Cryptanalyse en temps polynomial du schéma de McEliece basé sur les codes géométriques

A. Couvreur[1] I. Márquez-Corbella [1] R. Pellikaan [2]

[1] INRIA Saclay & LIX

[2] Department of Mathematics and Computing Science, TU/e.

Journées Codage et Cryptographie 2014

McEliece cryptosystem

- Let $\mathcal{F}$ be any family of linear codes with an **efficient decoding algorithm**.
- Every element of $\mathcal{F}$ is represented by the triple $(\mathcal{C}, \mathcal{A}_{\mathcal{C}}, t)$

$\mathcal{A}_{\mathcal{C}}$ = A decoding algorithm for $\mathcal{C} \in \mathcal{F}$ which corrects up to t errors.

Key Generation

Given:

- Any element $(\mathcal{C}, \mathcal{A}_{\mathcal{C}}, t) \in \mathcal{F}$.
- G a **non structured** gen. matrix of $\mathcal{C}$.

1. **McEliece Public Key:** $\mathcal{K}_{pub} = (G, t)$
2. **McEliece Private Key:** $\mathcal{K}_{secret} = (\mathcal{A}_{\mathcal{C}})$

Encryption

Encrypt a message $\mathbf{m} \in \mathbb{F}_q^k$ as

$$\mathbf{y} = \mathbf{m}G + \mathbf{e}$$

where $\mathbf{e}$ is a random error vector of weight at most t.

Decryption

Using $\mathcal{K}_{secret}$, the receiver obtain $\mathbf{m}$.

- McEliece introduced the first PKC based on Error-Correcting Codes in 1978.
- Advantages:
  1. Interesting candidate for post-quantum cryptography.
  2. Fast encryption (matrix-vector multiplication) and decryption functions.
- Drawback: Large key size.

R. J. McEliece.

A public-key cryptosystem based on algebraic coding theory.
DSN Progress Report, 42-44:114-116, 1978.

How to reduce the key size

➜ There have been many attempts on how to reduce the key size while keeping the same level of security.

M. Baldi and F. Chiaraluce. (2007).

Cryptanalysis of a new instance of McEliece cryptosystem based on QC-LDPC Codes.
In: ISIT 2007, pp. 2591-2595.

T. Berger, P. Cayrel, P. Gaborit and A. Otmani. (2009).
Reducing key length of the McEliece cryptosystem,
In AFRICACRYPT 2009, pp. 77-97.

P. Gaborit. (2005).

Shorter keys for code based cryptography.
In WCC 2005, pp. 81-90.

R. Misoczki and P.S. Barreto. (2009).

Compact McEliece Keys from Goppa Codes,
In Selected Areas in Cryptography, pp. 376ï¿ $\frac{1}{2}$ 392.

C. Monico, J. Rosenthal and A. Shokrollahi. (2000).

Using low density parity check codes in the McEliece cryptosystem
In IEEE International Symposium on Information Theory, p. 215

A. Otmani, J.-P. Tillich and L. Dallot. (2008).

Cryptanalysis of Two McEliece Based on Quasi-Cyclic Codes
In Symbolic Computation and Cryptography (SCC 2008).

➜ There are other public-key primitives based on the theory of Error-Correcting codes:

D. Augot, M. Finiasz and N. Sendrier. (2005).

A family of fast syndrome based cryptographic hash functions.
In Mycrypt 2005, pp. 64-83.

N. Courtois, M. Finiasz and N. Sendrier. (2001).

How to achieve a McEliece-based digital signature scheme.
In ASIACRYPT 2001, pp. 157-174.

P. Gaborit, C. Laudaroux and N. Sendrier. (2007).

SYND: a Very Fast Code-Based Stream Cipher with a Security Reduction.
In: ISIT 2007, pp. 186-190.

Attacks on the McEliece PKC I

➜ We have mainly two different ways of cryptanalyzing the McEliece cryptosystem:

Generic decoding attacks:

The best known techniques:: Information Set Decoding (ISD).

➜ ISD needs exponential time in the code length.

