Post-Quantum Cryptography - Codes

Exercise sheet #2 – Cryptanalysis

August 1-5, 2022

Exercice 1 (McEliece cryptosystem based on GRS codes)**.** This exercise is widely inspired by [CGG+13].

Notation 1. *Let k be a non-negative integer. We denote by $\mathbb{F}_q[X]_{\leqslant k}$ the space of polynomials with coefficients in $\mathbb{F}_q$ of degree less than or equal to k.*

Recall the definition of a Generalized Reed-Solomon code (GRS):

Definition 1 (Generalized Reed-Solomon Code)**.** *Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$ be an n-tuple of pairwise distinct elements of $\mathbb{F}_q$ (in particular it entails $n \leqslant q$), and let $k \leqslant n$. Let $\mathbf{y} = (y_1, \ldots, y_n) \in (\mathbb{F}_q^\times)^n$ be an n-tuple of* non zero *elements, not necessarily distinct. The $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$ code is defined as*

$$\mathbf{GRS}_k(\mathbf{x}, \mathbf{y}) \stackrel{def}{=} \{(y_1 P(x_1), \ldots, y_n P(x_n)) \mid P \in \mathbb{F}_q[X]_{<k}\}.$$

$\mathbf{x}$ and $\mathbf{y}$ are respectively called the support *and* multiplier *vectors of $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$.*

Remark 1. *When $\mathbf{y} = (1, \ldots, 1)$, this code is known as the Reed-Solomon code $\mathbf{RS}_k(\mathbf{x})$.*

Q1. Show that $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$ has length n, dimension k and minimum distance $n - k + 1$. In particular, it reaches the Singleton bound (such a code is called an *MDS code*).

Q2. Recall the decoding algorithm seen in Lecture 1. How many errors can be decoded ?

Q3. Show that $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^\perp = \mathbf{GRS}_{n-k}(\mathbf{x}, \mathbf{y}^\perp)$ where $y_i^\perp = \dfrac{1}{y_i \prod_{i \neq j}(x_i - y_j)}$. In particular, the dual of a GRS code of dimension k, is a GRS code of dimension $n - k$, with same support.

GRS codes have been introduced in cryptography by Niederreiter in [Nie86] who proposed to use them to instantiate McEliece cryptosystem in order to reduce the size of the keys. However, in [SS92] Sidelnikov and Shestakov proved that such an instantiation was insecure. The goal of this exercise is to give another attack on GRS-based McEliece cryptosystem, using a very versatile tool called the *star product* of codes. Although being slower than the historical attack of Sidelnikov and Shestakov, this tool proved itself very useful to cryptanalyse McEliece cryptosystem based on many families of algebraic codes such as Algebraic-Geometry codes (of any genus), which

are generalizations of Reed-Solomon codes, Wild Goppa codes over quadratic extensions, some subspace subcodes of GRS codes etc...

In other words, given access to a public GRS code $\mathscr{C}_{pub} = \mathbf{GRS}_k(\mathbf{x}, \mathbf{y})$ (through a generator matrix for instance), the aim is to recover the secret parameters $\mathbf{x}$ and $\mathbf{y}$. More precisely, the aim is to recover some $\mathbf{x}'$, $\mathbf{y}'$ such that $\mathscr{C}_{pub} = \mathbf{GRS}_k(\mathbf{x}', \mathbf{y}')$ (The parameters are not unique).

Q4. Let $\alpha, \gamma \in \mathbb{F}_q^\times$ and $\mathbf{b} = (b, \ldots, b) \in \mathbb{F}_q^n$.

(a) Show that $\mathbf{GRS}_k(\mathbf{x}, \mathbf{y}) = \mathbf{GRS}_k(\alpha\mathbf{x} + \mathbf{b}, \gamma\mathbf{y})$.

(b) Deduce that we can assume $x_1 = 0$ and $x_2 = 1$.

We introduce the *star-product*, also known as the Schur product (or coordinate-wise product):

Definition 2 ($\star$-product)**.**

- *Let* $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^n$*. We define the* $\star$*-product of* $\mathbf{a}$ *and* $\mathbf{b}$ *as*

$$\mathbf{a} \star \mathbf{b} \stackrel{def}{=} (a_1 b_1, \ldots, a_n b_n).$$

- *Let* $\mathscr{A}, \mathscr{B} \subset \mathbb{F}_q^n$ *be two linear codes. We define their* $\star$*-product as*

$$\mathscr{A} \star \mathscr{B} \stackrel{def}{=} \mathrm{Span}\{\mathbf{a} \star \mathbf{b} \mid \mathbf{a} \in \mathscr{A}, \mathbf{b} \in \mathscr{B}\},$$

(note the presence of Span here, to ensure that $\mathscr{A} \star \mathscr{B}$ *is still a linear code).*

When $\mathscr{A} = \mathscr{B}$*, we denote by* $\mathscr{A}^2 \stackrel{def}{=} \mathscr{A} \star \mathscr{A}$ *the* square *of the code* $\mathscr{A}$*.*

Q5. Let $\mathscr{C} \in \mathbb{F}_q^n$ be a linear code of dimension k.

