The GPT Cryptosystem

Siddhant Kar

Supervisor: Dr. Anirban Ghatak

Abstract

The paradigm of code-based cryptography (CBC) is of growing significance in the scenario of post-quantum cryptography. Initial CBC proposals, like the McEliece cryptosystem, were based on the Hamming metric. In this report, we discuss the evolution of a rank-metric CBC system, the Gabidulin-Paramonov-Tretjakov (GPT) cryptosystem. We begin with an outline of the generalized version of the GPT system and then present the details of a structural attack by R. Overbeck. The final sections discuss proposed modifications to resist Overbeck's attack and conclude with the present status of the GPT-based proposals.

Contents

Chapter 1

Introduction

The theory of error-correcting codes (cf. [1]) is a well-studied discipline that involves, among other techniques, devising algorithms which embed vector spaces, consisting of information-bearing vectors, into larger dimensional vector spaces. In conjunction, there should exist efficient inverse algorithms capable of recovering each original vector from a possibly perturbed - up to a reasonable level - version of the forward output. The first operation is termed *encoding*, and the inverse operation - *decoding*.

Several computationally hard problems have been defined in the context of decoding known error-correcting codes (cf. [2], [3]), some of them have reductions to other known hard problems. Current research in post-quantum public-key cryptography has witnessed several interesting attempts to formulate new cryptographic primitives based on these problems. Of these, the earliest, and of enduring significance, is the proposal by R. J. McEliece [4] based on Goppa codes. A parallel line of research is based on the so-called rank-metric codes (cf. [5]), and our article attempts to give a brief evolutionary overview of the code-based cryptosystem proposed by Gabidulin, Paramonov and Tretjakov [8].

The organization of this article is as follows.

We begin with an outline of the McEliece cryptosystem, both as an overview of how a code-based cryptosystem works and as a blueprint for the rank-metric based proposal. Chapter 3 introduces the rank metric along with a description of Gabidulin codes and discusses the GPT cryptosystem. We briefly mention the advantage of the rank-metric based proposal over the original McEliece cryptosystem in terms of key sizes involved. Finally, in chapter 4, we present an evolutionary discussion on the general GPT proposal. Specifically, we give the details of a structural attack by R. Overbeck which breaks the general GPT and many of its variants in polynomial time. We conclude with a discussion on the current status of the GGPT proposal.

Chapter 2

The McEliece Cryptosystem

The most well-known cryptosystems in code-based cryptography are the McEliece cryptosystem [4] and its variants. It is based on Goppa codes which are Hamming metric codes belonging to the class of generalized Reed-Solomon codes. We go over certain properties of these codes and briefly describe the McEliece cryptosystem.

2.1 Hamming Metric Codes

Hamming metric codes are linear codes over a finite field $\mathbb{F}$. An $[n, k]$ linear code over $\mathbb{F}$ is a k dimensional subspace of $\mathbb{F}^n$, where $n \in \mathbb{N}$. Any linear code can be written as $\mathcal{C} = row(G) = \{\mathbf{c} = \mathbf{m}G : \mathbf{m} \in \mathbb{F}^k\}$, where $G \in F^{k \times n}$ is the generator matrix of full rank. Over Hamming metric, we have the following.

Definition 2.1.1. *Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ in a code $\mathcal{C}$. The Hamming weight of x is defined as $|\{i : x_i \neq 0\}|$, and the Hamming distance between x and y is defined as the Hamming weight of $x - y$, $|\{i : x_i \neq y_i\}|$.*

Theorem 2.1.1 (Singleton Bound)**.** *In an $[n, k]$ Hamming metric code $\mathcal{C}$ with minimum distance d between any two distinct vectors, we have*

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

Proof. The distance between any two vectors is the weight of their difference and hence, d is also the minimum weight of a non-zero vector in $\mathcal{C}$. We have $row(G) = row(rref(G))$, where $rref(G)$ is the row-reduced echelon form of G. This means the codeword $c = (1, 0, ..., 0)rref(G)$ is simply the first row of $rref(G)$ and hence has at least $k - 1$ zeroes. Thus, the minimum weight cannot exceed $n - k + 1$. □

Decoding: A codeword $\mathbf{c} = \mathbf{m}G$ for a message $\mathbf{m}$ may develop an error $\mathbf{e}$ of weight r and change to $\mathbf{c} + \mathbf{e}$. Such errors can be corrected using efficient decoding algorithms to give back $\mathbf{m}$, if $r \leq \lfloor \frac{d-1}{2} \rfloor$. This bound ensures that an erroneous codeword $\mathbf{c} + \mathbf{e}$ uniquely corresponds to $\mathbf{c}$.

2.2 The McEliece Cryptosystem

The McEliece cryptosystem uses a particular instance of Hamming metric codes, namely Goppa codes. The system has proved to be secure against various structural attacks, unlike systems based on Reed-Solomon codes (cf. for instance, [6]). Successful attacks against the McEliece system and its variants are, therefore, generic attacks based on fine-tuning techniques like information-set decoding (see [7]).