A. Becker, A. Joux, A. May and A. Meurer. (2012).
Decoding random binary linear codes in $2^{n/20}$: How $1+1=0$ improves information set decoding.
In EUROCRYPT 2012. pp. 520-536.

D.J. Bernstein, T. Lange and C. Peters. (2008).
Attacking and Defending the McEliece Cryptosystem,
PQCrypto 2008. pp. 31-46.

D.J. Bernstein, T. Lange and C. Peters. (2011).
Smaller Decoding Exponents: Ball-Collision Decoding,
CRYPTO 2011, pp. 743-760.

A. Canteaut and F. Chabaud. (1998).
A new algorithm for finding minimum-weight words in a linear code: application to McElieceï¿ $\frac{1}{2}$ s cryptosystem and to narrow-sense BCH codes of length 511.
IEEE Transactions on Information Theory 44, 367-378.

M. Finiasz and N. Sendrier. (2009).
Security Bounds for the Design of Code-Based Cryptosystems.
In: ASIACRYPT 2009, pp. 88-105.

P.J. Lee and E.F. Brickell. (1988).
An observation on the security of McElieceï¿ $\frac{1}{2}$ s public-key cryptosystem,
in: EUROCRYPTï¿ $\frac{1}{2}$ 88, pp. 275-280.

A. May, A. Meurer and E. Thomae. (2011).
Decoding random linear codes in $\mathcal{O}(2^{0.054n})$
in: ASIACRYPT, pp. 107-124.

E. Prange. (1962).
The use of information sets in decoding cyclic codes.
IRE Transactions on Information Theory 8, 5-9.

C. Peters. (2011).
Curves, Codes and Cryptography.
Ph.D. thesis. Technische Universiteit Eindhoven.

J. Stern. (1989).
A method for finding codewords of small weight.

Attacks on the McEliece PKC II

Structural attacks:

Retrieve the code structure rather than use an unspecific decoding algorithm.

- ➜ Distinguishing a prescribed structured code from a random one.
- ➜ Structural attacks were efficiently applied to: Reed Solomon codes (Sidelnikov-Shestakov, 1992), Reed-Müller codes (Minder-Shokrollahi, 2007), Concatenated codes (Sendrier, 1994) and some special cases of AG codes.

A. Couvreur, P. Gaborit, V. Gauthier-Umaï¿ $\frac{1}{2}$a and J.P. Tillich, J.P. (2013).
Distinguisher-Based Attacks on Public-Key Cryptosystems using Reed-Solomon Codes,
in: WCC 2013, pp. 180-193.

J.C. Faugère, A. Otmani, L. Perret and J.P. Tillich. (2010).
Algebraic Cryptanalysis of McEliece Variants with Compact Keys,
in: EUROCRYPT 2010, pp. 279-298.

J.C. Faugère, V. Gauthier-Umana, A. Otmani, L. Perret and J.P. Tillich. (2013).
A Distinguisher for High-Rate McEliece Cryptosystems.
IEEE Transactions on Information Theory 59, 6830-6844.

V. Gauthier-Umana and G. Leander, G. (2009).
Practical key recovery attacks on two McEliece variants.
IACR Cryptology ePrint Archive 509.

P. Loidreau and N. Sendrier, N. (2001).
Weak keys in the McEliece public-key cryptosystem.
IEEE Trans. Inform. Theory 47, 1207-1211.

I. Márquez-Corbella and R. Pellikaan. (2012).
Error-correcting pairs for a public-key cryptosystem.
arXiv preprint arXiv:1205.3647.

L. Minder and A. Shokrollahi. (2007).
Cryptanalysis of the Sidelnikov cryptosystem,
in: EUROCRYPT 2007, pp. 347-360.

N. Sendrier. (1994).
On the structure of a randomly permuted concatenated code.
In: EUROCODE 94, pp. 169-173.

N. Sendrier. (2000).
Finding the permutation between equivalent linear codes: the support splitting algorithm.
IEEE Trans. Inform. Theory 46, 1193-1203.

V.M. Sidelnikov and S.O. Shestakov. (1992).
On the insecurity of cryptosystems based on generalized Reed-Solomon codes.
Discrete Math. Appl. 2, 439-444.