(a) Show that

$$\dim \mathscr{C}^2 \leqslant \min\left(n, \binom{k+1}{2}\right).$$

(b) Show that the complexity of computing a basis of $\mathscr{C}^2$ given a basis of $\mathscr{C}$ is $O(k^2 n^2)$ operations in $\mathbb{F}_q$.

In reality, this inequality is sharp for random linear codes. Indeed, it has been proven in [**?**] that when $\binom{k+1}{2} \leqslant n$ then $\dim \mathscr{C}^2 = \binom{k+1}{2}$ with overwhelming probability. In particular, the dimension of the square of a random code is *quadratic* in the dimension of the code.

Q6. (a) Show that for $k \leqslant (n+1)/2$,

$$\mathbf{GRS}_k(\mathbf{x}, \mathbf{y})^2 = \mathbf{GRS}_{2k-1}(\mathbf{x}, \mathbf{y} \star \mathbf{y}).$$

(b) Deduce a way to distinguish between small rate Generalized Reed-Solomon codes and random linear codes using the $\star$-product.

(c) Show that high rate GRS codes (when $2k-1 > n$) are also distinguishable from random codes.

Q7. (Bonus.) Can you give a very simple way (not using the $\star$-product) to distinguish between $\mathbf{RS}_k(\mathbf{x})$ (*i.e.* the multiplier is the all 1 vector) and a random linear code ?

So far we have found a *distinguisher* between GRS and random linear codes. It can undermine the security, but it is not yet an attack on the cryptosystem. There is still work to do to recover the secret parameters $(\mathbf{x}, \mathbf{y})$.

Filtration attack. In order to recover the secret parameters, we will build a *filtration* of codes, *i.e.* a sequence $(\mathscr{C}_i)$ of codes such that

$$\mathscr{C}_{pub} = \mathscr{C}_0 \supset \mathscr{C}_1 \supset \mathscr{C}_2 \supset \cdots \supset \mathscr{C}_i \supset \dots$$

where $\mathscr{C}_i \star \mathscr{C}_j \subset \mathscr{C}_{i+j}$.
From now on, we assume that $x_1 = 0, x_2 = 1$ and $k \leqslant \frac{n-1}{2}$.
In order to build this filtration, we will need a new operation on codes, namely the *shortening*.

Definition 3 (Shortened Code)**.** *Let $\mathscr{C}$ be a linear code, and $\mathcal{I} \subset \{1, \dots, n\}$ a set of positions. We define the shortening of $\mathscr{C}$ at $\mathcal{I}$ as the code $S_{\mathcal{I}}(\mathscr{C})$:*

$$S_{\mathcal{I}}(\mathscr{C}) \stackrel{def}{=} \{c \in \mathscr{C} \mid c_i = 0 \quad \forall i \in \mathcal{I}\}.$$

Remark 2. *This definition is slighly different as the one usually used. Indeed, with this definition $S_{\mathcal{I}}(\mathscr{C})$ contains codewords which are 0 on the same coordinates, and one usually delete those entries, yielding a code of length $n - |\mathcal{I}|$. However, the $\star$-product is only well defined for vectors of same length, therefore it is easier to keep those zero components.*

Q8. Given a code $\mathscr{C}$ and a set of positions $\mathcal{I}$, how can we compute a basis of $S_{\mathcal{I}}(\mathscr{C})$?

For $i, j > 0$ and $i + j < k$, we denote by $\mathscr{C}(i, j)$ the subcode of $\mathscr{C}_{pub} = \mathbf{GRS}(\mathbf{x}, \mathbf{y})$ given by the evaluation of polynomials vanishing at 0 with multiplicity at least i and at 1 with multiplicity at least j. We also set $\mathscr{C}(0, 0) \stackrel{\text{def}}{=} \mathscr{C}_{pub}$.

Q9. a) Give an interpretation of $\mathscr{C}(1, 0), \mathscr{C}(0, 1)$ and $\mathscr{C}(1, 1)$ as shortenings of $\mathscr{C}_{pub}$.
Hint: Recall that $x_1 = 0$ and $x_2 = 1$.

b) Deduce that they can be easily computed.

Q10. Assume $k \leqslant n/2$, and let i, j be integers such that $1 \leqslant i \leqslant k-2$ and $i + j \leqslant k - 2$.

a) Show that

$$\mathscr{C}(i+1, j) \star \mathscr{C}(i-1, j) = \mathscr{C}(i, j)^2 \quad \text{and} \quad \mathscr{C}(i, j+1) \star \mathscr{C}(i, j-1) = \mathscr{C}(i, j)^2.$$

b) Deduce an algorithm that, given generator matrices of $\mathscr{C}(i, j)$ and $\mathscr{C}(i-1, j)$, can recover a basis of $\mathscr{C}(i+1, j)$ in time $O(k^2 n^3 + k^3 n^2)$ operations over $\mathbb{F}_q$.

We are now ready to compute the filtration: For $i \leqslant k - 1$, set $\mathscr{C}_i \stackrel{\text{def}}{=} \mathscr{C}(i, 0)$.

Q11. Check that it indeed defines a filtration with the wanted properties, and that each term can be actually computed from previous ones.

Q12. What is the dimension of $\mathscr{C}_{k-1}$? What is the shape of a basis ?

