The McEliece Cryptosystem and Error-Correcting Codes

Midshipman 1/C Mario Nonog Jr.

Associate Professor Susan Margulies, Mathematics Department

1 Introduction

RSA was broken using Shor's Algorithm in 1994. However, it took 22 years for the National Institute of Standards and Technology (NIST) to begin the search for a new Post-Quantum Cryptography standard, in order to protect our secrets against adversaries. Since 2016, NIST has evaluated over 50 cryptosystems. As of July 2020, only four are left standing. One of these is the McEliece Cryptosystem, based on error-correcting codes. The goal of this project is to understand the McEliece Cryptosystem, which utilizes Linear Codes, Cyclic Codes, BCH Codes and Goppa Codes.

2 The McEliece Cryptosystem

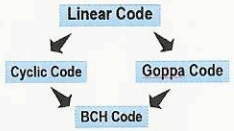
Bob: Chooses $k \times n$ generating matrix G, $k \times k$ invertible matrix S, and $n \times n$ permutation matrix P, calculates public key $G_1 = SGP$.
Alice: Chooses $1 \times k$ message $x \in \mathbb{Z}_2^k$, and error vector e of weight t, length n

Alice	Public	Bob
	$G_1 \leftarrow$	$G_1 = SGP$
$xG_1 + e = y$	$\rightarrow y \rightarrow$	1) $r_1 = yP^{-1} = (xG_1 + e)P^{-1}$ $= x(SGP)P^{-1} + eP^{-1}$ $= (xS)G + eP^{-1}$
		2) Compute syndrome $r_1 H^T$
		3) Lookup the codeword $c_1 = (xS)G = x_1 G$ associated with received word r_1
		4) Extract message $x_1 = (xS)$ associated with code word c_1 (first k bits)
	Eve	5) Compute $x_1 S^{-1} = (xS)S^{-1} = x$

The McEliece Cryptosystem uses an $[n, k, d] = [1024, 512, 101]$ Goppa code. In this case, Eve has $\binom{1024}{50} \approx 3x10^{85}$ possible locations of the errors.

3 Linear Code

- **Definition:** An $[n, k, d]$ linear code is a vector space with dimension k and length n over a field $\mathbb{F}$, where the combination of any two codewords is always a codeword, with minimum distance d between two codewords.



- **Parity Check Matrix (PCM):** Given generating matrix $G = [I_k, P] \in C$, then $H = [-P^T, I_{n-k}]$ is a parity check matrix for C if and only if $vH^T = 0$
- **Coset:** Given linear code C, and n-dimensional vector v, a coset is the set of $v + C$. The vector with minimum Hamming weight is the coset leader.
- Given a linear code C, s errors detected if minimum distance $d(C) \geq s + 1$
- Given a linear code C, t errors corrected if $d(C) \geq 2t + 1$
- **Syndrome:** This is defined as $S(v) = vH^T$

- **How to DECODE a received word r?**

1. Calculate the syndrome.
2. Find which coset the syndrome belongs.
3. Look for the coset leader.
4. message $= r -$ coset leader.

Example: Consider a $[4, 2, 2]$ linear code with G, H.

$$\text{generating matrix } G = \begin{bmatrix} 1&0&0&1 \\ 0&1&0&1 \end{bmatrix}, \qquad \text{parity check matrix } H = \begin{bmatrix} 0&0&1&0 \\ 1&1&0&1 \end{bmatrix}$$

Here is the lookup table for decoding received words.

$$\begin{array}{c} (0,0,0,0)(1,0,0,1)(0,1,0,1)(1,1,0,0) \\ (1,0,0,0)(0,0,0,1)(1,1,0,1)(0,1,0,0) \\ (0,0,1,0)(1,0,1,1)(0,1,1,1)(1,1,1,0) \\ (0,0,1,1)(1,0,1,0)(0,1,1,0)(1,1,1,1) \end{array}$$

Alice encodes message $x = [1, 1]$ by computing $xG = [1, 1, 0, 0]$. She sends to Bob through a noisy channel, and Bob receives $r = (1, 1, 1, 0)$.

Next, Bob must DECODE, by calculating the syndrome $S(r) = rH^T$.

Coset Leader	Syndrome
$(0,0,0,0)$	$(0,0)$
$(1,0,0,0)$	$(0,1)$
$(0,0,1,0)$	$(1,0)$
$(0,0,0,1)$	$(1,1)$

$$S(r) = (1,1,1,0)\begin{bmatrix} 0&1 \\ 0&1 \\ 1&0 \\ 0&1 \end{bmatrix} = (1,0).$$

Therefore, the code word $= \underbrace{(1,1,1,0)}_{\text{received word}} - \underbrace{(0,0,1,0)}_{\text{coset leader}} = (\underbrace{1,1}_{\text{message}}, 0, 0).$

4 Cyclic Code

A code is said to be a cyclic code if it contains the property

$$(c_1, c_2, ..., c_n) \in C \iff (c_n, c_1, c_2, \ldots, c_{n-1}) \in C$$

Given $g(X) = g_0 + g_1 X + \cdots + g_{n-1}X^{n-1} + g_{n-k}X^{n-k}$ (where $g(X) \mid X^n - 1$), we formulate $k \times n$ generating matrix G and $(n-k) \times n$ parity check matrix H.

