A digital signature scheme mCFS$^{\mathrm{QC-LDPC}}$ based on QC-LDPC codes

Ernesto Dominguez Fiallo

Institute of Cryptography
Havana University

September 17, 2020

Table of contents

Table of contents

Shor's quantum algorithm, 1994:

- a quantum **polynomial time** algorithm for solving the Integer Factorization Problem and Discrete Logarithm Problem.
- the two fundamental computational problems underpinning current asymmetric cryptography.

Shor, P. W. (1994). Polynomial-time algorithm for prime factorization and discrete logarithms on a quantum computer: proc. In *35th Annual Symp. on the Foundations of Computer Science* (Vol. 124).

- In 2015, the National Security Agency (NSA) announced a transition to quantum-resistant algorithms.
- In 2016, the National Institute of Standards and Technology (NIST) published a standardization plan for ***post-quantum cryptography***.
- There are others: Europe, Japan...

Chen, L., Chen, L., Jordan, S., Liu, Y. K., Moody, D., Peralta, R., ... & Smith-Tone, D. (2016). *Report on post-quantum cryptography.* US Department of Commerce, National Institute of Standards and Technology.

When will quantum computers with enough computation arrive?

- nobody knows.
- there are reports of considerable progress.
- there are prototypes of quantum computers, even in the market.

Transition towards post quantum algorithms:

- it takes time to analyze.
- maturity in cryptographic schemes.
- contains different branches of mathematics: coding, lattices and multivariate polynomials.

Code-based cryptography became a serious candidate to replace the current asymmetric cryptography:

- decoding a random linear code is a NP-complete problem.
- McEliece cryptosystem: almost as old as asymmetric cryptography, very sure, large key sizes.
- well established to encrypt and exchange keys, not so for digital signatures.

none of the finalists to post quantum standard for digital signature, is based on codes

Table of contents

The classic idea of hash-based signature schemes is as follows:

- apply a hash function $\mathcal{H}$ to the message msg getting $\mathcal{H}(msg)$.
- consider the hash value as the ciphertext and decrypt it with the private key $\mathcal{D}_{k_{priv}}(\mathcal{H}(msg))$.
- conform the signature as the pair $\left(msg, \mathcal{D}_{k_{priv}}(\mathcal{H}(msg))\right)$.

It's very hard to accomplish the second step:

$$d(\mathcal{H}(msg), c) \nleq t$$

c: codeword.
t: error correction capacity of the code.

The CFS: the main proposal of digital signature algorithm based on code:

- it's a probabilistic algorithm.
- apply a hash function to the message repeatedly until a valid syndrome has been found.
- it's uses an increment counter to tag the number of decoding attempts.
- this number is $t!$ on average.
- this number could grow relatively fast and it's the reason for the main inefficiency.

Courtois, N. T., Finiasz, M., & Sendrier, N. (2001, December). How to achieve a McEliece-based digital signature scheme. In *International Conference on the Theory and Application of Cryptology and Information Security* (pp. 157-174). Springer, Berlin, Heidelberg.

The mCFS: based on CFS signature algorithms:

- much secure.
- replace the counter by a random value uniformly distributed over $\{1, \ldots, 2^{n-k}\}$.
- it does not solve the problem of inefficiency.

Dallot, L. (2007, July). Towards a concrete security proof of Courtois, Finiasz and Sendrier signature scheme. In *Western European Workshop on Research in Cryptology* (pp. 65-77). Springer, Berlin, Heidelberg.

The mCFS$_c$: an improvement to the mCFS:

- uses the Merkle-Damgard principle to design hash functions.
- the output of the hash function satisfies the important condition

$$d(\mathcal{H}(msg), c) \leq t$$

- solves the inefficiency problems of the mCFS scheme without reducing security.
- use Goppa codes, the sizes of keys are very large.

Ren, F., Zheng, D., & Wang, W. (2017). An Efficient Code Based Digital Signature Algorithm. IJ Network Security, 19(6), 1072-1079.

Table of contents

In this paper we propose to replace the Goppa codes with QC-LDPC codes:

- this family of codes ***was*** one of the main alternatives proposed to NIST as a post quantum standard. (LEDAcrypt)
- we obtain a considerable reduction in public key sizes (the main problem in code-based cryptography) whithout losing security
- solves the inefficiency problems of the mCFS scheme without reducing security.

We describe how to design the base compression function of the hash function and the process of generating and verifying the signature using this family of codes.

