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

1. Post-Quantum Cryptography
2. Decoding in rank metric
3. Complexity issues: decoding random rank codes
4. Encryption/Authentication in rank metric
5. Signature in rank metric

P. Gaborit\textsuperscript{1}, O. Ruatta\textsuperscript{1}, J. Schrek\textsuperscript{2}, J. P. Tillich\textsuperscript{3} and G. Zemor\textsuperscript{4}: Rank based cryptography: a credible post-quantum alternative to classical cryptography.
Post-quantum cryptography

Motivations
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.
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, \(x = (0 \ldots 0, A^{-1}s, 0 \ldots 0)\).
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...

⇒ Multivariatate cryptography
General interest of post-quantum cryptography

- 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 $GF(q)$.

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

$$v(v_1, \ldots, v_n) \rightarrow V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}$$
Definition (Rank weight of word)

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

The determinant of $V$ does not depend on the basis.

Definition (Rank distance)

Let $x, y \in \mathbb{GF}(q^m)^n$, the rank distance between $x$ and $y$ is defined by $d_R(x, y) = \text{Rank}(x - y)$. 
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 $x \in GF(q^m)^n$: $\text{Rank}(x) \leq w_H(x)$: 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
  $x(x_1, x_2, \cdots, x_n)$: set of positions $x_i \neq 0$

- support of a word of $GF(q)^n$ in rank metric
  $x(x_1, x_2, \cdots, x_n)$: 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 \( \mathbb{GF}(q) \) in the space of dimension \( n \) \( \mathbb{GF}(q^m) \): Gaussian binomial \( \left[ \begin{array}{c} n \\ t \end{array} \right]_q \sim q^{tn} \)
Decoding 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)
→ 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 $e(e_1, \ldots, e_n)$ be an error vector of weight $r$, i.e.: $\forall e_i : e_i \in E$, and $\dim(E) = r$. Suppose $H.e^t = s = (s_1, \ldots, s_{n-k})^t$.

$$e_i \in E \langle E_1, \ldots, E_r \rangle, \quad h_{ij} \in F \langle F_1, F_2, \ldots, F_d \rangle$$

$$\Rightarrow s_k \in \langle E_1 F_1, \ldots, E_r F_d \rangle$$

$$\Rightarrow \text{if } n - k \text{ is large enough, it is possible to recover the product space } \langle E_1 F_1, \ldots, E_r F_d \rangle$$
Decoding LRPC codes

**Syndrome** $s(s_1, \ldots, s_{n-k}) : S = \langle s_1, \ldots, s_{n-k} \rangle \subset \langle E_1 F_1, \ldots, E_r F_d \rangle$

Suppose $S = \langle E.F \rangle \Rightarrow$ possible to recover $E$.

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

$$S = \langle E.F \rangle = \langle F_i E_1, F_i E_2, \ldots, F_i E_r, \ldots \rangle \Rightarrow E \subset S_i$$

$$E = S_1 \cap S_2 \cap \cdots \cap S_d$$
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 \( S = \langle s_1, \cdots, s_{n-k} \rangle \).

2. **Recovering the support \( E \) of the error**
   \( S_i = F_i^{-1}S, \quad E = S_1 \cap S_2 \cap \cdots \cap S_d \),

3. **Recovering the error vector \( e \)**
   Write \( e_i(1 \leq i \leq n) \) in the error support as
   \[ e_i = \sum_{j=1}^{n} e_{ij}E_j, \]
   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**
  - $S = F.E \implies r \leq n-k$.
  - Possibility that $\dim(S) = n - k \implies$ probabilistic decoding with error failure in $q^{-(n-k - r)}$.
  - 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:
  $e_E = (e_{11}, \ldots, e_{nr})$: $rn$ unknowns in $GF(q)$, the syndrome $s$ is written in the symbolic basis $\{E_1F_1, \ldots, E_rF_d\}$, $H$ is written in $h_{ij} = h_{ijk}F_k$, $nr \times m(n - k)$ 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 rank codes
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 \( GF(Q)^n \)

\[
\psi_\alpha : \quad GF(q)^n \rightarrow GF(Q)^n \quad x = (x_1, \ldots, x_n) \mapsto x = (x_1 \alpha_1, \ldots, x_n \alpha_n)
\]

and for any \( GF(q) \)-linear code \( C \) in \( GF(q)^n \), define \( C = C(C, \alpha) \) 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 \(\alpha\) random, then for convenient \(m\), with a very strong probability:

\[d_H(C) = d_R(C).\]

**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: \( O(q^{(r-1)\frac{(k+1)m}{n}}) \)
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 → 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 $q^r$
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:

<table>
<thead>
<tr>
<th>general problem</th>
<th>size of key</th>
<th>proof of NP-hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>factorization</td>
<td>$\Omega(n^3)$</td>
<td>no</td>
</tr>
<tr>
<td>discrete log (large car.)</td>
<td>$\Omega(n^3)$</td>
<td>no</td>
</tr>
<tr>
<td>ECDL</td>
<td>$\Omega(n)$</td>
<td>no</td>
</tr>
<tr>
<td>SVP ideal lattices</td>
<td>$\Omega(n)$</td>
<td>no</td>
</tr>
<tr>
<td>SD cyclic-codes</td>
<td>$\Omega(n)$</td>
<td>no</td>
</tr>
<tr>
<td>SD</td>
<td>$\Omega(n^2)$</td>
<td>yes</td>
</tr>
<tr>
<td>SVP</td>
<td>$\Omega(n^2)$</td>
<td>yes</td>
</tr>
<tr>
<td>RSD</td>
<td>$\Omega(n^{1.5})$</td>
<td>yes</td>
</tr>
</tbody>
</table>
ENCRIPTION 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 (∼ 5000b)
- inherent structural weakness from Gabidulin codes
- → many attacks (Overbeck ’05), many reparations
- last reparations: Loidreau PQC ’10, Gabidulin et al’09.
  → all parameters broken in 2012 by Gaborit et al.

→ 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 \((A|B) \rightarrow (I|H)\)
  - \(A\) and \(B\) : cyclic with 0 and 1, over \(\mathbb{Z}/q\mathbb{Z}\) (small weight) \((q=256), N \sim 300\)

- **MDPC**
  - double circulant matrix \((A|B) \rightarrow (I|H)\)
  - \(A\) and \(B\) : cyclic with 0 and 1, 45 1, (small weight) \(N \sim 4500\)

- **LRPC**
  - double circulant matrix \((A|B) \rightarrow (I|H)\)
  - \(A\) and \(B\) : cyclic with small weight (small rank)

→ weak structure, more difficult to attack (some specific structural attacks exist but are easy to counter Gentry ’02, Hauteville-Tillich 2015)
LRPC codes for cryptography (Gaborit et al. 2013)

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>m</th>
<th>q</th>
<th>d</th>
<th>r</th>
<th>failure</th>
<th>public key</th>
<th>security</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>41</td>
<td>41</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>-22</td>
<td>1681</td>
<td>80</td>
</tr>
<tr>
<td>106</td>
<td>53</td>
<td>53</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>-24</td>
<td>2809</td>
<td>128</td>
</tr>
<tr>
<td>74</td>
<td>37</td>
<td>23</td>
<td>$2^4$</td>
<td>4</td>
<td>4</td>
<td>-88</td>
<td>3404</td>
<td>110</td>
</tr>
</tbody>
</table>
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 $P$ chooses $x \in V_n$, $P \in GL_n(GF(q))$, and $Q \in GL_m(q)$. He sends $c_1, c_2, c_3$ such that:

$$c_1 = \text{hash}(Q|P|Hx^t), c_2 = \text{hash}(Q \times xP), c_3 = \text{hash}(Q \times (x + s)P)$$

2 [Challenge step] The verifier $V$ sends $b \in \{0, 1, 2\}$ to $P$.

3 [Answer step] there are three possibilities:
- if $b = 0$, $P$ reveals $x$ and $(Q|P)$
- if $b = 1$, $P$ reveals $x + s$ and $(Q|P)$
- if $b = 2$, $P$ reveals $Q \times xP$ and $Q \times sP$

4 [Verification step] there are three possibilities:
- if $b = 0$, $V$ checks $c_1$ and $c_2$.
- if $b = 1$, $V$ checks $c_1$ and $c_3$.
- if $b = 2$, $V$ checks $c_2$ and $c_3$ and that $\text{rank}(Q \times sP) = r$.  

**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 $GF(q^m) \sim 820$ bits.

→ 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 $\sim 1$
- 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

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

- implementation results

<table>
<thead>
<tr>
<th>n</th>
<th>n-k</th>
<th>m</th>
<th>q</th>
<th>d</th>
<th>signature time (ms)</th>
<th>verification time (ms)</th>
<th>security (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>8</td>
<td>18</td>
<td>$2^{8}$</td>
<td>2</td>
<td>2.75</td>
<td>4.4</td>
<td>80</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>24</td>
<td>$2^{8}$</td>
<td>2</td>
<td>6.13</td>
<td>12</td>
<td>104</td>
</tr>
</tbody>
</table>

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

P. Gaborit\(^1\), O. Ruatta\(^1\), J. Schrek\(^2\), J. P. Tillich\(^3\) and G. Zemor\(^4\): Rank based cryptography: a credible post-quantum alternative to classical cryptography and Post-Quantum Cryptography.
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 community
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