Post-quantum cryptography based on error-correcting codes

Dutto Simone

Ph.D. Student in Pure and Applied Mathematics
Università degli Studi di Torino & Politecnico di Torino, in collaboration with Telsy S.p.A.

La De Cifris incontra Torino - 14 Ottobre 2019

Introduction
PQC standardization

In 2017, the National Institute of Standard and Technologies (NIST) started a process to find standard post-quantum public-key cryptosystems.

In 2019, the 69 initial proposals were reduced to 26 alternatives based on:

- lattices (9 PKEs/KEMs + 3 signatures)
- error-correcting codes (7 PKEs/KEMs)
- multivariate polynomials (4 signatures)
- symmetric-key (2 signatures)
- isogenies on supersingular EC (1 PKE/KEM)

Error-correcting codes

Linear codes

The code-based alternatives for PQC rely on linear error-correcting codes.

Definition

A linear error-correcting code of length n and rank k is a linear vector subspace with dimension k, $C \subseteq \mathbb{F}_q^n$, where $\mathbb{F}_q$ is the finite field with q elements.

The elements $\mathbf{c} \in C$ are called codewords.

Generator and check matrices

Definition

The codewords in a basis of $C \subseteq \mathbb{F}_q^n$ can be collocated in the rows of a matrix $\mathbf{G} \in \mathbb{F}_q^{k,n}$ called generator matrix, which verifies $\forall \mathbf{m} \in \mathbb{F}_q^k$, $\mathbf{m} \cdot \mathbf{G} \in C$.

Definition

Given $C \subseteq \mathbb{F}_q^n$, the matrix $\mathbf{H} \in \mathbb{F}_q^{n-k,n}$ that verifies

$$\mathbf{x} \cdot {}^T\mathbf{H} = \mathbf{0} \Leftrightarrow \mathbf{x} \in C$$

is called parity-check matrix of C.

Hamming and rank distances

Definition

The Hamming distance between two words $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$ is the number of non-zero entries of $\mathbf{x} - \mathbf{y}$.

Definition

If $C \subseteq \mathbb{F}_{q^N}^n$ and $\{u_1, \ldots, u_N\}$ is a basis of $\mathbb{F}_{q^N}$ over $\mathbb{F}_q$, $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^N}^n$ has $x_j = x_{1,j}u_1 + \cdots + x_{N,j}u_N \, \forall j$, so that $\mathbf{x}$ can be seen as a matrix $\mathbf{X} = (x_{i,j}) \in \mathbb{F}_q^{N,n}$.
The rank distance between $\mathbf{x}, \mathbf{y} \in \mathbb{F}_{q^N}^n$ is $rank(\mathbf{X} - \mathbf{Y})$.

The minimum distance between distinct codewords of a code C is called distance of C and indicated as d.

Error correction

When an error $\mathbf{e} \in \mathbb{F}_q^n$ occurs, $\mathbf{c}' = \mathbf{c} + \mathbf{e}$ is received.
If $d(\mathbf{0}, \mathbf{e}) < \lfloor \frac{d-1}{2} \rfloor$ then $\mathbf{c}$ is the closest codeword to $\mathbf{c}'$, otherwise the correction fails.

The best correction strategy exploits that:

$$\mathbf{c}' \cdot {}^T\mathbf{H} = (\mathbf{c} + \mathbf{e}) \cdot {}^T\mathbf{H} = \mathbf{0} + \mathbf{e} \cdot {}^T\mathbf{H} = \mathbf{s} \neq \mathbf{0} \,.$$

The vector $\mathbf{s}$ is called syndrome of the error $\mathbf{e}$.

Syndrome decoding consists in precompute a table with syndromes and relative minimum-distance causing error, so that a simple look-up can correct an error.

Code-based cryptography

Security basic problems

The security of code-based cryptography relies on the hardness of the problem behind the syndrome decoding.

Definition

The (decisional) Maximum Likelihood Decoding (MLD) problem is defined as: given $\mathbf{H} \in \mathbb{F}_q^{m,n}, \mathbf{s} \in \mathbb{F}_q^m$ and $t \in \mathbb{N}$, does exists $\mathbf{x} \in \mathbb{F}_q^n \,|\, \mathbf{x} \cdot {}^T\mathbf{H} = \mathbf{s}$ and $d(\mathbf{0}, \mathbf{x}) = t$?

