A new rank metric codes based encryption scheme

Pierre Loidreau

A new rank metric codes based encryption scheme

Pierre Loidreau

DGA MI and Université de Rennes 1

Abstract. We design a new McEliece-like rank metric based encryption scheme from Gabidulin codes. We explain why it is not affected by the invariant subspace attacks also known as Overbeck's attacks. The idea of the design mixes two existing approaches designing rank metric based encryption schemes. For a given security our public-keys are more compact than for the same security in the Hamming metric based settings.

1 Introduction

The security of the main *post-quantum* (PQ) primitives relies on the difficulty of solving decoding problems in some metrics (Hamming metric for codes, Euclidean metric for lattices). The security of the encryption schemes is generally evaluated relatively to the best existing algorithms solving the considered problems.

At the beginning of the 90's another type of code-based cryptography emerged whose security was based on an alternative metric, the so-called rank metric [GPT91]. The difference with McEliece cryptosystem consists in the choice of the family of codes and in the choice of the metric. Originally, there was only one family of codes with an efficient algebraic polynomial-time decoding algorithm up to some bound, the family of Gabidulin codes [Gab85]. The initial proposals were attacked by Gibson who was able to recover a decoder from the public-key in polynomial time, [Gib95,Gib96]. Then, there was a succession of reparations and attacks, the latter being usually devastating. Overbeck in 2005 proposed a framework which could be adapted to all variants of Gabidulin codes based encryption schemes, [Ove05]. The structural weakness of the scheme came from the fact that Gabidulin codes contain a huge vector space invariant under the Frobenius automorphism. Exploiting this weakness lead to the complete cryptanalysis of all the previous Gabidulin codes based cryptosystems. From this date on, evolutions were proposed claiming to be secure against the existing attacks, [Gab08,GRH09,RGH10]. However, it was recently shown in [OKN16], that all existing variants could be reformulated as instances of the original problem, thus breakable in polynomial-time. Until now the common idea was that although rank metric would be a good candidate for designing code-based primitives with compact keys, a cryptosystem could not be designed from Gabidulin codes.

In the paper we argue against this idea. We show that Gabidulin codes can be used to design effective and secure code-based cryptosystems, moreover with compact keys. By *secure* we mean that the complexity of an attack consisting in recovering a polynomial-time decoder for the public code is exponential. The point is to scramble sufficiently the structure of Gabidulin codes to avoid existing attacks. Concerning Goppa codes that are subfield subcodes of Generalized Reed-Solomon codes (GRS) on scrambles the structure by keeping the subcode formed with the binary vectors of the parent GRS code. Though GRS codes are unsuitable for use in cryptosystems, Goppa codes are widely admitted as being suitable and even the best choice for the design of secure code-based primitives and even PQ primitives, [AB315]. Unfortunately, this idea does not work for Gabidulin codes since subfield subcodes of Gabidulin codes are isomorphic to the direct product of Gabidulin codes over smaller fields, [GL08]. We propose a new approach mixing original ideas such as the structure of the encryption scheme and more recent ideas who led to the design of *Low Rank Parity-check Codes* (LRPC) based encryption schemes. This idea can also be considered as an adaptation to rank metric of an idea in Hamming metric whose interesting instances were broken, [BBC$^+$16,COTG15]. For a given security of 128 and 256 bits, and a PQ security of 128 Mcbits, we propose a public-key size 20 times smaller than the proposition for *long term post-quantum systems*, in [AB315] relying on Goppa codes. The parameters being versatile, a designer can tune the parameters according to its needs (smaller key and larger ciphertext expansion or larger key and smaller ciphertext expansion for instance).

The structure of the paper is the following: First we define rank metric, the related decoding problems and we emphasize the fact that the complexity of generic decoding in rank metric is exponentially more difficult than in Hamming metric for the same settings. We evaluate the consequence of Grover algorithm on the generic decoding complexity in a PQ world. In a second part, we present how rank metric is commonly used in the design of encryption schemes. We also briefly detail the reason why Gabidulin codes based encryption schemes were broken. Finally, we show how to hide the structure of Gabidulin codes in a very simple manner, avoiding thus the main weaknesses of Gabidulin codes based cryptosystems. We also analyze the security of the encryption scheme against various attacks and propose sets of parameters.

2 Rank metric decoding problems

In this section, we show that for the same settings, rank metric decoding problems are exponentially more difficult to solve than their counterparts in Hamming metric.

2.1 Rank metric

As an ambient space we consider vectors of length n over a finite field $GF(2^m)$. Given basis $\mathfrak{B} = \{\beta_1, \ldots, \beta_m\}$ of $GF(2^m)$ regarded as a $GF(2)$–vector space, and a vector $\mathbf{x} = (x_1, \ldots, x_n) \in GF(2^m)^n$, we consider the transformation:

$$\mathbf{x} \mapsto \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$

where $(x_{11}, \ldots, x_{m1})^T$ is the binary expansion vector of x_i in the basis $\mathfrak{B}$, *i.e.*

$$x_i = \sum_{j=1}^{m} x_{ji} \beta_j .$$

The rank weight $\mathrm{Rk}(\mathbf{x})$ of vector $\mathbf{x}$ is: $\mathrm{Rk}(\mathbf{x}) \overset{def}{=} \mathrm{Rk}(\mathbf{X})$, where Rk is the usual rank of a binary matrix. Rank metric is independent of the chosen basis and in the following of the paper, we will consider that a binary basis is fixed.

