The Blockwise Rank Syndrome Learning problem and its applications to cryptography

Adrien Vinçotte

In collaboration with Nicolas Aragon, Pierre Briaud, Victor Dyseryn and Philippe Gaborit

Motivation

An easy problem:
Consider a public matrix $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$. Given $\mathbf{s} \in \mathbb{F}^{n-k}$, find $\mathbf{x} \in \mathbb{F}^{n}$ such that $\mathbf{H}\mathbf{x}^{\mathsf{T}} = \mathbf{s}^{\mathsf{T}}$.

How to make this problem difficult:
Add a constraint on $\mathbf{x}$: $\mathbf{x}$ must be of small weight for a particular metric:

- Hamming distance: code-based cryptography
- Euclidian distance: lattice-based cryptography
- Rank distance: rank-based cryptography

Background on rank metric

Definition: Rank metric over $\mathbb{F}_{q^m}^n$

For a vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$, we define the support: $\mathrm{Supp}(\mathbf{x}) = \langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$.
The rank weight of $\mathbf{x}$ is equal to: $\|\mathbf{x}\| = \dim(\mathrm{Supp}(\mathbf{x}))$.

Definition: $\mathbb{F}_{q^m}$-linear code

An $\mathbb{F}_{q^m}$-linear code of parameters $[n, k]_{q^m}$ is an $\mathbb{F}_{q^m}$-subspace of $\mathbb{F}_{q^m}^n$ of dimension k.
Such a code $\mathcal{C}$ can be represented by a full-rank generator matrix $\mathbf{G} \in \mathbb{F}_{q^m}^{k \times n}$ or by a full-rank parity-check matrix $\mathbf{H} \in \mathbb{F}_{q^m}^{(n-k) \times n}$.

Classic problems in rank metric

Definition: RD Problem

Given $(\mathbf{G}, \mathbf{y}) \in \mathbb{F}_{q^m}^{k \times n} \times \mathbb{F}_{q^m}^{n}$, the Rank Decoding problem $\mathrm{RD}(n, k, r)$ asks to compute $\mathbf{e} \in \mathbb{F}_{q^m}^{n}$ such that $\mathbf{y} = \mathbf{xG} + \mathbf{e}$ and $\|\mathbf{e}\| \leq r$.

There exists a probabilistic reduction to the SD problem [GZ14].

We will write RSD for the equivalent version written with a parity-check matrix.

Definition: RSD Problem

Given $(\mathbf{H}, \mathbf{s}) \in \mathbb{F}_{q^m}^{(n-k) \times n} \times \mathbb{F}_{q^m}^{n-k}$, the Rank Support Learning Problem $\mathrm{RSL}(n, k, r)$ asks to find $\mathbf{e} \in \mathbb{F}_{q^m}^{n}$ of rank $\|\mathbf{e}\| \leq r$ such that $\mathbf{He}^{\mathsf{T}} = \mathbf{s}^{\mathsf{T}}$.

Gabidulin codes [Gab85]

Definition: Gabidulin code

Let $k \leq n \leq m$ integers. Let $\mathbf{g} = (g_1, ..., g_n) \in \mathbb{F}_{q^m}^n$ an $\mathbb{F}_q$-linearly independent family. The Gabidulin code $[n, k]_{q^m}$ is defined by:

$$\mathcal{G}_{\mathbf{g}}(n, k, m) = \left\{ (P(g_1), ..., P(g_n)) \mid \deg_q(P) < k \right\}$$

It disposes of an efficient decoding algorithm allowing to decode until $\left\lfloor \frac{n-k}{2} \right\rfloor$ errors.

Original RQC scheme [AABB+14]

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 sk $= (\mathbf{x}, \mathbf{y})$, Public key: pk $= (\mathbf{h}, \mathbf{s} = \mathbf{x} + \mathbf{h}.\mathbf{y})$

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

Adaptation of HQC scheme in rank metric.
No need to mask a code.

LRPC codes [GMRZ13]

Definition: LRPC code

An $[n, k]_{q^m}$-linear code $\mathcal{C}$ is said to be LRPC of dual weight d if it admits a parity-check matrix $\mathbf{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$ such that: $\dim \mathrm{Supp}(\mathbf{H}) = d$

They dispose of an efficient syndrome decoding algorithm: Rank Support Recovery Algorithm.

