Ph.D. DISSERTATION

RM Code-Based Post Quantum Cryptosystems

RM 부호 기반 포스트 양자 암호시스템

BY

LEE WIJIK

FEBRUARY 2018

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Ph.D. DISSERTATION

RM Code-Based Post Quantum Cryptosystems

RM 부호 기반 포스트 양자 암호시스템

BY

LEE WIJIK

FEBRUARY 2018

DEPARTMENT OF ELECTRICAL ENGINEERING AND
COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

RM Code-Based Post Quantum Cryptosystems

RM 부호 기반 포스트 양자 암호시스템

지도교수 노 종 선
이 논문을 공학박사 학위논문으로 제출함

2018년 2월

서울대학교 대학원

전기 컴퓨터 공학부

이 위 직

이위직의 공학박사 학위 논문을 인준함

2018년 2월

위 원 장: _____
부위원장: _____
위 원: _____
위 원: _____
위 원: _____

Abstract

In this dissertation, Reed-Muller (RM) code-based cryptosystems and two families of p-ary sequences are considered. Three main contributions are given as follows.

First, McEliece cryptosystems based on punctured RM codes are proposed. It is shown that the already known attacks, such as the Minder-Shokrollahi's attack, the Chizhov-Borodin's attack, and the square code attack, do not work for the proposed RM code-based McEliece cryptosystems. We find an optimal puncturing scheme to prevent the previously known attacks for the proposed RM code-based cryptosystems in a sense that the exact locations of puncturing positions with the minimum number of punctured columns of the generator matrix should be found for attacking. It is important to carry out the minimum number of puncturing since the modification of codes by puncturing can reduce security level. In addition, the square code attack can also be prevented in the proposed RM code-based McEliece cryptosystems by using both the proposed puncturing and random insertion methods.

Second, a new signature scheme based on a punctured Reed–Muller (RM) code with random insertion is proposed. The proposed signature scheme improves the Goppa code-based signature scheme developed by Courtois, Finiasz, and Sendrier (CFS). The CFS signature scheme has certain drawbacks in terms of scaling of the parameters and a lack of existential unforgeability under adaptive chosen message attacks (EUF-CMA) security proof. Further, the proposed modified RM code-based signature scheme can use complete decoding, which can be implemented using a recursive decoding method and thus syndromes for errors larger than the error correctability can be decoded for signing, which improves the probability of successful signing and reduces the signing time. Using the puncturing and insertion methods, the proposed RM code-based signature scheme can avoid some known attacks for RM code-based cryptosystems. The parameters of the proposed signature scheme such as error weight pa-

rameter w and the maximum signing trial N, can be adjusted in terms of signing time and security level and it is also proved that the proposed signature scheme achieves EUF-CMA security.

Last, for an odd prime p such that $p \equiv 3 \bmod 4$ and an odd positive integer n, two new families of p-ary sequences of period $N = \frac{p^n-1}{2}$ are constructed by two decimated p-ary m-sequences $m(2t)$ and $m(dt)$, where $d = 4$ and $d = (p^n+1)/2 = N+1$. The upper bound on the magnitude of correlation values of two sequences in the family is derived by using Weil bound. Their upper bound is derived as $\frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}$ and the family size is $4N$, which is four times the period of the sequence.

keywords: Code-based cryptosystems, Courtois, Finiasz, and Sendrier (CFS) signature scheme, McEliece cryptosystem, m-sequences, p-ary sequences, post-quantum cryptosystem, public key cryptography, puncturing, Reed-Muller (RM) codes, Weil bound.