Q13. Consider the code $\mathscr{C}(k-2,1)$. What is its dimension ? What is the shape of a basis ?

Q14. Show that $\mathbf{x}$ can be easily recovered from a basis of $\mathscr{C}(k-1,0)$ and $\mathscr{C}(k-2,1)$.

Q15. Conclude the attack. What is the overall time complexity ?

Exercice 2 (Rank metric codes)**.**

Introduction. In this course, you have been introduced to codes endowed with the Hamming metric. From an Information Theory point of view, it makes sense as they were originally designed to effectively correct errors that happen "in real life". For instance, Reed-Solomon codes are used in QR codes. However, other kind of errors can happen, and other ways of quantifying (and correcting) errors have been developped. In particular, rank metric codes have been introduced by Delsarte [Del78] from a combinatoric point of view, and Gabidulin [Gab85] with decoding algorithms, and have found applications in network communications [SK11], or in data storage [RKSV14]. From a Cryptography point of view, we actually do not even care if the metric makes sense in Information Theory, as long as the underlying decoding problem (of a general instance) is hard, and that we can define a trapdoor (for McEliece cryptosystem it is a decoding algorithm for the public code).

Matrix Codes. Let k, m, m be non negative integers, and let $\mathbb{F}_q$ denote the finite field with q elements. A (linear) matrix code of length $m \times n$ is nothing else than a subspace of $\mathcal{M}_{m,n}(\mathbb{F}_q)$ of dimension k, endowed with the following distance:

Definition 4 (Rank distance)**.** *Let* $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$*. The rank distance between* $\mathbf{X}$ *and* $\mathbf{Y}$ *is defined as*

$$d(\mathbf{X}, \mathbf{Y}) \stackrel{def}{=} \mathrm{Rank}(\mathbf{X} - \mathbf{Y})$$

Q1. Show that it indeed defines a distance on the space of $m \times n$ matrices $\mathcal{M}_{m,n}(\mathbb{F}_q)$.

Consider a matrix code $\mathscr{C} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$, and denote by k its dimension. Therefore it is generated by k matrices $\mathbf{M}_1, \ldots, \mathbf{M}_k \in \mathcal{M}_{m,n}(\mathbb{F}_q)$. The decoding problem at distance r is now defined as follows: Given a matrix $\mathbf{Y} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$, recover (if exists) an element $\mathbf{C} = x_1\mathbf{M}_1 + \cdots + x_k\mathbf{M}_k \in \mathcal{C}$ (where $x_i \in \mathbb{F}_q$) at rank distance $\leqslant r$ from $\mathbf{Y}$. This is precisely the MinRank problem also used in *Multivariate Cryptography*:

Problem 1 (MinRank)**.** *Given an integer* $r \in \mathbb{N}$ *and* $k+1$ *matrices* $\mathbf{Y}, \mathbf{M}_1, \ldots, \mathbf{M}_k \in \mathcal{M}_{m,n}(\mathbb{F}_q)$*, output field elements* $x_1, \ldots, x_k \in \mathbb{F}_q$ *such that*

$$\mathrm{Rank}\left(\mathbf{Y} - \sum_{i=1}^{k} x_i \mathbf{M}\right) \leqslant r.$$

As it is the case for the Hamming metric, the decisional version of MinRank Problem is known to be NP-complete, and the best known algorithm solving it have exponential complexity on average, namely when inputs are drawn uniformly at random.

$\mathbb{F}_{q^m}$-linear rank metric codes. Few families of codes endowed with this rank metric have efficient decoding algorithms, and most of them are in fact linear over the extension field $\mathbb{F}_{q^m}$, *i.e.* are isometric (for a suitable definition of rank distance) to $\mathbb{F}_{q^m}$-subspaces of dimension k of $\mathbb{F}_{q^m}^n$. The idea is to see any vector $\mathbf{x} \in \mathbb{F}_{q^m}^n$ as an $m \times n$ matrix over $\mathbb{F}_q$ and define the rank distance naturally.

Definition 5 (Rank distance over $\mathbb{F}_{q^m}^n$.)**.** *Let $\mathcal{B} = (\beta_1, \ldots, \beta_m)$ be a basis of the extension field $\mathbb{F}_{q^m}/\mathbb{F}_q$. For any vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ we define its extension with respect to $\mathcal{B}$ as the matrix*

$$\mathrm{Ext}_{\mathcal{B}}(\mathbf{x}) \stackrel{def}{=} \begin{pmatrix} x_1^{(1)} & \ldots & x_n^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(m)} & \ldots & x_n^{(m)} \end{pmatrix} \in \mathcal{M}_{m,n}(\mathbb{F}_q).$$

where $x_i = \sum_{j=1}^{m} x_i^{(j)} \beta_j$ is the decomposition of x_i along the basis $\mathcal{B}$.

