Two attacks on rank metric code-based schemes: RankSign and an IBE scheme

Thomas Debris-Alazard and Jean-Pierre Tillich

Inria Saclay,
EPI GRACE

Results

Results of the paper:

- Attack on a code-based "hash-and-sign" scheme RankSign [GRSZ14] submitted to the NIST PQC Standardization;

  $\longrightarrow$ Can not be thwarted by changing the parameters.

- Attack on the first code-based Identity-Based-Encryption (IBE) [GHPT17] in rank-metric;

  $\longrightarrow$ Parameters can be chosen to avoid it.

- IBE: moving Rank $\rightarrow$ Hamming metric no go.

Rank vs Hamming in Cryptography

- Advantages:
  - In rank metric: alphabet size q^m has an impact on the metric
    $\rightarrow$Useful for security reductions
  - Smaller key sizes than Hamming.
- Disadvantage:
  - Rank metric: security less understood (algebraic attacks)

Code-Based Cryptography

$\mathbb{F}$ finite field.

Syndrome Decoding Problem.

- Given: a matrix $\mathrm{H} \in \mathbb{F}^{r \times n}$ with $r \leq n$, a vector $\mathrm{s} \in \mathbb{F}^r$, an integer w;
- Goal: find $\mathrm{e} \in \mathbb{F}^n$, $\begin{cases} \mathrm{H}\mathrm{e}^\intercal = \mathrm{s}^\intercal \\ \text{weight}(\mathrm{e}) = w \end{cases}$

Hamming: $\text{weight}(\cdot) = \#$ non-zero components and usually $\mathbb{F} = \mathbb{F}_2$

Rank: $\text{weight}(\cdot) =$ Rank metric and $\mathbb{F} = \mathbb{F}_{q^m}$

$\longrightarrow$ Probabilistic polynomial reduction (Gaborit & Zémor) to the decoding problem in Hamming metric

Rank Metric over $\mathbb{F}_{q^m}$

- $\mathbb{F}_{q^m}$ is a $\mathbb{F}_q$-space of dimension m
- $\mathrm{x} = (x_1, \cdots, x_n) \in \mathbb{F}_{q^m}^n$, its rank is defined as:

$$\text{Support of x :} \quad \langle x_1, \cdots, x_n \rangle_{\mathbb{F}_q} \triangleq \left\{ \sum_i \lambda_i x_i : \lambda_i \in \mathbb{F}_q \right\} \subseteq \mathbb{F}_{q^m}$$

$$\mathrm{rank}(\mathrm{x}) = \dim_{\mathbb{F}_q} \left(\langle x_1, \cdots, x_n \rangle_{\mathbb{F}_q} \right)$$

Some History...

- Gabidulin codes: first rank-codes with a polynomial decoder

  $\rightarrow$ Strong algebraic structure... and a zillion attacks (Overbeck'05...)

- LRPC-codes: decoder introduced in [GMRZ13]

  $\rightarrow$ Finding the underlying structure is close to solving the syndrome decoding problem.

LRPC-codes [GMRZ13]

- Random Code: Given some random matrix $\mathsf{H}_{\mathrm{Rand}} \in \mathbb{F}_{q^m}^{(n-k)\times n}$

$$\{\mathsf{c} : \mathsf{H}_{\mathrm{Rand}}\mathsf{c}^{\intercal} = 0\}$$

- LRPC Code: Given $\mathsf{H}_{\mathrm{LRPC}} = (h_{i,j}) \in \mathbb{F}_{q^m}^{(n-k)\times n}$ s.t

$$\dim\left(\langle h_{i,j} : i,j\rangle_{\mathbb{F}_q}\right) = \text{ small}$$

then,

$$\{\mathsf{c}_{\mathrm{LRPC}} : \mathsf{H}_{\mathrm{LRPC}}\mathsf{c}_{\mathrm{LRPC}}{}^{\intercal} = 0\}$$

When $\mathsf{H}_{\mathrm{Rand}} = (h_{i,j}) \in \mathbb{F}_{q^m}^{(n-k)\times n}$ is random, typically when $m < n(n-k)$:

$$\langle h_{i,j} : i,j\rangle_{\mathbb{F}_q} = \mathbb{F}_{q^m}.$$

LRPC-codes in RankSign[GRSZ14]

LRPC-codes come in RankSign with a decoder [GRSZ14]:

$$\forall \mathrm{s}, \text{ it computes polynomially e s.t } \begin{cases} \mathrm{H}_{\mathrm{LRPC}}\mathrm{e}^{\intercal} = \mathrm{s}^{\intercal} \\ \mathrm{rank}(\mathrm{e}) = w \end{cases}$$

- Constraint RankSign: $\mathrm{H}_{\mathrm{LRPC}} = (h_{i,j}) \in \mathbb{F}_{q^m}^{(n-k)\times n}$ s.t

$$(n-k)\dim\left(\langle h_{i,j} : i,j\rangle_{\mathbb{F}_q}\right) = n$$

Problem: Rows of $\mathrm{H}_{\mathrm{LRPC}}$ gives words of low weight...