Decoding algorithm for LRPC codes

Consider the RSD instance with an LRPC code: $\mathbf{e}\mathbf{H}^{\mathsf{T}} = s$.

We consider the following spaces:

- $E = \text{Supp}(\mathbf{e})$ the error support (to retrieve)
- $F = \text{Supp}(\mathbf{H})$
- $S = E \cdot F$ the syndrome space

Objective: retrieve E from F and S

Denoting $\{f_1, \dots f_d\}$ a basis of F, compute $E = \bigcap_{j=1}^{d} f_j^{-1} S$.

DFR: the Decoding Failure Rate increases with the dimension of S.

Original LRPC scheme

Alice		Bob
$\mathbf{x}, \mathbf{y} \xleftarrow{\$} \mathcal{S}_d^{2n}(\mathbb{F}_{q^m})$, $\mathbf{h} \leftarrow \mathbf{x}^{-1}\mathbf{y} \mod P$		
$F \leftarrow \text{Supp}(\mathbf{x}, \mathbf{y})$	$\xrightarrow{\mathbf{h}}$	$(\mathbf{e}_1, \mathbf{e}_2) \xleftarrow{\$} \mathcal{S}_r^{2n}(\mathbb{F}_{q^m})$
		$E \leftarrow \text{Supp}(\mathbf{e}_1, \mathbf{e}_2)$
$\mathbf{s} \leftarrow \mathbf{x}\mathbf{c}$	$\xleftarrow{\mathbf{c}}$	$\mathbf{c} \leftarrow \mathbf{e}_1 + \mathbf{e}_2\mathbf{h} \mod P$
$E \leftarrow \text{RSR}(F, \mathbf{s}, r)$		
Hash (E)	Shared Secret	Hash (E)

Figure: Informal description of ROLLO-I. **h** constitutes the public key.

Multi-syndrome approach [BBBG22]

The Multi-syndrome approach consists on giving to the decoder several syndromes associated to errors of same support.

Definition: RSL Problem

Given $(\mathbf{H}, \mathbf{S}) \in \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{N\times(n-k)}$, the Rank Support Learning Problem $\mathrm{RSL}(n, k, r, N)$ asks to compute a subspace $E \subset \mathbb{F}_{q^m}$ of dimension r for which there exists a matrix $\mathbf{V} \in E^{\ell\times n}$ such that $\mathbf{H}\mathbf{V}^{\mathsf{T}} = \mathbf{S}^{\mathsf{T}}$.

Increases capacity of decoding.

Augmented-Gabidulin codes

Definition: Augmented-Gabidulin code

Let $k \leq n' \leq m < n$ integers. Let $\mathbf{g} = (g_1, ..., g_{n'}) \in \mathbb{F}_{q^m}^{n'}$ an $\mathbb{F}_q$-linearly independent family.
The Gabidulin code $[n, k]_{q^m}$ is defined by:

$$\mathcal{G}_{\mathbf{g}}(n, n', k, m) = \left\{(P(g_1), ..., P(g_{n'}), 0, ..., 0) \mid \deg_q(P) < k\right\}$$

Advantage:
We immediatly deduce from the $n - n'$ last coordinates a part of the support (called the support erasure). If it has dimension ε , the decoding capacity is equal to $\left\lfloor \frac{n-k+\varepsilon}{2} \right\rfloor$.

Blockwise errors [SZHW23]

Definition: Blockwise ℓ-error

Let $\mathbf{n} = (n_1, \dots, n_\ell) \in \mathbb{N}^\ell$, $\mathbf{r} = (r_1, \dots, r_\ell) \in \mathbb{N}^\ell$ and $n \stackrel{\text{def}}{=} \sum_{i=1}^{\ell} n_i$.
An error $\mathbf{e} \in \mathbb{F}_{q^m}^n$ is said to be an ℓ-error with parameters $\mathbf{n}$ and $\mathbf{r}$ if $\mathbf{e} = (\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_\ell)$ such that:

- for all $i \in \{1, \dots, \ell\}$, $\|\mathbf{e}_i\| = r_i$,
- for all $i \neq j$, $\text{Supp}(\mathbf{e}_i) \cap \text{Supp}(\mathbf{e}_j) = \{0\}$.

We talk about a homogeneous error in the case of 1-error, i.e. standard error.