student number: 2012-20839

Contents

Abstract **i**

Contents **iii**

List of Tables **vi**

List of Figures **vii**

1 INTRODUCTION **1**

1.1 Background 1

1.2 Overview of Dissertation 6

2 Preliminaries **7**

2.1 RM Codes 7

2.2 Conventional Code-Based Cryptosystems 10

2.2.1 McEliece Cryptosystem 10

2.3 Attacks on McEliece Cryptosystems 12

2.3.1 Minder-Shokrollahi's Attack 12

2.3.2 Chizhov-Borodin's Attack 13

2.3.3 Square Code Attack 13

2.4 Conventional Code-Based Signature Scheme 14

2.4.1 Niederreiter Cryptosystem 14

2.4.2 CFS Signature Scheme 15

2.4.3 RM Code and Its Modification 17
2.5 Sequences 19

3 Punctured Reed–Muller Code-Based McEliece Cryptosystems 21
3.1 Modifications of RM Code-Based Cryptosystem 21
3.1.1 Modification by Puncturing 21
3.1.2 Modification With Puncturing and Insertion 23
3.2 Security of the Proposed Cryptosystems 24
3.2.1 Secure Against Minder-Shokrollahi's Attack 24
3.2.2 Secure Against Chizhov-Borodin's Attack 27
3.2.3 Secure Against Square Code Attack 27
3.2.4 Information Set Decoding Attack 29

4 A New Signature Scheme Based on Punctured Reed–Muller Code With Random Insertion 32
4.1 New Signature Scheme Using Punctured RM Code With Random Insertion 32
4.1.1 Proposed Signature Scheme 32
4.1.2 Preprocessing for Error Weight Parameter 36
4.1.3 Additional Modification of the Algorithm 38
4.2 Implementation of the Proposed Code-Based Signature Scheme . . . 41
4.2.1 List of Parameter Sets 41
4.2.2 Description of Platform 41
4.2.3 Time 42
4.2.4 Space 42
4.2.5 How Parameters Affect Performance 42
4.3 Security Analysis of the Proposed Code-Based Signature Scheme . . 43
4.3.1 EUF-CMA 43
4.3.2 Forgery Attack 51

4.3.3 Information Set Decoding Attack 52

5 New Families of p-ary Sequences of Period $\frac{p^n-1}{2}$ With Low Maximum Correlation Magnitude **54**

5.1 Known Sequences With Low Correlations 54

5.2 Characters and Weil Bound 57

5.3 New Sequence Families and Their Correlation Bound 58

6 Conclusion **63**

Abstract (In Korean) **70**

List of Tables

3.1 Required complexity for operations of the square code attack on RM$(r, 2r)$ code 28

3.2 Comparison of the proposed cryptosystems with original cryptosystems in terms of information set decoding 30

3.3 Comparison of the proposed cryptosystems with original McEliece cryptosystems in terms of public key size and security level 31

4.1 The probability of successful signing for parameters N and w in RM$(5, 10)$ 38

4.2 An error weight parameter w and their signing time for 10,000 random syndromes 39

4.3 CPU cycles of pqsigRM with —Intel(R) Xeon(R) CPU E5-2698 v4 2.20GHz— 42

4.4 Public key and secret key size of pqsigRM (byte) 43

4.5 The security of the proposed signature scheme for $N = 10,000$ . . . 51

4.6 The WF for each RM code 53

5.1 Comparison with some well-known sequence families 56

5.2 Simulation results of $C_{\max}$ and number of correlation values for some p and n 62

List of Figures

4.1 Modified parity check matrix of the proposed signature scheme. . . . 35
4.2 Signing process of the proposed signature scheme. 35
4.3 Distribution of Hamming weights of coset leaders among 10^7 in RM$(5, 10)$. 37

Chapter 1

INTRODUCTION

1.1 Background

It has been known that most of the conventional public key cryptosystems such as RSA cryptosystem, elliptic curve cryptosystem, and so on, can be broken by sophisticated operations on quantum computers. Thus, lots of researches have been devoted for the cryptosystems robust to the attack by quantum computers, called post-quantum cryptosystems. In 1978, McEliece [1] first proposed a code-based cryptosystem using a generator matrix of binary Goppa code, and later Niederreiter [14] suggested another version of code-based cryptosystem using a parity check matrix based on a syndrome decoding problem known as an NP-complete problem [16]. Although encryption and decryption of the McEliece cryptosystem are usually faster than those of the conventional cryptosystems such as RSA and elliptic curve cryptosystems, it requires very large public and private key sizes. Thus, there have been many works to reduce the key sizes of the McEliece cryptosystems.

