QC-MDPC-McEliece:
A public-key code-based encryption scheme based on quasi-cyclic moderate density parity check codes

NIST Workshop on Cybersecurity in a Post-Quantum World

Nicolas Sendrier and Jean-Pierre Tillich
(joint work with Rafael Misoczki and Paulo Barreto)
Error Correcting Codes for Public-Key Encryption

If a random linear expansion is used, no one can decode efficiently.

If a “good” error correcting code is used for the expansion, anyone who knows the structure has access to a fast decoder.

Assuming that the knowledge of the linear expansion does not reveal the code structure:

- The linear expansion is public and anyone can encrypt
- The decoder is known to the legitimate user who can decrypt
- For anyone else, the public linear expansion looks random
McEliece Public-key Encryption Scheme – Overview

\( \mathcal{F} \) a family of \( t \)-error correcting binary linear \([n,k]\) code

**Key generation:**

\[ \mathcal{C} \in \mathcal{F} \rightarrow \begin{cases} 
\text{Public Key: } G \in \{0,1\}^{k \times n}, \text{ a generator matrix} \\
\text{Secret Key: } \Phi : \{0,1\}^n \rightarrow \mathcal{C}, \text{ a decoder correcting } t \text{ errors} 
\end{cases} \]

**Encryption:**

\[ E_G : \{0,1\}^k \rightarrow \{0,1\}^n \]

\[ x \mapsto xG + e \]

with \( e \) random of weight \( t \)

**Decryption:**

\[ D_\Phi : \{0,1\}^n \rightarrow \{0,1\}^k \]

\[ y \mapsto \Phi(y)G^* \]

where \( GG^* = 1 \)

[McEliece, 1978] \( \mathcal{F} \) is a family of binary Goppa codes

\( n = 1024, k = 524, t = 50 \)
Hardness of Decoding

[Berlekamp, McEliece, & van Tilborg, 78]

Syndrome Decoding

**Instance:** $H \in \{0, 1\}^{(n-k) \times n}$, $s \in \{0, 1\}^{n-k}$, $w$ integer

**Question:** Is there $e \in \{0, 1\}^n$ such that $\text{wt}(e) \leq w$ and $eH^T = s$?

[Alekhnovich, 03]
Conjectured difficult on average for $w = n^\varepsilon$ and any $\varepsilon > 0$

Best known decoder for $w$ errors in an $[n, k]$ code has complexity

$$W_{SD}(n, k, w) = 2^{(c+o(1))w \log_2 \frac{n}{n-k}}$$

[Prange, 62] Information Set Decoding, $c \approx 1.1$ when $w = \theta(n)$
[Becker & Joux & May & Meurer, 12] $c \approx 0.9$ when $w = \theta(n)$

When $w = o(n)$ then $c = 1$ for all classical variants of ISD

[Bernstein, 09] quantum computing → $c$ divided by 2 (at most)
Security Reduction

For given parameters $n$, $k$, and $t$

$\mathcal{K} = \{0, 1\}^{k \times n}$ the “apparent” key space

$\mathcal{G} \subset \mathcal{K}$ the set of all public keys

Theorem

If there exists an efficient adversary against McEliece then

- either there exists an efficient distinguisher for $\mathcal{G}$ versus $\mathcal{K}$
- or there exists an efficient generic decoder for $t$ errors in $[n, k]$ codes

In other words, if we assume that

1. $\mathcal{G}$ is pseudorandom
2. decoding is hard on average

then McEliece’s scheme (with public keys in $\mathcal{G}$) is secure “on average”

+ a semantically secure conversion $\rightarrow$ any desirable security level
More on Semantic Security

Because the scheme is malleable (replay attack [Berson, 97], reaction attack [Kobara & Imai, 00]) a semantically secure conversion is mandatory.