We also propose a set of parameters for different security levels in the scheme.

Table of contents

Binary linear code

A $[n, k]$ binary linear code $\mathcal{C}$ of length n and dimension k is a k-dimensional subspace of $\mathbb{F}_2^n$.

$\mathcal{C}$ can be represented by two matrices:

- a $k \times n$ generator matrix G, such that $\mathcal{C} = \{mG,\ m \in \mathbb{F}_2^k\}$.
- a $(n-k) \times n$ parity check matrix H, such that $\mathcal{C} = \{c \in \mathbb{F}_2^n,\ cH^T = 0\}$

c is a codeword of $\mathcal{C}$.

The Hamming weight

$$w(x) = |\{i : x_i \neq 0\}|$$

The Hamming distance

$$d(x, y) = |\{i : x_i \neq y_i\}|$$

The minimum distance of a code

$$d_{\min} = \min\{d(x, y) : x, y \in \mathcal{C}, x \neq y\}$$

Circulant matrix

A $p \times p$ circulant matrix is obtained by cyclically right shifting of the first row as follows: $A = \begin{pmatrix} a_0 & a_1 & \dots & a_{p-1} \\ a_{p-1} & a_0 & \dots & a_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_0 \end{pmatrix}$

If we consider the ring $\mathbb{F}_2[x]/(x^p + 1)$, the map

$$A \rightsquigarrow a(x) = \sum_{i=0}^{p-1} a_i x^i$$

is an isomorphism.

A circulant matrix is completely described by only its first row.

A Quasi-Cyclic (QC) code

Is a linear code of length n and dimension k where $k = k_0 \cdot p;\ n = n_0 \cdot p$ and its parity check matrix has the form

$$H = \begin{pmatrix} H_{00} & H_{01} & \cdots & H_{0(n_0-1)} \\ H_{10} & H_{11} & \cdots & H_{1(n_0-1)} \\ \vdots & \vdots & \ddots & \vdots \\ H_{(r_0-1)0} & H_{(r_0-1)1} & \cdots & H_{(r_0-1)(n_0-1)} \end{pmatrix}$$

where each submatrix $H_{ij},\ 0 \leq i \leq r_0 - 1,\ 0 \leq j \leq n_0 - 1$ is a circulant matrix of order p.

The main property of QC codes is that each cyclic shift of a codeword by p positions is also a codeword.

A Low Density Parity Check (LDPC) code

A Low Density Parity Check (LDPC) code is a linear code admitting a parity-check matrix with constant row weight $w = \mathcal{O}(1)$ when $n \to \infty$.

Gallager, R. (1962). Low-density parity-check codes. *IRE Transactions on information theory*, 8(1), 21-28.

A Quasi-Cyclic Low-Density Parity-Check (QC-LDPC) codes

A Quasi-Cyclic Low-Density Parity-Check (QC-LDPC) codes is a particular class of QC codes that are characterized by low-density-parity-check matrices.

Table of contents

Table of contents

It is considered a binary QC-LDPC code with length $n = n_0 p$, dimension $k = k_0 p$ and redundancy $r = p$, where $k_0 = n_0 - 1$.

The private key

is formed by two matrices:
(1) the the full-rank sparse parity-check matrix H, randomly chosen, having the following form

$$H = \begin{pmatrix} H_0 & H_1 & \dots & H_{n_0-1} \end{pmatrix}$$

where each H_i, $i \in [0, n_0 - 1]$ is a circulant $p \times p$ matrix with weight d_v in each row or column.

The private key

(2) the sparse $n_0 p \times n_0 p$ non-singular transformation matrix Q

$$Q = \begin{pmatrix} Q_{00} & Q_{01} & \cdots & Q_{0(n_0-1)} \\ Q_{10} & Q_{11} & \cdots & Q_{1(n_0-1)} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{(n_0-1)0} & Q_{(n_0-1)1} & \cdots & Q_{(n_0-1)(n_0-1)} \end{pmatrix}$$

where each Q_{ij}, $i, j \in [0, n_0 - 1]$ is a circulant $p \times p$ matrix. The row/column weight of Q is constant and equal to $m = \sum_{i=0}^{n_0-1} m_{ij}$ for some fixed $j \in [0, n_0 - 1]$ where m_{ij} is the row/column weight of Q_{ij}.

