[2], *J.P.Tillich*[3] *and* G. Zémor[4]
1: University of Limoges, France
2: Telecom Bretagne, France
3: Inria, France
4: University of Bordeaux, France

NIST Workshop on Cybersecurity in a Post-Quantum World
2015

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 $R(GF(q), Z/qZ, GF(q^m))$
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, $x = (0 \dots 0, A^{-1}s, 0 \dots 0)$.

How to make this problem difficult ?

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

- metric = Hamming distance ⇒ **code-based cryptography**
- metric = Euclidean distance ⇒ **lattice-based cryptography**
- metric = Rank distance ⇒ **rank-based cryptography**

⇒ only difference : the metric considered, and its associated properties ! !

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

⇒ 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

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 $GF(q^m)$.

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

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

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

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 $d_R(x, y) = Rank(x - y)$.

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.

Analogy : counting subspaces

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 $\begin{bmatrix} n \\ t \end{bmatrix}_q (\sim q^{tn})$

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 matrix construction (Silva et al. 2008)
- LRPC codes (Gaborit et al. 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 $GF(q^m)$ 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 $GF(q^m)$ 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' dimensional vector space $F < F_1, F_2, \cdots, F_d >$ of $GF(q^m)$ 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 $H.e^t = s = (s_1, ..., s_{n-k})^t$.

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

$$\Rightarrow s_k \in < E_1 F_1, .., E_r F_d >$$

$\Rightarrow$ if $n - k$ is large enough, it is possible to recover the product space $< E_1 F_1, .., E_r F_d >$

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 $dim(S) = n - k \Rightarrow$ 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 $h_{ij} = \quad h_{ijk}F_k$, $\rightarrow nr \times m(n-k)$ 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.

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 $H^t x = \sum_{i=1}^{n} x_i h_i = s$?

Computational complexity of the RSD problem

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

Definition (embedding strategy)

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

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

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 :

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

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 (Randomized reduction)

If there exists a polynomial time algorithms which solves RSD with a strong probability ($RSD \in RP$) then NP=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 : $\leq (k+r)^3 r^3 q^{(r-1)(k+1)}$
- last improvement : Gaborit et al. '12 : adaptation of the ISD algorithm in the rank metric
  - Support attack : $\mathcal{O}(q^{(r-1)\frac{\lfloor (k+1)m \rfloor}{n}})$

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 $r+1$
- Gaborit et *al.* '12 : annulator polynomial $\rightarrow$ degree q^r

$\rightarrow$ 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 :

general problem	size of key	proof of NP-hardness
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 IN RANK METRIC

- Gabidulin *et al.* '91 : first encryption scheme based on rank metric - adaptation of McELiece scheme, with Gabidulin codes and rank metric
- small size of keys ($\sim$ 5000b)
- inherent structural weakness from Gabidulin codes
- $\rightarrow$ many attacks (Overbeck '05) , many reparations
- last reparations : Loidreau PQC '10, Gabidulin et al'09.
  $\rightarrow$ all parameters broken in 2012 by Gaborit et al.
- $\rightarrow$ similar situation to RS codes in Hamming metric : seems hard to hide a very structured family of codes (Gabidulin codes) - new systems proposed ?

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)

$\rightarrow$ weak structure, more difficult to attack (some specific structural attacks exist but are easy to counter Gentry '02, Hauteville-Tillich 2015)

Parameters

LRPC codes for cryptography (Gaborit et al. 2013)

n	k	m	q	d	r	failure	public key	security
82	41	41	2	5	4	-22	1681	80
106	53	53	2	6	5	-24	2809	128
74	37	23	2^4	4	4	-88	3404	110

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 $rank(Q * sP) = r$.

FIGURE: Rank-SD protocol

- **Public matrix H :** $(n-k) \times k \times m = 2691$ bits
- **Public key i :** $(n-k)m = 299$ bits
- **Secret key s :** $r(m+n) = 360$ bits
- **Average number of bits exchanged in one round :** 2 hash + one word of $\mathrm{GF}(q^m) \sim 820$ bits.

$\rightarrow$ security based on a general instance of the RSD problem

Signature with rank metric

RankSign : general idea

General idea : Inverting a random syndrome with mixed errors/erasure decoding

- Possible to adapt the LRPC decoding algo, with a few constraints
- Possible to find parameters for which unique decoding for erasure is obtained beyond RGV with proba ~ 1
- Matrices cannot be used directly for crypto and need a masking.
- best results : $d = 2$ anyway
- security proof for leaking information

Parameters

- examples of parameters

n	n-k	m	q	d	t	r'	r	GV	Sg	pk	sign	LP	Dual	DS	DA
16	8	18	2^{40}	2	2	4	6	5	8	57600	8640	**130**	1096	400	776
16	8	18	2^{8}	2	2	4	6	5	8	11520	1728	110	233	**80**	168
16	8	18	2^{16}	2	2	4	6	5	8	23040	3456	**120**	448	160	320
20	10	24	2^{8}	2	3	5	8	6	10	24960	3008	190	370	**104**	226
27	9	20	2^{6}	3	2	3	5	4	7	23328	1470	170	187	**120**	129
48	12	40	2^{4}	4	5	3	8	6	10	78720	2976	> 600	340	164	**114**
50	10	42	2^{4}	5	2	2	7	5	9	70560	2800	> 600	240	180	**104**

- implementation results

n	n-k	m	q	d	signature time (ms)	verification time (ms)	security (bits)
16	8	18	2^{8}	2	2.75	4.4	80
20	10	24	2^{8}	2	6.13	12	104

TABLE: Non optimized implementation time on a Intel Core i5-4200U CPU 1.60GHz processor with MPFQ library

GENERAL CONCLUSION

- rank metric is fun with a rich algebraic structure and many fascinating objects like q-polynomials (polynomials/matrices)
- cryptosystems with small parameters (encryption / signature / authentication) exist
- Rank metric has a very strong potential for PQ crypto since small parameters $\rightarrow$ strong resistance to best known attacks (analogy DL/ECDL with Hamming/rank).
- LRPC codes -weak structure-, similar to NTRU or MDPC offer many advantages
- needs more scrutiny from the communauty

Open problems

- Deterministic reduction to SD rather than only probabilistic ?
- Is it possible to have worst case - average case reduction ?
- Finding new primitives, in the standard model ?
- Better security reduction (although cryptosystems exist directly based on RSD) ?
- Attacks improvements : on rank ISD / algebraic settings ?
- Implementations ?
- homomorphic - FHE (be crazy !)

THANK YOU