Code-Based Post-Quantum Cryptography

Wijik Lee[1], Young-Sik Kim[2], and Jong-Seon No[1]

[1]Department of ECE, INMC, Seoul National University, Seoul, Korea
[2]Chosun University, Gwangju, Korea

September 07, 2017

Outline

Outline

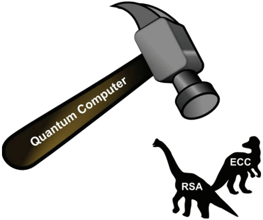
Quantum Computers

- Practical large quantum computers are just around the corner, which are developed by government(NSA), EU, and large companies (Google, IBM).

- A 50 qubit quantum computer can do computation in 2^{50} states at one time. (almost same as supercomputer)
- Recently, a 22 qubit quantum computer has been developed by Google.
- It is known that it can solve many hard problems for cryptography.

After Quantum Computers

- Google says that quantum computer is expected to be used within 10 to 20 years from now.
- In quantum computer,
  - Factoring is easy (Shor's algorithm).
    - Some researcher in Google says that 1024 bit RSA will be broken by quantum computer in 10 years (2027).
  - Search is also easy (Grover's algorithm).
    - Can search 2^n elements in time $2^{n/2}$.
- After quantum computer, conventional cryptosystems are all dead.
  - RSA, DSA, ECDSA
  - ECC, HECC etc.



Post-Quantum Cryptography

- In general, cryptosystem is a mathematical algorithm.
- Quantum cryptography uses physical techniques instead of mathematical algorithm (function).
- Recently, one of quantum cryptography is implemented for a secret key distribution algorithm (quantum key distribution, QKD).
- Quantum cryptography needs direct connection between the quantum cryptography hardwares via optical fiber and satellite.
- Quantum cryptosystem generates kB of keystream per second on special hardware costing $50,000.
    - Conventional cryptosystem generates GB of keystream per second on a $200 CPU.
- Post-quantum cryptography(PQC) is different from quantum cryptography.
- PQC is a mathematical algorithm, which is robust from quantum computer (quantum-resistant).

Post-Quantum Cryptography

Types of post-quantum cryptography

- Code-based cryptography
  - 1978 McEliece; hidden Goppa-code public-key encryption system.
- Hash-based cryptography
  - 1979 Merkle; hash-tree public-key signature system.
- Multivariate-quadratic equation cryptography
  - 1996 Patarin; "HFEV-" public key signature system.
- Lattice-based cryptography
  - 1998 "NTRU"
  - 1996 "SIS" (SVP)
  - 2005 "LWE" (CVP)

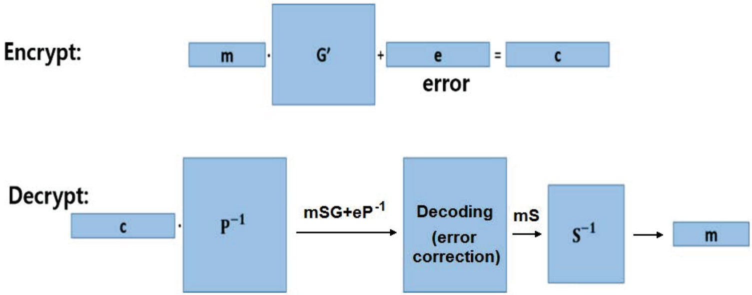
Call for Proposal for Post-Quantum Cryptosystems

NIST announced Call for Proposal for post-quantum cryptosystems on August 2016.

- Deadline for proposals; November 2017
- In the following three areas:
  1. Encryption Algorithm
  2. Digital Signature Algorithm
  3. Key Encapsulation Mechanism (KEM)
- First selection of the proposals for evaluation on March 2018.
- Popular PQCs
  - Lattice-based post-quantum cryptography
  - Code-based post-quantum cryptography

Code-Based Post-Quantum Cryptosystem

- Code-based cryptosystem is one of the well-known post-quantum cryptosystems by McEliece (1978).
- $G' = SGP$, G: generator matrix



Code-Based Post-Quantum Cryptosystem

- Encryption
  - Generator matrix $G' = SGP$
  - $c = mG' + e$
- Decryption
  - $cP^{-1} = mSG + eP^{-1}$
  - mS is obtained by decoding.
  - $mSS^{-1} = m$
- There are many variant versions of code-based cryptosystem.
- We proposed the modification methods for the McEliece cryptosystems based on the punctured RM codes (Sidelnikov).