2.2 Decoding problems

A rank code $\mathcal{C} \subset GF(2^m)^n$ is a set of vectors of $GF(2^m)^n$, together with the distance induced by rank metric.

In applications, it is usual to consider $\mathcal{C}$ as being an additive code. In that case, the minimum rank distance of $\mathcal{C}$ is the minimum rank weight of any non-zero codeword:

$$d_r(\mathcal{C}) \overset{def}{=} \min_{\mathbf{x} \in \mathcal{C} \setminus \{0\}} \mathrm{Rk}(\mathbf{x}).$$

If the code is $GF(2^m)$-linear of dimension k, it is called a $[n, k, d_r]$-code over $GF(2^m)$

Problem 1 (Bounded distance binary rank decoding ($BDR_2(\mathcal{C}, t, \mathbf{y})$))

- Instance:
  - *A* 2^m*-ary code* $\mathcal{C} =< \mathbf{g}_1, \ldots, \mathbf{g}_K >_{GF(2)}$,
  - *An integer* t
  - $\mathbf{y} \in GF(2^m)^n$
- Problem: *Find if it exists* $\lambda_1, \ldots, \lambda_K \in GF(2)^K$ *and* $\mathbf{e} \in GF(2^m)^n$ *of rank weight* t*, such that*

$$\mathbf{y} = \sum_{i=1}^{K} \lambda_i \mathbf{g}_i + \mathbf{e}$$

Solving $BDR_2(\mathcal{C}, t, \mathbf{y})$ is *NP*-hard. If one considers the matricial form of the problem by expanding every element of $GF(2^m)$ into a m-dimensional vector over $GF(2)$, then the associated decisional problem is an evolution of the *NP*-complete *MinRank* problem, [Cou01].

Though the complexity of this problem gives some arguments about the difficulty of decoding additive codes in rank metric, a designer will more probably consider linear codes over an extension field $GF(2^m)$. Therefore, it is more adequate to study the following decoding problem, for $GF(2^m)$-linear codes.

Problem 2 (Bounded distance 2^m-ary rank decoding ($BDR(\mathcal{C}, t, \mathbf{y})$))

- Instance:
  - *A 2^m-ary code $\mathcal{C} =< \mathbf{g}_1, \ldots, \mathbf{g}_k >_{GF(2^m)}$,*
  - *An integer t*
  - $\mathbf{y} \in GF(2^m)^n$
- Problem: *Find if it exists $\mu_1, \ldots, \mu_k \in GF(2^m)^k$ and $\mathbf{e} \in GF(2^m)^n$ of rank weight t, such that*

$$\mathbf{y} = \sum_{i=1}^{k} \mu_i \mathbf{g}_i + \mathbf{e}$$

It is not known if the decisional version of this latter problem is *NP*-complete. However when one considers the dual problem called *Rank Syndrome Decoding problem* (*RSD*) a nice result from [GZ14] establishes that if *RSD* is in *ZPP* then this would imply that *ZPP* = *NP*. The *ZPP* class is the class of decisional problems solvable by a Turing machine such that:

- The machine runs in polynomial-time of the size of the input
- Answers YES, NO or ? ;
- The answer YES or NO is the correct answer ;
- It answers ? with probability at most 1/2.

This statement backs up the feeling that decoding in rank metric is a hard problem.

2.3 Hamming metric vs rank metric

In the design of encryption schemes whose security relies on a *difficult* problem, it is worthwhile to have precise estimation of the best effective complexity of the algorithms solving the problem for randomly and uniformly chosen parameters in a given space.

Decoding in Hamming metric The complexity of decoding up to some bound a *random code* in Hamming metric is an old problem, [Pra62]. Since the seminal work by Prange a lot of work was done to improve the asymptotic complexity or the effective complexity. The most efficient algorithms are smart refinements of the so-called *Information Set Decoding* (ISD).

The basics of ISD are: Suppose one wants to decode δn errors in a k dimensional code of length n over $GF(2^m)$, where δ is less than *Varshamov-Gilbert* (GV) bound to ensure the uniqueness of the solution. Then one chooses k columns of the generator matrix of the code. If these positions are error-free, and if the $k \times k$ matrix has full rank, then decoding consists in making some linear algebra computations and check if the obtained vector has Hamming weight $\leq \delta n$. If this fails then one chooses randomly another set of k positions and proceeds as before, until it works. If any k columns of the generator matrix form a non-singular matrix (MDS code), then after

$$\frac{\binom{n}{k}}{\binom{n-\delta n}{k}}$$

attempts, the probability of success is greater than 1/2. The average complexity of ISD is

$$n^3\frac{\binom{n}{k}}{\binom{n-\delta n}{k}} \text{ operations in } GF(2^m).$$

For constant rate codes, *i.e.* $k = Rn$, provided that $0 < R < 1/2$, approximations of the Newton binomial gives a running time of: $\approx n^3 2^{n[H(R)-H(R-\delta)]}$ binary operations, where $H(R) = -R\log_2 R - (1-R)\log_2(1-R)$ is the binary entropy function. There has been many refinements of ISD. Still, the best decoding algorithms derive from ISD and have a complexity of

$$2^{c_{algo}(n+o(1))} \text{ ops. in the code alphabet,}$$

where c_{algo} is depends on the chosen algorithm, [BLP11,BJMM12,CTS16,MO15].

Decoding in rank metric The first paper giving a precise estimation of the complexity of solving $BDR(\mathcal{C}, t, \mathbf{y})$ was published in the, 90's, [CS96] and was later improved in [OJ02]. Recently a survey unifying different approaches was published [GRS16].

