Polynomial-Time Key Recovery Attack on the Lau-Tan Cryptosystem Based on Gabidulin Codes

Wenshuo Guo* and Fang-Wei Fu†

Abstract

This paper presents a key recovery attack on the cryptosystem proposed by Lau and Tan in a talk at ACISP 2018. The Lau-Tan cryptosystem uses Gabidulin codes as the underlying decodable code. To hide the algebraic structure of Gabidulin codes, the authors chose a matrix of column rank n to mix with a generator matrix of the secret Gabidulin code. The other part of the public key, however, reveals crucial information about the private key. Our analysis shows that the problem of recovering the private key can be reduced to solving a multivariate linear system, rather than solving a multivariate quadratic system as claimed by the authors. Apparently, this attack costs polynomial time, and therefore completely breaks the cryptosystem.

Keywords Post-quantum cryptography · Code-based cryptography · Gabidulin codes · Key recovery attack

1 Introduction

In post-quantum era, most public key cryptosystems based on number theoretic problems will suffer serious security threat. To resist quantum computer attacks, people have paid much attention to seek alternatives in the future. Among these alternatives, code-based cryptography is one of the most promissing candidates. The security of these cryptosystems rely on the difficulty of decoding general linear codes. The first code-based cryptosystem was the one proposed by McEliece in 1978, which is now called the McEliece cryptosystem [8]. Although this scheme still remains secure, it

*Wenshuo Guo is with the Chern Institute of Mathematics and LPMC, Nankai University, Tianjin 300071, China. E-mail:ws_guo@mail.nankai.edu.cn

†Fang-Wei Fu is with the Chern Institute of Mathematics and LPMC, Nankai University, Tianjin 300071, China. E-mail:fwfu@nankai.edu.cn

had never been used in practical situations due to the drawback of large key size. To tackle this problem, various improvements for McEliece's original scheme were proposed one after another. Generally these improvements can be divided into two categories: one is to substitute Goppa codes used in the McEliece system with other families of codes endowed with special structures, the other is to use codes endowed with the rank metric.

In 1991, Gabidulin et al. proposed an encryption scheme based on rank metric codes, which is now known as the GPT cryptosystem [1]. An important advantage of rank-based cryptosystems lies in their compact representation of public keys. Some representative variants based on the rank metric Gabidulin codes can be found in [2–6]. Unfortunately, most of these variants, including the original GPT cryptosystem, have been completely broken due to the inherent structural weakness of Gabidulin codes. Specifically, Gabidulin codes contain a large subspace invariant under the Frobenius transformation, which provides the feasibility for one to distinguish Gabidulin codes from general ones. Based on this observation, various structural attacks [9–13] on the GPT cryptosystem and some of their variants were designed.

In [7], Lau and Tan proposed a public key encryption scheme based on Gabidulin codes. In their cryptosystem, the published information consists of the generator matrix of a disturbed Gabidulin code by a random code that admits maximum rank weight n and a random vector of column rank n. This technique of masking the structure of Gabidulin codes, as claimed by Lau and Tan, can prevent some existing attacks [9–12]. Additionally, the recent Coggia-Couvreur attack [13] and Ghatak's attack [15] do not work on this cryptosystem either.

Our contributions. We mainly investigate the security of the Lau-Tan cryptosystem and present a simple but efficient algorithm for recovering the private key of this cryptosystem. Additionally, our analysis shows that all the generating vectors of a Gabidulin code, together with the zero vector, form a 1-dimensional linear space over $\mathbb{F}_{q^m}$. In other words, for a fixed generating vector $\boldsymbol{g}$ of a Gabidulin code $\mathcal{G} \subseteq \mathbb{F}_{q^m}^n$, any other generating vector of $\mathcal{G}$ must be of the form $\gamma\boldsymbol{g}$ for some $\gamma \in \mathbb{F}_{q^m}^*$. This suggests that there are totally $q^m - 1$ generating vectors for a Gabidulin code over $\mathbb{F}_{q^m}$. Meanwhile, we also introduce a different approach from the one proposed in [10] to compute a generating vector of a Gabidulin code when an arbitrary generator matrix is given.

The rest of this paper is organized as follows. Section 2 introduces some basic notations used throughout this paper, as well as the concept of Moore matrices and Gabidulin codes. Section 3 gives a simple description of the Lau-Tan cryptosystem. In Section 4, we mainly describe the principle of our attack. To do this, we first introduce some further results about Gabidulin codes that will be helpful for explaining why this attack works. Following this, we present this attack in two steps. Additionally we also give a complexity analysis of this attack and some experimental results using

Magma. In Section 5, we will make a few concluding remarks.

2 Preliminaries

In this section, we first introduce some notations in finite field and coding theory used throughout this paper. After that, we will recall some basic concepts about Gabidulin codes and some related results.

2.1 Notations and basic concepts

For a prime power q, denote by $\mathbb{F}_q$ the finite field with q elements, and $\mathbb{F}_{q^m}$ an extension field of $\mathbb{F}_q$ of degree m. Note that $\mathbb{F}_{q^m}$ can be seen as a linear space over $\mathbb{F}_q$ of dimension m. A vector $\boldsymbol{a} \in \mathbb{F}_{q^m}^m$ is called a basis vector if components of $\boldsymbol{a}$ form a basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. In particular, we call $\boldsymbol{a}$ a normal basis vector if $\boldsymbol{a}$ is of the form $(\alpha^{q^{m-1}}, \alpha^{q^{m-2}}, \ldots, \alpha)$ for some $\alpha \in \mathbb{F}_{q^m}^* = \mathbb{F}_{q^m} \backslash \{0\}$. For two positive integers k and n, denote by $\mathcal{M}_{k,n}(\mathbb{F}_q)$ the space of all $k \times n$ matrices over $\mathbb{F}_q$, and by $GL_n(\mathbb{F}_q)$ the set of all invertible matrices in $\mathcal{M}_{n,n}(\mathbb{F}_q)$. For a matrix $M \in \mathcal{M}_{k,n}(\mathbb{F}_q)$, denote by $\langle M \rangle_q$ the linear space spanned by rows of M over $\mathbb{F}_q$.