First semantically secure conversion: [Kobara & Imai, 01]

With a semantic security layer the public key can be in systematic form [Biswas & S., 08]

\[
G = \begin{pmatrix}
1 \\
\downarrow \\
1
\end{pmatrix}
\]

→ smaller key size, easier encryption
Quasi-Cyclic instances of McEliece’s Scheme (1/2)

(similar to NTRU, Ring LWE, ideal lattices)

The public key is formed of circulant blocks, for instance:

\[ G = \begin{bmatrix}
  1 & \quad & g \\
  1 & \quad & \\
\end{bmatrix} \]

\[ G = \begin{bmatrix}
  1 & 1 & g_{0,0} & g_{0,1} & g_{0,2} \\
  1 & 1 & g_{1,0} & g_{1,1} & g_{1,2} \\
\end{bmatrix} \]

Advantage: much smaller key size

Difficulty: hide the code structure (i.e. the secret decoder)
Quasi-Cyclic instances of McEliece’s Scheme (2/2)

- Goppa (or alternant) codes, initiated by [Gaborit, 05]
  Too much algebraic structure, some attempts have failed, to be used with care

- “Disguised” LDPC (Low Density Parity Check) codes
  [Baldi & Chiaraluce, 07]
  Less structure but still no convincing security reduction

- MDPC (Moderate Density Parity Check) codes
  [Misoczki & Tillich & S. & Barreto, 13]
  Even less structure, a security reduction

  [Misoczki & Barreto, 09]
  Also possible with dyadic blocks instead of circulant blocks
MDPC McEliece
QC-MDPC-McEliece Scheme (1/2)

Parameters: \( n, k, w, t \)
(for instance \( n = 9601, k = 4801, w = 90, t = 84 \))

**Key generation:** (rate 1/2, \( n = 2p, k = p \))

Pick a (sparse) vector \( (h_0, h_1) \in \{0, 1\}^p \times \{0, 1\}^p \) of weight \( w \)

\[
H_{\text{secret}} = \begin{bmatrix}
    \circlearrowleft & h_0 \\
    \circlearrowleft  & h_1
\end{bmatrix}
\]

with \( h_0(x) \) invertible in \( \mathbb{F}_2[x]/(x^p - 1) \)
(circulant binary \( p \times p \) matrices are isomorphic to \( \mathbb{F}_2[x]/(x^p - 1) \))

Publish \( h(x) = h_1(x)h_0^{-1}(x) \mod x^p - 1 \) or \( g(x) = \overline{h(x)}/x \)

\[
H = \begin{bmatrix}
    1 & \circlearrowleft \\
    1 & h
\end{bmatrix}
\quad \text{or} \quad
G = \begin{bmatrix}
    g & \circlearrowleft & 1
\end{bmatrix}
\]

\( H \) a parity check matrix, \( G \) a generator matrix
QC-MDPC-McEliece Scheme (2/2)

**Encryption:** (rate 1/2, $n = 2p$, $k = p$)

\[
\begin{align*}
F_2[x]/(x^p - 1) & \rightarrow F_2[x]/(x^p - 1) \times F_2[x]/(x^p - 1) \\
m(x) & \mapsto (m(x)g(x) + e_0(x), m(x) + e_1(x))
\end{align*}
\]

The error $e(x) = (e_0(x), e_1(x))$ has weight $t$

**Decryption:**

Iterative decoding (as for LDPC codes) which only requires the sparse parity check matrix. For instance the “bit flipping” algorithm

Parameters are chosen such that the decoder fails to correct $t$ errors with negligible probability

Each iteration has a cost proportional to $w \cdot (n - k)$, the number of iterations is small (3 to 5 in practice)
QC-MDPC-McEliece Security Reduction

\[
H = \begin{bmatrix} 1 & h \end{bmatrix} \quad \text{with} \quad h(x) = \frac{h_1(x)}{h_0(x)} \mod x^p - 1
\]

Secure under two assumptions

1. Pseudorandomness of the public key
   Hard to decide whether there exists a sparse vector in the code spanned by \( H \) (the dual of the MDPC code)

2. Hardness of generic decoding of QC codes
   Hard to decode in the code of parity check matrix \( H \) (for an arbitrary value of \( h \))
Sparse Polynomial Problems