$\rightarrow$ A masking is needed!

Masking LRPC-codes in RankSign

In RankSign [GRSZ14]:

- Increase the weight of rows: $[\mathrm{H}_{\mathrm{LRPC}}|\mathrm{R}]$ for R random;
- Change the code: $[\mathrm{H}_{\mathrm{LRPC}}|\mathrm{R}]\mathrm{P}$ for P invertible in $\mathbb{F}_q$.
- Change the basis: $\mathrm{Q}[\mathrm{H}_{\mathrm{LRPC}}|\mathrm{R}]\mathrm{P}$ for Q invertible;

$$\mathrm{H}_{\mathrm{pub}} \stackrel{\triangle}{=} \mathrm{Q}[\mathrm{H}_{\mathrm{LRPC}}|\mathrm{R}]\mathrm{P} : \text{ public key}$$

Idea of the Attack

To look for low weight codewords... where?

- Suspect: $\mathscr{C}_{\mathsf{pub}}^{\perp} \stackrel{\triangle}{=} \{\mathsf{m}\mathsf{H}_{\mathsf{pub}} : \mathsf{m} \in \mathbb{F}_{q^m}\}$;
- Real Problem: $\mathscr{C}_{\mathsf{pub}} \stackrel{\triangle}{=} \{\mathsf{c} : \mathsf{H}_{\mathsf{pub}}\mathsf{c}^{\intercal} = 0\}$.

Low Rank Codewords in an LRPC?

$$\mathbf{H}_{\mathrm{LRPC}} = (h_{i,j}) \in \mathbb{F}_{q^m}^{(n-k)\times n} \quad \text{with} \quad \langle h_{i,j} : i,j \rangle_{\mathbb{F}_q} = F$$

$$\mathbf{c} = (c_j) \in \mathbb{F}_{q^m}^{n}$$

$$\mathbf{H}_{\mathrm{LRPC}}\mathbf{c}^{\mathsf{T}} = 0 \iff \forall i \in [\![1, n-k]\!], \quad \sum_{j=1}^{n} h_{i,j} c_j = 0$$

Low Rank Codewords in an LRPC?

$$\mathrm{H}_{\mathrm{LRPC}} = (h_{i,j}) \in \mathbb{F}_{q^m}^{(n-k)\times n} \quad \text{with} \quad \langle h_{i,j} : i,j \rangle_{\mathbb{F}_q} = F$$
$$\mathrm{c} = (c_j) \in \mathbb{F}_{q^m}^{n}$$

$$\mathrm{H}_{\mathrm{LRPC}}\mathrm{c}^{\mathsf{T}} = 0 \iff \forall i \in [\![1, n-k]\!], \quad \sum_{j=1}^{n} h_{i,j}c_j = 0$$

$$\text{Suppose that } \langle c_1, \cdots, c_n \rangle_{\mathbb{F}_q} = F'$$

$$\forall i \in [\![1, n-k]\!], \quad \sum_{j=1}^{n} h_{i,j}c_j \in F' \cdot F \stackrel{\triangle}{=} \langle f'f : f' \in F', f \in F \rangle_{\mathbb{F}_q}$$

This gives a linear system in $\mathbb{F}_q$ with

- $(n-k)\dim_{\mathbb{F}_q}(F \cdot F')$ equations;
- $n\dim_{\mathbb{F}_q}(F')$ unknowns.

$\rightarrow$ We would like $\#\mathit{Unknowns} > \#\mathit{Equations}$ to ensure the existence of solutions

... But How to Choose F'?

What we want:

$$n \dim_{\mathbb{F}_q}(F') > (n-k) \dim_{\mathbb{F}_q}(F \cdot F')$$

What we typically have:

$$n \dim_{\mathbb{F}_q}(F') = (n-k) \dim_{\mathbb{F}_q}(F \cdot F')$$

Because,

$$\begin{cases} \dim_{\mathbb{F}_q}(F \cdot F') = \dim_{\mathbb{F}_q}(F) \dim_{\mathbb{F}_q}(F') \text{ (typically)} \\ (n-k) \dim(F) = n \text{ (RankSign)}. \end{cases}$$