Provided t is less than rank metric GV bound (ensuring uniqueness of the decoding), there exists an algorithm solving $BDR(\mathcal{C}, t, \mathbf{y})$ with probability $> 1/2$ running in:

$$m^3 2^{(t-1)\lfloor (k \min(m,n))/n \rfloor} \text{ binary ops.}$$

This implies in particular that if $m \geq n$ the running time is lower bounded by

$$m^3 2^{\delta R n^2} \text{ binary ops.}$$

Compared to the $n^3 2^{n[H(R)-H(R-\delta)]}$ complexity of generic ISD for Hamming metric, for the same set of parameters, decoding in rank metric is exponentially more difficult than in Hamming metric.

In **Table 1**, we fix some decoding complexity. In the second column and third column, we give parameters for codes whose average decoding complexity is approximately equal to the corresponding decoding complexity, in Hamming metric (2nd column) and rank metric (3rd column). The subscripts correspond to the size of the field alphabet, *i.e.* 2^m, if the considered field is $GF(2^m)$. Near the parameters we write the minimum size in bytes of the necessary information sufficient to characterize the corresponding code.

The complexity evaluations in Hamming metric are for binary codes and borrowed from [CTS16]. The chosen Hamming weight is close to the GV bound. This implies that these are the best possible codes, meaning that any other code satisfying the decoding complexity and rate requirements is necessarily longer than the proposed codes. Concerning rank metric, since for $m = n$, GV bound corresponds to $n(1-\sqrt{k/n})$ [Loi14] we chose parameters relatively close to this bound to express the decoding complexity.

Dec. Complex.	Ham. Met. Gen. Mat.	Rank Met. Gen. Mat.
2^{128}	$[2400, 2006, 58]_2 \approx 100\ KBytes$	$[48, 39, 4]_{2^{48}} \approx 2.2\ KBytes$
2^{256}	$[4150, 3307, 132]_2 \approx 350\ KBytes$	$[70, 50, 5]_{2^{70}} \approx 8.7\ KBytes$

Table 1: Comparisons of the decoding complexity of codes on GV bound for Hamming metric and for rank metric $[n, k, w]_q$ correspond to a q-linear code of length n, dimension k, correcting w errors in the considered metric

Remark 1. The comparison between rank metric and Hamming metric is not fair when one considers binary codes. Namely, if one fixes the alphabet of the field, the rate of the code in Hamming metric can be kept constant when the lenght goes to infinity. In the rank metric case however this has no sense to do this. The alphabet of the code has to grow to infinity. Therefore the comparison has a sense only when the alphabet size grows.

2.4 Post-Quantum security

A study of the PQ security of solving $BDR(\mathcal{C}, t, \mathbf{y})$ was already investigated in [GHT16]. Since the results are straightforward, we recall how to evaluate this PQ security.

In [Ber10] it is shown that the use of Grover's algorithm implies that the exponential term in the decoding complexity of ISD should be *square-rooted*. On our previous estimation of the decoding complexity this gives:

$$\approx n^3 2^{\frac{n}{2}[H(R)-H(R-\delta)]}$$

The most efficient algorithm solving $BDR(\mathcal{C}, t, \mathbf{y})$ solves the equivalent dual problem *RSD*, [GRSZ14]: Given a parity-check matrix $\mathbf{H}$ of an $[n, k, d]_r$ code $\mathcal{C}$ over $GF(2^m)$, find $\mathbf{e} \in GF(2^m)^n$ of rank weight t such that

$$\mathbf{y}\mathbf{H}^T = \mathbf{e}\mathbf{H}^T. \tag{1}$$

To stress how Grover's algorithm can be employed to improve the decoding complexity we need to recall the principles of the algorithm

- $\mathbf{e} = (e_1, \ldots, e_n)$ has rank weight $t \Rightarrow$ for all i, $e_i \in \mathcal{E}$, a t-dimensional binary subspace of $GF(2^m)$;
- Let $\mathfrak{B} = (\beta_1, \ldots, \beta_{t'})$, a basis of some $\mathcal{E}'$ such that $\mathcal{E} \subset \mathcal{E}'$;
- System (1) becomes: $\mathbf{y}\mathbf{H}^T = \mathfrak{B}\mathbf{T}\mathbf{H}^T$, where $\mathbf{T}$ and $\mathcal{E}'$ are unknown ;

- W.l.o.g we can suppose $\beta_1 = 1$. Thus solving (1) consists in enumerating $t' - 1$ dimensional binary vector subspaces of $GF(2^m)$ and trying to solve the linear system of $(n-k)m$ equations and $t'n$ unknowns.

An assumption is that if the system is overdefined ($t'n \leq m(n-k)$) then the solution is unique. Therefore the average number of tries to find a suitable vector space is $2^{(t-1)\lfloor (k \min(m,n))/n \rfloor}$. Remaining linear algebra can be implemented with circuits in $O(n^3)$ size. We can apply Grover's algorithm. A lower bound estimation of the PQ complexity of solving $BDR(\mathcal{C}, t, \mathbf{y})$ is thus

$$m^3 2^{(t-1)\lfloor (k \min(m,n))/(2n) \rfloor} \text{ ops.}$$

3 Rank metric based cryptography

In code-based cryptography, the security is estimated by the decoding complexity of random codes in the considered metric. This estimation requires that the public-key must look like *a random code*. This implies that the family of codes used as private-key space cannot be distinguished from a family of *randomly constructed codes*. Given a family $\mathcal{F}$ of $[n,k,d]$ codes over a finite field $GF(2^m)$ with known decoding algorithm up to errors of rank weight t, the original and general procedure to design the pair public/private key pair for a McEliece type cryptosystem is:

1. Select *randomly* a code in $\mathcal{F}$. The code is given by generator or parity-check matrix $\mathbf{G}$ under a form enabling an efficient decoding.
2. Publish a scrambled structure of $\mathbf{G} \to \mathbf{G}_{pub}$ such that $\mathbf{G}_{pub}$ looks like random. The scrambling procedure has to be a linear isometry of the metric.