Definition 2.2.1. *Let q be a prime and $m \in \mathbb{N}$. Define $\nu = (\nu_1, ..., \nu_n)$, where $\nu_i \in \mathbb{F}_{q^m}^*, i = 1, ..., n$ and $\alpha = (\alpha_1, ..., \alpha_n)$, where $\alpha_i \in \mathbb{F}_{q^m}, i = 1, ..., n$ are distinct. Then the $[n, k]$ generalized Reed-Solomon code is defined as*

$$\mathrm{GRS}_k(\alpha, \nu) = \{(\nu_1 f(\alpha_1), ..., \nu_n f(\alpha_n)) : f(z) \in \mathbb{F}_{q^m}[z], \deg(f(z)) \leq k - 1\}.$$

An $[n, k, d]$ GRS code satisfies $d = n - k + 1$. Goppa codes are restrictions of GRS codes over $\mathbb{F}_{q^m}$ to the base field $\mathbb{F}_q$.

Definition 2.2.2. *Let q be a prime and $m \in \mathbb{N}$. Define a Goppa polynomial $G(z) \in \mathbb{F}_{q^m}[z]$ and a support set $L = \{\alpha_1, ..., \alpha_n\}$, where $\alpha_i, i = 1, ..., n \in \mathbb{F}_{q^m}$ such that $G(\alpha_i) \neq 0$ for all i. Further, for $\mathbf{a} = (a_1, ..., a_n) \in \mathbb{F}_{q^m}^n$, define the corresponding rational function*

$$R_{\mathbf{a}}(z) = \sum_{i=1}^{n} \frac{a_i}{z - \alpha_i}.$$

Then the Goppa code $\Gamma(L, G)$ is defined as

$$\Gamma(L, G) = \{\mathbf{a} \in \mathbb{F}_{q^m}^n : R_{\mathbf{a}}(z) \equiv 0 \bmod G(z)\}.$$

2.2.1 Key Generation

We generate a $k \times n$ matrix of the form $G_{pub} = SGT$ where $S \in \mathbb{F}^{k \times k}$, $T \in \mathbb{F}^{n \times n}$ are invertible and G is the generator of an $[n, k, d]$ Goppa code $\mathcal{C}$. We then randomly choose $r \in \mathbb{N}$ such that $r \leq \lfloor \frac{d-1}{2} \rfloor$. The public key is the tuple (G_{pub}, r) and the private key is (S, G, T).

2.2.2 Encryption

Our plaintext $\mathbf{p}$ is from $\mathbb{F}^k$. We randomly choose an error $\mathbf{z} \in \mathbb{F}^n$ of weight r, so that decoding is ensured. The ciphertext is given by

$$\mathbf{c} = \mathbf{p}G_{pub} + \mathbf{z}.$$

2.2.3 Decryption

To decrypt $\mathbf{c}$, we multiply it by T^{-1} to get

$$\mathbf{c}T^{-1} = \mathbf{c}SG + \mathbf{z}T^{-1}.$$

As the weight of the error $\mathbf{z}T^{-1}$ is unchanged, we then apply a decoding algorithm for $\mathcal{C}$ to obtain $\mathbf{c}S$. Multiplying the decoder output by S^{-1} returns $\mathbf{c}$.

Remarks:
The original McEliece cryptosystem based on Goppa codes was broken by Bernstein *et al.* [7], with an optimized generic decoding attack. Most of the Hamming metric McEliece cryptosystems are resistant to structural attacks, but the large size of the public key have prevented them from being adopted in practice.

Chapter 3

The GPT Cryptosystem

3.1 Rank Metric Codes

In this chapter, we introduce rank metric codes and as an instantiation, describe the so-called Gabidulin codes [5]. Rank metric codes are linear codes over a finite field $\mathbb{F}_{q^m}$, where q is (power of) a prime and $m \in \mathbb{N}$. We start by defining rank distance.

3.1.1 Rank Metric

Definition 3.1.1. *Let $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{F}_{q^m}^n$ and $\mathbf{b} = (b_1, ..., b_m)$ be a basis of the vector space $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. Each x_i can be written as $x_i = \sum_{j=1}^{m} x_{ij} b_j$ with $x_{ij} \in \mathbb{F}_q$. The rank weight of a vector $\mathbf{x} \in \mathbb{F}_{q^m}^n$ is defined as:*

$$|\mathbf{x}| = rank((x_{ij})_{1 \leq i \leq n, 1 \leq j \leq m} \mid \mathbb{F}_q).$$

For two vectors $\mathbf{x} \in \mathbb{F}_{q^m}^n$ and $\mathbf{y} \in \mathbb{F}_{q^m}^n$, the rank distance between x and y is defined as:

$$|\mathbf{x} - \mathbf{y}| = rank((x_{ij} - y_{ij})_{1 \leq i \leq n, 1 \leq j \leq m} \mid \mathbb{F}_q)$$

where x_{ij} and y_{ij} are defined with respect to a fixed basis $\mathbf{b}$.