One of the approaches to reduce the key size is adopting other error correcting codes instead of the Goppa code [2]–[4] and utilizing their mathematical structures. For example, the generalized Reed-Solomon (GRS) code [2], the polar code [3], and the Reed-Muller (RM) code [4] have been used for the code-based cryptosystems. In

the GRS code-based cryptosystem, the private key matrix is determined by two vectors α, v, and thus the key size of the McEliece cryptosystem can be dramatically reduced. Further, it is known that McEliece cryptosystem using RM codes can add much larger number of errors than the minimum distance of the RM codes and thus the the matrix size can also be reduced with the same security level [4].

However, the McEliece cryptosystem based on RM codes is proved to be insecure due to the Minder-Shokrollahi's attack [5] and later the Chizhov-Borodin's attack. While the structure of error correcting codes helps us to reduce the key size, the code structure may also reveal information on the private key to attackers. In order to avoid these attacks, the McEliece cryptosystems based on the GRS or RM codes with random column insertion for the generator matrix are proposed [7], [8]. However, it turns out that they can be broken by the square code attack [9].

Many code-based cryptosystems have since been proposed by replacing the binary Goppa code with other error correcting codes. However, no valid code-based digital signature scheme were proposed for more than two decades after that point. In 2001, Courtois, Finiasz, and Sendrier introduced the first code-based digital signature scheme, called the Courtois, Finiasz, and Sendrier (CFS) signature scheme [17]. The CFS signature scheme is based on Niederreiter cryptosystem. In the signing process of the CFS scheme, the hash of message M, $h(h(M)|i)$, is generated, and is considered as a syndrome of the given code, where $h(\cdot)$ is a cryptographic hash function from $\{0,1\}^*$ to $\{0,1\}^{n-k}$, and i is a counter value used to adjust the hashed message corresponding to the syndrome for a valid error. Therefore, to generate a valid signature, we need to search the corresponding error vector e whose Hamming weight is less than or equal to the error correctability t of the code such that $He^T = h(h(M)|i)$. This error vector and counter i are the signature of the given message M. Clearly, finding e is a syndrome decoding problem, and it is known that, to find a valid signature e, $t!$ trials (increase i, recalculate $h(h(M)|i)$, and apply syndrome decoding to find e) are expected on average for the case of Goppa codes in the CFS signature scheme [17],

which requires a tremendously large number of signing trials for a large t. Therefore, the CFS signature scheme is only applicable for a relatively small t and is thus based on a high rate Goppa code. However, it is known that the generator matrix of a high rate Goppa code can be distinguished from a random matrix [18], and thus the CFS scheme is not robust or existentially unforgeable against a chosen message attack (EUF-CMA) [18].

Thus, there have been many efforts to relieve the security problem of the CFS signature scheme. One approach is to adopt other codes in place of the Goppa code. For example, low-density generator matrix (LDGM) [19] and convolutional code-based signature schemes have been proposed [20]. However, the LDGM code-based signature scheme was recently proved to be insecure [21]. After collecting a large number of signatures, attackers can find the correlation of signatures and can then decompose the public key $H' = SHQ$ into the private keys S, H, and Q. Another approach to relieve the security problem of the CFS signature scheme is utilizing a modified Goppa code. Because a high rate Goppa code-based signature scheme is not secure against EUF-CMA [18], they proposed a modified CFS scheme using an $(n, k-1)$ expurgated Goppa code to evade the Goppa code distinguishing problem [22]. In general, the CFS signature scheme has a small value of t, which causes a vulnerability to birthday attacks [23]. It is noted that if we can use complete decoding, the decoding for errors larger than t can be possible, which improves the security and successful signing probability.