The public key

is obtained by multiplying the matrices H and $\left(Q^T\right)^{-1}$

$$L = H\left(Q^T\right)^{-1}$$

Due to the QC structure of matrix L, it is only necessary to store the first row of it. Therefore, **the public key size** is $n_0 p$ bits.

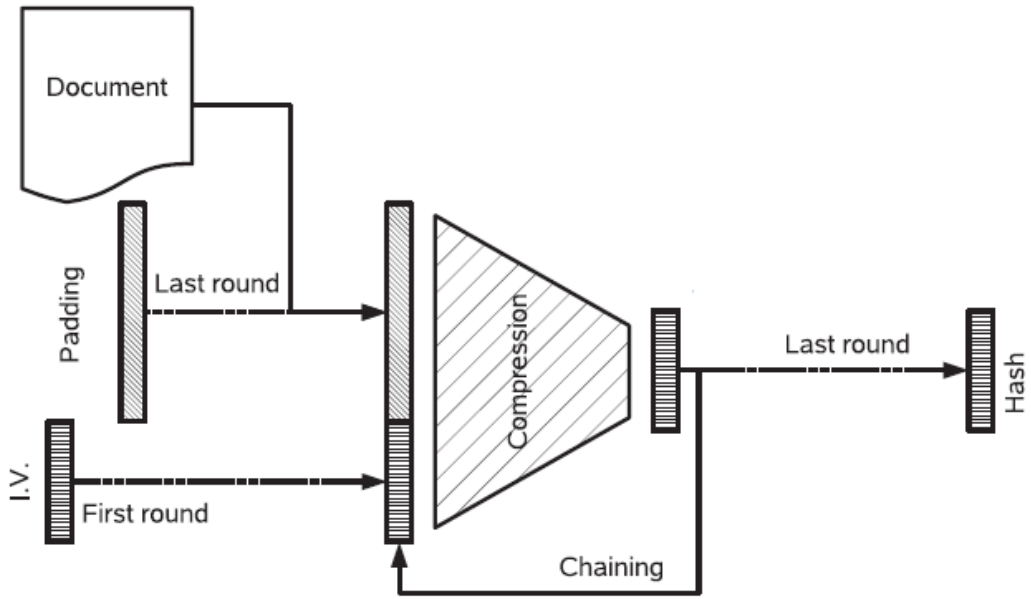
Table of contents

Follow the ideas proposed in

Augot, D., Finiasz, M., & Sendrier, N. (2005, September). A family of fast syndrome based cryptographic hash functions. In *International Conference on Cryptology in Malaysia* (pp. 64-83). Springer, Berlin, Heidelberg.

based on Merkle-Damgard design principle.

Let $f : \mathbb{F}_2^s \to \mathbb{F}_2^r,\ r < s$ be the compression function. The hash function $\mathcal{H}$ is obtained by doing the following:



The compression function

n_0 is selected such that $n_0 \leq t$. Any codeword can then be divided into n_0 blocks, each block of $n/n_0 = p$ bits.

Regular codeword

A codeword of weight n_0 is *regular* if it has exactly one non-zero position in each of the n_0 intervals $\left[(i-1)\frac{n}{n_0}, i\frac{n}{n_0}\right]_{i=1,\ldots,n_0}$.

The compression function

The public key matrix L can be divided into n_0 matrices $L = (L_1, L_2, \ldots, L_{n_0})$ of size $r \times \frac{n}{n_0}$ where

$$L_i = (l_{(i-1)\frac{n}{n_0}+1}, l_{(i-1)\frac{n}{n_0}+2}, \ldots, l_{i\frac{n}{n_0}}), \quad i = 1, \ldots, n_0$$

and $l_j,\ j = 1, \ldots, n$ is the jth column of L.

The compression function $f : \mathbb{F}_2^s \to \mathbb{F}_2^r$ where $s = n_0 \log_2 \frac{n}{n_0}$ and $r = p$ it is constructed by the following algorithm.