Thus, the rank distance is nothing but the rank of the difference of two vectors over the base field $\mathbb{F}_q$. Let X denote the matrix $(x_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ for any vector $\mathbf{x} \in \mathbb{F}_{q^m}^n$. Now, we prove that rank distance is a metric on the set of vectors $\mathbb{F}_{q^m}^n$.

Definition 3.1.2. *A metric on a set S is a function*

$$d : S \times S \to [0, \infty)$$

that satisfies the following for all $x, y, z \in S$:
1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) \leq d(x, y) + d(y, z)$

We will need the following result.

Lemma 3.1.1. *Given two subspaces S, T of a finite dimensional vector space V, we have*

$$dim(S) + dim(T) = dim(S + T) + dim(S \cap T)$$

where $S + T$ is the subspace $\{s + t : s \in S, t \in T\}$.

Theorem 3.1.1. *Rank distance is a metric on the set of vectors $\mathbb{F}_{q^m}^n$.*

Proof. It is easy to see that rank distance is a function from $\mathbb{F}_{q^m}^n \times \mathbb{F}_{q^m}^n$ to $[0, \infty)$. (1) holds since rank of a matrix is always non-negative. (3) is again obvious since rank of $X - Y$ is the same as that of $Y - X$.

Now, for two vectors $\mathbf{x}$ and $\mathbf{y}$, if rank of the matrix $X - Y$ is zero, then $X - Y$ itself must be zero and hence, $\mathbf{x} = \mathbf{y}$. Conversely, if $\mathbf{x} = \mathbf{y}$ for two vectors $\mathbf{x}$ and $\mathbf{y}$, then the matrix X-Y is zero and hence, its rank is zero. This proves (2).

To prove (4), consider $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{F}_{q^m}$ and corresponding matrices X, Y, Z over $\mathbb{F}_q$. In Lemma 3.1.1, let the subspaces, in $\mathbb{F}_{q^m}^n$, be $S = row(X - Y)$ and $T = row(Y - Z)$, where $row(M)$ denotes the linear span of the rows of M. S and T have dimensions $rank(X - Y)$ and $rank(Y - Z)$ respectively. Now, $S + T$ is nothing but $row(X - Z)$ and thus has dimension $rank(X - Z)$. It follows that $rank(X - Z) \leq rank(X - Y) + rank(Y - Z)$. □

3.1.2 Gabidulin Codes

An $[n, k]$ rank distance code $\mathcal{C}$ is simply a k dimensional subspace of the set of vectors $\mathbb{F}_{q^m}^n$. It can thus be generated by some $k \times n$ matrix G over $\mathbb{F}_{q^m}$. Since rank distance is a metric on $\mathcal{C}$, we can define the minimum rank distance (or weight equivalently) between any two distinct vectors in $\mathcal{C}$:

$$d = min\{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in \mathcal{C} \subseteq \mathbb{F}_{q^m}, \mathbf{x} \neq \mathbf{y}\}.$$

Theorem 3.1.2 (Singleton bound)**.** *For an $[n, k, d]$ rank distance code $\mathcal{C}$ generated by some matrix G, we have*

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

Proof. Any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}_{q^m}^n$ are also related by their Hamming distance, which is the number of positions at which $\mathbf{x}$ and $\mathbf{y}$ differ:

$$d_H = |\{i : x_i \neq y_i\}|.$$

So, any position i which does not contribute to the Hamming distance, i.e., for which $x_i = y_i$, must also not contribute to the rank distance of $\mathbf{x}$ and $\mathbf{y}$, since the ith column vanishes in the matrix $X - Y$. Thus, the rank distance between $\mathbf{x}$ and $\mathbf{y}$ must be smaller than their Hamming distance. The result follows from the Singleton bound for Hamming distance codes. □

Definition 3.1.3. *An $[n, k, d]$ rank distance code $\mathcal{C}$ is called a maximum rank distance (MRD) code if $d = n - k + 1$.*

We now define Gabidulin codes, which are a class of rank distance codes.

Definition 3.1.4. *Let $\mathbf{g} = (g_1, ..., g_n) \in \mathbb{F}_{q^m}^n$ be a vector such that its components $g_i, i = 1, ..., n$ are linearly independent over $\mathbb{F}_q$. Let*

$$G = \begin{pmatrix} g_1 & g_2 & \dots & g_n \\ g_1^q & g_2^q & \dots & g_n^q \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{q^{k-1}} & g_2^{q^{k-1}} & \dots & g_n^{q^{k-1}} \end{pmatrix} \in \mathbb{F}_{q^m}^{k \times n}.$$

The set $\mathcal{G} = \{\mathbf{m}G : \mathbf{m} \in \mathbb{F}_{q^m}^k\}$ is called the $[n, k]$ Gabidulin code with generator matrix G. The vector $\mathbf{g}$ is called the generator vector of $\mathcal{G}$.

We skip the proof of the following theorems.