An $[n, k]$ linear code $\mathcal{C}$ over $\mathbb{F}_{q^m}$ is a k-dimensional subspace of $\mathbb{F}_{q^m}^n$. The dual code of $\mathcal{C}$, denoted by $\mathcal{C}^\perp$, is the orthogonal space of $\mathcal{C}$ under the usual inner product over $\mathbb{F}_{q^m}^n$. A $k \times n$ matrix G is called a generator matrix of $\mathcal{C}$ if its row vectors form a basis of $\mathcal{C}$ over $\mathbb{F}_{q^m}$. A generator matrix H of $\mathcal{C}^\perp$ is called a parity-check matrix of $\mathcal{C}$. For a codeword $\boldsymbol{c} \in \mathcal{C}$, the rank support of $\boldsymbol{c}$, denoted by $\mathrm{Supp}(\boldsymbol{c})$, is the linear space spanned by components of $\boldsymbol{c}$ over $\mathbb{F}_q$. The rank weight of $\boldsymbol{c}$ with respect to $\mathbb{F}_q$, denoted by $\mathrm{rk}(\boldsymbol{c})$, is defined to be the dimension of $\mathrm{Supp}(\boldsymbol{c})$ over $\mathbb{F}_q$. The minimum rank distance of $\mathcal{C}$, denoted by $\mathrm{rk}(\mathcal{C})$, is defined to be the minimum rank weight of all nonzero codewords in $\mathcal{C}$. For a matrix $M \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$, the rank support of M, denoted by $\mathrm{Supp}(M)$, is defined to be the linear space spanned by entries of M over $\mathbb{F}_q$. Similarly, the rank weight of M with respect to $\mathbb{F}_q$, denoted by $\mathrm{rk}(M)$, is defined as the dimension of $\mathrm{Supp}(M)$ over $\mathbb{F}_q$.

2.2 Gabidulin codes

In this section, we will recall the concept of Gabidulin codes. Before doing this, we first introduce of the definition of Moore matrices and some related results.

Definition 1 (Moore matrices). For an integer i and $\alpha \in \mathbb{F}_{q^m}$, we define $\alpha^{[i]} = \alpha^{q^i}$ to be the i-th Frobenius power of α. For a vector $\boldsymbol{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{F}_{q^m}^n$, we define $\boldsymbol{a}^{[i]} = (\alpha_1^{[i]}, \alpha_2^{[i]}, \ldots, \alpha_n^{[i]})$ to be the i-th Frobenius power of $\boldsymbol{a}$. For positive integers $k \leqslant n$, a $k \times n$ Moore matrix induced by

$\boldsymbol{a}$ is defined as

$$\mathrm{Mr}_k(\boldsymbol{a}) = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^{[1]} & \alpha_2^{[1]} & \cdots & \alpha_n^{[1]} \\ \vdots & \vdots & & \vdots \\ \alpha_1^{[k-1]} & \alpha_2^{[k-1]} & \cdots & \alpha_n^{[k-1]} \end{pmatrix}.$$

For a positive integer l and a matrix $M = (M_{ij}) \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$, we denote by $M^{[l]} = (M_{ij}^{[l]})$ the l-th Frobenius power of M. For a set $V \subseteq \mathbb{F}_{q^m}^n$, we denote by $V^{[l]} = \{\boldsymbol{v}^{[l]} : \boldsymbol{v} \in V\}$ the l-th Frobenius power of V. Particularly, for a linear code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$, it is easy to verify that $\mathcal{C}^{[l]}$ is also a linear code over $\mathbb{F}_{q^m}$.

The following proposition presents simple properties about Moore matrices.

Proposition 1. (1) *For two $k \times n$ Moore matrices $A, B \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$, the sum $A + B$ is also a $k \times n$ Moore matrix.*

(2) *For a Moore matrix $M \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ and a matrix $Q \in \mathcal{M}_{n,l}(\mathbb{F}_q)$, the product MQ forms a $k \times l$ Moore matrix.*

(3) *For a vector $\boldsymbol{a} \in \mathbb{F}_{q^m}^n$ with $\mathrm{rk}(\boldsymbol{a}) = l$, there exist $\boldsymbol{a}' \in \mathbb{F}_{q^m}^l$ with $\mathrm{rk}(\boldsymbol{a}') = l$ and $Q \in GL_n(\mathbb{F}_q)$ such that $\boldsymbol{a} = (\boldsymbol{a}', \boldsymbol{0})Q$. Furthermore, let $A = \mathrm{Mr}_k(\boldsymbol{a})$ and $A' = \mathrm{Mr}_k(\boldsymbol{a}')$, then $A = [A'|0]Q$.*

(4) *For positive integers $k \leqslant n \leqslant m$, let $\boldsymbol{a} \in \mathbb{F}_{q^m}^n$ be a vector such that $\mathrm{rk}(\boldsymbol{a}) = n$, then the Moore matrix $\mathrm{Mr}_k(\boldsymbol{a})$ has rank k.*

Proof. Statements (1),(2) and (3) are trivial and therefore the proof is omitted here.

