Gröbner Bases in Cryptography through the example of the Rank Decoding Problem

# Maxime Bros

# Algebraic Rewriting Seminar

April 15th, 2021



### Introduction

- (A)symmetric Cryptography
- First Asymmetric Scheme
- Hard Problems and Quantum Threat
- Decoding Problem

### 2 Rank Decoding Problem

- Introduction to Coding Theory
- Rank Metric

### 3 Algebraic Attacks

- Definition
- Gröbner Basis
- Linearization
- Our Modeling for Rank Decoding
  - Modeling
  - Complexity
- 6 Conclusion
  - Comparison with previous Attacks
  - Summary of our Contributions

### Our Attacks



Magali Bardet, Pierre Briaud, Maxime Bros, Philippe Gaborit, Vincent Neiger, Olivier Ruatta, and Jean-Pierre Tillich. An algebraic attack on rank metric code-based cryptosystems. In *EUROCRYPT 2020*, pages 64–93. Springer, 2020.

Magali Bardet, Maxime Bros, Daniel Cabarcas, Philippe Gaborit, Ray Perlner, Daniel Smith-Tone, Jean-Pierre Tillich, and Javier Verbel. Improvements of algebraic attacks for solving the rank decoding and minrank problems.

In ASIACRYPT 2020, pages 507-536. Springer, 2020.

A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

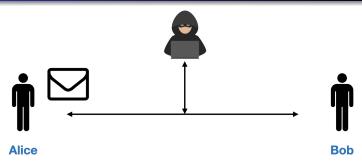
### Notation

- $\mathbb{F}_q$ : finite field with q elements (usually q = 2, sometimes q = p'),
- $\mathbb{F}_{q^m}$  is its extension of degree m,
- $\mathbb{F}_{q^m}$  is also a finite field with  $q^m$  elements,
- $\mathbb{F}_{q^m}$  can be seen as an  $\mathbb{F}_q$ -vector space of dimension m,

• 
$$a \stackrel{\$}{\longleftarrow} \llbracket 1, 2, \dots, N \rrbracket.$$

Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

# Cryptography

(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem



### Many limitations due to keys' exchanges

### Assumption

There exists such a symmetric cryptographic scheme (e.g. AES).

Maxime Bros Gröbner Bases in Cryptography through the RSD problem

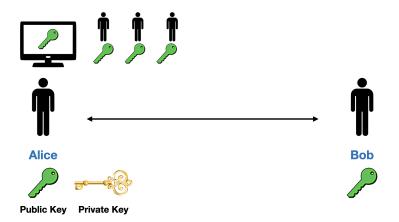
Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion

#### (A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Thi Decoding Problem

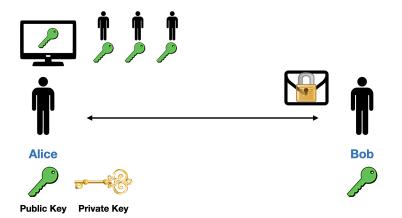


Public Key Private Key

Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion (A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Three Decoding Problem



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion (A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Three Decoding Problem



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion (A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Thre Decoding Problem



Rank Decoding Problem Algebraic Attacks Our Modeling for Rank Decoding Conclusion (A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Thre Decoding Problem



(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

### • Merkle: 1975

Revolution dates

• First protocol in 1976:

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-22,

# New Directions in Cryptography

Invited Paper

WHITFIELD DIFFIE AND MARTIN E. HELLMAN, MEMBER, IEEE

The First Protocol

(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threa Decoding Problem

Let p be a (big) prime, let us consider

$$G = \{1, g, g^2, \dots, g^{p-2}\} \quad (= (\mathbb{Z}/p\mathbb{Z})^{\times})$$

• Alice: 
$$a \stackrel{\$}{\longleftarrow} \llbracket 0, 1, \dots, p-2 \rrbracket$$
 (long term)

• Bob:  $r \xleftarrow{\single} [0, 1, \dots, p-2]$  (single use)

(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

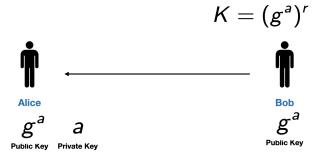
### The First Protocol

Let p be a (big) prime, let us consider

$$G = \{1, g, g^2, \dots, g^{p-2}\} \quad (= (\mathbb{Z}/p\mathbb{Z})^{\times})$$