Two types of decoding algorithms are considered:

- Algebraic decoding: It is used for Goppa codes in Hamming metric [McE78]. This family is recommended for *long-term* PQ security under well chosen parameters. In rank metric, only Gabidulin codes are of this kind.
- Probabilistic decoding: MDPC or QC-MDPC in Hamming metric [MTSB12], or LRPC in rank metric, [GMRZ13].

Since our interest concerns rank metric, we present how Gabidulin codes and LRPC are used in the design of code-based encryption schemes. We explain the reason why, until now, Gabidulin codes cannot be used in the design of secure encryption schemes. We also present the idea sustaining the design of the family of LRPC since this seminal idea is a natural path which leads us to propose a new families of codes with algebraic decoding to be used in the design of rank metric codes based cryptosystems.

3.1 Algebraic decoding based cryptosystems

Gabidulin codes Let $n \leq m$ and let $\mathbf{g} = (g_1, \dots, g_n) \in GF(2^m)$, where the $g_i's$ are linearly independent over $GF(2)$. Let $[i] = 2^i$ such that $x \mapsto x^{[i]}$ is the ith power of the Frobenius automorphism $x \mapsto x^2$. The code $Gab_k(\mathbf{g})$, is the linear code with generator matrix

$$\mathbf{G} = \begin{pmatrix} g_1 & \cdots & g_n \\ g_1^{[1]} & \cdots & g_n^{[1]} \\ \vdots & \ddots & \vdots \\ g_1^{[k-1]} & \cdots & g_n^{[k-1]} \end{pmatrix}, \tag{2}$$

i.e.

$$Gab_k(\mathbf{g}) = \{\mathbf{xG} \mid \mathbf{x} \in GF(2^m)^k\}.$$

These codes can be decoded in polynomial-time for errors of rank weight up to $\lfloor (n-k)/2 \rfloor$, see [Gab85].

Invariant subspace attack From the origins, see [GPT91], numerous designs of Gabidulin codes based encryption schemes were proposed relying on the model of McEliece cryptosystem. However, all these proposals were broken by derivations of the so-called *invariant subspace attack*. The reason is the inherent structure of the family of Gabidulin codes. A detailed analysis can be found in [Ksh07,OKN16].

We present the principle of the attacks. This is essential to understand where lies the weakness and how to get rid of it. In every proposed Gabidulin codes based encryption scheme, the public-key $\mathbf{G}_{pub}$ can be rewritten under the form

$$\mathbf{G}_{pub} = \mathbf{S}(\mathbf{X} \mid \mathbf{G})\mathbf{P}, \tag{3}$$

where $\mathbf{P}$ is a binary $(n+t) \times (n+t)$ invertible matrix, $\mathbf{G}$ is a matrix generating an $[n, k, d_r]$ Gabidulin code under the form (2), and $\mathbf{S}$ is an $u \times k$-matrix with entries in $GF(2^m)$, where $u \leq k$. Now consider the action of $x \to x^{2^i} \overset{def}{=} x^{[i]}$ on the entries of $\mathbf{G}_{pub}$ denoted by $\mathbf{G}_{pub}^{[i]}$. We have

$$\mathbf{G}_{pub}^{[i]} = \mathbf{S}^{[i]}(\mathbf{X}^{[i]} \mid \mathbf{G}^{[i]})\mathbf{P}^{[i]}.$$

Since $\mathbf{P}$ is binary this implies

$$\mathbf{G}_{pub}^{[i]} = \mathbf{S}^{[i]}(\mathbf{X}^{[i]} \mid \mathbf{G}^{[i]})\mathbf{P}.$$

Let $\mathcal{C}_{pub}$ be the code generated by $\mathbf{G}_{pub}$ and let $\mathcal{C}_{pub}^{[i]}$ be the code obtained by raising the codewords of $\mathcal{C}_{pub}$ (resp. $\mathcal{C}$) to the ith power of the Frobenius automorphism. From the structure of the public code, we have

$$\dim\left(\mathcal{C}_{pub} + \cdots + \mathcal{C}_{pub}^{[i]}\right) \leq \min(n, k+i+t). \tag{4}$$

If $\mathcal{C}_{pub}$ were a random u-dimensional code one would expect the dimension of $\mathcal{C}_{pub} + \cdots + \mathcal{C}_{pub}^{[i]}$ to be equal to $\min(n, (i+1)u)$ with a high probability. Hence the previous property provides a distinguisher of the public code. Moreover, if $k+i+t = n-1$, and if the dimension is exactly equal to 1 then a polynomial-time decoder for the public code can be recovered by simple elementary linear algebra.

3.2 Probabilistic decoding based cryptosystems

Low Rank Parity-Check Codes The principle consists in

- selecting randomly a λ-dimensional vector space $\mathcal{V} \subset GF(2^m)$.
- constructing an $(n-k) \times n$ matrix $\mathbf{H} = (h_{ij})$, where $h_{ij} \in \mathcal{V}$ are randomly selected.