(4) Let $\boldsymbol{a} = (\alpha_1, \cdots, \alpha_n) \in \mathbb{F}_{q^m}^n$. If $\mathrm{Rank}(\mathrm{Mr}_k(\boldsymbol{a})) < k$, then there exists $\boldsymbol{\lambda} = (\lambda_0, \cdots, \lambda_{k-1}) \in \mathbb{F}_{q^m}^k \backslash \{\boldsymbol{0}\}$ such that $\boldsymbol{\lambda}\mathrm{Mr}_k(\boldsymbol{a}) = \boldsymbol{0}$. Let $f(x) = \sum_{j=0}^{k-1} \lambda_j x^{[j]} \in \mathbb{F}_{q^m}[x]$, then $f(\alpha_i) = 0$ holds for any $1 \leqslant i \leqslant n$. It follows that $f(\alpha) = 0$ for any $\alpha \in \langle \alpha_1, \cdots, \alpha_n \rangle_q$, which conflicts with the fact that $f(x) = 0$ admits at most q^{k-1} roots.

□

The following proposition states a fact that a Moore matrix can be decomposed as the product of a specific Moore matrix and a matrix over the base field. This fact was once exploited by Loidreau in [14] to cryptanalyze an encryption scheme [3] based on Gabidulin codes.

Proposition 2 (Moore matrix decomposition)**.** *Let $\boldsymbol{a}$ be a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. For positive integers $k \leqslant m$, let $M \in \mathcal{M}_{k,m}(\mathbb{F}_{q^m})$ be a Moore matrix generated by $\boldsymbol{a}$. Then for any $k \times n$ Moore matrix $M' \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$, there exists $Q \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ such that $M' = MQ$.*

Now we formally introduce the definition of Gabidulin codes.

Definition 2 (Gabidulin codes). For positive integers $k \leqslant n \leqslant m$, let $\boldsymbol{a} \in \mathbb{F}_{q^m}^n$ such that $\mathrm{rk}(\boldsymbol{a}) = n$. The $[n, k]$ Gabidulin code generated by $\boldsymbol{a}$, denoted by $\mathrm{Gab}_{n,k}(\boldsymbol{a})$, is defined as the linear space spanned by rows of $\mathrm{Mr}_k(\boldsymbol{a})$ over $\mathbb{F}_{q^m}$. $\mathrm{Mr}_k(\boldsymbol{a})$ is called a standard generator matrix of $\mathrm{Gab}_{n,k}(\boldsymbol{a})$, and $\boldsymbol{a}$ a generating vector respectively.

Remark 1. Gabidulin codes can be seen as a counterpart of generalized Reed-Solomon (GRS) codes in the rank metric, both of which admit good algebraic properties. An $[n, k]$ Gabidulin code has minimum rank distance $n - k + 1$ [16] and can therefore correct up to $\lfloor \frac{n-k}{2} \rfloor$ rank errors in theory. Efficient decoding algorithms for Gabidulin codes can be found in [17–19].

Definition 3 (Partial circulant matrices). For a vector $\boldsymbol{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{F}_{q^m}^n$, the circulant matrix induced by $\boldsymbol{a}$, denoted by $\mathrm{Cir}_n(\boldsymbol{a})$, is defined to be a matrix whose first row is $\boldsymbol{a}$ and i-th row is obtained by cyclically right shifting the $i-1$-th row for $2 \leqslant i \leqslant n$. The $k \times n$ partial circulant matrix induced by $\boldsymbol{a}$, denoted by $\mathrm{Cir}_k(\boldsymbol{a})$, is defined to be the first k rows of $\mathrm{Cir}_n(\boldsymbol{a})$.

Remark 2. For a normal basis vector $\boldsymbol{a}$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$, it is easy to verify that the $k \times n$ partial circulant matrix induced by $\boldsymbol{a}$ is exactly the $k \times n$ Moore matrix generated by $\boldsymbol{a}$. In other words, mathematically we have $\mathrm{Cir}_k(\boldsymbol{a}) = \mathrm{Mr}_k(\boldsymbol{a})$.

3 Lau-Tan cryptosystem

In this section, we mainly give a formal description of the Lau-Tan cryptosystem that uses Gabidulin codes as the underlying decodable code. For a given security level, choose positive integers $m > n > k > k' \geqslant 1$ and r, such that $k' = \lfloor \frac{k}{2} \rfloor$ and $r \leqslant \lfloor \frac{n-k}{2} \rfloor$. The Lau-Tan cryptosystem consists of the following three algorithms.

- Key Generation

  Let $\mathcal{G}$ be an $[n, k]$ Gabidulin code over $\mathbb{F}_{q^m}$, and $G \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ be a generator matrix of $\mathcal{G}$ of standard form. Randomly choose matrices $S \in GL_k(\mathbb{F}_{q^m})$ and $T \in GL_n(\mathbb{F}_q)$. Randomly choose $\boldsymbol{u} \in \mathbb{F}_{q^m}^n$ such that $\mathrm{rk}(\boldsymbol{u}) = n$ and set $U = \mathrm{Cir}_k(\boldsymbol{u})$. Let $G_{pub} = SG + UT$, then we publish $(G_{pub}, \boldsymbol{u})$ as the public key, and keep (S, G, T) as the private key.

- Encryption

  For a plaintext $\boldsymbol{m} \in \mathbb{F}_{q^m}^{k'}$, randomly choose $\boldsymbol{m}_s \in \mathbb{F}_{q^m}^{k-k'}$ such that $\mathrm{rk}((\boldsymbol{m}, \boldsymbol{m}_s)U) > \lceil \frac{3}{4}(n - k) \rceil$. Randomly choose $\boldsymbol{e}_1, \boldsymbol{e}_2 \in \mathbb{F}_{q^m}^n$ such that $\mathrm{rk}(\boldsymbol{e}_1) = r_1 \leqslant \frac{r}{2}$ and $\mathrm{rk}(\boldsymbol{e}_2) = r_2 \leqslant \frac{r}{2}$.