• Alice: 
$$a \xleftarrow{\ } [0, 1, \dots, p-2]$$
 (long term)

• Bob: 
$$r \xleftarrow{\$} [0, 1, \dots, p-2]$$
 (single use)



(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

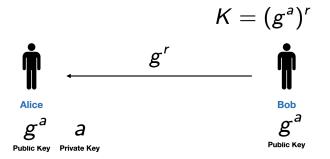
### The First Protocol

Let p be a (big) prime, let us consider

$$G = \{1, g, g^2, \dots, g^{p-2}\} \quad (= (\mathbb{Z}/p\mathbb{Z})^{\times})$$

• Alice: 
$$a \stackrel{\$}{\longleftarrow} \llbracket 0, 1, \dots, p-2 \rrbracket$$
 (long term)

• Bob: 
$$r \xleftarrow{\$} [[0, 1, \dots, p-2]]$$
 (single use)

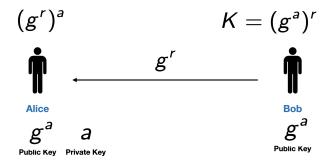


(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

### The First Protocol

Let p be a (big) prime, let us consider

$$G = \{1, g, g^2, \dots, g^{p-2}\} \quad (= (\mathbb{Z}/p\mathbb{Z})^{\times})$$



(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

Quantum Threat and a Worldwide Competition

- Diffie-Hellman key exchange protocol.
- It relies on the hardness of the Discrete Logarithm problem.
- RSA (integer factorization) and DH (discrete logarithm) are widely used nowadays.
- Similar complexities and most importantly: solved by Shor's algorithm (1994).
- **NIST Standardization Process** was announced in 2015-2016, first deadline at the end of 2017.
- 5 new families of problem: lattice, coding theory, multivariate systems, isogeny, hash functions.

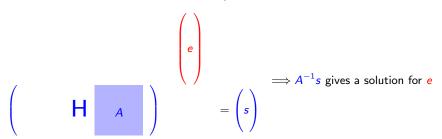
A simple problem in linear algebra

Let k < n be integers,  $H \in \mathbb{F}_q^{(n-k) \times n}$ ,  $e \in \mathbb{F}_q^{n \times 1}$ , and  $s \in \mathbb{F}_q^{(n-k) \times 1}$ .

 $\begin{pmatrix} e \\ H \end{pmatrix} = \begin{pmatrix} s \\ \end{pmatrix}$ 



Let k < n be integers,  $H \in \mathbb{F}_q^{(n-k) \times n}$ ,  $e \in \mathbb{F}_q^{n \times 1}$ , and  $s \in \mathbb{F}_q^{(n-k) \times 1}$ .



• One easily finds one or several solutions for *e*.



Let k < n be integers,  $H \in \mathbb{F}_q^{(n-k) \times n}$ ,  $e \in \mathbb{F}_q^{n \times 1}$ , and  $s \in \mathbb{F}_q^{(n-k) \times 1}$ .

н

• One easily finds one or several solutions for e.

е

• Therefore, one can not control the weight of e for a given metric!

Maxime Bros

= (s)

 $\implies$   $A^{-1}s$  gives a solution for e

Gröbner Bases in Cryptography through the RSD problem

(A)symmetric Cryptography First Asymmetric Scheme Hard Problems and Quantum Threat Decoding Problem

### Decoding Problem

Definition (Syndrome Decoding (SD) Problem - computational version)

**Input:** a parity-check matrix  $H \in \mathbb{F}_{q^m}^{(n-k) \times n}$  of a code C (i.e. a subspace of  $\mathbb{F}_{q^m}^n$ ), an integer  $r \in \mathbb{N}$  and a vector  $s \in \mathbb{F}_{q^m}^{n-k}$ . **Output:** a vector  $e \in \mathbb{F}_{q^m}^n$  such that  $He^T = s^T$  and  $w(e) \leq r$ .