Algebraic Geometry Codes I

- Let $\mathcal{X}$ be an algebraic curve of genus g over the finite field $\mathbb{F}_q$.
- $\mathcal{Q} = (Q_1, \ldots, Q_n)$ be an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$.

$D_{\mathcal{Q}}$ denotes the divisor $D_{\mathcal{Q}} = Q_1 + \cdots + Q_n$

- F be an $\mathbb{F}_q$-divisor of $\mathcal{X}$ such that

$$\mathrm{supp}(F) \cap \mathrm{supp}(D_{\mathcal{Q}}) = \emptyset$$

Since $\mathrm{supp}(F) \cap \mathrm{supp}(D_{\mathcal{Q}}) = \emptyset$ the following **evaluation map** is well defined:

$$\begin{array}{rccc} \mathrm{ev}_{\mathcal{Q}} : & L(F) & \longrightarrow & \mathbb{F}_q^n \\ & f & \longmapsto & \mathrm{ev}_P(f) = (f(Q_1), \ldots, f(Q_n)) \end{array}$$

Algebraic Geometry Code (AG code)

The **AG code** associated to the triplet $(\mathcal{X}, \mathcal{Q}, F)$ is:

$$C_L(\mathcal{X}, \mathcal{Q}, F) = \{\mathrm{ev}_{\mathcal{Q}}(f) = (f(Q_1), \ldots, f(Q_n)) \mid f \in L(F)\}$$

Algebraic Geometry Codes II

➜ If $\{f_1, \ldots, f_k\}$ is a basis of $L(F)$ then

$$G = \begin{pmatrix} f_1(Q_1) & \ldots & f_1(Q_n) \\ \vdots & \ddots & \vdots \\ f_k(Q_1) & \ldots & f_k(Q_n) \end{pmatrix} \in \mathbb{F}_q^{k \times n}$$

is a **generator** matrix of the code $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$

Notation: The dimension of a linear code $\mathcal{C}$ will be denoted by $k(\mathcal{C})$ and its minimum distance by $d(\mathcal{C})$.

Theorem I [Parameters of an AG code]

Let $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$. If $\deg(F) = m < n$ then

$$k(\mathcal{C}) \geq m + 1 - g \quad \text{and} \quad d(\mathcal{C}) \geq n - m$$

Moreover, if $n > m > 2g - 2$ then $k(\mathcal{C}) = m - g + 1$.

Algebraic Geometry Codes III

Dual of AG code

Let:

- ω be a **differential form** with a simple pole and residue 1 at Q_j for all $j = 1, \ldots, n$.
- K be the **canonical divisor** of ω.

Then $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)^\perp = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F^\perp)$ with

$$F^\perp = D_\mathcal{Q} - F + K \quad \text{and} \quad \deg(F^\perp) = n - m + 2g - 2$$

Theorem II [Parameters of the dual of an AG code]

Let $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$. If $\deg(F) = m > 2g - 2$ then

$$k(\mathcal{C}^\perp) \geq n - m - 1 + g \quad \text{and} \quad d(\mathcal{C}^\perp) \geq m - 2g + 2$$

Moreover, if $n > m > 2g - 2$ then $k(\mathcal{C}^\perp) = n - m - 1 + g$

GRS for Code-Based PKC I

➜ The class of **GRS** codes was proposed by **Niederreiter** for code-based PKC.

- **Sidelnikov-Shestakov** in 1992 introduced an algorithm that breaks this proposal in polynomial time.

H. Niederreiter. (1986).
Knapsack-type crypto system and algebraic coding theory.
Problems of Control and Information Theory.

V.M. Sidelnikov and S.O. Shestakov. (1992).
On the insecurity of cryptosystems based on generalized Reed-Solomon codes.
Discrete Math. Appl. 2, 439-444.