For $\mathbf{x}, \mathbf{y} \in \mathbb{F}_{q^m}^n$, their rank distance is defined as

$$d(\mathbf{x}, \mathbf{y}) \stackrel{def}{=} d\left(\mathrm{Ext}_{\mathcal{B}}(\mathbf{x}), \mathrm{Ext}_{\mathcal{B}}(\mathbf{y})\right) = \mathrm{Rank}\left(\mathrm{Ext}_{\mathcal{B}}(\mathbf{x}) - \mathrm{Ext}_{\mathcal{B}}(\mathbf{y})\right).$$

We also define the rank weight of a vector $|\mathbf{x}|_R$ (or simply $|\mathbf{x}|$ when the context is clear) as

$$|\mathbf{x}| \stackrel{def}{=} \mathrm{Rank}(\mathrm{Ext}_{\mathcal{B}}(\mathbf{x})) = d(\mathbf{x}, 0).$$

Q2. Show that this distance is well defined, that is to say does not depend on the choosen basis $\mathcal{B}$.

Definition 6 ($\mathbb{F}_{q^m}$-linear code)**.** *An $[n, k]$ $\mathbb{F}_{q^m}$-linear code is a k-dimensional linear subspace of $\mathbb{F}_{q^m}^n$, endowed with the above rank distance. As for usual codes, n and k are respectively called the length and the dimension of the code.*

Q3.
a) Show that an $\mathbb{F}_{q^m}$-linear code of dimension k induces a matrix code in $\mathcal{M}_{m,n}(\mathbb{F}_q)$ of dimension km.
b) How many bits are needed to represent the basis of a matrix code $\mathscr{C} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ of dimension $K = km$?
c) What about an $\mathbb{F}_{q^m}$-linear code of dimension k ?
d) Conclude that $\mathbb{F}_{q^m}$-linear codes can be represented more compactly.

For $\mathbb{F}_{q^m}$-linear rank metric codes, the decoding problem is more look-alike to the decoding problem in the Hamming metric:

Problem 2 (Rank Decoding Problem (RDP))**.** *Given a generator matrix $\mathbf{G}$ of a rank metric code $\mathscr{C} \subset \mathbb{F}_{q^m}^n$, and $\mathbf{y} \in \mathbb{F}_{q^m}^n$ such that $\mathbf{y} = \mathbf{mG} + \mathbf{e}$ where $\mathbf{m} \in \mathbb{F}_{q^m}^k, \mathbf{e} \in \mathbb{F}_{q^m}^n$ and $|\mathbf{e}| = t$, find $\mathbf{e}$.*

Q4. Recall that the field $\mathbb{F}_{q^m}$ has a $\mathbb{F}_q$ basis $\mathcal{B}$ of the form $(1, \alpha, \alpha^2, \ldots, \alpha^{m-1})$ [α is nothing else than a generator of the cyclic group $(\mathbb{F}_{q^m}^\times)$]. Let $\mu_\alpha \stackrel{\text{def}}{=} \sum_{i=0}^{m-1} a_i X^i + X^m$ be its monic minimal polynomial, and consider its companion matrix

$$\mathbf{C}_\mu \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \ddots & \vdots & -a_1 \\ 0 & 1 & \ddots & 0 & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 1 & -a_{m-1} \end{pmatrix} \in \mathcal{M}_{m,n}(\mathbb{F}_q).$$

a) Show that $\mathbf{C}_\mu$ represents the $\mathbb{F}_q$ linear map $\varphi\colon y \in \mathbb{F}_{q^m} \mapsto \alpha \cdot y$ in the basis $\mathcal{B}$.

b) Consider the $\mathbb{F}_q$ vector subspace $\mathscr{A}$ of $\mathcal{M}_{m,n}(\mathbb{F}_q)$, generated by all the $(\mathbf{C}_\mu^j)_{j\in\mathbb{N}}$. Show that it is an $\mathbb{F}_q$ algebra of dimension m.

c) Let $\mathscr{C}$ be an $[n,k]$ $\mathbb{F}_{q^m}$-linear code, and let $\widehat{\mathscr{C}} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ be the km dimensional associated matrix code. Show that $\widehat{\mathscr{C}}$ is stable by left multiplication by elements of $\mathscr{A}$.

d) Deduce that RDP is a very specific case of MinRank.

The above question shows that $\mathbb{F}_{q^m}$-linear codes have a quite strong algebraic structure, and explains why we can win a factor m by using such codes compared to usual matrix codes, or even $\mathbb{F}_{q^m}$ codes endowed with the Hamming metric.

Similarly to what happens in the Hamming metric, the parameter of a code $\mathscr{C}$ that quantifies the maximum number of rank errors that can be uniquely decoded is its minimum distance defined as the minimum distance between two distinct codewords. Alternatively, due to the linearity, it is also the minimum rank weight of a non zero codeword:

$$d_{min}(\mathscr{C}) \stackrel{\text{def}}{=} \min\{|\mathbf{c}| \mid \mathbf{c} \in \mathscr{C}\backslash\{0\}\}.$$

Most Hamming metric quantities (*e.g.* bounds), have a rank metric counterpart.

Q5. Recall the Singleton bound: For an $[n,k]$-linear code $\mathscr{C}$ endowed with the Hamming metric, its minimum distance d is upper bounded by $n-k+1$. Prove its rank metric counterpart:

For an $\mathbb{F}_{q^m}$-linear code $\mathscr{C} \in \mathbb{F}_{q^m}^n$ of dimension k, its minimum rank distance d satisfies

$$d \leqslant n - k + 1.$$

Hint: $\mathscr{C}$ can be endowed with the Hamming metric !