Definition: ℓ-LRPC code

An ℓ-LRPC code is a code such that its parity check matrix $\mathbf{H} = (\mathbf{H}_1 \mid \cdots \mid \mathbf{H}_\ell) \in \mathbb{F}_{q^m}^{(n-k)\times n}$ is such that:

- $\mathbf{H}_i \in \mathbb{F}_{q^m}^{(n-k)\times n_i}$ has its coefficients in a subspace F_i, with $d_i = \dim F_i$.
- For $i \neq j$: $F_i \cap F_j = \{0\}$.

If $\mathbf{e} = (\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_\ell)$ is an ℓ-error, then:

$$\mathbf{s}^\mathsf{T} = \mathbf{H}\mathbf{e}^\mathsf{T} = \sum_{i=1}^{\ell} \mathbf{H}_i \mathbf{e}_i^\mathsf{T}$$

Advantage: For a 2-LRPC code $[2n, n]$, with $r_1 = r_2 = \frac{r}{2}$ and $d_1 = d_2 = \frac{d}{2}$.

- With 2-errors: syndrome of weight $r_1 d_1 + r_2 d_2 = \frac{rd}{2}$
- Classical LRPC with homogeneous error of weight r and LRPC code of dual weight d: syndrome of weight rd

Have block errors of smaller weight has a strong impact on decoding capacity.

Our contribution:
We show that it is possible to combine these previous improvements (Multiple Syndromes, Augmented Gabidulin codes and BlockWise errors).

Allows to obtain new versions of RQC and LRPC schemes, with very small size of keys and ciphertexts.

Description of our RQC-MS-AG scheme

$\mathcal{S}^{(n_2,n_2)}_{(r_x,r_y)}(\mathbb{F}_{q^m})$: set of 2-errors of size (n_2, n_2) and weight (r_x, r_y)

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

Alice		Bob
$\mathbf{h} \xleftarrow{\$} \mathbb{F}_{q^m}^n$ $(\mathbf{x}, \mathbf{y}) \xleftarrow{\$} \mathcal{S}^{(n_2,n_2)}_{(r_x,r_y)}(\mathbb{F}_{q^m}), \mathbf{s} \leftarrow \mathbf{x} + \mathbf{h}.\mathbf{y}$	$\xrightarrow{\mathbf{h},\mathbf{s}}$	$(\mathbf{R}_1, \mathbf{R}_2, \mathbf{E}) \xleftarrow{\$} \mathcal{S}^{n_1 \times 3n_2}_{(r_1,r_2,r_e)}(\mathbb{F}_{q^m})$ $\mathbf{U} \leftarrow \mathbf{R}_1 + \mathbf{h} \cdot \mathbf{R}_2,$ $\mathbf{V} \leftarrow \mathrm{Fold}(\mathbf{m}\mathbf{G}) + \mathbf{s} \cdot \mathbf{R}_2 + \mathbf{E}$
$\mathbf{m} \leftarrow \mathcal{G}.\mathrm{Decode}(\mathrm{Unfold}(\mathbf{V} - \mathbf{y} \cdot \mathbf{U}))$	$\xleftarrow{\mathbf{U},\mathbf{V}}$	

$$\mathbf{V} - \mathbf{y} \cdot \mathbf{U} = \mathbf{mG} + \text{Unfold}(\mathbf{x} \cdot \mathbf{R}_2 - \mathbf{y} \cdot \mathbf{R}_1 + \mathbf{E})$$

Structure of errors:

- $(\mathbf{x}|\mathbf{y})$: 2-error of weight $\mathbf{r} = (r_x, r_y)$ and size $\mathbf{n} = (n_2, n_2)$.
- $(\mathbf{R}_1|\mathbf{R}_2|\mathbf{E})$: collection of 3-errors of weight $\mathbf{r} = (r_1, r_2, r_e)$ and size $\mathbf{n} = (n_2, n_2, n_2)$.

The weight of error to decode is equal to $r_x r_2 + r_y r_1 + r_e$.

Assume that $r_x = r_y = r_1 = r_2 = r_e = r$.

As blockwise errors, the error to correct would have weight $2r^2 + r$.

If $(\mathbf{x}|\mathbf{y})$ and $(\mathbf{R}_1|\mathbf{R}_2|\mathbf{E})$ were considered as homogeneous errors, their weight would be $2r$ and $3r$, and the error to correct would have weight $6r^2$.