Another problem on which security can be based is the distinguishing problem, since some particular codes are difficult to differentiate from random linear codes.

Basic cryptosystems

The code-based proposals in NIST selection rely on:

- McEliece cryptosystem:
  1. Classic McEliece
  2. NTS-KEM
- similar Learning-With-Errors cryptosystem:
  1. BIKE
  2. HQC
  3. LEDAcrypt

     } Hamming distance
  4. ROLLO
  5. RQC

     } rank distance

McEliece cryptosystem

Robert McEliece introduced this public-key encryption algorithm in 1978, but it remained unused until now.

The main requirement is an efficiently decodable linear code, generated by $\mathbf{G} \in \mathbb{F}_q^{k,n}$ and with distance d.

The original algorithm and the post-quantum proposals use Goppa codes. They are algebraic geometric linear codes constructed from non-singular projective curves over $\mathbb{F}_q$. Their efficient decoding algorithm was discovered in 1975 by Nicholas J. Patterson.

Key generation. $pk_{\mathcal{A}} = (\hat{\mathbf{G}}, t)$ and $sk_{\mathcal{A}} = (\mathbf{S}, \mathbf{G}, \mathbf{P})$, where:

- $\mathbf{G} \in \mathbb{F}_q^{k,n}$ generates an efficiently decodable linear code able to correct $t = \lfloor \frac{d-1}{2} \rfloor$ errors
- $\mathbf{S} \in \mathbb{F}_q^{k,k}$ is a non-singular matrix
- $\mathbf{P} \in \mathbb{F}_q^{n,n}$ is a permutation matrix
- $\hat{\mathbf{G}} = \mathbf{S} \cdot \mathbf{G} \cdot \mathbf{P}$

Message encryption. To send $\mathbf{m} \in \mathbb{F}_q^k$ to $\mathcal{A}$, $\mathcal{B}$ has to:

- obtain the codeword $\mathbf{c} = \mathbf{m} \cdot \hat{\mathbf{G}}$
- send $\mathbf{c}' = \mathbf{c} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{F}_q^n$ is an error of weight t

Message decryption. $\mathcal{A}$ obtains $\mathbf{m}$ by:

- computing $\mathbf{c}' \cdot \mathbf{P}^{-1} = \mathbf{m} \cdot \mathbf{S} \cdot \mathbf{G} + \mathbf{e} \cdot \mathbf{P}^{-1}$
- efficiently decoding to $\mathbf{m} \cdot \mathbf{S}$ (the error has weight t)

Niederreiter cryptosystem

The dual version of the McEliece cryptosystem, called Niederreiter cryptosystem, is also important. It uses the check matrix $\mathbf{H} \in \mathbb{F}_q^{n-k,n}$ instead of the generator $\mathbf{G}$.

Key generation. $pk_{\mathcal{A}} = (\hat{\mathbf{H}}, t)$ and $sk_{\mathcal{A}} = (\mathbf{S}, \mathbf{H}, \mathbf{P})$, where all is as before except $\hat{\mathbf{H}} = \mathbf{S} \cdot \mathbf{H} \cdot \mathbf{P}$.

Message encryption. $\mathbf{m} \in \mathbb{F}_q^k$ has weight at most t and $\mathcal{B}$ sends $\mathbf{c} = \hat{\mathbf{H}} \cdot^T \mathbf{m}$ to $\mathcal{A}$.

Message decryption. $\mathcal{A}$ obtains $\mathbf{m}$ by:

- computing $\mathbf{S}^{-1} \cdot \mathbf{c} = \mathbf{H} \cdot \mathbf{P} \cdot^T \mathbf{m}$
- efficiently syndrome decoding to $\mathbf{P} \cdot^T \mathbf{m}$

PQ cryptosystems based on McEliece

Classic McEliece is based on the Niederreiter cryptosystem. Security is based on the MLD and on the distinguishing problem for Goppa codes.
Pros: short ciphertexts, good performance, no failures.
Cons: large public key size.

NTS-KEM exploits both McEliece and Niederretier cryptosystems. As before, security is based on the MLD and on the distinguishing problem for Goppa codes.
Pros: short ciphertexts, good performance.
Cons: large public key size, possible failures.

Similar Learning-With-Errors cryptosystem

Learning-With-Error (LWE) cryptosystems are lattice-based, another branch of PQ cryptography. They rely on the difficulty of distinguish a particular distribution from a random one.

