What is... the McEliece system?

Violetta Weger

University of Zurich

Zurich Graduate Colloquium

20 November 2018
Outline

1. Coding Theory
2. Public Key Cryptography
3. McEliece cryptosystem
4. Research

What is... the McEliece system?
Repetition Code:

\[
\begin{align*}
\text{Me} & \rightarrow \text{you} \\
1 & \rightarrow 0
\end{align*}
\]
Repetition Code:

\[
\begin{align*}
\text{Me} & \rightarrow \text{you} \\
1 & \rightarrow 0 \\
111111 & \rightarrow 111010
\end{align*}
\]

We can correct 2 errors and detect 3 errors.
Let $\mathbb{F}_q$ be a finite field.

**Definition (Linear Code)**

An $[n, k]$-linear code $C$ over $\mathbb{F}_q$ is a $k$-dimensional linear subspace of $\mathbb{F}_q^n$. $c \in C$ is called a codeword.

The toy example of the repetition code was a $[6, 1]$-linear code over $\mathbb{F}_2$, with the codewords $\{\text{000000}, \text{111111}\}$. 

Violetta Weger  

What is... the McEliece system?
Let $C$ be an $[n, k]$-linear code over $\mathbb{F}_q$.

**Definition (Generator Matrix)**

There exists an $k \times n$ generator matrix $G$ of $C$ defined by:

$$C = \left\{ uG \mid u \in \mathbb{F}_q^k \right\}.$$ 

**Definition (Parity Check Matrix)**

There exists an $(n - k) \times n$ parity check matrix $H$ of $C$ defined by:

$$C = \left\{ x \in \mathbb{F}_q^n \mid Hx^T = 0 \right\}.$$
Let $C$ be an $[n, k]$-linear code over $\mathbb{F}_q$. Let $G$ be its $k \times n$ generator matrix.

**Definition (Information Set)**

A set of $k$ coordinates $I \subset \{1, \ldots, n\}$, for which the columns of $G$ are linearly independent is called an information set.

**Definition (Systematic Form)**

If $G$ is of the form

$$ (Id_k \mid A), $$

we say $G$ is of systematic form and then $H$ is given by

$$ (-A^T \mid Id_{n-k}). $$
Let $x, y \in \mathbb{F}_q^n$.

**Definition (Hamming Distance)**

The Hamming distance of $x, y$ is defined as

$$d(x, y) = | \{ i \in \{1, \ldots, n\} | x_i \neq y_i \} | .$$

**Definition (Hamming Weight)**

The Hamming weight of $x$ is defined as

$$wt(x) = | \{ i \in \{1, \ldots, n\} | x_i \neq 0 \} | .$$
Let $\mathcal{C}$ be an $[n, k]$-linear code over $\mathbb{F}_q$.

**Definition (Minimum Distance)**

We define the minimum distance of $\mathcal{C}$ to be

$$d(\mathcal{C}) = \min \{ d(x, y) \mid x, y \in \mathcal{C}, x \neq y \} = \min \{ \text{wt}(x) \mid x \in \mathcal{C}, x \neq \mathbf{0} \}.$$

In our toy example of the $[6, 1]$-Repetition code we have $d(\mathcal{C}) = 6$.

**Theorem (Singleton Bound)**

Let $\mathcal{C}$ be an $[n, k]$-linear block code. Then $d(\mathcal{C}) \leq n - k + 1$. 
Theorem

Let $C$ be an $[n, k]$-linear code over $\mathbb{F}_q$ with minimum distance $d$. Then $C$ can correct up to $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ errors.
Let $\mathbb{F}_q$ be a finite field and $1 \leq k < n \leq q$ integers.