- **Berger and Loidreau** in 2005 propose another version of the Niederreiter scheme designed to resist the Sidelnikov-Shestakov attack.
  - ➜ **Main idea:** work with subcodes of the original GRS code.
- **Attacks:**
  1. **Wieschebrink:** (2006, 2010)
     - Presents the first feasible attack to the Berger-Loidreau cryptosystem but is impractical for small subcodes.
     - Notes that if the square code of a subcode of a GRS code of parameters $[n, k]$ is itself a GRS code of dimension $2k - 1$ then we can apply Sidelnikov-Shestakov attack.
  2. **M-Mártinez-Pellikaan:** (2012) Give a characterization of the possible parameters that should be used to avoid attacks on the Berger-Loidreau cryptosystem.

GRS for Code-Based PKC II

T. Berger and P. Loidreau.

How to mask the structure of codes for a cryptographic use.
Designs, Codes and Cryptography, 35: 63-79, 2005.

I. Márquez-Corbella, E. Martínez-Moro and R. Pellikaan.
The non-gap sequence of a subcode of a generalized Reed-Solomon code.
Designs, Codes and Cryptography, 66, 317-333.

M. Baldi, M. Blanchi, F. Chiaraluce, J. Rosenthal, D. Schipani. (2011).
Enhanced public key security for the McEliece cryptosystem. ArXiv Preprint.

A. Couvreur, P. Gaborit, V. Gauthier-Umaï¿ $\frac{1}{2}$ a and J.P. Tillich, J.P. (2013).
Distinguisher-Based Attacks on Public-Key Cryptosystems using Reed-Solomon Codes,
in: WCC 2013, pp. 180-193.

C. Wieschebrink.

An attack on the modified Niederreiter encryption scheme.
In PKC 2006, Lecture Notes in Computer Science, volume 3958, 14-26, Berlin, 2006. Springer.

C. Wieschebrink.

Cryptoanalysis of the Niederreiter public key scheme based on GRS subcodes.
In Post-Quantum Cryptography, Lecture Notes in Computer Science, volume 6061, 6-72, Berlin, 2010. Springer.

- **Wieschebrick** (2006) and **Baldi et al.** (2011) proposed other variants of the Niederreiter scheme.
- **Attacks: Couvreur et al.** (2013) provide a cryptanalysis of these schemes.

C. Wieschebrink.

Two NP-complete problems in coding theory with an application in code based cryptography.
In IEEE Inf. Theory, pp. 1733-1737.

AG codes for code-based PKC

- In 1996 **Janwa and Moreno** propose to use AG codes for the McEliece cryptosystem.
- This system was broken for:
  1. Codes on curves of genus $g = 0$ by the **Sidelnikov-Shestakov** attack.

     GRS codes are Algebraic Geometry codes on the projective line.
  2. Codes on curves of genus $g \leq 2$ by **Faure and Minder**.
  3. We can retrieve the structure of the code (in polynomial time) of **VSAG** codes by **M-Martínez-Pellikaan-Ruano**

Very Strong Algebraic-Geometric (VSAG) codes

A code $\mathcal{C}$ has a VSAG representation if $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, P, E)$ where the curve $\mathcal{X}$ has genus g, P consists of n points and E has degree m such that

$$2g + 2 < m < \tfrac{1}{2}n \quad \text{or} \quad \tfrac{1}{2}n + 2g - 2 < m < n - 4$$

C. Faure and L. Minder.
Cryptanalysis of the McEliece cryptosystem over hyperelliptic codes.
Proceedings 11th Int. Workshop on Algebraic and Combinatorial Coding Theory, 2008.

H. Janwa and O. Moreno.
McEliece public crypto system using algebraic-geometric codes.
Designs, Codes and Cryptography, 1996.

I. Márquez-Corbella, E. Martínez-Moro and R. Pellikaan.
On the unique representation of very strong algebraic geometry codes.
Designs, Codes and Cryptography, 2013.

I. Márquez-Corbella, E. Martínez-Moro, R. Pellikaan and D. Ruano.
Computational aspects of retrieving a representation of an algebraic geometry code.
Journal of Symbolic Computation.