Hardness of RDP. While both MinRank and the Decoding Problem in Hamming metric are known to be NP-complete, this is not the case for $\mathbb{F}_{q^m}$-linear rank metric codes for which there only exists a *randomized* reduction due to Gaborit and Zémor [GZ16]. However, NP-hardness is only a *worst-case* hardness, while we are most interested as an *average-case* hardness. Therefore, such a randomized reduction is still relevant for cryptography. Moreover, best known generic decoding algorithms have a time complexity exponential in the rank weight of the error(see [AGHT18, BBC$^+$20] for more information).

As a consequence, rank metric codes are interesting for cryptographic applications, because the underlying decoding problem remains hard, while the codes can be described more compactly: The $\mathbb{F}_{q^m}$-linearity structure allows to win a factor m in the size of the description of a code compared to the Hamming metric.

Rank metric McEliece cryptosystem. Recall that McEliece cryptosystem requires codes that

- Have an efficient decoding algorithm.
- Are indistinguishable from random codes of same parameters.

Contrary to what happens in the Hamming metric, in the rank metric we know very few codes that come with efficient decoders, and most of them are $\mathbb{F}_{q^m}$-linear codes. We can distinguish two families:

- Codes with a probabilistic decoder, such as LRPC codes [GMRZ13] whose dual have a basis of small rank weight codewords. They can be thought as rank metric analogues of LRPC/MDPC codes.
- Codes with a deterministic decoder, such as Gabidulin codes [Gab85], rank metric analogues of Reed-Solomon codes.

ROLLO [ABD+19], an instantiation of McEliece cryptosystem with LRPC codes, made to the second round of the NIST competition, but the first cryptosystem that used rank metric codes is due to Gabidulin, Paramonov and Tretjakov ([GPT91]) who proposed to instantiate McEliece cryptosystem with Gabidulin codes. They are rank metric analogues of Reed-Solomon codes, and both families of codes share many properties. Gabidulin codes are evaluation codes of a class of polynomials called q-polynomials, *i.e.* polynomials of the form

$$P(X) = p_0 X + p_1 X^q + \cdots + p_r X^{q^r}, \quad \text{with } p_i \in \mathbb{F}_{q^m}, \quad p_r \neq 0$$

The integer r is called the q-degree of P and is denoted by $\deg_q(P)$. A q-polynomial induces an $\mathbb{F}_q$-linear map $\mathbb{F}_{q^m} \to \mathbb{F}_{q^m}$ whose kernel has dimension bounded by $\deg_q(P)$. The set of all q-polynomials is a graded ring, endowed with the usual addition, and the composition operation.

Remark 3. *Note that the ring of q-polynomials is not commutative. Indeed, let $a \in \mathbb{F}_{q^m} \backslash \mathbb{F}_q$. Then*

$$aX \cdot X^q = aX^q, \qquad \textit{while } X^q \cdot aX = a^q X^q.$$

Although being not commutative, this ring has a very rich arithmetic structure. Indeed, it is both left and right Euclidean, and those Euclidean divisions can be efficiently computed.

Proposition 1.

- For any q-polynomials (A, B), there exists a unique (Q, R) such that

$$A = B \circ Q + R, \quad \text{and } \deg_q(R) < \deg_q(B).$$

- For any q-polynomials (A, B), there exists a unique (S, T) such that

$$A = S \circ B + T, \quad \text{and } \deg_q(T) < \deg_q(B).$$

As a consequence, one needs to be careful on which side they work, but they can adapt most of efficient algorithms for polynomials to q-polynomials. We are now ready to define Gabidulin codes.

Definition 7 (Gabidulin code.)**.** *Let* $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_n) \in \mathbb{F}_{q^m}^n$ *be a vector of* linearly independent *elements over* $\mathbb{F}_q$ *(In particular, it entails that* $n \leqslant m$*). The Gabidulin code of dimension* k *and evaluation vector* $\mathbf{g}$ *is defined as*

$$\mathbf{Gab}_k(\mathbf{g}) = \left\{(P(g_1), \dots, P(g_n)) \mid \deg_q(P) < k\right\} \subset \mathbb{F}_{q^m}^n.$$

Q6. Show that $\mathbf{Gab}_k(\mathbf{g})$ has length n, dimension k and minimum distance $n - k + 1$. In particular, it reaches the Singleton bound.

Most of efficient decoding algorithms for Reed-Solomon codes can be translated to give efficient decoders for Gabidulin codes up to $\lfloor \frac{n-k}{2} \rfloor$.

Q7. (Bonus) Adapting the Berlekamp-Welch algorithm for decoding Reed-Solomon codes (see Lecture notes 1), propose an algorithm to decode $[n, k]$ Gabidulin codes up to $\lfloor \frac{n-k}{2} \rfloor$ rank errors.

However, contrary to Reed-Solomon codes for which there exist polynomial time decoders slightly beyond this unique decoding radius, for Gabidulin codes the situation is completely different ([RWZ15]).

Overbeck's distinguisher. If they share many interesting properties, Gabidulin codes and Reed-Solomon codes also share their flaws. Indeed, Gabidulin codes are also very easily distinguishable from random $\mathbb{F}_{q^m}$-linear codes.