The private-key consists of the knowledge of $\mathbf{H}$ and the public key is $\mathbf{G}_{pub}$, the generator matrix of the code with parity-check matrix $\mathbf{H}$ under systematic form. The key idea behind the decoding procedure is: Suppose one receives a ciphertext $\mathbf{y} = \mathbf{x}\mathbf{G}_{pub} + \mathbf{e}$, where $\mathbf{G}_{pub}\mathbf{H}_{pub}^T = 0$. Then

$$\mathbf{y}\mathbf{H}^T = \underbrace{\mathbf{x}\mathbf{G}_{pub}\mathbf{H}^T}_{\mathbf{0}} + \mathbf{e}\mathbf{H}^T.$$

Since $\mathbf{e}$ has rank t its entries belong to a binary vector space $\mathcal{E}$ of dimension t. This implies that the entries of $\mathbf{y}\mathbf{H}^T$ belong to the binary vector space

$$\mathcal{E} * \mathcal{V} = \{ev \mid e \in \mathcal{E},\ v \in \mathcal{V}\}.$$

The dimension of $\mathcal{E} * \mathcal{V}$ is upper bounded by $t\lambda$. If $\dim(\mathcal{E} * \mathcal{V}) = t\lambda$ a basis for $\mathcal{E}$ can be recovered with an estimated error probability of $2^{-(n-k+1-t\lambda)}$, [GRSZ14].

- The main strength of this scheme is that the private key is randomly selected with entries in a secret λ-dimensional vector space. Thus it prevents all types of attacks attempting to use some algebraic properties to break the scheme.
- Concerning the weaknesses of the scheme, the first one is that the estimated residual error decoding probability is non-negligible for small parameters. The second main weakness comes from the fact that the decoding is probabilistic. This could induce attacks on the model of [GSJ16] consisting in guessing the secret vector space by observing the behavior of a decoder.

4 The new cryptosystem

In the case of Gabidulin codes, the strategy, which works for GRS codes, consisting of scrambling their structure by considering a subfield subcode is a dead-end. The reason is that a subfield subcode of a Gabidulin code is essentially isomorphic to the direct sum of Gabidulin codes over the subfield, [GL08].

Our approach consists in scrambling the codes via the choice of a randomly selected vector space of $GF(2^m)$ of fixed dimension. The essential idea comes from the rank multiplication property used to show that the LRPC decoding procedure works. This could also be interpreted as a rank metric equivalent of the idea in [BBC$^+$16] which, for short, replaces the permutation matrix in McEliece cryptosystem by a matrix multiplying the Hamming weight of the vectors.

Proposition 1 (Rank multiplication). *Let $\mathbf{P} \in M_n(\mathcal{V})$ be an invertible matrix with entries in a binary λ-dimensional vector space $\mathcal{V} \subset GF(2^m)$. For all $\mathbf{x} \in GF(2^m)^n$, $Rk(\mathbf{xP}) \leq \lambda Rk(\mathbf{x})$.*

Proof. Consider $\mathbf{x} = (x_1, \ldots, x_n) \in GF(q^m)$ of rank weight r. Let $\mathcal{X} =< x_1, \ldots, x_n >$ be generated by $< y_1, \ldots, y_t >$. Suppose moreover that $\mathcal{V} =< \alpha_1, \ldots, \alpha_\lambda >$, then the entries of $\mathbf{xP}$, belong to the vector space $< y_i\alpha_j >_{i,j}$ which has dimension $\leq \lambda t$.

4.1 Design of the encryption scheme

The key generation procedure is the following:

- Private key:
  - A Gabidulin code of length n over $GF(2^m)$, dimension k with generator matrix $\mathbf{G}$ under the form (2) ;
  - A non-singular $k \times k$- matrix $\mathbf{S}$ with entries in $GF(2^m)$;
  - A λ-dimensional subspace of $GF(2^m)$, denoted by $\mathcal{V}$;
  - A non-singular matrix $\mathbf{P}$ with entries in $\mathcal{V}$, *i.e.* $\mathbf{P} \in M_n(\mathcal{V})$.
- Public key: $\mathbf{G}_{pub} = \mathbf{SGP}^{-1}$. The public code $\mathcal{C}_{pub}$ is generated by $\mathbf{G}_{pub}$.

The encryption and decryption procedures are:

- Encryption of $\mathbf{x} \in GF(2^m)^k$:
  - Choose a random vector $\mathbf{e} \in GF(2^m)^n$ of rank weight $\lfloor (n-k)/(2\lambda) \rfloor$;
  - Compute $\mathbf{y} = \mathbf{xG}_{pub} + \mathbf{e}$;
  - Send the encrypted message $\mathbf{y}$ to the receiver.
- Decryption of $\mathbf{y}$:
  - Compute $\mathbf{yP} = \mathbf{xSG} + \mathbf{eP}$;
  - Since $\mathbf{P}$ has entries in $\mathcal{V}$ and from **Proposition 1** $\mathbf{eP}$ has rank weight $\leq \lambda \lfloor (n-k)/(2\lambda) \rfloor \leq \lfloor (n-k)/2 \rfloor$, and can be decoded with $\mathbf{G}$;
  - Recover $\mathbf{xS}$ and $\mathbf{eP}$ by decoding and recover $\mathbf{x}$ by multiplying with $\mathbf{S}^{-1}$.