Notation

- For all $\mathbf{a}, \mathbf{b} \in \mathbb{F}_q^n$ we define:
  - **Star Multiplication:** $\mathbf{a} * \mathbf{b} = (a_1 b_1, \ldots, a_n b_n) \in \mathbb{F}_q^n$
  - **Standard Inner Multiplication:** $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i=1}^{n} a_i b_i$

- For all subsets $A,\ B \subseteq \mathbb{F}_q^n$ we define:
  - $A * B = \langle \{ \mathbf{a} * \mathbf{b} \mid \mathbf{a} \in A \text{ and } \mathbf{b} \in B \} \rangle$

    For $B = A \Longrightarrow A * A$ is denoted as $A^{(2)}$
  - $A \perp B \Longleftrightarrow \langle \mathbf{a}, \mathbf{b} \rangle = 0 \quad \forall\, \mathbf{a} \in A \text{ and } \mathbf{b} \in B$

Decoding by Error Correcting Pairs

Error-correcting pairs (ECP)

Let $\mathcal{C}$ be an $\mathbb{F}_q$ linear code of length n. The pair (A, B) of $\mathbb{F}_{q^m}$-linear codes of length n is a t-ECP for $\mathcal{C}$ over $\mathbb{F}_{q^m}$ if the following properties hold:

- E.1 $(A * B) \perp \mathcal{C}$.
- E.2 $k(A) > t$.
- E.3 $d(B^{\perp}) > t$.
- E.4 $d(A) + d(\mathcal{C}) > n$.

An $[n, k]$ code which has a t-ECP over $\mathbb{F}_{q^m}$ has a decoding algorithm with complexity $\mathcal{O}\left((nm)^3\right)$.

R. Pellikaan

On decoding by error location and dependent sets of error positions.
Discrete Math., 106–107: 369–381 (1992).

R. Kötter.

A unified description of an error locating procedure for linear codes.
In Proceedings of Algebraic and Combinatorial Coding Theory, 113–117. Voneshta Voda (1992).

t-ECP for AG codes

Consider the AG code $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)^\perp$ where

- $\mathcal{X}$ is an algebraic curve of genus g over $\mathbb{F}_q$
- $\mathcal{Q} = (Q_1, \ldots, Q_n)$ is an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$.
- F is an $\mathbb{F}_q$-divisor of $\mathcal{X}$ such that

$$\deg(F) = m > 3g - 1 \quad \text{and} \quad \operatorname{supp}(F) \cap \operatorname{supp}(D_{\mathcal{Q}}) = \emptyset$$

Theorem [Pellikaan - 1992]

The pair of codes (A, B) defined by

$$A = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, E) \quad \text{and} \quad B = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - E)$$

with $m > \deg(E) \geq t + g$ is a t-ECP for $\mathcal{C}$.
Such a pair always exists whenever $m > 2g - 2$ and $t = \left\lfloor \frac{d^* - 1 - g}{2} \right\rfloor$.

Corollary [Main Corollary]

Let us define $A_0 = (B * \mathcal{C})^\perp$. Then (A_0, B) is a t-ECP for $\mathcal{C}$

In order to compute a t-ECP for $\mathcal{C}$, it suffices to compute a code of type $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - E)$ for some divisor E with $\deg(E) \geq t + g$.

Context of the Cryptosystem

From now on:

- $\mathcal{X}$ is an algebraic curve of genus g over $\mathbb{F}_q$
- $\mathcal{Q} = (Q_1, \ldots, Q_n)$ is an n-tuple of mutually distinct $\mathbb{F}_q$-rational points of $\mathcal{X}$.
- F is an $\mathbb{F}_q$-divisor of $\mathcal{X}$ such that

$$\deg(F) = m > 3g - 1 \quad \text{and} \quad \operatorname{supp}(F) \cap \operatorname{supp}(D_{\mathcal{Q}}) = \emptyset$$

Public Key:

$$\mathcal{K}_{\text{pub}} = G \quad \text{and} \quad t = \left\lfloor \frac{d^* - g - 1}{2} \right\rfloor$$

where:

- G is a generator matrix of the **public code**:

$$\mathcal{C}_{\text{pub}} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)^{\perp}$$

- $d^* = m - 2g + 2$ is the designed minimum distance of $\mathcal{C}_{\text{pub}}$

- Our t seems reasonable if $\mathcal{K}_{\text{secret}}$ is based on ECP.

$$t = \left\lfloor \frac{d^* - g - 1}{2} \right\rfloor \leq \left\lfloor \frac{d^* - 1}{2} \right\rfloor = \text{actual error-correction capability of } \mathcal{C}$$

- **Future work!!!**

The P-Filtration

- Let P be a point of the n-tuple $\mathcal{Q}$.
- We focus on the sequence of codes:

$$\mathcal{B}_i := (\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - iP))_{i \in \mathbb{N}}$$

Which Elements of the Sequence do We know?