Compute $\boldsymbol{c}_1 = (\boldsymbol{m}, \boldsymbol{m}_s)U + \boldsymbol{e}_1$ and $\boldsymbol{c}_2 = (\boldsymbol{m}, \boldsymbol{m}_s)G_{pub} + \boldsymbol{e}_2$. Then the ciphertext is $\boldsymbol{c} = (\boldsymbol{c}_1, \boldsymbol{c}_2)$.

- Decryption

  For a ciphertext $\boldsymbol{c} = (\boldsymbol{c}_1, \boldsymbol{c}_2) \in \mathbb{F}_{q^m}^{2n}$, compute $\boldsymbol{c}' = \boldsymbol{c}_2 - \boldsymbol{c}_1 T = (\boldsymbol{m}, \boldsymbol{m}_s)SG + \boldsymbol{e}_2 - \boldsymbol{e}_1 T$. Note that $\mathrm{rk}(\boldsymbol{e}_2 - \boldsymbol{e}_1 T) \leqslant \mathrm{rk}(\boldsymbol{e}_2) + \mathrm{rk}(\boldsymbol{e}_1 T) \leqslant r$, decoding $\boldsymbol{c}'$ with the existing decoder of $\mathcal{G}$ will lead to $\boldsymbol{m}' = (\boldsymbol{m}, \boldsymbol{m}_s)S$, then by computing $\boldsymbol{m}' S^{-1}$ one can recover the plaintext $\boldsymbol{m}$.

4 Key recovery attack

In this section, we will describe a method of how to efficiently recover an equivalent private key in the Lau-Tan cryptosystem. We point out that the privacy of T is of great importance for the security of the whole cryptosystem. Specifically, if one can find the secret T, then one can recover everything he needs to decrypt an arbitrary ciphertext. Before formally describing this attack, we introduce some further results about Gabidulin codes.

4.1 Further results about Gabidulin codes

Similar to GRS codes in the Hamming metric, Gabidulin codes admit good algebraic structure. For instance, if $\mathcal{G}$ is a Gabidulin code over $\mathbb{F}_{q^m}$, then its l-th Frobenius power is still a Gabidulin code. Formally, we introduce the following lemma.

Lemma 4. *Let $\mathcal{G}$ be an $[n, k]$ Gabidulin code over $\mathbb{F}_{q^m}$, with $G \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ as a generator matrix. For any positive integer l, $\mathcal{G}^{[l]}$ is also an $[n, k]$ Gabidulin code and has $G^{[l]}$ as a generator matrix.*

Proof. Trivial from a straightforward verification. □

For a proper positive integer l, the intersection of a Gabidulin code and its l-th Frobenius power is still a Gabidulin code. Formally, we introduce the following proposition.

Proposition 3. *For an $[n, k]$ Gabidulin code $\mathcal{G}$ over $\mathbb{F}_{q^m}$, let $\boldsymbol{g} \in \mathbb{F}_{q^m}^n$ be a generating vector of $\mathcal{G}$. For a positive integer $l \leqslant \min\{k-1, n-k\}$, the intersection of $\mathcal{G}$ and its l-th Frobenius power is an $[n, k-l]$ Gabidulin code with $\boldsymbol{g}^{[l]}$ as a generating vector. In other words, we have the following equality*

$$\mathcal{G} \cap \mathcal{G}^{[l]} = \mathrm{Gab}_{n,k-l}(\boldsymbol{g}^{[l]}).$$

Proof. By the definition of Gabidulin codes, $\mathcal{G}$ is a linear space spanned by $\boldsymbol{g}, \ldots, \boldsymbol{g}^{[k-1]}$ over $\mathbb{F}_{q^m}$, i.e. $\mathcal{G} = \langle \boldsymbol{g}, \ldots, \boldsymbol{g}^{[k-1]} \rangle_{q^m}$. By Lemma 4, we have $\mathcal{G}^{[l]} = \langle \boldsymbol{g}^{[l]}, \ldots, \boldsymbol{g}^{[k+l-1]} \rangle_{q^m}$. Note that $l \leqslant \min\{k-1, n-k\}$, then $k+l \leqslant n$ and $\boldsymbol{g}, \ldots, \boldsymbol{g}^{[k+l-1]}$ are linearly independent over $\mathbb{F}_{q^m}$. It follows that $\mathcal{G} \cap \mathcal{G}^{[l]} = \langle \boldsymbol{g}^{[l]}, \ldots, \boldsymbol{g}^{[k-l-1]} \rangle_{q^m}$ forms an $[n, k-l]$ Gabidulin code, having $\boldsymbol{g}^{[l]}$ as a generating vector. This completes the proof. □

Proposition 4. *For positive integers $k < n$, let $\mathcal{G} \subset \mathbb{F}_{q^m}^n$ be an $[n, k]$ Gabidulin code, and $A \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ be a nonzero Moore matrix. If all the row vectors of A are codewords in $\mathcal{G}$, then A must be a generator matrix of $\mathcal{G}$.*