$$G = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-k} & 0 & \cdots & 0 \\ 0 & g_0 & g_1 & \ldots & g_{n-k} & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_0 & g_1 & \cdots & g_{n-k} \end{bmatrix} \quad H = \begin{bmatrix} h_k & h_{k-1} & \cdots & h_0 & 0 & \cdots & 0 \\ 0 & h_k & h_{k-1} & \cdots & h_0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_k & h_{k-1} & \cdots & h_0 \end{bmatrix}$$

Example: Here is an $[n, k, d] = [7, 3, 4]$ cyclic code C.

$$X^7 - 1 = g(X)h(X) = \underbrace{(X^4 + X^2 + X + 1)}_{g(X)}\underbrace{(X^3 + X + 1)}_{h(X)}.$$

Then 3×7 generating matrix G and 4×7 parity check matrix H are:

$$G = \begin{bmatrix} 1&1&1&0&1&0&0 \\ 0&1&1&1&0&1&0 \\ 0&0&1&1&1&0&1 \end{bmatrix} \qquad H = \begin{bmatrix} 1&1&0&1&0&0&0 \\ 0&1&1&0&1&0&0 \\ 0&0&1&1&0&1&0 \\ 0&0&0&1&1&0&1 \end{bmatrix}$$

5 BCH (Bose-Chaudhuri-Hocquenhem) Code

Theorem 1 *Let C be an $[n, k, d]$ cyclic code over $\mathbb{F}_{q=p^m}$, where $p \nmid n$. Let α be a primitive n-th root of unity, and let $g(X)$ be a generating polynomial for C. Suppose there exist integers ℓ and δ such that*

$$g(\alpha^\ell) = g(\alpha^{\ell+1}) = \cdots = g(\alpha^{\ell+\delta}) = 0$$

Then the minimum distance $d \geq \delta + 2$.

Recall:

$$X^n - 1 = \begin{cases} f_1(X)f_2(X)\cdots f_r(X) & \text{over } \mathbb{F}_p \\ (X-1)(X-\alpha)(X-\alpha^2)\cdots(X-\alpha^{n-1}) & \text{over extension field } \mathbb{F}_q \end{cases}$$

Let $q_j(x) = f_i(\alpha^j) = 0$. Then, the generating polynomial of a BCH code is

$$g(x) = \text{LCM}\big(q_{k+1}(X), q_{k+2}(X), \ldots, q_{k+d-1}(X)\big)$$

for some integer k. The parity check matrix H is

$$H = \begin{bmatrix} 1 & \alpha^{k+1} & \alpha^{2(k+1)} & \cdots & \alpha^{(n-1)(k+1)} \\ 1 & \alpha^{k+2} & \alpha^{2(k+2)} & \cdots & \alpha^{(n-1)(k+2)} \end{bmatrix}$$

How to DECODE a received word r for one error?

1. Calculate $rH^T = (s_1, s_2)$.
2. If $s_1 = 0$, then no error (r is a codeword).
3. If $s_1 \neq 0$, compute $\frac{s_2}{s_1} = \alpha^{j-1}$, where j is position of error.
4. $r - e =$ codeword

Example: Consider a $[7, 1, 7]$ BCH code with the generating polynomial $g(X) = X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$. There are two codewords: $(0,0,0,0,0,0,0)$ and $(1,1,1,1,1,1,1)$. Suppose Bob receives $r = (1,1,1,0,1,1,1)$. Now, detect and correct the error.

Solution: Since $rH^T = (s_1, s_2)$, we see

$$rH^T = (1,1,1,0,1,1,1)\begin{bmatrix} 1 & 1 \\ \alpha & \alpha^2 \\ \alpha^2 & \alpha^4 \\ \vdots & \vdots \\ \alpha^6 & \alpha^{12} \end{bmatrix} = (1+\alpha+\alpha^2+\alpha^4+\alpha^5+\alpha^6, 1+\alpha^2+\alpha^4+\alpha^8+\alpha^{10}+\alpha^{12})$$

$$rH^T = (\alpha^3, \alpha^6) = (s_1, s_2)$$

Since $s_1 \neq 0$, we calculate $\frac{s_2}{s_1} = \frac{\alpha^6}{\alpha^3} = \alpha^3$. Therefore, $j - 1 = 3$, so the error position is at $j = 4$. Finally, we see

$$r - e = \underbrace{(1,1,1,0,1,1,1)}_{\text{received word}} - \underbrace{(0,0,0,1,0,0,0)}_{\text{error vector}} = \underbrace{(1,1,1,1,1,1,1)}_{\text{codeword}}$$

6 References

- *Introduction to Cryptography with Coding Theory*, by W. Trappe, L. Washington, Pearson, 2nd edition, 2006
- *Fundamentals of Error-Correcting Codes*, by V. Pless, W.C. Huffman, Cambridge University Press, 2003