1. From a generator matrix of $\mathcal{C}_{\text{pub}} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)^{\perp}$ one can compute $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$
   - Computed by **Gaussian elimination**.
2. $\mathcal{B}_0 =$ the code $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$ **punctured** at position P.
3. $\mathcal{B}_1 =$ the code $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$ **shortened** at position P.
   - Computed by **Gaussian elimination**.

The codes $\mathcal{B}_0$ and $\mathcal{B}_1$ are **known**.

Effective computation - Algorithm I

How to compute $\mathcal{B}_2$?

- If $\frac{n+1+g}{2} > m$, then $\mathcal{B}_1^{(2)} \subsetneq \mathbb{F}_q^n$.
- If $\deg(F-P) = m-1 \geq 2g+1$, then

$$\mathcal{B}_1^{(2)} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F-P)^{(2)} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, 2F-2P)$$

Therefore if $\frac{n+1+g}{2} > m \geq 2g+1$ then $\mathcal{B}_2$ is the solution space of the following problem

$$\mathbf{z} \in \mathcal{B}_1 \quad \text{and} \quad \mathbf{z} * \mathcal{B}_0 \subseteq (\mathcal{B}_1)^{(2)} . \tag{1}$$

Proposition

Let F, G two divisors on $\mathcal{X}$ such that $\deg(F) \geq 2g$ and $\deg(G) \geq 2g+1$. Then,

$$\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F) * \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, G) = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F+G)$$

Effective computation - Algorithm I

Theorem I

If $s \geq 1$ and $\frac{n+1+g}{2} > m \geq 2g + s + 1$, then $\mathcal{B}_{s+1}$ is the solution space of the following problem

$$\mathbf{z} \in \mathcal{B}_s \quad \text{and} \quad \mathbf{z} * \mathcal{B}_{s-1} \subseteq (\mathcal{B}_s)^{(2)} . \tag{2}$$

$(t + g)$ repeated applications of **Theorem I** determines the code $\mathcal{B}_{t+g}$.

Algorithm 5.1: Let $\frac{n+1+g}{2} > m \geq 3g + t + 1$

Data: Generator matrices for the codes $\mathcal{B}_0$ and $\mathcal{B}_1$
Result: A generator matrix for the code $\mathcal{B}_{t+g}$
for $s = 2, \ldots, t + g$ **do**
  Compute $\mathcal{B}_s$ from the codes $\mathcal{B}_{s-1}$ and $\mathcal{B}_{s-2}$ using **Theorem I**.

Algorithm complexity: We solve $(t + g)$ systems of linear equations of type (2).

Effective computation - Algorithm II

We can do **better** by **decreasing** the number of iterations and **relaxing** the parameters conditions $\Rightarrow$ Algorithm II

➜ **Algorithm I:**

$$\mathcal{B}_0 \supseteq \mathcal{B}_1 \supseteq \mathcal{B}_2 \supseteq \mathcal{B}_3 \supseteq \ldots \supseteq \mathcal{B}_{t+g-1} \supseteq \mathcal{B}_{t+g}$$

Solve $(t+g)$ systems of linear equations

➜ **Algorithm II:**

$$\mathcal{B}_0 \supseteq \mathcal{B}_1 \supseteq \mathcal{B}_2 \supseteq \mathcal{B}_4 \supseteq \ldots \supseteq \mathcal{B}_{\frac{t+g}{2}} \supseteq \mathcal{B}_{t+g}$$

