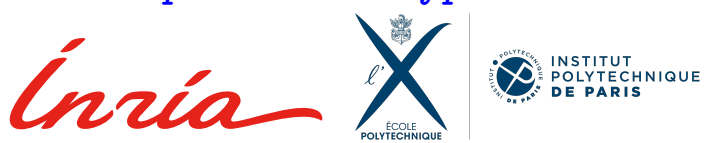
Code-based cryptography: 101

Class 1

Gustavo Banegas

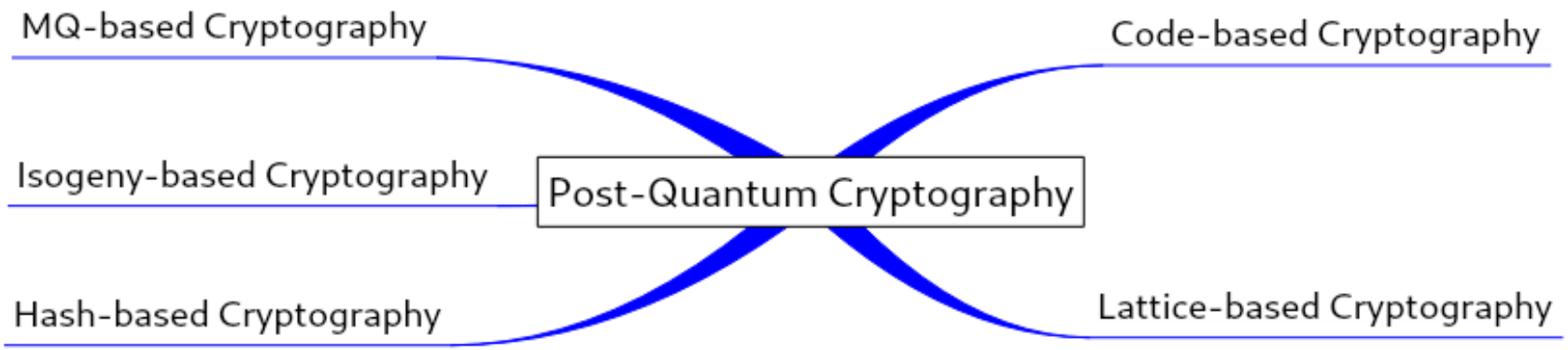
Inria & Laboratoire d'Informatique de l'Ecole polytechnique, France
gustavo@cryptme.in
https://www.cryptme.in



May 25, 2022

Outline

What is Post-quantum Cryptography?



First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$$1 + 1 = ?$$

First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$$1 + 1 = ?$$

$$1 + 1 = 0$$

First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$$1 + 1 = ?$$

$$1 + 1 = 0$$

$$1 + 0 = 1$$

First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$$1 + 1 = ?$$
$$1 + 1 = 0$$
$$1 + 0 = 1$$
$$0 + 0 = 0$$

First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$$1 + 1 = ?$$
$$1 + 1 = 0$$
$$1 + 0 = 1$$
$$0 + 0 = 0$$

Arithmetic in $\mathbb{F}_{2^m}$

First let's remember some arithmetic

Arithmetic in $\mathbb{F}_2$

$1 + 1 = ?$

$1 + 1 = 0$

$1 + 0 = 1$

$0 + 0 = 0$

Arithmetic in $\mathbb{F}_{2^m}$

It is just "vector" of 0's and 1's

it is the same arithmetic.

Introduction to error correction

First a little bit of theory in error correction

- Enable data recovery after noisy transmission.
- In general, k bits of data get stored in n bits, adding redundancy.
- If no error occurred, these n bits satisfy $n - k$ parity check equations; else can correct some errors from the error pattern.
- Check equations can be represented by a matrix.
- Good codes can correct many errors without blowing up storage too much; offer guarantee to correct t errors (often can correct or at least detect more).