**Definition (Generalized Reed-Solomon Code)**

Let $\alpha \in \mathbb{F}_q^n$ be an $n$-tuple of distinct elements and $\beta \in \mathbb{F}_q^n$, be an $n$-tuple of nonzero elements.

\[
GRS_{n,k}(\alpha, \beta) = \{ (\beta_1 p(\alpha_1), \ldots, \beta_n p(\alpha_n)) \mid p \in \mathbb{F}_q[x], \ deg(p) < k \}.
\]

We can write the generator matrix of $GRS_{n,k}(\alpha, \beta)$ as

\[
G = \begin{pmatrix}
\beta_1 & \cdots & \beta_n \\
\beta_1 \alpha_1 & \cdots & \beta_n \alpha_n \\
\vdots & \ddots & \vdots \\
\beta_1 \alpha_1^{k-1} & \cdots & \beta_n \alpha_n^{k-1}
\end{pmatrix}.
\]
Difference between Coding and Cryptography

Coding

\[
\begin{align*}
&\text{encoding} & m & \rightarrow & c & \text{noisy channel} & & c + e & \rightarrow & m \\
\end{align*}
\]

Public Key Cryptography

\[
\begin{align*}
&m & \rightarrow & c & \text{encryption} & \rightarrow & c & \text{send} & \rightarrow & c & \text{decryption} & \rightarrow & m \\
\end{align*}
\]

What is... the McEliece system?
We consider two people: Bob and Alice.

**Key generation:**
Bob constructs a private key and a public key, which he publishes.

**Encryption:**
Alice uses the public key to encrypt the message \( m \) to get the cipher \( c \) and sends \( c \) to Bob.

**Decryption:**
Bob uses the private key to decrypt the cipher \( c \) and recover the message \( m \).
What is... the McEliece system?
Public-Key Cryptography

What is the McEliece system?
Public-Key Cryptography

**Example: RSA**

Let $p, q$ be primes. Compute $n = pq$ and the Euler-totient function $\phi(n) = (p - 1)(q - 1)$. Choose $e < \phi(n)$, s.t. $\gcd(e, \phi(n)) = 1$.

Public Key  =  $(n, e)$

Private Key  =  $(p, q)$

Encryption: Let $m$ be the message. The cipher is computed as

$$c = m^e \mod n.$$  

Decryption: Compute $d$ and $b$ s.t.

$$de + b\phi(n) = 1.$$  

Then by computing $c^d \mod n$ we recover the message, since

$$c^d = (m^e)^d = m^{1-b\phi(n)} = m(m^{\phi(n)})^{-b} \equiv m1^{-b} = m \mod n.$$
The PKC systems, which we currently use are: RSA, DLP over elliptic curves or finite fields, ...

NSA and NIST believe that a quantum computer will be available in 2030.

Shor’s Algorithm and Grover’s Algorithm are quantum algorithms and will break those systems.

Cryptosystems which will be resistant against attacks on a quantum computer are called post-quantum cryptosystems.

Promising candidates for post-quantum cryptography are: lattice-based cryptosystems, multivariate cryptography and code-based cryptography.

What is... the McEliece system?
Choose an \([n, k] \)-linear code \(C\) over \(\mathbb{F}_q\), which can correct upto \(t\) errors and has an efficient decoding algorithm. \(C\) has a generator matrix \(G\) of size \(k \times n\). Choose a \(k \times k\) invertible matrix \(S\) and a \(n \times n\) permutation matrix \(P\) and compute \(G' = SGP\).

\[
\begin{align*}
\text{Public Key} & \quad = \quad (G', t) \\
\text{Private Key} & \quad = \quad (S, G, P)
\end{align*}
\]
Encryption: Let \( m \in \mathbb{F}_q^k \) be the message and \( e \in \mathbb{F}_q^n \) the error vector, s.t. \( \text{wt}(e) \leq t \), then the cipher is computed as

\[
c = mG' + e.
\]

Decryption: Compute

\[
cP^{-1} = mSG + eP^{-1},
\]

then \( mSG \) is a code word of \( \mathcal{C} \) and since \( \text{wt}(eP^{-1}) \leq t \), we can apply the decoding algorithm and get \( mS \) and by multiplication with the inverse of \( S \) we get the message \( m \).
Choose an \([n, k]\)-linear code \(C\), that can correct up to \(t\) errors and has an efficient decoding algorithm. \(C\) has a parity check matrix \(H\) of size \((n - k) \times n\). Choose a \((n - k) \times (n - k)\) invertible matrix \(S\) and a \(n \times n\) permutation matrix \(P\) and compute \(H' = SHP\).