Lattice-Based Post-Quantum Cryptosystem

- Features of Lattice-Based Cryptography
  - Based on NP-hard problem
    - SVP (shortest vector problem)
    - CVP (closest vector problem)
  - Seemingly very different assumptions from factoring, discrete log, and elliptic curves.
  - Simple descriptions and implementations.
  - Very parallelizable.
  - Seems to resist quantum attacks.
  - Security based on worst-case problems.
- Great Advantages
  - Very strong security proofs.
  - The schemes are fairly simple.
  - Relatively efficient.
- Major Drawback
  - Schemes have very large key size.

Outline

Code-Based Post-Quantum Cryptography

- Code-based post-quantum cryptosystems
  - McEliece cryptosystem by generator matrix of Goppa code, 1978
  - Niederreiter cryptosystem by parity check matrix of Goppa code, 1986
- Code-based signature scheme
  - CFS signature scheme (Courtois, Finiasz, Sendrier, 2001)

McEliece Cryptosystem

- In 1978, McEliece introduced a public key cryptosystem based on error correcting codes.
- The cracking problem for McEliece cryptosystem is the problem of syndrome decoding.

Syndrome decoding problem

Given parity check matrix H and syndrome s, find the minimum Hamming weight e, such that $He^T = s$.

- The problem of syndrome decoding is proven to be NP-hard.

Goppa Code

- Goppa code is a special case of alternant code.

Definition. Alternant code

A q-ary alternant code of order r associated with $\mathbf{x} = (x_1, \cdots, x_n) \in F_{q^m}^n$ where all x_i's are distinct and $\mathbf{y} = (y_1, \cdots, y_n) \in (F_{q^m}^*)^n$ is defined as

$$\mathcal{A}_r(\mathbf{x}, \mathbf{y}) = \{c \in F_q^n | V_r(\mathbf{x}, \mathbf{y})c^T = 0\},$$

where

$$V_r(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} y_1 & \cdots & y_n \\ y_1 x_1 & \cdots & y_n x_n \\ \vdots & & \vdots \\ y_1 x_1^{r-1} & \cdots & y_n x_n^{r-1} \end{pmatrix}$$

Goppa Code

Definition. Goppa codes

A q-ary Goppa code $\mathcal{G}(\mathbf{x}, \gamma)$ associated with a polynomial $\gamma(z) = \sum_{i=0}^{r} \gamma_i x^i$ of degree r over F_{q^m} and an n-tuple $\mathbf{x} = (x_1, \cdots, x_n)$ of distinct elements of F_{q^m} satisfying $\gamma(x_i) \neq 0$ for all $i, 1 \leq i \leq n$, is the q-ary alternant code $\mathcal{A}_r(\mathbf{x}, \mathbf{y})$ with $y_i = \gamma(x_i)^{-1}$.

- In McEliece cryptosystem, binary Goppa code is used ($q = 2$).

McEliece Cryptosystem

- Based on binary Goppa code
  - Let C be a length n binary Goppa code Γ of dimension k with minimum distance $2t+1$, where $t = \frac{n-k}{\log_2 n}$.
  - Original parameters: $n = 1024, k = 524, t = 50$.
- There are no efficient structural attacks distinguishable between a permuted Goppa code used by McEliece and a random code.
  - Original parameter designed for 2^{64} security.
  - Recently, it is known that it should be 2^{128} security.
  - Easily scale up for higher security.

McEliece Cryptosystem

- **Key Generation**
- Private key
  - G: $k \times n$ generator matrix of error correcting code (Goppa code).
  - S: $k \times k$ scrambling matrix
  - P: $n \times n$ permutation matrix.
  - An efficient t-error correcting decoding algorithm for Goppa code.
- Public key
  - $G' = SGP$
  - An error correcting capability t
- Key size is very large.

McEliece Cryptosystem

- **Encryption**

Encryption algorithm

Input: message m, G'
Output: ciphertext c

1. Choose a random $e \in \{0,1\}^n$ with Hamming weight at most t
2. Compute the ciphertext $c = mG' + e$ and send c.

- Need efficient implementation for matrix multiplication.
- Need an appropriate random number generator.

McEliece Cryptosystem

- **Decryption**

Decryption algorithm

Input: ciphertext c, S, G, P, decoding algorithm
Output: message m

1. Multiply P^{-1} as $cP^{-1} = mSG + eP^{-1}$
2. Use decoding algorithm to decode cP^{-1} to mS
3. Recover m by multiplying S^{-1}

- Require operations in binary extension fields.

McEliece Cryptosystem

- Advantages
  - Robust to quantum computer (NP-hard problem).
  - The encryption and decryption processes are fast.
  - The encryption and decryption processes have a low complexity.
- Disadvantages
  - The private and public keys are large matrices.
  - The public key size is 100 kB to several MB.