The security reduction and the attacks can be stated in terms of polynomials

1. **Key Security**

   Given $h(x)$, find non-zero $(h_0(x), h_1(x))$ such that
   \[
   \begin{cases}
   h_0(x) + h(x)h_1(x) = 0 \mod x^p - 1 \\
   \text{wt}(h_0) + \text{wt}(h_1) \leq w
   \end{cases}
   \]

   or simply decide the existence of a solution $\rightarrow$ distinguisher

2. **Message Security**

   Given $h(x)$ and $S(x)$, find $e_0(x)$ and $e_1(x)$ such that
   \[
   \begin{cases}
   e_0(x) + h(x)e_1(x) = S(x) \mod x^p - 1 \\
   \text{wt}(e_0) + \text{wt}(e_1) \leq t
   \end{cases}
   \]

   In both cases, best known solutions use generic decoding algorithms
Practical Security – Best Known Attacks

Let $W_{SD}(n, k, t)$ denote the cost for the generic decoding of $t$ errors in a binary $[n, k]$ code.

We consider a QC-MDPC-McEliece instance with parameters $n, k, w, t$ and circulant blocks of size $p$.

1. **Key Attack:** find a word of weight $w$ in a quasi-cyclic binary $[n, n - k]$ code

   $W_K(n, k, w) \geq \frac{W_{SD}(n, n - k, w)}{n - k}$

   (there are $n - k$ words of weight $w$)

2. **Message Attack:** decode $t$ errors in a quasi-cyclic binary $[n, k]$ code

   $W_M(n, k, t, p) \geq \frac{W_{SD}(n, k, t)}{\sqrt{p}}$

   (Decoding One Out of Many [S., 11] → factor $\sqrt{p}$)
Parameter Selection

Choose a code rate $k/n$ and a security exponent $S$ (for instance 80 or 128). Then increase the block size until the following succeeds:

- find $w$ the smallest integer such that $W_K(n, k, w) \geq 2^S$
- find $t$ the error correcting capability of the corresponding MDPC code
- check that $W_M(n, k, t, p) \geq 2^S$

<table>
<thead>
<tr>
<th>80 bits of security</th>
<th>128 bits of security</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 9602$</td>
<td>$n = 19714$</td>
</tr>
<tr>
<td>$k = 4801$</td>
<td>$k = 9857$</td>
</tr>
<tr>
<td>$p = 4801$</td>
<td>$p = 9857$</td>
</tr>
<tr>
<td>$w = 90$</td>
<td>$w = 142$</td>
</tr>
<tr>
<td>$t = 84$</td>
<td>$t = 134$</td>
</tr>
</tbody>
</table>

N. Sendrier and J.-P. Tillich – Cryptosystems based on MDPC codes 13/15
A binary $[n, k]$ code with $n - k$ parity equations of weight $w$ will correct $t$ errors with an LDPC-like decoding algorithm as long as $t \cdot w \lesssim n$.

For LDPC codes, we have essentially $w = O(1)$. For MDPC codes we have $w = O(\sqrt{n})$ and thus $t = O(\sqrt{n})$.

The optimal trade-off between the key size ($K$) and the security ($S$) is obtained for codes of rate 1/2 and

$$K \approx cS^2 \quad \text{with} \quad c < 1$$

For Goppa code, the optimal code rate is $\approx 0.8$ and

$$K \approx c(S \log_2 S)^2 \quad \text{with} \quad c \approx 2$$
Conclusion

QC-MDPC-McEliece is a promising variant which enjoys

- a reasonable key size
- good security arguments (very little structure)
- secure against quantum computers
- easy implementation (including lightweight implementation)

[Heyse & von Maurich & Güneysu, 13]
Thank you for your attention
Bit-Flipping Decoding

Parameter: a threshold $T$

input: $y \in \{0, 1\}^n$, $H \in \{0, 1\}^{(n-k) \times n}$

Repeat

Compute the syndrome $H y^T$

for $j = 1, \ldots, n$

if more than $T$ parity equations involving $j$ are violated then

flip $y_j$

$H y^T = \begin{pmatrix} s_1 \\ \vdots \\ s_{n-k} \end{pmatrix}$, if $s_i \neq 0$ the $i$-th parity equation is violated

If $H$ is sparse enough and $y$ close to the code of parity check matrix $H$ then the algorithm finds the closest codeword after a few iterations.