\[
\begin{align*}
\text{Public Key} &= (H', t) \\
\text{Private Key} &= (S, H, P)
\end{align*}
\]
Encryption: Let $m \in \mathbb{F}_q^n$ be the message, s.t. $\text{wt}(m) \leq t$, then the cipher is computed as

$$c^T = H' m^T.$$ 

Decryption: Compute

$$S^{-1} c^T = HPm^T = H(mP^T)^T.$$ 

Since $\text{wt}(mP^T) \leq t$, we can apply syndrome decoding to get $mP^T$ and by multiplication with the inverse of $P^T$ we get the message $m$. 
The underlying problem of decoding a random linear code is an
NP-complete problem, this makes it a quantum-secure
cryptosystem.
Nevertheless, the codes we use are not random, hence there
might exist structural attacks.
There also exists a nonstructural attack called Information Set
Decoding (ISD), which has to be considered for the choice of
secure parameters. The complexity of the best algorithms so far
is $O(2^{n/20})$. 
The easiest version of the ISD algorithm is given by Lee-Brickell over the binary:
We denote by $e_I$, $c_I$, $G_I$ its $k$ columns indexed by the information set.

Input: $G \in \mathbb{F}_2^{k \times n}$, $c = mG + e$, where $e \in \mathbb{F}_2^n$ of weight $t \in \mathbb{N}$, $p < t$.
Output: $e \in \mathbb{F}_2^n$.

1. Choose an information set $I \subset \{1, \ldots, n\}$ of size $k$.
2. Choose $e_I$ with $wt(e_I) = p$.
3. If $wt(c + (c_I + e_I)G_I^{-1}G) = t$:
   Output $e = c + (c_I + e_I)G_I^{-1}G$.
4. Else: go back to 1.
To picture how this algorithm works, assume that $G$ is given in systematic form and hence $I = \{1, \ldots, k\}$ and

$$G = (\text{Id}_k \mid A).$$

Hence if we have chosen $e_I$ correctly, i.e. the correct error distribution in the first $k$ bits, then $c_I + e_I = mG_I$ and hence $(c_I + e_I)G_I^{-1} = m$ and $c + (c_I + e_I)G_I^{-1}G = c + mG = e.$
Advantages and Disadvantages of McEliece Cryptosystem

Although the McEliece system is quantum secure, there is the major drawback of large key sizes:

<table>
<thead>
<tr>
<th>Security Level</th>
<th>Key Size RSA</th>
<th>Key Size original McEliece</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2^{80}$</td>
<td>1248</td>
<td>520047</td>
</tr>
<tr>
<td>$2^{128}$</td>
<td>3248</td>
<td>1537536</td>
</tr>
<tr>
<td>$2^{256}$</td>
<td>15424</td>
<td>7667855</td>
</tr>
</tbody>
</table>
The main idea to bring down the key sizes is to use another family of codes.

<table>
<thead>
<tr>
<th>Proposal</th>
<th>Idea</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niederreiter</td>
<td>GRS codes</td>
<td>Sidelnikov-Shestakov</td>
</tr>
<tr>
<td>Berger, Loidreau</td>
<td>Subcodes of GRS codes</td>
<td>Wieschebrink</td>
</tr>
<tr>
<td>Gabidulin et al.</td>
<td>Gabidulin codes</td>
<td>Overbeck</td>
</tr>
<tr>
<td>Sidelnikov</td>
<td>Reed-Muller codes</td>
<td>Minder-Shokrollahi</td>
</tr>
<tr>
<td>Baldi et al.</td>
<td>LDPC codes</td>
<td>Couvreur et al.</td>
</tr>
<tr>
<td>Rosenthal et al.</td>
<td>GRS, new scrambling</td>
<td>Couvreur et al.</td>
</tr>
</tbody>
</table>
New proposals:

<table>
<thead>
<tr>
<th>Proposal</th>
<th>Idea</th>
<th>Attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baldi et al.</td>
<td>QC-MDPC codes</td>
<td></td>
</tr>
<tr>
<td>Baldi et al.</td>
<td>MDPC codes</td>
<td></td>
</tr>
<tr>
<td>Khathuria, Rosenthal, W.</td>
<td>GRS, weight two matrix</td>
<td></td>
</tr>
<tr>
<td>Horlemann-Trautmann, W.</td>
<td>Ring linear codes</td>
<td></td>
</tr>
</tbody>
</table>
Thank you!