This distinguisher is due to Overbeck [Ove05] and is based on the Frobenius operator. Given a vector $\mathbf{x} \in \mathbb{F}_{q^m}^n$ and a non negative integer s, we denote by $\mathbf{x}^{[s]}$ the vector:

$$\mathbf{x}^{[s]} \stackrel{\text{def}}{=} (x_1^{q^s}, \dots, x_n^{q^s}).$$

Similarly, given a code $\mathscr{C} \subset \mathbb{F}_{q^m}^n$, the code $\mathscr{C}^{[s]}$ is defined as

$$\mathscr{C}^{[s]} \stackrel{\text{def}}{=} \{\mathbf{c}^{[s]} \mid \mathbf{c} \in \mathscr{C}\}.$$

Q8. Let $\mathbf{Gab}_k(\mathbf{g})$ be a Gabidulin code of dimension k, and let $s \in \mathbb{N}$.

a) Show that $\mathbf{Gab}_k(\mathbf{g}) + \mathbf{Gab}_k(\mathbf{g})^{[1]} + \cdots \mathbf{Gab}_k(\mathbf{g})^{[s]} = \mathbf{Gab}_{k+s}(\mathbf{g})$.

b) Deduce that $\dim_{\mathbb{F}_{q^m}} \left(\mathbf{Gab}_k(\mathbf{g}) + \mathbf{Gab}_k(\mathbf{g})^{[1]} + \cdots \mathbf{Gab}_k(\mathbf{g})^{[s]}\right)$ grows linearly in s.

This has to be compared with the behaviour of a random $\mathbb{F}_{q^m}$-linear code of dimension k.

Let $\mathscr{C}$ be a subspace of $\mathbb{F}_{q^m}^n$ chosen uniformly at random among all k dimensional subspaces, *i.e.* a uniformly random $\mathbb{F}_{q^m}$-linear code of dimension k. The aim of the following questions is to estimate the dimension of $\mathscr{C} + \mathscr{C}^{[1]} + \cdots + \mathscr{C}^{[s]}$. More precisely, we want to prove that

$$\mathbb{P}\left(\dim_{\mathbb{F}_{q^m}} \left(\mathscr{C} + \mathscr{C}^{[1]} + \cdots + \mathscr{C}^{[s]}\right) \leqslant \min(n, (s+1)k) - \varepsilon\right) = O(q^{-m\varepsilon}).$$

In order to do that we will follow [CC20], which presents an elegent proof of this result.

Q9. Let $\begin{bmatrix} n \\ k \end{bmatrix}_q$ denote the Gaussian binomial coefficient, which counts the number of k dimensional subspaces of $\mathbb{F}_q^n$.

a) Show that

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{t=0}^{k-1} \frac{q^n - q^t}{q^k - q^t} = q^{k(n-k)} \prod_{t=0}^{k-1} \frac{1 - q^{t-n}}{1 - q^{t-k}}.$$

b) Show that there exists a positive constant C such that for any pair of positive integers $n \geqslant k$, we have

$$q^{k(n-k)} \leqslant \begin{bmatrix} n \\ k \end{bmatrix}_q \leqslant C \cdot q^{k(n-k)}.$$

Consider the following map Ψ.

$$\Psi \colon \begin{cases} \mathscr{C} \times \cdots \times \mathscr{C} \to \mathbb{F}_{q^m}^n \\ (\mathbf{c}_0, \ldots, \mathbf{c}_s) \mapsto \mathbf{c}_0 + \mathbf{c}_1^{[1]} + \cdots + \mathbf{c}_s^{[s]}. \end{cases}$$

The dimension we want to estimate is related to the dimension of $\ker \Psi$ through the rank–nullity theorem, therefore we will first estimate $\mathbb{E}(|\ker \Psi|)$.

Q10. Let $\mathscr{A}$ be a subspace of $\mathbb{F}_{q^m}^n$ of dimension $t \leqslant k$. Show that

$$\mathbb{P}(\mathscr{A} \subset \mathscr{C}) \leqslant C \cdot q^{-mt(n-k)}.$$

Q11. Show that

$$\mathbb{E}(|\ker \Psi|) = \sum_{\substack{(\mathbf{x}_0, \ldots, \mathbf{x}_s) \in (\mathbb{F}_{q^m}^n)^s \\ \mathbf{x}_0 + \mathbf{x}_1^{[1]} + \cdots + \mathbf{x}_s^{[s]} = \mathbf{0}}} \mathbb{P}\left(\mathbf{x}_0, \ldots, \mathbf{x}_s \in \mathscr{C}\right).$$

For $0 \leqslant t \leqslant s+1$, let

$$E_t \stackrel{\text{def}}{=} \left\{ (\mathbf{x}_0, \ldots, \mathbf{x}_s) \in \left(\mathbb{F}_{q^m}^n\right)^s \;\middle|\; \begin{array}{c} \mathbf{x}_0 + \mathbf{x}_1^{[1]} + \cdots + \mathbf{x}_s^{[s]} = \mathbf{0} \\ \dim_{\mathbb{F}_{q^m}} \operatorname{Span}(\mathbf{x}_0, \ldots, \mathbf{x}_s) = t \end{array} \right\}.$$

Q12. Show that

$$\mathbb{E}(|\ker \Psi|) \leqslant C \cdot \sum_{t=0}^{s+1} q^{-mt(n-k)} |E_t|.$$

We are now reduced to estimate $|E_t|$ for all $0 \leqslant t \leqslant s+1$. Let $(\mathbf{x}_0, \ldots, \mathbf{x}_s) \in E_t$.