With an analogous scheme, a cryptosystem can be based on the distinguishing problem of a error-correcting code. These alternatives exploit Quasi-Cyclic (QC) codes ($C \in \mathbb{F}_q^n$ is quasi-cyclic if it is closed with respect to a left shift of b places, where b is coprime to n).

Key generation. $pk_{\mathcal{A}} = (\mathbf{G}, \mathbf{a}, \mathbf{b})$ and $sk_{\mathcal{A}} = (\mathbf{s})$, where $\mathbf{G} \in \mathbb{F}_q^{k,n}$ generates an efficiently decodable linear code able to correct t errors, $\mathbf{a}, \mathbf{s}, \mathbf{r} \in \mathbb{F}_q^n$ and $\mathbf{b} = \mathbf{a} * \mathbf{s} + \mathbf{r}$.

Message encryption. To send $\mathbf{m} \in \mathbb{F}_q^k$ to $\mathcal{A}$, $\mathcal{B}$ has to:

- generate $\mathbf{s}', \mathbf{r}_1, \mathbf{r}_2 \in \mathbb{F}_q^n$
- send $\mathbf{b}' = \mathbf{a} * \mathbf{s}' + \mathbf{r}_1$ and $\mathbf{c} = \mathbf{m} \cdot \mathbf{G} + \mathbf{b} * \mathbf{s}' + \mathbf{r}_2$

Message decryption. $\mathcal{A}$ obtains $\mathbf{m}$ by efficiently decoding

$$\begin{aligned}\mathbf{c} - \mathbf{b}' * \mathbf{s} &= (\mathbf{m} \cdot \mathbf{G} + \mathbf{b} * \mathbf{s}' + \mathbf{r}_2) - (\mathbf{a} * \mathbf{s}' + \mathbf{r}_1) * \mathbf{s} \\ &= \mathbf{m} \cdot \mathbf{G} + (\mathbf{r} * \mathbf{s}' + \mathbf{r}_2 - \mathbf{r}_1 * \mathbf{s}) = \mathbf{m} \cdot \mathbf{G} + \mathbf{e}\,.\end{aligned}$$

The decoding fails unless the weight of $\mathbf{e}$ is less than t. There are restrictions for the weights of $\mathbf{s}, \mathbf{r}, \mathbf{s}', \mathbf{r}_1$ and $\mathbf{r}_2$, but the failure rate is not zero.

PQ cryptosystems similar to LWE (Hamming metric)

BIKE exploits QC Moderate-Density-Parity-Check codes ($\mathbf{H}$ has row weight $w = O(\sqrt{n})$).
Pros: good key and ciphertexts sizes and performance.
Cons: possible failures.

HQC is based on Syndrome Decoding for QC codes.
Pros: good performance, lower failure rate.
Cons: larger key and ciphertexts sizes.

LEDAcrypt relies on QC Low-Density-Parity-Check codes (constructed using a sparse bipartite graph).
Pros: good key and ciphertexts sizes and performance.
Cons: possible failures.

PQ cryptosystems similar to LWE (rank metric)

ROLLO collects and refines some parameters of three similar schemes based on Low-Rank-Parity-Check codes (similar to LDPC but with the rank metric).
Pros: good key and ciphertexts sizes and performance.
Cons: possible failures.

RQC exploits the Ideal Rank Syndrome Decoding (as the one with QC codes but based on the rank).
Pros: good key size, no failures.
Cons: larger ciphertexts, slower decryption.

Conclusions

Type	Public Key	Ciphertext/Signature
Lattice	medium	medium
Goppa Code	large	small
QC Code	medium	medium
Multivariate HFE	large	small
Multivariate UOV	medium	small
Multivariate MQ	small	large
Hash	small	large
Isogeny	small	small
ZKP	small	large

Figure 1: Sizes of data in post-quantum types.

Type	Key Generation	Encryption/Verification	Decryption/Signing
Lattice	fast	fast	fast
Code	slow	fast	medium
Multivariate	slow	fast	medium
Hash	slow	fast	slow
Isogeny	slow	slow	slow
ZKP	medium	slow	slow

Figure 2: Performance speed of subroutines in post-quantum types.

Thank you for your attention!

Cryptography and Number Theory group, DISMA, PoliTo
https://crypto.polito.it/

Personal mail address
simone.dutto@polito.it