Introduction to error correction

Hamming Code

Parity check matrix ($n = 7$, $k = 4$):

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

An error-free string of 7 bits $\mathbf{b} = (b_0, b_1, b_2, b_3, b_4, b_5, b_6)$ satisfies these three equations:

$$\begin{array}{lllllll} b_0+ & b_1+ & & b_3+ & b_4 & & = 0 \\ b_0+ & & b_2+ & b_3+ & & b_5 & = 0 \\ & b_1+ & b_2+ & b_3+ & & b_6 & = 0 \end{array}$$

If one error occurred at least one of these equations will not hold.

Introduction to error correction

Hamming Code

For example, if we have as the result $1, 0, 1$:

$$\begin{array}{lllllll} b_0+ & b_1+ & & b_3+ & b_4 & & = 1 \\ b_0+ & & b_2+ & b_3+ & & b_5 & = 0 \\ & b_1+ & b_2+ & b_3+ & & b_6 & = 1 \end{array}$$

What does it mean?

Introduction to error correction

Hamming Code

For example, if we have as the result $1, 0, 1$:

$$\begin{array}{cccccccc} b_0+ & b_1+ & & b_3+ & b_4 & & & = 1 \\ b_0+ & & b_2+ & b_3+ & & b_5 & & = 0 \\ & b_1+ & b_2+ & b_3+ & & & b_6 & = 1 \end{array}$$

What does it mean?
A: The bit b_1 flipped.

Introduction to error correction

Linear Code

A binary linear code C of length n and dimension k is a k-dimensional subspace of $\mathbb{F}_2^n$.
C is usually specified as

- the row space of a generating matrix $G \in \mathbb{F}_2^{k \times n}$

$$C = \{\boldsymbol{m}G | \boldsymbol{m} \in \mathbb{F}_2^k\}$$

- the kernel space of a parity-check matrix $H \in \mathbb{F}_2^{(n-k) \times n}$

$$C = \{\boldsymbol{c} | H\boldsymbol{c}^T = 0, \boldsymbol{c} \in \mathbb{F}_2^n\}$$

Introduction to error correction

Example of Linear Code

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$\boldsymbol{c} = (111)G = (10011)$ is a codeword.

Introduction to error correction

Example of Linear Code

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$\boldsymbol{c} = (111)G = (10011)$ is a codeword.
Linear codes are linear:
The sum of two codewords is a codeword:

Introduction to error correction

Example of Linear Code

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$\boldsymbol{c} = (111)G = (10011)$ is a codeword.
Linear codes are linear:
The sum of two codewords is a codeword:
$\boldsymbol{c}_1 + \boldsymbol{c}_2 = \boldsymbol{m}_1 G + \boldsymbol{m}_2 G = (\boldsymbol{m}_1 + \boldsymbol{m}_2)G$

Introduction to error correction

Example of Linear Code

$$G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

$\boldsymbol{c} = (111)G = (10011)$ is a codeword.
Linear codes are linear:
The sum of two codewords is a codeword:
$\boldsymbol{c}_1 + \boldsymbol{c}_2 = \boldsymbol{m}_1 G + \boldsymbol{m}_2 G = (\boldsymbol{m}_1 + \boldsymbol{m}_2)G$
Same with parity-check matrix:
$H(\boldsymbol{c}_1 + \boldsymbol{c}_2) = H\boldsymbol{c}_1 + H\boldsymbol{c}_2 = 0 + 0 = 0.$

Introduction to error correction

A little bit of more concepts

- Hamming weight of a word is the number of nonzero elements:

$$wt(1,0,0,1,1) = 3$$

.

Introduction to error correction

A little bit of more concepts

- Hamming weight of a word is the number of nonzero elements:

$$wt(1,0,0,1,1) = 3$$

.

- The Hamming distance between two words in $\mathbb{F}_2^n$ is the number of coordinates in which they differ:

Introduction to error correction

A little bit of more concepts

- Hamming weight of a word is the number of nonzero elements:

$$wt(1,0,0,1,1) = 3$$

.

- The Hamming distance between two words in $\mathbb{F}_2^n$ is the number of coordinates in which they differ:

$$d((1,1,0,1,1),(1,0,0,1,0)) = 2$$

The hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ equals to the Hamming weight of $\boldsymbol{x} + \boldsymbol{y}$:

$$d((1,1,0,1,1),(1,0,0,1,0)) = wt(0,1,0,0,1)$$

Code-based cryptography 101

Key Generation:

- Choose ω-error correcting code $\mathcal{C}$.
- SK: nice code description for $\mathcal{C}$.
- PK: pertubed code description given by generator matrix G.