The public-key is a randomly chosen generator matrix of the code

$$\mathcal{C}_{pub} \overset{def}{=} \mathcal{C}\mathbf{P}^{-1} = \{\mathbf{cP}^{-1} \mid \mathbf{c} \in \mathcal{C}\}.$$

A corollary of **Proposition 1** gives:

Corollary 1. *Let $\mathcal{C}$ be a $[n,k,d]_r$ code over $GF(q^m)$. Let $\mathcal{V}$ be a λ-dimensional subspace of $GF(q^m)$ seen as a $GF(q)$-vector space. And let $\mathbf{P} \in M_n(\mathcal{V})$. Then*

$$\mathcal{C}\mathbf{P}^{-1} \overset{def}{=} \{\mathbf{cP}^{-1} \mid \mathbf{c} \in \mathcal{C}\}$$

has dimension k and minimum rank distance $d' \geq \lfloor d/\lambda \rfloor$.

Proof. Since $\mathbf{P}$ is invertible $\mathcal{C}$ and $\mathcal{C}\mathbf{P}^{-1}$ have the same dimension. Concerning the minimum distance, suppose that $d' < d/\lambda$. Then let $\mathbf{c} \in \mathcal{C}\mathbf{P}^{-1} \neq \mathbf{0}$ with rank weight d'. By construction $\mathbf{cP} \in \mathcal{C}$. From proposition 1, $\text{Rk}(\mathbf{cP}) \leq d'\lambda < d$, which implies that $\mathbf{cP} = \mathbf{0}$. Thus $\mathbf{c} = \mathbf{0}$, which contradicts the hypothesis.

4.2 Security arguments

We analyze the security of the scheme.

1. The first type of attacks consists in decoding the ciphertext in the public code. We suppose that the public code cannot be distinguished from a *random* code. Therefore the complexity of recovering a plaintext from a ciphertext corresponds to the complexity of solving $BDR(\mathcal{C}_{pub}, \lambda \mathrm{Rk}(\mathbf{e}), \mathbf{y})$. From sections 2.3 and 2.4, we have:
   - Decoding complexity: $m^3 2^{(\lambda r-1)\lfloor (k \min(m,n))/n \rfloor}$ binary operations.
   - PQ-security: $m^3 2^{\frac{1}{2}(\lambda r-1)\lfloor (k \min(m,n))/n \rfloor}$ operations.
2. The question of the distinguishability of the public-code from a random code is raised. The arguments presented in section 3.1 do not work. Namely, raising the public-key to the ith power of the Frobenius gives:

   $$\mathbf{G}_{pub}^{[i]} = \mathbf{S}^{[i]}\mathbf{G}^{[i]}(\mathbf{P}^{-1})^{[i]}.$$

   Matrix $\mathbf{P}$ has entries in $\mathcal{V}$, but the entries of $\mathbf{P}^{-1}$ have no reason to belong to some strict subspace of $GF(2^m)$. Therefore (4) is not satisfied and the usual distinguisher for a Gabidulin code does not work.
   An attacker could try to recover a decoder for the public code by solving

   $$\mathbf{H}_{pub} = \mathbf{H}\mathbf{P}, \tag{5}$$

   where $\mathbf{H}_{pub}$ is a $(n-k)\times n$ parity-check matrix of the public code $\mathcal{C}_{pub}$ under systematic form, $\mathbf{H} = (h_j^{[i]})$ is a parity-check matrix of a Gabidulin code, and $\mathbf{P}$ has entries in a λ-dimensional vector space.
   W.l.o.g, we suppose that $\mathbf{H}$ is known. This hypothesis might seem very strong but if we consider the case $m = n$ this does not remove security. In that let $< h'_1, \ldots, h'_n >_2$ be a basis of $GF(2^m)$ regarded as a $GF(2)$–vector space, and let a matrix $\mathbf{H}' = ((h'_j)^{[i]})$ under the form (2). There exists a binary invertible matrix $\mathbf{M}$ such that

   $$\mathbf{H} = \mathbf{H}'\mathbf{M}.$$

   System (5) becomes $\mathbf{H}_{pub} = \mathbf{H}'\underbrace{\mathbf{M}\mathbf{P}}_{\mathbf{P}'}$. Since $\mathbf{M}$ is binary, $\mathbf{P}'$ has entries in $\mathcal{V}$. Under this setting, we investigate two ways of solving (5), which gives us a lower bound on the estimation of the complexity of recovering a decoder from the public-key.
   - System (5) is an underdefined affine system with $n \times n$ unknowns (the entries of $\mathbf{P}$) and $n(n-k)$ equations. Given a solution $\mathbf{P}_0$ of the system, an attacker has to search for a matrix in the coset $\mathbf{P}_0 + \mathcal{P}$ whose entries belong to a λ-dimensional vector space. A solution is to enumerate the coset of size $2^{m(n^2-n(n-k))}$ and if the matrices belong to a common λ-dimensional vector space.

- Another approach consists in decomposing the entries of $\mathbf{P} = (p_{ij})$ under the form $p_{ij} = \sum_{u=1}^{\lambda} \mu_{ij}^{(u)} \alpha_u$, where $\alpha_1, \ldots, \alpha_\lambda$ are candidates to be the basis of $\mathcal{V}$, and the $\mu_{ij}^{(u)}$ are binary elements. Once the equations are projected on the binary field, we obtain a system with $mn(n-k)$ equations and $m\lambda + \lambda n^2$ unknowns. If $\alpha_1, \ldots, \alpha_\lambda$ is fixed then the system is linear and overdefined ($\lambda < n \leq m$) and can thus be solved in polynomial time.