Solve $2\lceil \log_2(t+g) \rceil + 2$ systems of linear equations

Effective computation - Algorithm II

Theorem II

If $\frac{n+1+g}{2} > m \geq 2g + \lfloor \frac{s}{2} \rfloor + 1$, then $\mathcal{B}_s$ is the solution space of the following problem

$$\mathbf{z} \in \mathcal{B}_{\lfloor s/2 \rfloor} \quad \text{and} \quad \mathbf{z} * \mathcal{B}_0 \subseteq \mathcal{B}_{\lfloor s/2 \rfloor} * \mathcal{B}_{\lfloor s/2 \rfloor + 1}. \tag{3}$$

Algorithm 5.2: Let $\frac{n+1+g}{2} > m \geq \frac{5g+t}{2} + 1$

Data: Generator matrices for the codes $\mathcal{B}_0$ and $\mathcal{B}_1$
Result: A generator matrix for the code $\mathcal{B}_{t+g}$
Let $b = \lfloor \log_2 n \rfloor + 1$, then n satisfies: $2^{b-1} \leq n < 2^b$. Therefore, $2 \leq n/2^{b-2} < 4$.

Compute $\mathcal{B}_2$ and $\mathcal{B}_3$ using **Theorem I**;
for $s = b - 2, \ldots, 1$ **do**
  Compute the codes $\mathcal{B}_{\lfloor (t+g)/2^{s-1} \rfloor}$ and $\mathcal{B}_{\lfloor (t+g)/2^{s-1} \rfloor + 1}$ from the codes $\mathcal{B}_{\lfloor (t+g)/2^{s} \rfloor}$ and $\mathcal{B}_{\lfloor (t+g)/2^{s} \rfloor + 1}$ by applying twice **Theorem II**

Algorithm complexity: We solve $2\lceil \log_2(t+g) \rceil + 2$ systems of linear equations of type (2) and (3).

Extending the attack I

We have been working under the assumption that

$$m \leq \frac{n+1+g}{2}$$

On the following we see **how this conditions can be weakened!!!**.

The code $C(J)$

Let $J \subseteq \{1, \ldots, n\}$. We define the code $C(J)$ as:

$$C(J) = \{\mathbf{c} \in \mathcal{C} \mid c_j = 0, \text{ for all } j \in J\}$$

Lemma I §5.2

Let $I \subseteq \{2, \ldots, n\}$ with $|I| = i$ elements and $Q_1, \ldots, Q_n$ be n different rational points. Take the code

$$\mathcal{B}_1(I) = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - Q_1 - \sum_{j \in I} Q_j).$$

If $\frac{n+i+g}{2} > m \geq 2g+1$, then $(\mathcal{B}_1(I))^{(2)} \subsetneq \mathbb{F}_q^{n-i-1}$.

Extending the attack II

Lemma II §5.2

Let:

- $Q_1, \ldots, Q_n$ be n different rational points.
- $\mathcal{C} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - Q_1)$.
- $I_1, \ldots, I_s$ be different subsets of $\{2, \ldots, n\}$ such that

$$\cap_{j=1}^{s} I_j = \emptyset \text{ and } k(\mathcal{C}) - |I_j| \geq |\cap_{i=j+1}^{s} I_i| - |\cap_{i=j}^{s} I_i|$$

Then $\mathcal{C} = \mathcal{C}(I_1) + \cdots + \mathcal{C}(I_s) = \sum_{j=1}^{s} \mathcal{C}(I_j)$.

Extending the attack III

Suppose there exists $i > 1$ such that

$$\frac{n+i+g}{2} > m > \frac{n+1+g}{2}.$$

Then:

1. Find different subsets $I_1, \ldots, I_s \subseteq \{2, \ldots, n\}$ with at least i elements each satisfying the assumptions of **Lemma II §5.2**
2. Apply any Algorithm from §5.1 to obtain the codes

$$\mathcal{B}_{t+g}(I_j) = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - (t+g)\left(Q_1 - \sum_{i \in I_j} Q_i\right)) \text{ with } j = 1, \ldots, s$$

3. Obtain $\mathcal{B}_{t+g}$ from **Lemma II §5.2**

From degenerate to non degenerate

Unfortunately the codes $\mathcal{B}_i$ are **degenerated** for $i > 0$.