The compression function

Algorithm 1: The compression function

Data: $s = n_0 \log_2 \frac{n}{n_0}$ bits of the message msg

Result: a binary string of length p

1 For all $x \in \mathbb{F}_2^s$, x is divided into n_0 blocks of equal length: $x = (x_1, \ldots, x_{n_0})$, $x_i \in \mathbb{F}_2^{\log_2 \frac{n}{n_0}}$, $i = 1, \ldots, n_0$;

2 Convert each x_i to an integer between 0 and $\frac{n}{n_0} - 1$;

3 Choose the corresponding $(x_i + 1)$th column in each L_i, that is $l_{(i-1)\frac{n}{n_0}+x_i+1}$;

4 Calculate $f(x) = \bigoplus_{i=1}^{n_0} l_{(i-1)\frac{n}{n_0}+x_i+1}$;

The essential question: $f(x)$ is exactly the syndrome of a regular vector y of length n and weight n_0, that is, $f(x) = Ly^T$ and $w(y) = n_0$. **The output of the hash function $\mathcal{H}$ is the syndrome of a regular vector y of length n and weight n_0.**

The compression function

- msg: q bits.
- each iteraion: $n_0 p = n$ XORs of bits.
- bits read in the document at each iteration:

$$s - r = n_0 \log_2 \frac{n}{n_0} - p$$

- iterations: $\left\lceil \frac{q}{n_0 \log_2 \frac{n}{n_0} - p} \right\rceil$

We can approximate the number of binary XORs by

$$N_{XOR} \approx \frac{qn}{n_0 \log_2 p - p}$$

Table of contents

Let

- *msg* the message to sign.
- $\mathcal{H}$ the hash functionn previously proposed.
- *sign* the signature.

The following algorithm shows the process of signature generation:

Algorithm 2: Signature generation

Data: msg, L, $\mathcal{H}$
Result: $sign$

1 Choose a one-time random number $R \in \{1, 2, \ldots, 2^p\}$;
2 Calculate $x = \mathcal{H}\left(\mathcal{H}(msg) \| R\right)$;
3 Decode $v = \mathcal{D}_{ec}\left(x\right)$;
4 Calculate $y = vQ$;
5 $sign = (msg, R \| y)$;

Let

- $sign' = (msg, R' \| u)$ the signed message received.
- the matrices H and Q the private key.

The following algorithm shows the process of signature verification:

Algorithm 3: Signature verification

Data: $sign' = (msg, R' \| u)$, L
Result: Accept or Reject
1 Calculate $a = \mathcal{H}(\mathcal{H}(msg) \| R')$;
2 Calculate $b = Lu^T$;
3 Accept if $a = b$, Reject if otherwise;

Proof that signature verification works

$$b = Lu^T = \left(H\left(Q^T\right)^{-1}\right)(vQ)^T = H\left(\left(Q^T\right)^{-1}Q^T\right)v^T = Hv^T$$

that is, b is a syndrome vector. If $R' = R$ and b corresponds to the output of the hash function $\mathcal{H}$, we have that

$$a = \mathcal{H}\left(\mathcal{H}(msg)\|R'\right) = \mathcal{H}\left(\mathcal{H}(msg)\|R\right) = Hv^T = b$$

Table of contents

In our construction, the Goppa codes have been replaced by the QC-LDPC codes. This means two fundamental differences:

- the random hash function h is replaced by the QC-LDPC code based hash function $\mathcal{H}$.
- the relationship between the signer's public an private keys is now subject to the security of the McEliece variant based on QC-LDPC codes (LEDAcrypt specifically).

Theorical security model

The theorical security model of our signature scheme is equivalent to the mCFS and mCFS$_c$ schemes.

Table of contents

For the hash function used:

- **inversion:** is reduced to solve the Regular Syndrome Decoding (RSD) problem.
- **collision finding:** is reduced to solve the 2-Regular Null Syndrome Decoding (2-RNSD) problem.

Both problems are NP-complete.

Augot, D., Finiasz, M., & Sendrier, N. (2005, September). A family of fast syndrome based cryptographic hash functions. In *International Conference on Cryptology in Malaysia* (pp. 64-83). Springer, Berlin, Heidelberg.

From the practical point of view, there are two kinds of algorithms to attacks the hash function:

- Information Set Decoding (ISD): have computacional complexity

$$WF_{Pre} = \frac{p^3 2^p}{\left(\frac{p}{n_0}\right)^{n_0}} \qquad WF_{Col} = 2^{\frac{p}{3.3}}$$

- Wagner's Generalized Birthday Paradox: have computacional complexity

$$WF_{Col}^{Wagner} = p 2^a 2^{p/(a+1)}$$

where the parameter $a = 1, 2, \ldots$ is subject to the following restriction for its selection: $\frac{2^a}{a+1} \leq \frac{n_0}{p} \log_2 p$

Wagner, D. (2002, August). A generalized birthday problem. In *Annual International Cryptology Conference* (pp. 288-304). Springer, Berlin, Heidelberg.