Encryption:

- Message is a vector $m \in \mathbb{F}_{q^m}^k$.
- Select random error vector $e \in \mathbb{F}_{q^m}^n$ of weight ω.
- $c = mG + e$.

Decryption:

- Map c to equivalent vector c' in nice code representation.
- Set $m =$Decode(c) and return m.

Introduction to Code-based cryptography

McEliece cryptosystem

- Use Goppa codes for public-key cryptography.
- Oldest (1978) code-based cryptosystem.
- Easily scale up for higher security.
- Big public key: at least $\approx 256KB$.

Alternative Codes

Other codes that can be used

- Quasi-cyclic codes (QC).
- Quasi-Dyadic Codes (Misoczki, Barreto '09).
- Generalized Srivastava (Persichetti '11).

Use subfield subcode construction to encrypt in the subcode and decrypt using parent code.

$\mathbb{F}_{q^m}$ - Decryption

$\mathbb{F}_q$ - Encryption

Code-based cryptography 101

The McEliece cryptosystem in more details

- Let C be a length-n binary Goppa code Γ of dimension k with minimum distance $2t+1$ where $t \approx (nk)/\log_2(n)$; original parameters (1978) $n = 1024$, $k = 524$, $t = 50$.
- The McEliece secret key consists of a generator matrix G for Γ, an efficient t-error correcting decoding algorithm for Γ; an $n \times n$ permutation matrix P and a nonsingular $k \times k$ matrix S.
- n, k, t are public; but Γ. P, S are randomly generated secrets.
- The McEliece public key is $k \times n$ matrix $G' = SGP$.

Code-based cryptography 101

The McEliece cryptosystem in more details

- Encrypt: compute $\boldsymbol{m}G'$ and add a random vector $\boldsymbol{e}$ of weight t and length n. Send $\boldsymbol{y} = \boldsymbol{m}G' + e$.
- Decryption: compute $\boldsymbol{y}P^{-1} = \boldsymbol{m}G'P^{-1} + \boldsymbol{e}P^{-1} = (\boldsymbol{m}S)G + \boldsymbol{e}^{P^{-1}}$. This works because $\boldsymbol{e}P^{-1}$ has the same weright as $\boldsymbol{e}$ because of P is a permutation matrix. Use fast decoding to find $\boldsymbol{m}S$ and $\boldsymbol{m}$.

Code-based cryptography 101

How to decode?

- The sydrome of $\boldsymbol{x} \in \mathbb{F}_2^n$ is $\boldsymbol{s} = H\boldsymbol{x}$.
  $H\boldsymbol{x} = H(\boldsymbol{c} + \boldsymbol{e}) = H\boldsymbol{c} + H\boldsymbol{e} = H\boldsymbol{e}$ depends only on $\boldsymbol{e}$.
- To decode $\boldsymbol{x}$ with syndrome decoder, compute $\boldsymbol{e}$ from $H\boldsymbol{x}$, then $\boldsymbol{c} = \boldsymbol{x} + \boldsymbol{e}$.
  To expand syndrome, assume $H = (Q^T | I_{nk})$. Then $\boldsymbol{x} = (00\ldots0)||\boldsymbol{s}$ satisfies $\boldsymbol{s} = H\boldsymbol{x}$.

Code-based cryptography 101

One last step...

Binary Goppa code:
Let $q = 2^m$. A binary goppa code is often define by

- A list $L = (a_1, \ldots, a_n)$ of n disting elements in $\mathbb{F}_q$, called support.
- a square-free polynomial $g(x) \in \mathbb{F}_q[x]$ of degree t such that $g(a) \neq 0$ for all $a \in L$. $g(x)$ is called the Goppa polynomial.
- e.g. choose $g(x)$ irreducible over $\mathbb{F}_q$.

Questions

Thank you for your attention.
Questions?
gustavo@cryptme.in