Theorem 3.1.3. *Let $\mathcal{G}$ be an $[n, k, d]$ Gabidulin code defined as above. Then*
1. $\mathcal{G}$ is a k dimensional subspace of $\mathbb{F}_{q^m}^n$.
2. $\mathcal{G}$ is an MRD code, i.e., $d = n - k + 1$.

Theorem 3.1.4 (Check matrix)**.** *An $[n, k, d]$ Gabidulin code $\mathcal{G}$ can also be defined in terms of a full rank $n \times n - k$ matrix H: $\mathcal{G} = \{\mathbf{x} \in \mathbb{F}_{q^m}^n : \mathbf{x}H = 0\}$. The matrix H is called the check matrix of $\mathcal{G}$ and satisfies $GH = 0$, where G is the generator matrix of $\mathcal{G}$. Further, $H^\top$ has full column rank n over $\mathbb{F}_q$ and satisfies*

$$H^\top = \begin{pmatrix} h_1 & h_2 & \dots & h_n \\ h_1^q & h_2^q & \dots & h_n^q \\ \vdots & \vdots & \ddots & \vdots \\ h_1^{q^{n-k-1}} & h_2^{q^{n-k-1}} & \dots & h_n^{q^{n-k-1}} \end{pmatrix} \in \mathbb{F}_{q^m}^{(n-k) \times n}$$

where $h_i, i = 1, ..., n$ are linearly independent over $\mathbb{F}_q$.

Decoding Gabidulin codes: Any code with minimum distance d can correct errors of weight up to $\lfloor \frac{d-1}{2} \rfloor$. In case of Gabidulin codes, $d = n - k + 1$. A message can thus be recovered from its erroneous codeword only if weight of the error is less than or equal to $\lfloor \frac{n-k}{2} \rfloor$. The errors can be corrected using an efficient decoding algorithm $\mathcal{D}_{\mathcal{G}}$.

3.2 The GPT Cryptosystem

The original GPT cryptosystem was proposed by Gabidulin, Paramonov and Tretjakov in 1991 [8]. It is essentially similar to the McEliece cryptosystem. In what follows, we describe a generalized version of GPT (GGPT), which was proposed in response to initial attacks on the original.

3.2.1 Key Generation

System Parameters: q, $k < n \leq m$, $t < n - k - 1$ and $s \leq min\{t, k\} \in \mathbb{N}$.

Generate the following matrices:

- $G \in \mathbb{F}_{q^m}^{k \times n}$: Generator matrix of an $[n, k, d]$ Gabidulin code,
- $X \in \mathbb{F}_{q^m}^{k \times t}$: Random, rank s over $\mathbb{F}_{q^m}$ and t over $\mathbb{F}_q$ (Distortion matrix),
- $S \in \mathbb{F}_{q^m}^{k \times k}$: Random, invertible (Row scrambler) and
- $T \in \mathbb{F}_q^{n \times n}$: Random, invertible (Column scrambler).

Finally, generate the $k \times n$ matrix

$$G_{pub} = S(\begin{bmatrix} X & 0 \end{bmatrix} + G)T = S[X + G_{.\{1,\ldots,t\}} | G_{.\{t+1,\ldots,n\}}]T$$

and choose $r \in \mathbb{N}$ such that $r \leq \lfloor \frac{n-t-k}{2} \rfloor$. This bound ensures that the ciphertext is decodable. Further, let $\mathcal{D}_{\mathcal{G}}$ be an efficient decoding algorithm for the Gabidulin code generated by $G_{.\{t+1,\ldots,n\}}$.

Public Key: (G_{pub}, r).
Private Key: $(\mathcal{D}_{\mathcal{G}}, S, T)$ or (G, S, T).

3.2.2 Encryption

To encrypt a plaintext $\mathbf{p} \in \mathbb{F}_{q^m}^k$, choose a random error $\mathbf{z} \in \mathbb{F}_{q^m}^n$ of rank weight r. The ciphertext $\mathbf{c}$ is given by

$$\mathbf{c} = \mathbf{p}G_{pub} + \mathbf{z}.$$

3.2.3 Decryption

To decrypt the ciphertext $\mathbf{c}$, we apply the decoding algorithm $\mathcal{D}_{\mathcal{G}}$ to the last $n-t$ columns of $\mathbf{c}T^{-1}$, where

$$\mathbf{c}T^{-1} = \mathbf{p}G_{pub}T^{-1} + \mathbf{z}T^{-1} = \mathbf{p}S\left[X + G_{.\{1,\ldots,t\}} \quad G_{.\{t+1,\ldots,n\}}\right] + \mathbf{z}'.$$

We apply the decoding algorithm $\mathcal{D}_{\mathcal{G}}$ only to

$$\mathbf{c}T^{-1}_{.\{t+1,\ldots,n\}} = \mathbf{p}SG_{.\{t+1,\ldots,n\}} + \mathbf{z}'_{.\{t+1,\ldots,n\}}.$$

The error $\mathbf{z}' = \mathbf{z}T^{-1}$ still has rank $r \leq \lfloor \frac{n-t-k}{2} \rfloor$ since T is invertible and hence, it can be corrected. Thus, we can get $\mathbf{p}S$ by applying the algorithm, and then simply multiply by S^{-1} to get the plaintext $\mathbf{p}$ back.