Niederreiter Cryptosystem

- Proposed by Niederreiter in 1986, based on parity check matrix.
- Niederreiter cryptosystem is also based on the nature of the syndrome decoding problem being NP-hard.
- McEliece cryptosystem and Niederreiter cryptosystem are proven to be equivalent.

Niederreiter Cryptosystem

- **Key Generation:**
  - H: $k \times n$ parity check matrix
  - S: $k \times k$ scrambling matrix
  - P: $n \times n$ permutation matrix
  - Private key: H, S, P
  - Public key: $H' = SHP$, error correcting capability t
- **Encryption:** Message m is converted into a vector with Hamming weight less than or equal to t, called an error vector e in F_2^n. Alice sends the ciphertext $s' = H'e^T$ to Bob.
- **Decryption:** When Bob receives the ciphertext s' and he multiply S^{-1} as $S^{-1}s' = HPe^T$.
  Using decoding algorithm, Bob finds Pe^T and then recovers e by multiplying P^{-1}. From the known algorithm, e is converted into m.

Code-Based Signature Scheme

- CFS signature scheme (Courtois, Finiasz, Sendrier, 2001)
  - CFS signature scheme is based on Niederreiter cryptosystem.
  - Message is treated as a syndrome and signature is treated as an error.
    - $h(m)$: hashed massage.
    - Find signature z such that $H'z = h(m)$, where H' is a parity check matrix.
- Advantage
  - Signing time does not depend on n, k.
- Disadvantage
  - The probability of finding decodable syndrome is $\frac{1}{t!}$.
  - The private and public key sizes are large.
- Other signature schemes have been broken, such as KKS, KKS variants, and CFS based on LDGM codes.

Code-Based Signature Scheme

- Key generation
  - Private key: (S, H, P), where S is a scrambling matrix and P is a permutation matrix.
  - Public key: $H' = SHP$, hash function h.
- Signature
  - Find z such that $H'z = h(h(m)|i)$.
  - Initiallize $i = 0$.
  - Do
    - $s_i = Q^{-1}h(h(m)|i)$
    - $i \leftarrow i + 1$
  - Until s_i decodable in $H'z = s_i$.
  - $s \leftarrow s_i, z \leftarrow P^{-1}\mathrm{decode}(s)$, $\mathrm{decode}(s)$ means finding Pz from H and s.
  - Signature $\sigma = (m, z, i)$
- Verification
  - Check $\mathrm{wt}(z) \leq t$
  - Check $h(h(m)|i) = H'z^T$.

Outline

Variants of McEliece Cryptosystem

- To overcome the key size problem of McEliece cryptosystem,
- Use other codes
  - GRS code (broken)
  - RM code (broken)
  - LDPC, MDPC (still alive)
- Modify the code structures to be survived.
  - Quasi cyclic: QC-LDPC, QC-MDPC, QC-LRPC
  - Puncturing: punctured RM code (our work)

Variants of McEliece Cryptosystem (Our Work*)

- RM code-based McEliece cryptosystem
- We find the exact number and locations of puncturing of the generator matrix of the original RM code to prevent the various known attacks.
- Further, we also modify it by puncturnig and random column insertion of generator matrix.

$$\boldsymbol{G}' = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 1 & \cdots & 1 & \cdots & 0 & 1 \\ 0 & & 1 & & 0 & & 0 & & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & & 1 & & 0 & & 0 & & 1 & 0 \\ 0 & & 0 & & 1 & & 1 & & 1 & 1 \\ 1 & \cdots & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & 1 \end{bmatrix}$$

Column puncturing

Ciphertext c:

* Wijik Lee, Jong-Seon No, and Young-Sik Kim, "Punctured Reed-Muller code-based McEliece cryptosystems," IET Communications, vol. 11, no. 10, pp. 1543–1548, July 2017.

Variants of McEliece Cryptosystem (Our Work)

- The proposed modification of RM code-based McEliece cryptosystem can be presented by the following three algorithms.
- **Key Generation**
- Private key
  - Set of column indices L_D for puncturing
  - Delete columns with indices in L_D from G, which is denoted by G_D.
  - G: $k \times n$ generator matrix for Γ.
  - S: $k \times k$ scrambling matrix
  - P: $(n - |L_D|) \times (n - |L_D|)$ permutation matrix.
  - An efficient t-error correcting decoding algorithm for Γ.
- Public key
  - $G'_D = SG_DP$
  - An error correcting capability $t' = \lfloor t - |L_D|/2 \rfloor$ of G_D

Variants of McEliece Cryptosystem (Our Work)

- **Encryption**

Encryption algorithm

Input: message m, G'_D
Output: ciphertext c

1. Choose a random $e \in \{0,1\}^{n-|L_D|}$ with Hamming weight at most t'
2. Compute the ciphertext $c = mG'_D + e$.

Variants of McEliece Cryptosystem (Our Work)

- **Decryption**

Decryption algorithm

Input: ciphertext c, S, G, P, decoding algorithm
Output: message m

1. Multiply P^{-1} as $cP^{-1} = mSG_D + eP^{-1}$.
2. Insert the erasure mark '?' in the jth positions, where $j \in L_D$.
3. Use a decoding algorithm with erasures to decode cP^{-1} to mS.
4. Recover m by multiplying S^{-1}.