Proof. It suffices to prove $\mathrm{Rank}(A) = k$. Suppose that A is generated by $\boldsymbol{a} \in \mathbb{F}_{q^m}^n$, i.e. $A = \mathrm{Mr}_k(\boldsymbol{a})$. Let $l = \mathrm{rk}(\boldsymbol{a})$, then there exist $\boldsymbol{a}' \in \mathbb{F}_{q^m}^l$ with $\mathrm{rk}(\boldsymbol{a}') = l$ and $Q \in GL_n(\mathbb{F}_q)$ such that $\boldsymbol{a} = (\boldsymbol{a}', \boldsymbol{0})Q$. Let $A' \in \mathcal{M}_{k,l}(\mathbb{F}_{q^m})$ be a Moore matrix generated by $\boldsymbol{a}'$, then it follows immediately that $A = [A'|0]Q$. If $l > k$, then $\mathrm{Rank}(A) = \mathrm{Rank}(A') = k$ due to Proposition 1 and therefore the conclusion is proved. Otherwise, there will be $\langle A' \rangle_{q^m} = \mathbb{F}_{q^m}^l$. From this we can deduce that the minimum rank distance of $\mathcal{G}$ will be 1, which conflicts with the fact that $\mathrm{rk}(\mathcal{G}) = n - k + 1 \geqslant 2$. Hence $l > k$ and $\mathrm{Rank}(A) = k$. This completes the proof. □

By Definition 2, a Gabidulin code is uniquely determined by its generating vector. Naturally, it is important to make clear what all these vectors look like and how many generating vectors there exist for a Gabidulin code.

Proposition 5. *Let $\mathcal{G}$ be an $[n, k]$ Gabidulin code over $\mathbb{F}_{q^m}$, with $\boldsymbol{g} \in \mathbb{F}_{q^m}^n$ as a generating vector. Let $\boldsymbol{g}' \in \mathbb{F}_{q^m}^n$ be a codeword in $\mathcal{G}$, then $\boldsymbol{g}'$ forms a generating vector if and only if there exists $\gamma \in \mathbb{F}_{q^m}^*$ such that $\boldsymbol{g}' = \gamma \boldsymbol{g}$.*

Proof. Assume that $\boldsymbol{g} = (\alpha_1, \ldots, \alpha_n)$ and $\boldsymbol{g}' = (\alpha_1', \ldots, \alpha_n')$, let $G = \mathrm{Mr}_k(\boldsymbol{g})$ and $G' = \mathrm{Mr}_k(\boldsymbol{g}')$. The conclusion is trivial if $\boldsymbol{g} = \boldsymbol{g}'$. Otherwise, without loss of generality we assume that $\alpha_1' \neq \alpha_1$, then there exists $\gamma \in \mathbb{F}_{q^m}^* \backslash \{1\}$ such that $\alpha_1' = \gamma \alpha_1$. Let

$$S = \begin{pmatrix} \gamma & 0 & \cdots & 0 \\ 0 & \gamma^{[1]} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \gamma^{[k-1]} \end{pmatrix},$$

then $SG = \mathrm{Mr}_k(\gamma \boldsymbol{g})$. Let $\boldsymbol{g}^* = \gamma \boldsymbol{g} - \boldsymbol{g}' = (0, \gamma\alpha_2 - \alpha_2', \ldots, \gamma\alpha_n - \alpha_n')$ and $G^* = \mathrm{Mr}_k(\boldsymbol{g}^*)$, then $G^* = SG - G'$. Apparently all the row vectors of G^* are codewords in $\mathcal{G}$. If $\boldsymbol{g}^* \neq \boldsymbol{0}$, then G^* forms a generator matrix of $\mathcal{G}$ of standard form due to Proposition 4. Together with $\mathrm{rk}(\boldsymbol{g}^*) \leqslant n-1$,

easily we can deduce that $\mathrm{rk}(\boldsymbol{c}) \leqslant n-1$ for any $\boldsymbol{c} \in \mathcal{G}$. By the definition of Gabidulin codes, however, there exists at least one codeword having rank weight n. Therefore there must be $\boldsymbol{g}^* = \mathbf{0}$, or equivalently $\boldsymbol{g}' = \gamma \boldsymbol{g}$. The opposite is obvious from a straightforward verification. □

The following corollary is drawn immediately from Proposition 5.

Corollary 1. *An $[n,k]$ Gabidulin code over $\mathbb{F}_{q^m}$ admits q^m-1 generator matrices of standard form, or equivalently q^m-1 generating vectors.*

4.2 Recovering the secret T

In this section, we mainly describe an efficient algorithm for recovering the secret T. In summary, the technique we adopt here is to convert the problem of recovering T into solving a multivariate linear system, which clearly costs polynomial time. Before doing this, we introduce the so-called subfield expanding transform.

Subfield Expanding Transform. For $\beta_1, \ldots, \beta_n \in \mathbb{F}_{q^m}$, we construct an equation as

$$\sum_{j=1}^{n} x_j \beta_j = 0, \tag{1}$$

where x_j's are underdetermined variables in $\mathbb{F}_q$. Let $\boldsymbol{a}$ be a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. For each $1 \leqslant j \leqslant n$, there exists $\boldsymbol{b}_j \in \mathbb{F}_q^m$ such that $\beta_j = \boldsymbol{b}_j \boldsymbol{a}^T$. It follows that $\sum_{j=1}^n x_j \beta_j = \sum_{j=1}^n x_j (\boldsymbol{b}_j \boldsymbol{a}^T) = (\sum_{j=1}^n x_j \boldsymbol{b}_j)\boldsymbol{a}^T$, and moreover, (1) holds if and only if

$$\sum_{j=1}^{n} x_j \boldsymbol{b}_j = \mathbf{0}. \tag{2}$$

Obviously, the linear systems (1) and (2) share the same solution space. A transform that derives (2) from (1) is called a subfield expanding transform.