### Definition (Decoding Problem - computational version)

**Input:** a code C (i.e. a subspace of  $\mathbb{F}_{q^m}^n$ ), an integer  $r \in \mathbb{N}$  and a vector  $y \in \mathbb{F}_{q^m}^n$ . **Output:**  $c \in C$  such that  $w(y - c) = w(e) \leq r$ .

Decoding Problem \iff Syndrome Decoding Problem

- Euclidian metric  $\implies$  lattice-based cryptography
- Hamming metric  $\implies$  code-based cryptography
- Rank metric  $\implies$  rank-based cryptography

Introduction to Coding Theory Rank Metric

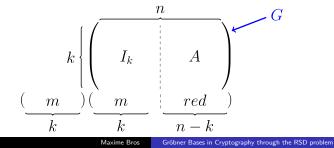
### Introduction to Coding Theory

Error correcting codes are used to transmit informations (satellites, DVD, ...) and also for cryptographic purpose!

### Definition (Code)

A code C is vector space of  $GF(q)^n$  of dimension k.

$$\begin{array}{cccc} \mathcal{E} \colon GF(q)^k & \longrightarrow & GF(q)^n \\ m & \longmapsto & mG \end{array}$$



Introduction to Coding Theory Rank Metric

Reminder about Error Correcting Codes

Definition (Parity Check Matrix)

H is a parity check matrix for the code  $\mathcal C$  if for every word  $c \in GF(q)^n$ :

$$c \in \mathcal{C} \iff Hc^T = \mathbf{0}_{n-k}.$$

### Summary

- G is the generator matrix, H the parity-check matrix of a code.
- Compute *mG* to encode the message *m*, send it.
- The receiver gets y = mG + e, e "small"

 $\Longrightarrow$  decoding instance.

Alternatively, he computes

$$Hy^T = H(mG + e) = Hc^T + He^T = s$$

 $\implies$  syndrome decoding instance.

Introduction to Coding Theory Rank Metric

### Rank Decoding Problem

Definition (Decoding Problem for  $\mathbb{F}_{q^m}$ -linear codes)

**Input:** a code C which is a subspace of  $(\mathbb{F}_{q^m})^n$  of dimension k, and a vector y = c + e where  $c \in C$  and  $\text{Rank}(e) = r \in \mathbb{N}$ . **Output:** c.

*Remark:* the metric considered here is the **Rank metric** in  $(\mathbb{F}_{q^m})^n$ .

### Rank metric on a toy example

Let  $B = \{1, \alpha, \alpha^2, \alpha^3\}$  be a basis of  $\mathbb{F}_{2^4}$  seen as an  $\mathbb{F}_2$ -vector space;  $\alpha^4 = \alpha + 1$ .

Definition Gröbner Basis Linearization

### Algebraic attack

- Algebraic Attack: one models a problem with a system of algebraic equations and solves it.
- In cryptanalysis, the (often unique) solution to this system of equations can be the **private key** or the **plaintext**.
- For the Rank Decoding problem, the solution is the small rank error e;
- Classic approaches:
  - generic Gröbner basis (GB) algorithms
  - specific Linearization techniques

Definition Gröbner Basis Linearization

### Gröbner Basis algorithms

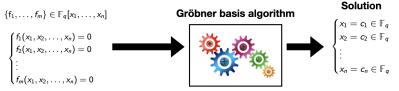
### System of equations

$$\{f_1, \dots, f_m\} \in \mathbb{F}_q[x_1, \dots, x_n]$$
  
 
$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ \vdots \\ f_m(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Definition Gröbner Basis Linearization

### Gröbner Basis algorithms

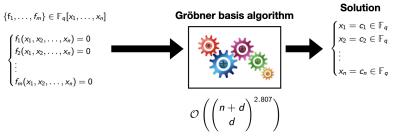
### System of equations



Definition Gröbner Basis Linearization

### Gröbner Basis algorithms

### System of equations



Definition Gröbner Basis Linearization

### Gröbner Basis Complexity

Let us consider this system with quadratic polynomials in  $\mathbb{F}_2[x, y, z]$ 

$$F := \begin{cases} f_1 = xy + xz \\ f_2 = y^2 + yz \\ f_3 = x^2 + yz + 1 \end{cases}$$

Definition Gröbner Basis Linearization

### Gröbner Basis Complexity

Let us consider this system with quadratic polynomials in  $\mathbb{F}_2[x, y, z]$ 

$$F := \begin{cases} f_1 = xy + xz \\ f_2 = y^2 + yz \\ f_3 = x^2 + yz + 1 \end{cases}$$

One wants to compute **S**-polynomials.