3.3 Comparison

It is evident that the GPT cryposystem can be interpreted as a *rank-metric Mc Eliece cryptosystem* based on Gabidulin codes. The main advantage of the GPT cryptosystem and its variants over the original McEliece cryptosystem, which is based on the Hamming metric, is the small size of the public key. For the same security level, the public key size for the former is roughly ten times smaller than the latter and is more useful for practical purposes. The table shows the public key sizes in both systems, with their work factors. We refer to the McEliece system presented in [11].

	q	m	n	t	s	$\lfloor \frac{d-1}{2} \rfloor$	Pk size (kBits)	Wf
McEliece	2	-	1024	-	-	41	614	2^{66}
GPT	2	46	52	6	2	9	57.8	2^{66}

However, most of the rank-metric based cryptosystems have been proven structurally insecure, whereas a few variants of the McEliece cryptosystem have still not been broken. This disadvantage may be explained in terms of the more "structured" description of the Gabidulin code as compared to the Goppa code or its variants. In chapter 4 we discuss how this inherent structure of the Gabidulin code has been exploited to define a distinguisher and hence formulate a plaintext-recovery attack on the GPT and its variants.

Chapter 4

Overbeck's Attack

In this chapter, we present a structural attack on the GGPT by R. Overbeck [9, 10]. We begin with the formulation of a distinguisher for the generator matrix of a Gabidulin code and prove that it is possible to construct an equivalent set of column and row scrambler matrices from the public key. Next we outline a proposed repair which changes the base field of the column scrambler and resists the previously described attack. We conclude by briefly discussing the current status of the GGPT proposal in the context of the recent attacks on it.

4.1 A Distinguisher

Definition 4.1.1. *Let $M = (m_{ij})_{1\leq i\leq l, 1\leq j\leq n}$ be an arbitrary $l \times n$ matrix over $\mathbb{F}_{q^m}$ and $f \in \mathbb{N}$. Then the operator Λ_f is defined as*

$$\Lambda_f : \mathbb{F}_{q^m}^{l\times n} \to \mathbb{F}_{q^m}^{((f+1)\cdot l)\times n}$$

$$\Lambda_f(M) = \begin{bmatrix} M \\ (M)^{[1]} \\ \vdots \\ (M)^{[f]} \end{bmatrix}$$

where $(M)^{[f]}$ denotes the Frobenius map $(m_{ij}^{q^f})_{1\leq i\leq l, 1\leq j\leq n}$ for $f \in \mathbb{N}$.

Lemma 4.1.1. *If $M \in \mathbb{F}_{q^m}^{k\times n}$ defines an $[n, k]$ Gabidulin code with generator vector $\mathbf{g}$ and $f \leq n - k - 1$, then the subspace formed by the rows of $\Lambda_f(M)$ is the $[n, k + f]$ Gabidulin code with generator $\mathbf{g}$.*

Proof. Let G be the canonical generator matrix of $\mathcal{G}$. Since both M and G span $\mathcal{G}$, the rows of M lie in the span of G. Hence, $M = SG$ for some

$S \in \mathbb{F}_{q^m}^{k\times k}$. It follows that $row(\Lambda_f(M)) = row(\Lambda_f(G))$. Now, consider $\Lambda_f(G)$. It is easy to see that the only row in $\Lambda_f(G)$ which is not spanned by $\Lambda_{f-1}(G)$ is the last row $\mathbf{g}^{p^{k-1}p^f} = \mathbf{g}^{p^{k+f-1}}$. Since $\Lambda_1(G)$ has rank k, it follows that $\Lambda_f(G)$ has rank $k+f$ and defines an $[n, k+f]$ Gabidulin code. □

Lemma 4.1.2. *Let $M \in \mathbb{F}_{q^m}^{l\times n}$ be a random matrix of full column rank over $\mathbb{F}_q$. Then $\Lambda_f(M)$ has rank $min(n, (f+1)\cdot l)$ i.e. full rank with probability $\geq 1 - 4q^{-m}$.*

Observe that when the field is large, this probability is very close to 1 for a random M, whereas it is zero when M defines a Gabidulin code. Thus, λ_f serves as a distinguisher between a generator matrix and a random matrix.

4.2 Attacking GGPT

We saw that depending on the connection between M and a Gabidulin code $\mathcal{G}$, the matrix $\Lambda_f(M)$ defines different subspaces of $\mathbb{F}_{q^m}^n$. The main idea is to analyze $\Lambda_f(G_{pub})$ and separate out the Gabidulin part from G_{pub}.