In the Lau-Tan cryptosystem, let $H \in \mathcal{M}_{n-k,n}(\mathbb{F}_{q^m})$ be a parity-check matrix of $\mathcal{G}$ of standard form. Let $M \in \mathcal{M}_{n-k,m}(\mathbb{F}_{q^m})$ be a Moore matrix generated by a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$, then there exists an underdetermined matrix $X \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ such that $H = MX$. Let $T^* \in GL_n(\mathbb{F}_q)$ be another underdetermined matrix such that $G_{pub} - \mathrm{Cir}_k(\boldsymbol{u})T^* = G_{pub} - UT^*$ forms a generator matrix of $\mathcal{G}$, or equivalently

$$(G_{pub} - UT^*)(MX)^T = G_{pub}X^TM^T - UT^*X^TM^T = 0. \tag{3}$$

We therefore obtain a system of $k(n-k)$ multivariate quadratic equations, with $n(m+n)$ variables in $\mathbb{F}_q$. This system admits at least q^m solutions. Specifically, we introduce the following proposition.

Proposition 6. *The linear system* (3) *has at least* q^m *solutions.*

Proof. If $T^* = T$, then we can deduce from (3) that

$$(G_{pub} - UT^*)(MX)^T = G_{pub}X^TM^T - UT^*X^TM^T = SGX^TM^T = SG(MX)^T = 0.$$

Note that $SG \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ forms a generator matrix of $\mathcal{G}$. By $SG(MX)^T = 0$, we conclude that all the row vectors of MX are contained in the dual code of $\mathcal{G}$, which is an $[n, n-k]$ Gabidulin code. On the other hand, it is easy to see that MX is an $(n-k)\times n$ Moore matrix. By Proposition 4, MX forms a standard generator matrix of $\mathcal{G}^\perp$ for a nonzero X. Then the conclusion is immediately proved from Corollary 1. Furthermore, we have that X is an $m \times n$ matrix of full rank. □

Solving this multivariate quadratic system, however, requires exponential time in general. To avoid solving this system directly, the technique we exploit here is to consider each entry of T^*X^T as a new variable in $\mathbb{F}_q$ and set $Y = XT^{*T}$. In other words, we write (3) into the following matrix equation

$$G_{pub}X^TM^T - UY^TM^T = 0. \tag{4}$$

We therefore obtain a linear system of $k(n-k)$ equations, with coefficients in $\mathbb{F}_{q^m}$ and $2mn$ variables in $\mathbb{F}_q$. Instead of solving (4) directly, we usually convert this problem into an instance over the base field $\mathbb{F}_q$. Applying the subfield expanding transform to (4) leads to a linear system of $mk(n-k)$ equations over $\mathbb{F}_q$, with $2mn$ variables to be determined. For a cryptographic use, generally we have $mk(n-k) \geqslant 2mn$.

Remark 3. For each solution (X, T^*) of (3), one can easily obtain a solution of (4) by computing $Y = XT^{*T}$, which implies that there are also at least q^m solutions for (4). Conversely, if (4) has exactly q^m solutions, then these solutions must correspond to solutions of (3) where $T^* = T$. In this situation, for each nonzero solution of (4), solving the matrix equation $Y = XT^{*T}$ will lead to the secret $T = T^*$.

As for whether or not the system (4) has other types of solutions, we make an assumption here as follows.

Assumption. Let G be an arbitrary generator matrix of an $[n, k]$ Gabidulin code and $\boldsymbol{u} \in \mathbb{F}_{q^m}^n$ such that $\mathrm{rk}(\boldsymbol{u}) = n$. There does not exist a nonzero matrix $T \in \mathcal{M}_{n,n}(\mathbb{F}_q)$ such that all the row vectors of $G + \mathrm{Cir}_k(\boldsymbol{u})T$ are contained in an $[n, k]$ Gabidulin code.

According to our experiments on Magma, this assumption holds with high probability. Specifically, we construct a linear system

$$(G + \mathrm{Cir}_k(\boldsymbol{u})T)(MX)^T = 0,$$

where $M = \mathrm{Mr}_{n-k}(\boldsymbol{a})$ for a basis vector $\boldsymbol{a}$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ and $X \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ is an underdetermined matrix. After that, we apply the subfield expanding transform to this system to obtain a new system over $\mathbb{F}_q$. By Remark 3, if this new system admits a solution space of dimension m, then there must be $T = 0$. Eventually we ran 1000 random tests for $q = 2, m = 25, n = 23$ and $k = 10$. It turns out that the assumption holds in all of these random instances.

Algorithm 1 : T-recovering algorithm

Input: The public key $(G_{pub}, \boldsymbol{u})$

Output: The secret T

1: Let $\boldsymbol{a}$ be a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ and set $M = \mathrm{Mr}_{n-k}(\boldsymbol{a})$

2: Let $X, Y \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ be two underdetermined matrices and construct a linear system

$$G_{pub}X^TM^T - \mathrm{Cir}_k(\boldsymbol{u})Y^TM^T = 0 \tag{5}$$

3: Applying the subfield expanding transform to (5) to obtain a linear system over $\mathbb{F}_q$

4: Solve this system for (X, Y)

5: For any nonzero (X, Y), solve the matrix equation $Y = XT^{*T}$ for T^*

6: **return** $T = T^*$

4.3 Finding an equivalent (S', G')

From a generating vector, we can deduce many characteristics about a Gabidulin code, such as an efficient decoding algorithm. Thus a natural question is how to derive a generating vector of a Gabidulin code from an arbitrary generator matrix. In [10] the authors proposed an iterative method of computing the generating vector. Here in this paper we present a different approach to do this, as described in the following.