The Subspace $F \cdot F'$

$$F \stackrel{\triangle}{=} \langle x_1, \cdots, x_d \rangle_{\mathbb{F}_q} \quad (F = \langle h_{i,j} : i, j \rangle_{\mathbb{F}_q})$$

$$\text{Let } F' \stackrel{\triangle}{=} \langle x_1, x_2 \rangle_{\mathbb{F}_q} \subseteq F.$$

$$F \cdot F' = \langle x_1^2, x_1 x_2, \cdots, x_1 x_d, x_2 x_1, x_2^2, \cdots, x_2 x_d \rangle_{\mathbb{F}_q}.$$

$$\Rightarrow \dim(F \cdot F') \leq 2d - 1$$

Therefore,

$$\begin{aligned} \#Unknowns - \#Equations &= n \dim_{\mathbb{F}_q}(F') - (n-k) \dim_{\mathbb{F}_q}(F \cdot F') \\ &= 2n - (n-k)(2d-1) \end{aligned}$$

The Subspace $F \cdot F'$

$$F \stackrel{\triangle}{=} \langle x_1, \cdots, x_d \rangle_{\mathbb{F}_q} \quad (F = \langle h_{i,j} : i,j \rangle_{\mathbb{F}_q})$$

$$\text{Let } F' \stackrel{\triangle}{=} \langle x_1, x_2 \rangle_{\mathbb{F}_q} \subseteq F.$$

$$F \cdot F' = \langle x_1^2, x_1 x_2, \cdots, x_1 x_d, x_2 x_1, x_2^2, \cdots, x_2 x_d \rangle_{\mathbb{F}_q}.$$

$$\Rightarrow \dim(F \cdot F') \leq 2d - 1$$

Therefore,

$$\begin{aligned} \#Unknowns - \#Equations &= n \dim_{\mathbb{F}_q}(F') - (n-k)\dim_{\mathbb{F}_q}(F \cdot F') \\ &= 2n - (n-k)(2d-1) \end{aligned}$$

Constraint in RankSign:

$$n = (n-k)d$$

which gives:

$$\begin{aligned} \#Unknowns - \#Equations &= 2(n-k)d - (n-k)(2d-1) \\ &= n - k > 0 \end{aligned}$$

Low Rank Codewords in RankSign

- Fact: $\mathrm{rank}(\mathsf{c}_{\mathsf{LRPC}}) = 2$ such that $\mathsf{H}_{\mathsf{LRPC}}\mathsf{c}_{\mathsf{LRPC}}{}^{\intercal} = 0$

$$\Rightarrow \begin{cases} \text{(i)} & \mathsf{H}_{\mathsf{pub}}\left((\mathsf{c}_{\mathsf{LRPC}}, 0)\mathsf{P}^{\intercal\,-1}\right)^{\intercal} = 0 \\ \text{(ii)} & \mathrm{rank}(\mathsf{c}_{\mathsf{LRPC}}, 0)\mathsf{P}^{\intercal\,-1} = 2. \end{cases}$$

Indeed, P invertible in $\mathbb{F}_q$ and:

$$\mathsf{H}_{\mathsf{pub}} = \mathsf{Q} \begin{array}{|c|c|} \hline \mathsf{H}_{\mathsf{LRPC}} & \mathsf{R} \\ \hline \end{array} \mathsf{P}$$

$$\begin{array}{|c|c|} \hline \mathsf{c}_{\mathsf{LRPC}} & 0 \\ \hline \end{array} \mathsf{P}^{-1\,\intercal}$$

Summary

We proved, whatever is the choice of parameters, there are codewords of rank 2 in the public key.

How to Effectively Find Them?

Low-rank codewords in public keys of RankSign. How to find them?

$\longrightarrow$ Gröbner basis techniques with a system of equations:

- Bilinear;
- Over-determined composed of $(\#Unknowns)^2$ equations;
- With an exponential number of solutions.

The attack is effective: we find low rank codewords in $20s$ for 128bits of security (with Magma)

Limits of the Attack

$(n-k)d = n$ is essential for the attack and

Generally $(n-k)d \neq n$ for other schemes based on LRPC codes;

$\rightarrow$ LRPC codes: be careful with the choice of parameters.

Attacks Against the Code-Based IBE [GHPT17]

One IBE in code-based cryptography: it used RankSign...

The problem is deeper: even without RankSign, we also broke the parameters in the encryption part of the IBE.

Still admissible parameters for the encryption part.

Attacks Against the Code-Based IBE [GHPT17]

One IBE in code-based cryptography: it used RankSign...

The problem is deeper: even without RankSign, we also broke the parameters in the encryption part of the IBE.

Still admissible parameters for the encryption part.

Changing Rank $\rightarrow$ Hamming metric in the IBE scheme [GHPT17]: we gave a polynomial attack against the encryption part.

Thank You!