 $\implies$  Macaulay matrix of a system at a given degree.

$$\mathcal{M}_{F,2} = \begin{array}{cccccc} x^2 & xy & xz & y^2 & yz & z^2 & 1 \\ f_1 & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ f_3 & \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

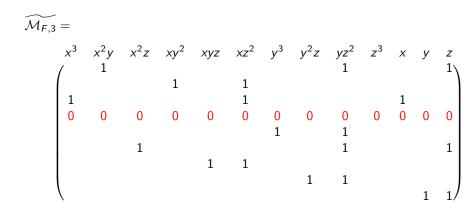
Definition Gröbner Basis Linearization

### Gröbner Basis Complexity

$\mathcal{M}_{F,3} =$													
	<i>x</i> <sup>3</sup>	$x^2y$	$x^2z$	xy <sup>2</sup>	xyz	xz <sup>2</sup>	$y^3$	$y^2z$	yz <sup>2</sup>	$z^3$	x	у	z
$xf_1$	1	1	1										
$xf_2$	1			1	1								
xf <sub>3</sub>	1				1						1		
$yf_1$				1	1								
$yf_2$							1	1					
yf <sub>3</sub>		1						1				1	
$zf_1$					1	1							
$zf_2$								1	1				
$zf_3$	/		1						1				1/

Definition Gröbner Basis Linearization

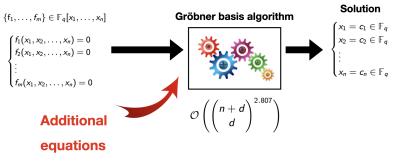
### Gröbner Basis Complexity



Definition Gröbner Basis Linearization

## **Our Previous Attack**

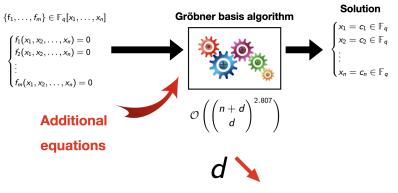
### System of equations



Definition Gröbner Basis Linearization

## **Our Previous Attack**

### System of equations



Definition Gröbner Basis Linearization

# Linearization

- Sometimes the number of equations is greater than the number of **distinct monomials** that appear in the system.
- This allows one to solve the system directly by linearization.
- Thus, one only has to solve a huge linear system and no longer requires generic GB algorithms.
- Moreover, one can take advantage of the sparsity of the system to use Wiedemann's algorithm instead of Strassen's.

## Linearization toy example

$$\begin{cases} f_1 = xz + yz + z \\ f_2 = yz + z + 1 \\ f_3 = xyz + xz + 1 \\ f_4 = xyz + z + 1 \end{cases}, \in \mathbb{F}_2[x, y, z].$$

- We want to find the only point  $(x_0, y_0, z_0) \in (\mathbb{F}_2)^3$  where all these polynomials vanish.
- 20 dictinct monomials of degree less than or equal to 3 in  $\mathbb{F}_2[x, y, z]$ .
- Nevertheless, only 5 of them appear in this system of 4 equations.  $\implies$  One looks for a vector in the right kernel of the form  $(c_1, c_2, c_3, c_4, 1)^\top$

### Linearization toy example

$$\begin{cases} f_1 = xz + yz + z \\ f_2 = yz + z + 1 \\ f_3 = xyz + xz + 1 \\ f_4 = xyz + z + 1 \end{cases}, \in \mathbb{F}_2[x, y, z].$$

- We want to find the only point  $(x_0, y_0, z_0) \in (\mathbb{F}_2)^3$  where all these polynomials vanish.
- 20 dictinct monomials of degree less than or equal to 3 in  $\mathbb{F}_2[x, y, z]$ .
- Nevertheless, only 5 of them appear in this system of 4 equations.  $\implies$  One looks for a vector in the right kernel of the form  $(c_1, c_2, c_3, c_4, 1)^{\top}$