A approach to compute the generating vector. For an $[n, k]$ Gabidulin code $\mathcal{G}$ over $\mathbb{F}_{q^m}$, let $G \in \mathcal{M}_{k,n}(\mathbb{F}_{q^m})$ be an arbitrary generator matrix of $\mathcal{G}$. First we compute a generator matrix of $\mathcal{G}$ of systematic form from G, say $[I_k|A]$, where I_k is the identity matrix of order k and $A \in \mathcal{M}_{k,n-k}(\mathbb{F}_{q^m})$. Let $H = [-A^T|I_{n-k}]$, then H forms a parity-check matrix of $\mathcal{G}$. Let $M \in \mathcal{M}_{k,m}(\mathbb{F}_{q^m})$ be a Moore matrix generated by a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$, then there exists an underdetermined matrix $X \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ such that MX forms a standard generator matrix of $\mathcal{G}$. From $(MX)H^T = 0$ we obtain a linear system of $k(n-k)$ equations, with coefficients in $\mathbb{F}_{q^m}$ and mn variables in $\mathbb{F}_q$. Applying the subfield expanding transform to this system leads to a new linear system over the base field $\mathbb{F}_q$, with $mk(n-k)$ equations and mn variables. For a cryptographic use, generally we have $mk(n-k) \geqslant mn$. By Corollary 1, this new system admits $q^m - 1$ nonzero solutions. And for each nonzero solution, say X, the first row of MX will be a generating vector

of $\mathcal{G}$.

Algorithm 2 : Finding an equivalent (S', G')

Input: $(G_{pub}, \boldsymbol{u}, T)$

Output: (S', G') such that G' forms a standard generator matrix of $\mathcal{G}$ and $S'G' = SG$

1: Let $\boldsymbol{a}$ be a basis vector of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ and construct a Moore matrix $M = \mathrm{Mr}_k(\boldsymbol{a})$

2: Compute $SG = G_{pub} - \mathrm{Cir}_k(\boldsymbol{u})T$ and $\mathcal{G} = \langle SG \rangle_{q^m}$

3: Let $H \in \mathcal{M}_{n-k,n}(\mathbb{F}_{q^m})$ be a parity-check matrix of $\mathcal{G}$

4: Let $X \in \mathcal{M}_{m,n}(\mathbb{F}_q)$ be an underdetermined matrix and construct a linear system as

$$(MX)H^T = 0 \tag{6}$$

5: Applying the subfield expanding transform to (6) to obtain a new system over $\mathbb{F}_q$

6: Solve this new system for a nonzero X and compute $G' = MX$

7: Compute $S' \in GL_k(\mathbb{F}_{q^m})$ such that $S'G' = SG$

8: **return** (S', G')

4.4 Complexity of the attack

Our attack consists of two procedures: firstly, we manage to recover the secret T from the published information, as described in Algorithm 1; secondly, with the knowledge of T and the public key, we compute a standard generator matrix G' of the secret Gabidulin code and an invertible matrix S', as described in Algorithm 2. Hence the complexity analysis is done in the following two parts.

Complexity of Algorithm 1. In Step 1 we construct a Moore matrix $M \in \mathcal{M}_{n-k,m}(\mathbb{F}_{q^m})$ whose first row vector forms a basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. To avoid executing the Frobenius operation, here we choose $\boldsymbol{a}$ to be a normal basis vector, then we construct $M = \mathrm{Cir}_{n-k}(\boldsymbol{a})$. In Step 2 we construct a multivariate linear system by performing matrix multiplication, requiring $\mathcal{O}(mn^3)$ operations in $\mathbb{F}_{q^m}$. The subfield expanding transform performed to (5) requires $\mathcal{O}(m^3n^3)$ operations in $\mathbb{F}_{q^m}$. Step 4 requires $\mathcal{O}(m^3n^3)$ operations to solve the linear system over $\mathbb{F}_q$ and Step 5 requires $\mathcal{O}(n^3)$ operations in $\mathbb{F}_q$. The total complexity of Algorithm 1 consists of $\mathcal{O}(m^3n^3+mn^3)$ operations in $\mathbb{F}_{q^m}$ and $\mathcal{O}(m^3n^3 + n^3)$ operations in $\mathbb{F}_q$.

Complexity of Algorithm 2. In Step 1 we still choose a normal basis vector to construct M. To compute SG, we perform matrix addition and multiplication with $\mathcal{O}(n^3)$ operations in $\mathbb{F}_{q^m}$. Step 3 computes a parity-check H of $\mathcal{G}$ from SG, requiring $\mathcal{O}(n^3)$ operations in $\mathbb{F}_{q^m}$. Then we construct a linear system in Step 4, which costs $\mathcal{O}(mn^3)$ operations in $\mathbb{F}_{q^m}$. In Step 5 we apply the subfield expanding transform to (6) to obtain a new system over $\mathbb{F}_q$, requiring $\mathcal{O}(m^3n^3)$ operations in $\mathbb{F}_{q^m}$. Solving this new system in Step 6 costs $\mathcal{O}(m^3n^3)$ operations in $\mathbb{F}_q$, and compute $G' = MX$ with

$\mathcal{O}(mn^2)$ operations in $\mathbb{F}_{q^m}$. In Step 7, we shall compute S' from $S'G'$ with $\mathcal{O}(n^3)$ operations. The total complexity of Algorithm 2 consists of $\mathcal{O}(m^3n^3 + mn^3 + n^3)$ operations in $\mathbb{F}_{q^m}$ and $\mathcal{O}(m^3n^3)$ operations in $\mathbb{F}_q$.

Finally, the total complexity of the attack is $\mathcal{O}(m^3n^3 + mn^3 + n^3)$ in $\mathbb{F}_{q^m}$ plus $\mathcal{O}(m^3n^3 + n^3)$ in $\mathbb{F}_q$.