For the previous reasons, we estimate that a lower bound on the complexity of recovering a decoder corresponds to the enumeration of $\lambda - 1$-dimensional $GF(2)$-subspaces of $GF(2^m)$. The choice of $\lambda - 1$ rather than λ is justified since if $\lambda = 1$, *i.e.* $\mathcal{V} =< \alpha >$, then for some element $\alpha \in GF(2^m)$, it is obvious that an attack can be achieved in polynomial time. Namely, $\mathbf{P} = (1/\alpha)\mathbf{P}'$ with $\mathbf{P}'$ has entries in $GF(2)$. Therefore, we suppose that $1 \in \mathcal{V}$. The lower bound on the complexity is thus $2^{(\lambda-1)m-(\lambda-1)^2}$.

4.3 Choice of parameters

Table 2 proposes some parameters for an expected security, and with a ciphertext expansion between 1.6 and 1.8. Since the parameters on can consider to decrease the key-size by increasing the expansion factor, but the designer has to note that case other types of decoding attacks can occur and should be taken into account, [GRSZ14].

For a 2^{128} bits security the key-size proposed in [AB315] for a McEliece encryption scheme using Goppa codes is approximately of 1 MB. For an equivalent rate our proposal gives a public-key 20 times smaller.

Param.	Dec. Sec.	PQ Dec. Sec.	K. Rec. Sec.	Key size
$m = n = 50,\ k = 32,\ \lambda = 3,\ t = 3$	$\approx 2^{81}$	$\approx 2^{49}$	$\approx 2^{96}$	3.6 KB
$m = 96,\ n = 64,\ k = 40,\ \lambda = 3,\ t = 4$	$\approx 2^{139}$	$\approx 2^{80}$	$\approx 2^{188}$	11.5 KB
$m = 128,\ n = 120,\ k = 80,\ \lambda = 5,\ t = 4$	$\approx 2^{261}$	$\approx 2^{141}$	$\approx 2^{496}$	51 KB

Table 2: Proposition of parameters for the family of codes used in the cryptosystem

5 Acknowledgments

The author expresses deep thanks to Alain Couvreur who pointed out the existence of a straightforward distinguisher if the parameters are not carefully chosen. The proposed parameters avoid the problem.

The author also wishes to thank the reviewers who made constructive comments to improve the quality of the paper and pointed out mistakes and misses in the original submission.

6 Conclusion

We proposed a new code-based public-key cryptosystem based on the derivation of Gabidulin codes. We did not consider security reductions but presented detailed arguments why we think that our proposal makes it possible to design secure code-based encryption schemes in rank metric. Security conversion exist that take as input One-way encryption schemes and convert it into a $IND-CCA2$ in the random oracle model, [KI01,BL04]. Our proposal is versatile and can be declined for finite fields of any characteristic since Gabidulin codes have the same structure over any finite field. To evaluate the security we need to replace 2 by the cardinality of the considered base field, say q if we consider Gabidulin codes over $GF(q^m)$.

References

[AB315] *Initial recommendations of long-term secure post-quantum systems*, Tech. report, 2015, http://pqcrypto.eu.org/docs/initial-recommendations.pdf.

[BBC$^+$16] M. Baldi, M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani, *Enhanced Public Key Security for the McEliece Cryptosystem*, J. Cryptology **29** (2016), no. 1.

[Ber10] D. J. Bernstein, *Grover vs. McEliece*, Post-Quantum Cryptography 2010 (Nicolas Sendrier, ed.), Lecture Notes in Comput. Sci., vol. 6061, Springer, 2010, pp. 73–80.

[BJMM12] A. Becker, A. Joux, A. May, and A. Meurer, *Decoding random binary linear codes in $2^{n/20}$: How $1+1=0$ improves information set decoding*, Advances in Cryptology - EUROCRYPT 2012, Lecture Notes in Comput. Sci., Springer, 2012.

[BL04] T. P. Berger and P. Loidreau, *Designing an efficient and secure public-key cryptosystem based on reducible rank codes*, Progress in Cryptology - INDOCRYPT 2004, Lecture Notes in Comput. Sci., vol. 2656, 2004, pp. 360–373.

[BLP11] D. J. Bernstein, T. Lange, and C. Peters, *Smaller decoding exponents: ball-collision decoding*, Advances in Cryptology - CRYPTO 2011, Lecture Notes in Comput. Sci., vol. 6841, 2011, pp. 743–760.

[COTG15] A. Couvreur, A. Otmani, J.-P. Tillich, and V. Gauthier-Umaña, *A polynomial-time attack on the BBCRS scheme*, Public-Key Cryptography - PKC 2015 (J. Katz, ed.), Lecture Notes in Comput. Sci., vol. 9020, Springer, 2015, pp. 175–193.

[Cou01] N. Courtois, *Efficient zero-knowledge authentication based on a linear algebra problem MinRank*, Advances in Cryptology - ASIACRYPT 2001, vol. 2248, 2001, pp. 402–421.

[CS96] F. Chabaud and J. Stern, *The cryptographic security of the syndrome decoding problem for rank distance codes*, Advances in Cryptology - ASIACRYPT 1996 (Kyongju, Korea), Lecture Notes in Comput. Sci., vol. 1163, Springer, November 1996, pp. 368–381.

[CTS16] R. Canto-Torres and N. Sendrier, *Analysis of information set decoding for a sub-linear error weight*, Post-Quantum Cryptography 2016 (Fukuoka, Japan), Lecture Notes in Comput. Sci., February 2016, pp. 144–161.