Modeling Complexity

Our Modeling for the Rank Decoding problem

- Recall that we receive the word y = c + e where  $c \in C$  and Rank (e) = r.
- New code:  $\widetilde{\mathcal{C}} = \mathcal{C} + \langle y \rangle$  contains all non-zero multiples  $\lambda e, \ \forall \lambda \in \mathbb{F}_{q^m}^{\times}$ .
- Let H be a parity-check matrix of  $\widetilde{\mathcal{C}}$ .
- Since all words of rank r in  $\tilde{C}$  are multiples of e, we want to solve the following equation:

$$(S_1 S_2 \ldots S_r) CH^{\top} = (0).$$

Remark: the entries of C are in  $\mathbb{F}_q$ 

$$\underbrace{\left(S_1 \ S_2 \ \dots \ S_r\right)}_{e'} \left(CH^{\top}\right) = (0).$$

- $e' \neq 0 \in \text{Ker}(CH^{\top}) \implies \text{Rank}(CH^{\top}) \leq r 1.$ Thus, all maximal minors of  $CH^{\top}$  vanish.
- This modeling (based on Ourivksi and Johansson's one) is considered in Bardet and al., EUROCRYPT 2020.

*Proof:* Cauchy-Binet's formula that generalizes the formula for determinant of square matrices: det(AB) = det(A) det(B)

• Fact 1: consider det(C)<sub>T</sub> as new variables c<sub>T</sub>'s

 $\implies$  It yields to a linear system in the  $c_T$ 's.

• Fact 2: specialization  $I_r$  in C.

$$C = \begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,(n-r)} \\ \vdots & \vdots & \vdots & \vdots \\ C_{r,1} & C_{r,2} & \dots & C_{r,(n-r)} \end{bmatrix}$$

 $\implies$  Those new variables include the coefficients of C!

- This is called the MaxMinors modeling.
- Note that we consider determinants (instead of monomials only) as new (linearization) variables!

Modeling Complexity

Complexity of our attack against Rank Decoding

Recall that we have a **linear system** in the variables  $c_T$ 's arising from the vanishing of maximal minors of  $CH^{\top}$ .

$$igg( igc( igc) _r ^n igc) -1 \qquad ext{variables } c_{\mathcal{T}} ext{'s (in } \mathbb{F}_q) \ igc( igcm( igcm n^{-k-1} igc) igcm) = ext{equations over } \mathbb{F}_q.$$

Complexity of our algorithm against RD: overdetermined case When  $m\binom{n-k-1}{r} \ge \binom{n}{r} - 1$ ,  $\mathcal{O}\left(m\binom{n-k-1}{r}\binom{n}{r}^{\omega-1}\right)$ .

Remark: this linear system is not dense at all, it is sparse, but not sparse enough to benefit from using the Wiedemann approach!

,

Modeling Complexity

Complexity of our attack against Rank Decoding

Recall that we have a **linear system** in the variables  $c_T$ 's arising from the vanishing of maximal minors of  $CH^{\top}$ .

$$\begin{cases} \binom{n}{r} - 1 & \text{variables } c_T \text{'s (in } \mathbb{F}_q), \\ m\binom{n-k-1}{r} & \text{equations over } \mathbb{F}_q. \end{cases}$$

### Super-overdetermined case

One chooses the biggest integer p so that

$$m\binom{n-k-1-p}{r} \ge \binom{n-p}{r} - 1$$
$$\implies \mathcal{O}\left(m\binom{n-k-1-p}{r}\binom{n-p}{r}\right)^{\omega-1}$$

Modeling Complexity

Complexity of our attack against Rank Decoding

Recall that we have a **linear system** in the variables  $c_T$ 's arising from the vanishing of maximal minors of  $CH^{\top}$ .

$$\begin{cases} \binom{n}{r} - 1 & \text{variables } c_T \text{'s (in } \mathbb{F}_q), \\ m\binom{n-k-1}{r} & \text{equations over } \mathbb{F}_q. \end{cases}$$

### Hybrid case

One chooses the smallest integer a so that

$$m\binom{n-k-1}{r} \ge \binom{n-a}{r} - 1$$
$$\implies \mathcal{O}\left(q^{ar}m\binom{n-k-1}{r}\binom{n-a}{r}\right)^{\omega-1}$$

Comparison with previous Attacks Summary of our Contributions

## Comparison with previous Attacks

- Attack in k bits  $\implies$  Require 2<sup>k</sup> bit-operations,
- $\bullet$  Personal computer  $\approx 2^{37}/\textit{second},$  and  $\approx 2^{62}/\textit{year},$
- Previous standard (DES, 1977): 56 bits, broken (late 90's),
- New standard (AES, 2001): from 128 to 256 bits (widely used today).