4.5 Implementation

This attack has been implemented on Magma and permits to recover the secret T. We tested this attack on a personal computer and succeeded for parameters as illustrated in Table 1. For each parameter set, the attack has been run 100 times and the last column gives the average timing (in seconds). Our implementation is just a proof of the feasibility of this attack and could be further optimised.

Table 1 These experiments were performed using Magma V2.11-1 on an 11th Gen Intel(R) Core(TM) i7-11700 @ 2.5GHz processor with 16 GB of memory.

q	m	n	k	$t(s)$
2	22	18	9	8.6
2	28	22	9	40.7
2	35	26	12	173.2

5 Conclusion

Our attack revealed the structural weakness in the Lau-Tan cryptosystem. Although the first part of the public key perfectly hid the structure of Gabidulin codes, the second part did reveal important information that can be used to design a key recovery attack. Specifically, we convert the problem of recovering the private key into solving a linear system over the base field $\mathbb{F}_q$. Although this system admits a solution space of dimension m, we are able to recover the secret T and then an equivalent (S', G') from any nonzero solution. Extensive experiments have been performed and the results show that our attack accords with the theoretical expectations. In summary, we found a polynomial-time key recovery attack on the Lau-Tan cryptosystem under reasonable assumptions.

Acknowledgements This research is supported by the National Key Research and Development Program of China (Grant No. 2018YFA0704703), the National Natural Science Foundation of China (Grant No. 61971243), the Natural Science Foundation of Tianjin (20JCZDJC00610), and the Fundamental Research

Funds for the Central Universities of China (Nankai University).

References

[1] Gabidulin, E.M., Paramonov, A.V., Tretjakov, O.V.: Ideals over a non-commutative ring and their application in cryptology. In: Davies, D.W. (Ed.): Proceedings of Advances in Cryptology-EUROCRYPT'91, LNCS, vol. 547, pp. 482–489. Springer (1991).

[2] Gabidulin, E.M., Ourivski, A.V., Honary, B., Ammar, B.: Reducible rank codes and their applications to cryptography. IEEE Trans. Inform. Theory 49(12), 3289–3293 (2003).

[3] Loidreau, P.: A new rank metric codes based encryption scheme. In: Lange, T., Takagi, T. (Eds.): Proceedings of PQCrypto 2017, LNCS, vol. 10346, pp. 3–17. Springer (2017).

[4] Faure, C., Loidreau, P.: A new public-key cryptosystem based on the problem of reconstructing p-polynomials. In: Ytrehus, $\varnothing$. (Ed.): Proceedings of WCC 2005, LNCS, vol. 3969, pp. 304–315. Springer (2005).

[5] Lau, T.S.C., Tan, C.H.: New rank codes based encryption scheme using partial circulant matrices. Des. Codes Cryptogr. 87(12), 2979–2999 (2019).

[6] Berger, T., Loidreau, P.: Designing an efficient and secure public-key cryptosystem based on reducible rank codes. In: Proceedings of INDOCRYPT 2004, LNCS, vol. 3348, pp. 218–229. Springer (2004).

[7] Lau, T.S.C., Tan, C.H.: A new encryption scheme based on rank metric codes. In: Proceedings of ACISP 2018, LNCS, vol. 10946, pp. 750–758. Springer (2018).

[8] McEliece, R.J.: A public-key cryptosystem based on algebraic coding theory. Jet Propuls. Lab. DSN Progr. Rep. 42-44, 114–116 (1978).

[9] Overbeck, R.: Structural attacks for public key cryptosystems based on Gabidulin codes. J. Cryptology 21(2), 280–301 (2008).

[10] Horlemann-Trautmann, A.-L., Marshall, K., Rosenthal, J.: Extension of Overbeck's attack for Gabidulin-based cryptosystems. Des. Codes Cryptogr. 86(2), 319–340 (2018).

[11] Otmani, A., Kalachi, H.T., Ndjeya, S.: Improved cryptanalysis of rank metric schemes based on Gabidulin codes. Des. Codes Cryptogr. 86(9), 1983–1996 (2018).

[12] Gaborit, P., Otmani, A., Kalachi, H.T.: Polynomial-time key recovery attack on the Faure–Loidreau scheme based on Gabidulin codes. Des. Codes Cryptogr. 86(7),1391–1403 (2018).

[13] Coggia, D., Couvreur, A.: On the security of a Loidreau rank metric code based encryption scheme. Des. Codes Cryptogr. 88(9), 1941–1957 (2020).

[14] Loidreau, P.: Analysis of a rank metric codes based encryption scheme. https://drive.google.com/file/d/1FuMgqm0NfGMJOxaZyrIrI1OWn0UICwPo/view. Accessed July 1, 2021.

[15] Ghatak, A: Extending Coggia–Couvreur attack on Loidreau's rank-metric cryptosystem. Des. Codes Cryptogr. (2021). https://doi.org/10.1007/s10623-021-00972-7.

[16] Horlemann-Trautmann, A.-L., Marshall, K.: New criteria for MRD and Gabidulin codes and some rank-metric code constructions. arXiv:1507.08641 [cs.IT] (2015).

[17] Gabidulin, E.M.: Theory of codes with maximum rank distance. Prob. Peredachi Inf. 21(1), 3–16 (1985).

[18] Loidreau, P.: A Welch-Berlekamp like algorithm for decoding Gabidulin codes. In: Ytrehus, ∅. (Ed.): Proceedings of WCC 2005, LNCS, vol. 3969, pp. 36–45. Springer (2005).

[19] Richter, G., Plass, S.: Error and erasure decoding of rank-codes with a modified Berlekamp-Massey algorithm. ITG FACHBERICHT, pp. 203–210 (2004).