Description of our ILRPC-block-MS scheme

Alice		Bob
$(\mathbf{x},\mathbf{y}) \xleftarrow{\$} \mathcal{S}^{(n_2,n_2)}_{(d_1,d_2)}(\mathbb{F}_{q^m})$, $\mathbf{h} \leftarrow \mathbf{x}^{-1}\mathbf{y} \mod P$ $F = (F_1, F_2) \leftarrow \mathrm{Supp}(\mathbf{x},\mathbf{y})$	$\xrightarrow{\mathbf{h}}$	Sample $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2)$ s.t. $\dim \mathcal{V}_i = r_i$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$ $\mathbf{e}_i = (\mathbf{e}_{i,1}, \mathbf{e}_{i,2}) \xleftarrow{\$} \mathcal{S}^{(n_2,n_2)}_{r}(\mathcal{V})$
$\mathbf{S} \leftarrow (\mathbf{x}\mathbf{c}_1, \dots, \mathbf{x}\mathbf{c}_{n_1})$ $\mathcal{V} \leftarrow \mathrm{RSR}(F, \mathbf{S}, r)$	$\xleftarrow{\mathbf{c}_i}$	$\mathbf{c}_i \leftarrow \mathbf{e}_{i,1} + \mathbf{e}_{i,2}\mathbf{h} \mod P$
Hash $(\mathcal{V})$	Shared Secret	Hash $(\mathcal{V})$

Figure: Informal description of ROLLO-I. $\mathbf{h}$ constitutes the public key.

Our structural attack against 2-LRPC codes

Let $\mathbf{H} = (\mathbf{H}_1|\mathbf{H}_2)$ a 2-LRPC parity check matrix of a code $\mathcal{C}$.
Objective: retrieve the structure of the code from a generator matrix $\mathbf{G}$ or an other parity check matrix $\mathbf{H}'$, by finding a blockwise vector of rank (r_1, r_2) in $\mathcal{C}^\perp$.

Let $(H_i)_{i \in \{1,\dots,n\}}$ the rows of $\mathbf{H}$.
There exists with high probability a word $\mathbf{x} = \sum_{i=1}^{n} a_i H_i \in \mathcal{C}^\perp$, whose the first $\lfloor n/d_1 \rfloor$ coefficients are equal to 0.

Truncate the $\lfloor n/d_1 \rfloor$ first coordinates of the code $\mathcal{C}^\perp$ and find this word.

Cryptanalysis of parameters given by SZHW23

Structural attack against 2-LRPC codes on an instance of the 2-RSD problem, a 2-LRPC code $[2n, n]$ on $\mathbb{F}_{q^m}$.

n	m	(d_1, d_2)	(r_1, r_2)	Claimed Sec.	Our Attack
67	61	(5,4)	(4,4)	145	**116**
79	71	(5,5)	(5,5)	225	**166**
89	79	(6,5)	(5,5)	281	**224**

Figure: Security of parameters for LOCKER given in SZHW23

Comparaison of parameters of different RQC schemes

Scheme	DFR	pk + ct
RQC-Block-MS-AG-128	-145	**1.4 kB**
RQC-Block-128	-	2.5 kB
RQC-NH-MS-AG-128	-158	2.7 kB
RQC-128	-	5.3 kB
RQC-Block-MS-AG-192	-206	**2.8 kB**
RQC-Block-192	-	5.3 kB
RQC-NH-MS-AG-192	-238	4.7 kB
RQC-192	-	8.3 kB

AG: Augmented Gabidulin
MS: Multiple Syndrome
Block: Blockwise errors

Scheme	DFR	pk + ct
ILRPC-Block-xMS-128	-128	1.7 kB
ILRPC-Block-xMS-192	-194	3.3 kB

Figure: Parameters of ILRPC schemes

Scheme	128 bits	192 bits
RQC-Block-MS-AG	**1.4 kB**	**2.8 kB**
KYBER	1.5 kB	2.2 kB
ILRPC-Block-MS	**1.7 kB**	**3.3 kB**
BIKE	3.1 kB	6.2 kB
HQC	6.7 kB	13.5 kB
Classic McEliece	261.2 kB	624.3 kB

Figure: Comparaison of different post-quantum encryption schemes

The sizes represent the sum of the key and the ciphertext.

Thank you for your attention