Q13. Show that there exists a unique $(s+1-t) \times (s+1)$ full rank matrix $\mathbf{M}$ in row reduced echelon form, with coefficients in $\mathbb{F}_{q^m}$, such that

$$\mathbf{M} \cdot \underbrace{\begin{pmatrix} \mathbf{x}_0 \\ \hline \vdots \\ \hline \mathbf{x}_s \end{pmatrix}}_{\in \mathcal{M}_{s+1,n}(\mathbb{F}_{q^m})} = 0.$$

Q14. Let us label the columns of $\mathbf{M}$ from 0 to s. Let $\mathcal{P} \subset \{0, \dots, s\}$ be the set of indices of columns which are pivots for $\mathbf{M}$ (*i.e.* which contain a leading 1 and 0 elsewhere), and denote by $\mathcal{P}^c$ its complement. Set a to be the smallest element of $\mathcal{P}^c$. Check that

$$|\mathcal{P}^c| = t \quad \text{and} \quad a \leqslant s + 1 - t.$$

Q15. Show that for any $1 \leqslant i \leqslant n$, there exists a q-polynomial Q of q-degree at most a such that $Q(x_{a,i}) = 0$.

Q16. a) Show that there are at most $q^{m(t-1)+a} \leqslant q^{m(t-1)+s+1-t}$ possible choices for the tuple $(x_{0,i}, \dots, x_{s,i})$ for each $1 \leqslant i \leqslant n$.

b) Deduce that

$$|E_t| \leqslant C \cdot q^{(mt+n)(s+1-t)+mn(t-1)}.$$

Hint: Full rank $(s+1-t) \times (s+1)$ matrices in row reduced echelon form, with entries in $\mathbb{F}_{q^m}$, are in one-to-one correspondance with t-dimensional linear subspaces of $\mathbb{F}_{q^m}^{s+1}$.

Q17. a) Show that

$$\mathbb{E}(|\ker \Psi|) \leqslant C^2 \cdot q^{m(k(s+1)-n)} \cdot \sum_{t=0}^{s+1} q^{(s+1-t)(mt+n-mk)}$$

b) Conclude that there exists a positive constant C' such that

$$\mathbb{E}(|\ker \Psi|) \leqslant C' \cdot q^{m(k(s+1)-n)}.$$

We are now able to finish the proof of our distinguisher. Assume that

$$\dim_{\mathbb{F}_{q^m}} \left(\mathscr{C} + \mathscr{C}^{[1]} + \cdots + \mathscr{C}^{[s]} \right) \leqslant \min\{n, (s+1)k\} - \varepsilon,$$

i.e.

$$\dim_{\mathbb{F}_{q^m}} \ker \Psi \geqslant \max\{0, k(s+1) - n\} + \varepsilon.$$

Q18. a) Show that

$$\mathbb{P}\left(|\ker \Psi| \geqslant q^{m\left(\max\{0, k(s+1)-n\}+\varepsilon\right)} \right) \leqslant C' \cdot q^{-m\varepsilon}$$

b) Conclude the proof.

We have now a very easy distinguisher between Gabidulin codes and random $\mathbb{F}_{q^m}$-linear codes. It is not enough to recover the secret, however it is enough to break McEliece cryptosystem instantiated with Gabidulin codes.

In the previous exercise it was shown that the $\star$-product distinguisher for GRS can be turned into a key recovery attack. For Gabidulin codes, it is also the case, but the exercise is long enough. The interested reader can see the original paper by Overbeck [Ove05].

In fact, it turns out that Overbeck idea to look at sum of iterated Frobenius has proved to be very fruitful to cryptanalyse cryptosystems based on Gabidulin codes and variants [GOTK18, CC20] (non exhaustive).

What more ? Obviously, rank metric cryptography does not stop here. One can define a rank metric analogue of HQC cryptosystem: RQC, which made to the second round of NIST competition [AAB$^+$20]. Notice that this cryptosystem also uses Gabidulin codes, the security does not rely on the hardness of distinguishing them from random codes. As it is the case for HQC, we only need a (public) code that can be efficiently decoded, and Gabidulin codes are suitable for that.

If no rank metric cryptosystems have made to the following rounds, partly due to algebraic attacks that improved generic decoding of $\mathbb{F}_{q^m}$-linear codes by solving the particular instance of MinRank that appeared in Eurocrypt 2020 and Asiacrypt 2020 [BBB$^+$20, BBC$^+$20], this does not mean the end of rank metric cryptography, and NIST insisted in their second round report that "The rank metric cryptosystems offer a nice alternative to traditional Hamming metric codes with comparable bandwith." [AJD$^+$20].