In the case of the use of the QC-LDPC codes:

Attack	**Conditions**	**Work Factor(WF)**
Linearization	$p \leq 2n_0$	p^3
Cyclic	$p \leq 4n_0$	$(p/4)^3$
Cyclic + Wagner	-	$(p/2)2^{a'}2^{(p/2)/(a'+1)}$

$a' = a$ or $a' = a - 1$

Finiasz, M., Gaborit, P., & Sendrier, N. (2007, May). Improved fast syndrome based cryptographic hash functions. In *Proceedings of ECRYPT Hash Workshop* (Vol. 2007, p. 155).

Saarinen, M. J. O. (2007, December). Linearization attacks against syndrome based hashes. In *International Conference on Cryptology in India* (pp. 1-9). Springer, Berlin, Heidelberg.

Fouque, P. A., & Leurent, G. (2008). Cryptanalysis of a hash function based on quasi-cyclic codes. In *Topics in Cryptology–CT-RSA 2008* (pp. 19-35). Springer, Berlin, Heidelberg.

Table of contents

The attacker needs to perform a key recovery attack against the McEliece variant based on QC-LDPC codes: dual code attack (better than a exhaustive key search).

Is based on the classic ISD algorithm for decoding.

In our case

- only the dual code is used both in the private key matrix (private code) and in the public key matrix (public code).
- the relationship between both matrices is

$$L = H\left(Q^T\right)^{-1}$$

where H is sparse but $\left(Q^T\right)^{-1}$ is not.

The worst possible case

$\left(Q^T\right)^{-1}$ it is also sparse of weight m.

This guarantees a security level below of the actual security against the attack.

The WF of this attack is

$$WF_{DUAL}^{QC-LDPC} = \frac{WF_{ISD}(n, p, n_0 \cdot d_v \cdot m)}{p}$$

where $n_0 \cdot d_v \cdot m$ is the weight of each row of L and d_v is the column weight of H.

$$WF_{ISD}(n, p, n_0 \cdot d_v \cdot m) = 2^{n_0 \cdot d_v \cdot m \cdot \log_2 \frac{n_0}{n_0 - 1}}$$

Torres, R. C., & Sendrier, N. (2016, February). Analysis of information set decoding for a sub-linear error weight. In *International Workshop on Post-Quantum Cryptography* (pp. 144-161). Springer, Cham.

$$WF_{DUAL}^{QC-LDPC} = \frac{2^{n_0 \cdot d_v \cdot m \cdot \log_2 \frac{n_0}{n_0 - 1}}}{p}$$

Table of contents

Proposed parameters

Security in bits	p	n_0	d_v	m	KRA security
80	557	110	9	7	83
128	937	192	11	9	134
256	1 949	256	17	11	260

To compare the public key sizes with the scheme mCFS$_c$, we have relied on the security update to the McEliece based on Goppa codes given in

Bernstein, D. J., Lange, T., & Peters, C. (2008). *Attacking and defending the McEliece cryptosystem.* In International Workshop on Post-Quantum Cryptography (pp. 31-46). Springer, Berlin, Heidelberg.

Security	**mCFS$_c$**	**mCFS$^{\mathrm{QC-LDPC}}$**
80 bits	454 839	61 270
128 bits	1 534 896	179 904
256 bits	7 685 340	498 944

Conclusions

- We have to replace the Goppa codes with the QC-LDPC codes in the digital signature scheme mCFS.
- Allowed to greatly reduce the public key sizes without losing security.
- The corresponding analysis was carried out on the hash function and on the public/private key setting.
- With the corresponding analysis and a consequent adjustment in the parameters, the scheme is safe against attacks on quantum computers.

Final remark

A very recent paper (after this investigation):

Apon, D., Perlner, R., Robinson, A., & Santini, P. (2020, August). *Cryptanalysis of LEDAcrypt*. In Annual International Cryptology Conference (pp. 389-418). Springer, Cham.

- weak key discovery.
- take advantage of the structure in the multiplication of matrices (selection of matrix Q).

Considerations

- With $Q = I$, the attack is completely evaded.
- NIST did not move LEDAcrypt to 3rd round for major changes to avoid attack.
- Other selections of matrix Q are the subject of future research (to take advantage of the Q-decoders).

Thank you for your attention
Sorry for my English