Comparison with previous Attacks Summary of our Contributions

## Comparison with previous Attacks

- Attack in k bits  $\implies$  Require 2<sup>k</sup> bit-operations,
- $\bullet$  Personal computer  $\approx 2^{37}/\textit{second},$  and  $\approx 2^{62}/\textit{year},$
- Previous standard (DES, 1977): 56 bits, broken (late 90's),
- New standard (AES, 2001): from 128 to 256 bits (widely used today).

	( <i>m</i> , <i>n</i> , <i>k</i> , <i>r</i> )	$\frac{m\binom{n-k-1}{r}}{\binom{n}{r}-1}$	а	р	Def.	Prev.	Last
ROLLO-I-128	(79, 94, 47, 5)	1.97	0	9	128	117	71
ROLLO-I-192	(89, 106, 53, 6)	1.06	0	0	192	144	87
ROLLO-I-256	(113, 134, 67, 7)	0.67	3	0	256	197	151*
ROLLO-II-128	(83, 298, 149, 5)	2.42	0	40	128	134	93
ROLLO-II-192	(107, 302, 151, 6)	1.53	0	18	192	<b>164</b>	111
ROLLO-II-256	(127, 314, 157, 7)	0.89	0	6	256	217	159*
ROLLO-III-128	(101, 94, 47, 5)	2.52	0	12	128	119	70
ROLLO-III-192	(107, 118, 59, 6)	1.31	0	4	192	148	88
ROLLO-III-256	(131, 134, 67, 7)	0.78	0	0	256	200	131*
RQC-I	(97, 134, 67, 5)	2.60	0	18	128	123	77
RQC-II	(107, 202, 101, 6)	1.46	0	10	192	156	101
RQC-III	(137, 262, 131, 7)	0.93	3	0	256	214	144

Comparison with previous Attacks Summary of our Contributions

# **Our Contributions**

• Improved significantly the best known attack against the Rank Decoding problem.

Comparison with previous Attacks Summary of our Contributions

# Our Contributions

- Improved significantly the best known attack against the Rank Decoding problem.
- New algebraic attack against MinRank as well.

Comparison with previous Attacks Summary of our Contributions

## **Our Contributions**

- Improved significantly the best known attack against the Rank Decoding problem.
- New algebraic attack against MinRank as well.
- 2 Rank-based cryptosystems (ROLLO and RQC) did not reach the Third Round of the celebrated NIST Post-Quantum Standardization Process... because of our attacks!

Comparison with previous Attacks Summary of our Contributions

## **Our Contributions**

- Improved significantly the best known attack against the Rank Decoding problem.
- New algebraic attack against MinRank as well.
- 2 Rank-based cryptosystems (ROLLO and RQC) did not reach the Third Round of the celebrated NIST Post-Quantum Standardization Process... because of our attacks!
- Nevertheless, in their report "NISTIR 8309" on the Second Round, NIST emphasized on the importance to keep studying Rank-based cryptography:



Comparison with previous Attacks Summary of our Contributions

# **Our Contributions**

- Improved significantly the best known attack against the Rank Decoding problem.
- New algebraic attack against MinRank as well.
- 2 Rank-based cryptosystems (ROLLO and RQC) did not reach the Third Round of the celebrated NIST Post-Quantum Standardization Process... because of our attacks!
- Nevertheless, in their report "NISTIR 8309" on the Second Round, NIST emphasized on the importance to keep studying Rank-based cryptography:



"Despite the development of algebraic attacks, NIST believes rank-based cryptography should continue to be researched. The rank metric cryptosystems offer a nice alternative to traditional hamming metric codes with comparable bandwidth."