Definition 4.2.1. *For a $k\times n$ matrix M, we define the dual matrix $M^\perp$ of M as the generator of $row(M)^\perp$. By abuse of notation, we may call $row(M)^\perp$ the dual space of M.*

Theorem 4.2.1. *Let $2r = n - t - k$ in a GGPT cryptosystem. For $0 \leq f \leq 2r-1$ there exists a dual matrix of $\Lambda_f(G_{pub})$ of the form*

$$\Lambda_f(G_{pub})^\perp = \begin{bmatrix} 0 & H_f^\top \\ B_1 & B_2 \end{bmatrix} \cdot (T^{-1})^\top \in \mathbb{F}_{q^m}^{(2r-f+l)\times n}$$

where $H_f \in \mathbb{F}_{q^m}^{(n-t)\times(2r-f)}$ is the check matrix of a $k+f$ dimensional Gabidulin code $\mathcal{G}_f$ of length $n-t$, B_1 is some $l\times t$ matrix with $0 \leq l \leq t$ and B_2 is some $l \times (n-t)$ matrix.

Proof. We can write

$$\Lambda_f(G_{pub})T^{-1} = \Lambda_f(S(\begin{bmatrix} X & 0 \end{bmatrix} + G)T)T^{-1}.$$

Observe that since T is over $\mathbb{F}_q$, $T^{[f]} = T$ for any $f \in \mathbb{N}$. Hence, it follows that

$$\begin{aligned} \implies \quad \Lambda_f(G_{pub})T^{-1} &= \Lambda_f(S(\begin{bmatrix} X & 0 \end{bmatrix} + G))TT^{-1} \\ \implies \quad \Lambda_f(G_{pub})T^{-1} &= \begin{bmatrix} \Lambda_f(SX + SG_{.\{1,\ldots,t\}}) & \Lambda_f(SG_{.\{t+1,\ldots,n\}}) \end{bmatrix}. \end{aligned}$$

By Lemma 4.1.1, the last $n-t$ columns of $\Lambda_f(G_{pub})T^{-1}$ define an $[n-t, k+f]$ Gabidulin code $\mathcal{G}_f$. The $\mathcal{G}_f$ check matrix $H_f \in \mathbb{F}_{q^m}^{(n-t)\times(n-t-k-f)}$ has full rank $n-t-k-f = 2r-f$ and satisfies $GH_f = 0$ where G is the generator matrix of $\mathcal{G}_f$. This implies that

$$\begin{aligned} \implies \quad & (\Lambda_f(G_{pub})T^{-1})_{.\{t+1,\ldots,n\}}(H_f^\top)^\top &&= 0 \\ \implies \quad & \Lambda_f(G_{pub})T^{-1}\begin{bmatrix}0 & H_f^\top\end{bmatrix}^\top &&= 0 \\ \implies \quad & \Lambda_f(G_{pub})(\begin{bmatrix}0 & H_f^\top\end{bmatrix}(T^{-1})^\top)^\top &&= 0. \end{aligned}$$

Hence, the subspace spanned by the $(2r-f)\times n$ matrix $\begin{bmatrix}0 & H_f^\top\end{bmatrix}(T^{-1})^\top$ lies in the dual space of $\Lambda_f(G_{pub})$.

Now, $\Lambda_f(G_{pub})T^{-1}$ has rank $\geq k+f$ due to the columns $\Lambda_f(SG_{.\{t+1,\ldots,n\}})$. Since T is invertible, $\Lambda_f(G_{pub})$ has rank $\geq k+f$ as well. Hence, by the rank-nullity theorem, its dual has rank $\leq n-k-f = 2r-f+t$ and since $\begin{bmatrix}0 & H_f^\top\end{bmatrix}(T^{-1})^\top$ already has rank $2r-f$, we need to add at most t more rows to span the whole dual space. Thus, the dual matrix of $\Lambda_f(G_{pub})$ has the given form, and B_1 and B_2 must have rank $\leq t$. □

The consequences of the above theorem are as follows.

- Repeated application of the Frobenius map on G_{pub} results in a matrix of $\mathbb{F}_{q^m}$-rank $\geq 2r-f$. Moreover, for the value of f where the rank of $\Lambda_f(G_{pub})$ exactly matches the lower bound, we have

$$\begin{bmatrix}0 & H_f^\top\end{bmatrix}(T^{-1})^\top = \Lambda_f(G_{pub})^\perp.$$

- It follows that a valid column scrambler T^{-1} will have an invertible submatrix which corresponds to the structure of $\begin{bmatrix}0 & H_f^\top\end{bmatrix}$.

These considerations lead to the formulation of an alternative column scrambler by simply working with the public key as shown in the next theorem.

Theorem 4.2.2. *If there exists an $f \leq 2r-1$ such that the rank of $\Lambda_f(G_{pub})^\perp$ is exactly $2r-f$, then the rank of $\Lambda_f(G_{pub})^\perp$ over $\mathbb{F}_q$ is $n-t$. Further, every invertible matrix $\widehat{T} \in \mathbb{F}_q^{n\times n}$ which maps $\Lambda_f(G_{pub})^\perp$ to a matrix whose first t columns are zero is a column scrambler for a valid private key.*

Proof. For such an f, rank of $\Lambda_f(G_{pub})^\perp$ and $H_f^\top$ are equal and hence we have

$$\Lambda_f(G_{pub})^\perp = \begin{bmatrix}0 & H_f^\top\end{bmatrix}(T^{-1})^\top.$$

Also, by Theorem 3.1.4, the column rank of $H_f^\top$ over $\mathbb{F}_q$ is $n-t$. Since T is invertible, the matrix $\Lambda_f(G_{pub})^\perp$ has $\mathbb{F}_q$-rank $n-t$ as well.