References

[AAB+20] Carlos Aguilar Melchor, Nicolas Aragon, Slim Bettaieb, Loïc Bidoux, Olivier Blazy, Maxime Bros, Alain Couvreur, Jean-Christophe Deneuville, Philippe Gaborit, Gilles Zémor, and Adrien Hauteville. Rank quasi cyclic (RQC). Second Round submission to NIST Post-Quantum Cryptography call, April 2020.

[ABD+19] Nicolas Aragon, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, Adrien Hauteville, Olivier Ruatta, Jean-Pierre Tillich, Gilles Zémor, Carlos Aguilar Melchor, Slim Bettaieb, Loïc Bidoux, Bardet Magali, and Ayoub Otmani. ROLLO (merger of Rank-Ouroboros, LAKE and LOCKER). Second round submission to the NIST post-quantum cryptography call, March 2019.

[AGHT18] Nicolas Aragon, Philippe Gaborit, Adrien Hauteville, and Jean-Pierre Tillich. A new algorithm for solving the rank syndrome decoding problem. In *2018 IEEE International Symposium on Information Theory, ISIT 2018, Vail, CO, USA, June 17-22, 2018*, pages 2421–2425. IEEE, 2018.

[AJD+20] Gorjan Alagic, Alperin-Sheriff Jacob, Apon Daniel, David Cooper, Quynh Dang, John Kelsey, Yi-Kai Liu, Carl Miller, Dustin Moody, Rene Peralta, Ray Perlner, Angela Robinson, and Daniel Smith-Tone. Status report on the second round of the NIST post-quantum cryptography standardization process. Technical Report NISTIR 8309, NIST, July 2020.

[BBB+20] Magali Bardet, Pierre Briaud, Maxime Bros, Philippe Gaborit, Vincent Neiger, Olivier Ruatta, and Jean-Pierre Tillich. An algebraic attack on rank metric code-based cryptosystems. In *Advances in Cryptology - EUROCRYPT 2020 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020. Proceedings*, 2020.

[BBC+20] Magali Bardet, Maxime Bros, Daniel Cabarcas, Philippe Gaborit, Ray Perlner, Daniel Smith-Tone, Jean-Pierre Tillich, and Javier Verbel. Improvements of algebraic attacks for solving the rank decoding and minrank problems. In *Advances in Cryptology - ASIACRYPT 2020, International Conference on the Theory and Application of Cryptology and Information Security, 2020. Proceedings*, pages 507–536, 2020.

[CC20] Daniel Coggia and Alain Couvreur. On the security of a Loidreau's rank metric code based encryption scheme. *Des. Codes Cryptogr.*, 88:1941–1957, 2020.

[CGG+13] Alain Couvreur, Philippe Gaborit, Valérie Gautier, Ayoub Otmani, and Jean-Pierre Tillich. Distinguisher-Based Attacks on Public-Key Cryptosystems Using Reed-Solomon Codes. In *International Workshop on Coding and Cryptography - WCC 2013*, pages 181–193, Bergen, Norway, April 2013.

[Del78] Philippe Delsarte. Bilinear forms over a finite field, with applications to coding theory. *J. Comb. Theory, Ser. A*, 25(3):226–241, 1978.

[Gab85] Ernest M. Gabidulin. Theory of codes with maximum rank distance. *Problemy Peredachi Informatsii*, 21(1):3–16, 1985.

[GMRZ13] Philippe Gaborit, Gaétan Murat, Olivier Ruatta, and Gilles Zémor. Low rank parity check codes and their application to cryptography. In *Proceedings of the Workshop on Coding and Cryptography WCC'2013*, Bergen, Norway, 2013.

[GOTK18] Philippe Gaborit, Ayoub Otmani, and Hervé Talé-Kalachi. Polynomial-time key recovery attack on the Faure-Loidreau scheme based on Gabidulin codes. *Des. Codes Cryptogr.*, 86(7):1391–1403, 2018.

[GPT91] Ernst M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a non-commutative ring and their applications to cryptography. In *Advances in Cryptology - EUROCRYPT'91*, number 547 in LNCS, pages 482–489, Brighton, April 1991.

[GZ16] Philippe Gaborit and Gilles Zémor. On the hardness of the decoding and the minimum distance problems for rank codes. *IEEE Trans. Inform. Theory*, 62(12):7245–7252, 2016.

[Nie86] Harald Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Problems of Control and Information Theory*, 15(2):159–166, 1986.

[Ove05] Raphael Overbeck. A new structural attack for GPT and variants. In *Mycrypt*, volume 3715 of *LNCS*, pages 50–63, 2005.

[RKSV14] Ankit S. Rawat, O. Ozan Koyluglu, Natalia Silberstein, and Sriram Vishwanath. Optimal locally repairable and secure codes for distributed storage systems. *IEEE Trans. Inform. Theory*, 60(1):212–236, 2014.

[RWZ15] Netanel Raviv and Antonia Wachter-Zeh. Some Gabidulin codes cannot be list decoded efficiently at any radius. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 6–10, 2015.

[SK11] Danilo Silva and Frank R. Kschischang. Universal secure network coding via rank-metric codes. *IEEE Trans. Inform. Theory*, 57(2):1124–1135, Feb 2011.

[SS92] Vladimir Michilovich Sidelnikov and S.O. Shestakov. On the insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discrete Math. Appl.*, 1(4):439–444, 1992.