- Our proposed McEliece cryptosystem is further modified by puncturnig and random column insertion of the generator matrix.

Outline

Security of the McEliece Cryptosystem

- Attack on the McEliece cryptosystem
  - Information set decoding
  - Finding low weight codeword
- Attacks on McEliece cryptosystem using some codes other than Goppa code.
  - GRS, RM, polar codes, etc.
- Semantic security
  - CCA2 (NIST requirement for encryption scheme)
  - EUF-CMA (NIST requirement for signature scheme)

CCA2: adaptive chosen ciphertext attack
EUF-CMA: existential unforgeability under chosen message attack

Security of the McEliece Cryptosystem (Information Set Decoding)

- Based on finding k-error free bits c_k of ciphertext randomly.
  - An adversary chooses k-columns of G' with error free indices of the ciphertext c_k, denoted by G'_k.
  - Then, $c_k = mG'_k + e_k$ with $e_k = 0$.
  - Decryption is done by $m = c_k \cdot (G'_k)^{-1}$.
- Probability of choosing k error free bits is given as:

$$\binom{n-t}{k} \Big/ \binom{n}{k}$$

- Security of the McEliece cryptosystem

(1024,524,50)	64
(2048,1751,27)	80
(6960,5413,119)	128

Security of the McEliece Cryptosystem (Finding Low Weight Codeword)

- The minimum weight codeword of the following $(k+1) \times n$ matrix

$$\begin{bmatrix} G' \\ c \end{bmatrix}$$

  is the error vector, where $c = mG' + e$.
- By using the Stern's algorithm, we can find the minimum weight codeword of the matrix.
- The original parameters $(n, k, t) = (1024, 524, 50)$ have the work factor of $2^{64.2}$.

Security of the McEliece Cryptosystem

Using some code other than Goppa code-based McEliece cryptosystem are almost broken as follows.

- GRS code (1992)
  - Sidelnikov's attack (1992)
  - Wieschebrink's attack (2010)
- RM code (1994)
  - Minder-Shokrollahi's attack (2007)
  - Chizhov-Borodin's attack (2013)
  - RM code with random insertion; square code attack (2015)
- Polar code (2014)
  - Bardet's attack (2016)
- Algebraic geometry codes and their subcodes (1996)
  - Couvreur's attack (2017)

Security for PQC by Modified RM Code (Our Work)

- By puncturnig method, we can prevent Minder-Shokrollahi's attack and Chizhov-Borodin's attack.
- By puncturnig and random insertion methods, we can also prevent square code attack.

Semantic Security

- It is required by NIST for proposed PQC encryption algorithms.
- Security for indistinguishability and non-malleability.
- CCA2 (indistinguishability under adaptive chosen ciphertext attack)

CCA2

1. Challenger runs KeyGen and obtain (private key, public key). Adversary obtains only public key.
2. The adversary can query polynomial number of decryption to decryption oracle (at any step).
3. The adversary submits two distinct chosen plaintexts m_0, m_1.
4. Challenger chooses $b \in \{0, 1\}$ and sends $c = Enc(m_b)$ to the adversary.
5. If adversary guesses the value b correctly without quering c to decryption oracle, attack is successful.

Semantic Security

- It is required by NIST for proposed PQC signature schemes.
- EUF-CMA is the signature version of CCA2.
- EUF-CMA (existential unforgeability under chosen message attack)

EUF-CMA

1. Challenger runs KeyGen and obtains (private key, public key). Forger obtains only public key.
2. Forger can query polynomial number of messages to signature oracle (and hash oracle).
3. If forger can generate message signature pair (m, σ), then attack is successful.

Outline

Conclusions

- With the development of quantum computers, conventional cryptosystems become vulnerable and thus post-quantum cryptosystems are required.
- Code-based cryptography is one of the post-quantum cryptosystems and we present some code-based cryptosystems and their security property.
- We proposed the secure modification methods for the McEliece cryptosystems based on the punctured RM codes.