We can thus choose a set N_1 of $n-t$ columns such that $\Lambda_f(G_{pub})^{\perp}_{.N_1}$ is of $\mathbb{F}_q$-rank $n-t$. Further, let N_2 denote the set of rows $\{t+1, ..., n\}$. We have

$$\Lambda_f(G_{pub})^{\perp}_{.N_1} = H_f^{\top}(T^{-1})^{\top}_{N_2N_1}.$$

Since both $\Lambda_f(G_{pub})^{\perp}_{.N_1}$ and $H_f^{\top}$ have full $\mathbb{F}_q$-rank, the submatrix $(T^{-1})^{\top}_{N_2N_1} = T^{-1}_{N_1N_2}$ does as well and is invertible. We may assume, without loss of generality, that $N_1 = \{t+1, ..., n\} = N_2$ and $T^{-1}_{N_1N_2} = I_{n-t}$. We then solve the system

$$\Lambda_f(G_{pub})^{\perp}_{.\{1,...,t\}} = H_f^{\top}\widetilde{T}^{\top}.$$

for $\widetilde{T} \in \mathbb{F}_q^{t\times(n-t)}$. Finally, we define

$$\widehat{T}^{-1} = \begin{bmatrix} I_t & \widetilde{T} \\ 0 & I_{n-t} \end{bmatrix} \in \mathbb{F}_q^{n\times n}.$$

We now have

$$\Lambda_f(G_{pub})\widehat{T}^{-1}\begin{bmatrix}0 & H_f^{\top}\end{bmatrix}^{\top} = 0.$$

Note that $row(G_{pub}) \subseteq \Lambda_f(G_{pub})$ and so $row(\Lambda_f(G_{pub})^{\perp}) \subseteq row(G_{pub}^{\perp})$. Also, it can be shown that $row(\begin{bmatrix}0 & H_0^{\top}\end{bmatrix}) \subseteq row(\begin{bmatrix}0 & H_f^{\top}\end{bmatrix})$. Hence, the above equation gives us

$$G_{pub}\widehat{T}^{-1}\begin{bmatrix}0 & H_0^{\top}\end{bmatrix}^{\top} = 0.$$

This means $\begin{bmatrix}0 & H_0^{\top}\end{bmatrix}$ lies in the dual space of $G_{pub}\widehat{T}^{-1}$. Hence, the last $n-t$ columns of $G_{pub}\widehat{T}^{-1}$ span an $[n-t, k]$ Gabidulin code. $\widehat{T}$ is thus a valid candidate for the column scrambler T. □

4.3 A Repair on the Column Scrambler

In the previous section, we saw that the attack is built on the fact that T is a matrix over $\mathbb{F}_q$ and not $\mathbb{F}_{q^m}$. Moreover, as T is defined over $\mathbb{F}_q$, the new error $\mathbf{z}T^{-1}$ in the decryption process still has the same rank as the original error $\mathbf{z}$.

We now look at a variant of the GGPT cryptosystem [12], where the elements of the column scrambler T can be from $\mathbb{F}_{q^m}$. This variant was proposed by H. Rashwan, E. Gabidulin and B. Honary. The rest of the encryption and decryption processes remain the same. This new proposal claimed to resist Overbeck's attack on the GGPT.

Choice of T: The matrix T is chosen such that for all errors $\mathbf{z}$ of rank weight $r' \leq r = \frac{n-t-k}{2}$, $\mathbf{z}T^{-1}$ has rank weight $\leq r$. It is clear that this must be true for successful decoding during decryption.

Lemma 4.3.1. *Let $T \in \mathbb{F}_{q^m}^{n\times n}$ such that T^{-1} is of the form*

$$T^{-1} = \begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{F}_{q^m}^{n\times n}$$

where $P \in \mathbb{F}_{q^m}^{n\times(r-r')}$ and $Q \in \mathbb{F}_q^{n\times(n-r+r')}$, and $r' \leq r = \frac{n-t-k}{2}$. Then, for any error $\mathbf{z} \in \mathbb{F}_{q^m}^n$ of rank weight r', $\mathbf{z}T^{-1}$ has rank weight $\leq r$.