[Gab85] E. M. Gabidulin, *Theory of codes with maximum rank distance*, Probl. Inf. Transm. **21** (1985), no. 1, 3–16.

[Gab08] ———, *Attacks and counter-attacks on the GPT public key cryptosystem*, Des. Codes Cryptogr. **48** (2008), no. 2, 171–177.

[GHT16] P. Gaborit, A. Hauteville, and J.-P. Tillich, *Ranksynd a PRNG based on rank metric*, Post-Quantum Cryptography 2016, February 2016, pp. 18–28.

[Gib95] K. Gibson, *Severely denting the Gabidulin version of the McEliece public key cryptosystem*, Des. Codes Cryptogr. **6** (1995), no. 1, 37–45.

[Gib96] ———, *The security of the Gabidulin public key cryptosystem*, Advances in Cryptology - EUROCRYPT '96 (Ueli Maurer, ed.), Lecture Notes in Comput. Sci., vol. 1070, Springer, 1996, pp. 212–223.

[GL08] E. M. Gabidulin and P. Loidreau, *Properties of subspace subcodes of Gabidulin codes*, Adv. in Math. of Comm. **2** (2008), no. 2, 147–157.

[GMRZ13] P. Gaborit, G. Murat, O. Ruatta, and G. Zémor, *Low rank parity check codes and their application to cryptography*, Proceedings of the Workshop on Coding and Cryptography WCC'2013 (Bergen, Norway), 2013, Available on `www.selmer.uib.no/WCC2013/pdfs/Gaborit.pdf`.

[GPT91] E. M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov, *Ideals over a non-commutative ring and their applications to cryptography*, Advances in Cryptology - EUROCRYPT'91 (Brighton), Lecture Notes in Comput. Sci., no. 547, April 1991, pp. 482–489.

[GRH09] E. Gabidulin, H. Rashwan, and B. Honary, *On improving security of GPT cryptosystems*, Proc. IEEE Int. Symposium Inf. Theory - ISIT, 2009, pp. 1110–1114.

[GRS16] P. Gaborit, O. Ruatta, and J. Schrek, *On the complexity of the rank syndrome decoding problem*, IEEE Trans. Information Theory **62** (2016), no. 2, 1006–1019.

[GRSZ14] P. Gaborit, O. Ruatta, J. Schrek, and G. Zémor, *New results for rank-based cryptography*, Progress in Cryptology - AFRICACRYPT 2014 - 7th International Conference on Cryptology in Africa, Marrakesh, Morocco, May 28-30, 2014. Proceedings, 2014, pp. 1–12.

[GSJ16] Q. Guo, P. Stankovski, and T. Johannson, *A key recovery attak on MDPC with CCA security using decoding errors*, Advances in Cryptology - ASIACRYPT 2016, 2016.

[GZ14] P. Gaborit and G. Zémor, *On the hardness of the decoding and the minimum distance problems for rank codes*, CoRR **abs/1404.3482** (2014).

[KI01] K. Kobara and H. Imai, *Semantically secure McEliece public-key cryptosystems-conversions for McEliece PKC*, Public-Key Cryptography - PKC 2001 (Cheju Island, Korea) (Kwangjo Kim, ed.), Lecture Notes in Comput. Sci., vol. 1992, Springer, February 2001, pp. 19–35.

[Ksh07] A. Kshevetskiy, *Security of GPT-like public-key cryptosystems based on linear rank codes*, 3rd International Workshop on Signal Design and Its Applications in Communications, IWSDA 2007, 2007.

[Loi14] P. Loidreau, *Asymptotic behaviour of codes in rank metric over finite fields*, Des. Codes Cryptography **71** (2014), no. 1, 105–118.

[McE78] R. J. McEliece, *A public-key system based on algebraic coding theory*, pp. 114–116, Jet Propulsion Lab, 1978, DSN Progress Report 44.

[MO15] A. May and I. Ozerov, *On Computing Nearest Neighbors with Applications to Decoding of Binary Linear Codes*, Advances in Cryptology - EUROCRYPT 2015 (E. Oswald and M. Fischlin, eds.), vol. 9056, 2015, pp. 203–228.

[MTSB12] R. Misoczki, J.-P. Tillich, N. Sendrier, and P. S. L. M. Barreto, *MDPC-McEliece: New McEliece variants from moderate density parity-check codes*, IACR Cryptology ePrint Archive, Report2012/409 (2012).

[OJ02] A. V. Ourivski and T. Johansson, *New technique for decoding codes in the rank metric and its cryptography applications*, Problems of Information Transmission **38** (2002), no. 3, 237–246 (English).

[OKN16] A. Otmani, H. T. Kalashi, and S. Ndjeya, *Improved cryptanalysis of rank metric schemes based on Gabidulin codes*, http://arxiv.org/abs/1602.08549v1, 2016.

[Ove05] R. Overbeck, *A new structural attack for GPT and variants*, Mycrypt, Lecture Notes in Comput. Sci., vol. 3715, 2005, pp. 50–63.

[Pra62] E. Prange, *The use of information sets in decoding cyclic codes*, IRE Transactions on Information Theory **8** (1962), no. 5, 5–9.

[RGH10] H. Rashwan, E. M. Gabidulin, and B. Honary, *A smart approach for GPT cryptosystem based on rank codes*, Proc. IEEE Int. Symposium Inf. Theory - ISIT, 2010, pp. 2463–2467.