➜ Them following result allows to compute a generator matrix of

$$\underbrace{\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - E')}_{\text{with } \mathrm{supp}(F-E')\cap\mathrm{supp}(D_{\mathcal{Q}})=\emptyset} \quad \text{from the codes } \mathcal{B}_{t+g} \text{ and } \mathcal{B}_{t+g+1}.$$

Theorem

Let G be a generator matrix of $\mathcal{B}_{t+g}$ of the form:

$$G = \left(\begin{array}{c|c} 0 & \mathbf{c}_1 \\ \hline \mathbf{0} & G_1 \end{array}\right), \text{ where}$$

$\left(\begin{array}{c|c} 0 & \mathbf{c}_1 \end{array}\right) \in \mathcal{B}_{t+g} \setminus \mathcal{B}_{t+g-1}$ and $\left(\begin{array}{c|c} \mathbf{0} & G_1 \end{array}\right)$ is a gen. matrix of $\mathcal{B}_{t+g+1}$

Then the following matrix is a generator matrix for a code of type

$$\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - E') \text{ with } \mathrm{supp}(F - E') \cap \mathrm{supp}(D_{\mathcal{Q}})$$

$$\hat{G} = \left(\begin{array}{c|c} 1 & \mathbf{c}_1 \\ \hline \mathbf{0} & G_1 \end{array}\right)$$

The Attack

Public Key: $\mathcal{K}_{\text{pub}} = G$ and $t = \left\lfloor \frac{d^* - g - 1}{2} \right\rfloor$ where:

- G is a generator matrix of the **public code**: $\mathcal{C}_{\text{pub}} = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)^{\perp}$
- $d^* = m - 2g + 2$ is the designed minimum distance of $\mathcal{C}_{\text{pub}}$

The Algorithm:

Suppose that $\frac{n+1+g}{2} \geq m$. Otherwise we apply techniques set on §5.2

Step 1. Determine the values g and m using the following Proposition.

Proposition

If $2g + 1 \leq m < \frac{1}{2}n$. Let k_1 and k_2 be the dimension of $\mathcal{C}$ and $\mathcal{C}^{(2)}$, respectively. Then,

$$m = k_2 - k_1 \quad \text{and} \quad g = k_2 - 2k_1 + 1 \quad .$$

Step 2. Compute $\mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F)$ by Gaussian elimination.
Step 3. Compute the code $B = \mathcal{C}_L(\mathcal{X}, \mathcal{Q}, F - (t + g)P_1)$, using one of the algorithms described in §5.1
Step 4. Deduce from B a non degenerated code $\hat{B}$ using §5.3
Step 5. Apply the following Corollary to deduce an ECP. from $\hat{B}$.

Corollary

Let us define $A_0 = (B * \mathcal{C})^{\perp}$. Then (A_0, B) is a t-ECP for $\mathcal{C}$.

Complexity

- The **costly part** of the attack is the computation of the code B

  $\Longrightarrow$ We can apply one of the algorithms of §5.1

  Computing a generator matrix of $\mathcal{C}^{(2)}$ and then apply Gaussian elimination to such matrix has complexity:

  $$O\left(\binom{k}{2} n + \binom{k}{2} n^2\right) \sim O\left(k^2 n^2\right).$$

- Roughly speaking the cost of our attack is about $O\left((\lambda+1)n^4\right)$ where:

  1. $\lambda =$ Linear systems to solve depending on the chosen algorithm from §5.1
  2. The term $(\lambda + 1)$ is due to §5.3

Conclusions

- We constructed a **polynomial-time** algorithm which breaks the **McEliece scheme based on AG codes** whenever

$$2 < t \leq \left\lfloor \frac{d^* - g - 1}{2} \right\rfloor$$

- **COMPLEXITY:** $O\left(n^4\right)$

- **Future work:** using the concept of Error-Correcting Arrays (ECA) or well-behaving sequence obtain an attack for

$$t = \left\lfloor \frac{d^* - 1}{2} \right\rfloor$$

Thank you for your attention!