Proof. We have

$$\mathbf{z}T^{-1} = \mathbf{z}\begin{bmatrix} P & Q \end{bmatrix} = \begin{bmatrix} \mathbf{z}P & \mathbf{z}Q \end{bmatrix}.$$

Since $\mathbf{z}$ has rank r', we can write $\mathbf{z} = \begin{bmatrix} w_1 & w_2 & \cdots & w_{r'} \end{bmatrix} A$, where $w_i \in \mathbb{F}_{q^m}, i = 1, 2, ..., r'$ are linearly independent over $\mathbb{F}_q$ and A is some $r' \times n$ matrix of rank r' over $\mathbb{F}_q$. Hence, $\mathbf{z}Q = \begin{bmatrix} w_1 & w_2 & \cdots & w_{r'} \end{bmatrix} AQ$ and since both A and Q have entries from $\mathbb{F}_q$, $\mathbf{z}Q$ has rank r'. Also, the $n \times (r - r')$ matrix $\mathbf{z}P$ clearly has rank $\leq (r - r')$ over $\mathbb{F}_q$. Thus, it follows that

$$rank(\mathbf{z}T^{-1}) \leq rank(\mathbf{z}P) + rank(\mathbf{z}Q) = r - r' + r' = r$$

over $\mathbb{F}_q$. □

This proposal, however, was proved to be reducible to the original proposal in [13] - it was shown that one could define an equivalent system with a new column scrambler defined on $\mathbb{F}_q$, so that it is still vulnerable to Overbeck's attack.

Another variant of the GPT system was proposed by H. Rashwan, E. Gabidulin and B. Honary, the so-called *Smart GPT* [14], which again resists Overbeck's attack. However, this repair was successfully attacked in [15], and this general attack has been able to break most variants of the GPT system.

4.4 Conclusion

We discussed a code-based cryptosystem, the GPT, which is similar to the McEliece cryptosystem, but based on rank metric codes instead of Hamming metric codes. Typically cryptosystems based on the rank metric usually offer smaller key sizes than those on the Hamming metric for comparable security. However, they were proved to be vulnerable to structural plaintext-recovery attacks. We have presented the details of such an attack by R. Overbeck, which broke many variants of the GPT system. There have been several proposals for repair, but as of now, the GGPT proposal has been proven to be structurally insecure for most parameter regimes of interest.

Bibliography

[1] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, Elsevier Science Publishing Company Inc., NY, USA, 1977.

[2] E. R. Berlekamp, R. J. McEliece and H. van Tilborg, "On the inherent intractability of certain coding problems", IEEE Trans. Information Theory, 1978.

[3] A. Kiayias A., M. Yung, "Cryptographic hardness based on the decoding of Reed-Solomon codes", IEEE Trans. Information Theory, vol. 54, no. 6, pp. 2752–2769, June 2008.

[4] R. J. McEliece, "A public-key cryptosystem based on algebraic coding theory". Jet Propulsion Lab. DSN Progress Report, 1978.

[5] E.M. Gabidulin. "Theory of codes with maximal rank distance". Problems of Information Transmission (Russian), 1985.

[6] D. Augot, M. Finiasz, "A public key encryption scheme based on the polynomial reconstruction problem", Proceedings of EUROCRYPT, 2003. LNCS, vol. 2656, Springer, Heidelburg, pp. 229 – 240.

[7] D. J. Bernstein, T. Lange, C. Peters, "Attacking and Defending the McEliece Cryptosystem". In: J. Buchmann, J. Ding (eds.) Post-Quantum Cryptography (PQCrypto) 2008. Lecture Notes in Computer Science, vol 5299. Springer, Berlin, Heidelberg.

[8] E.M. Gabidulin, A.V. Paramonov, O.V. Tretjakov, "Ideals over a non-commutative ring and their applications to cryptography", *in Proc. Eurocrypt '91*. LNCS, vol. 547 (Springer, Berlin, 1991).

[9] R. Overbeck, "Extending Gibson's attacks on the GPT cryptosystem", in *Proc. of WCC 2005*. LNCS, vol. 3969 (Springer, Berlin, 2006), pp. 178-188.

[10] R. Overbeck. "Structural attacks for public key cryptosystems based on Gabidulin codes", Journal of Cryptology (IACR), pp. 280 -301, 2008.

[11] A. Canteaut and F. Chabaud. "A new algorithm for finding minimum weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511". IEEE Trans. Inform. Theory, vol. 44, no. 1, pp. 367–378, January 1998.

[12] E. M. Gabidulin, H. Rashwan, B. Honary. "On improving security of GPT cryptosystems". *Proc. ISIT* (Seoul, South Korea, 2009).

[13] A. Otmani, H. T. Kalachi, S. Ndjeya, "Improved cryptanalysis of rank metric schemes based on Gabidulin codes", Des. Codes Cryptography, (2018) vol. 86. https://doi.org/10.1007/s10623-017-0343-7

[14] E.M. Gabidulin, A.V. Paramonov, O.V. Tretjakov. "A smart approach for GPT cryptosystem based on rank codes". ISIT 2010, Austin, U.S.A., June 2010.

[15] A-L. Horlemann-Trautmann, K. Marshall, & J. Rosenthal, "Extension of Overbeck's attack for Gabidulin-based cryptosystems", Des. Codes Cryptography, (2018) vol.86. https://doi.org/10.1007/s10623-017-0343-7.