In this dissertation, a modification method of McEliece cryptosystems based on the punctured RM codes with random insertion is proposed. In this modification, some columns of the generator matrix of the original RM codes are carefully punctured to prevent effective cryptanalysis. In fact, the puncturing of generator matrix in the McEliece cryptosystem was considered in the quasi cyclic-low density parity check (QC-LDPC) code-based McEliece cryptosystem [10]. However, in [10], they only analyzed the security of the QC-LDPC code-based cryptosystem in terms of information

set decoding. On the contrary, we focus on the effect of puncturing of the generator matrix and figure out how many columns should be punctured and where are the effective locations in order to hide mathematical structure of the codes against the known attacks.

Here, we focus on the McEliece cryptosystem based on RM codes, while the modification of the RM codes by puncturing can be applied to the other code-based McEliece cryptosystems. We will show that if public keys are properly modified by the proposed sophisticated puncturing of the generator matrix with or without randomly inserting random columns into the punctured generator matrix, all the known attacks for the McEliece cryptosystems based on RM codes do not work anymore. While an attack with randomly permuted and scrambled generator matrix is not possible to find the mathematical structure of the original code, the legitimate receiver can reconstruct the original message because they know the exact locations of punctured and inserted columns of the generator matrix. We carry out analysis on the security with respect to the known attacks for the proposed McEliece cryptosystem based on punctured RM codes with random insertion, which is proven to be secure.

In this dissertation, we propose a new variant of the CFS signature scheme based on punctured Reed–Muller (RM) code with random insertion. The modified RM code can perform complete decoding by utilizing a well-known and efficient recursive decoding [24], [25], called closest coset decoding, that is, for a given received vector, the closest codeword can be found. The closest coset decoding method does not guarantee an exact error correction, but finds an error vector (coset leader in the standard array) corresponding to the syndrome. However, the exact error correction is not essential for signing in code-based signature schemes, but we need to find the error vector with the smallest Hamming weight in the coset corresponding to the syndrome. In this respect, the RM code-based signature scheme can be considered as a solution to the small t constrained problem of the Goppa code-based signature scheme. Further, the proposed RM code-based signature scheme can compromise the signing time and security level

by adjusting the allowable maximum Hamming weight of error vectors, called the error weight parameter $w = t + \delta$.

However, the simple replacement of Goppa code with RM code in the CFS signature scheme results in vulnerability to several attacks. The RM code-based McEliece cryptosystem [4] is insecure owing to Minder–Shokrollahi attack [5] and Chizhov–Borodin attack [6]. With these two attacks, the private keys S, G, and Q can be revealed from the public key $G' = SGQ$. These attacks can similarly be applied to the RM code-based signature scheme. It is shown herein that the punctured RM codes with random insertion can be secure against these attacks and an optimal puncturing scheme for preventing Minder–Shokrollahi and Chizhov–Bordin attacks is proposed [26]. In addition, it is also shown that the punctured RM code with random insertion is secure from a square code attack [9], which can distinguish randomly inserted columns from the modified generator matrix. In this dissertation, it is also proved that the proposed modified RM code-based signature scheme is EUF-CMA secure under the assumption that the parity check matrix of the modified RM code is not distinguishable from a random matrix.

Pseudo random sequences with low correlation are widely used in random number generation and wireless communications, that is, code division multiple access, spread spectrum, cryptography, and error correcting codes.

In this dissertation, new p-ary sequence families with low correlation are constructed. For an odd prime $p \equiv 3 \pmod 4$ and an odd integer n, two new p-ary sequence families of period $N = \frac{p^n-1}{2}$ having the correlation magnitude upper bounded by $\frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}} + \frac{1}{2}$ are constructed. These sequence families can be obtained from shift and addition of two decimated p-ary m-sequences by 2 and d. One sequence family is obtained for $d = 4$ and the other sequence family is constructed for $d = N + 1$. The hybrid sum of Weil bound is used for the proof of the upper bound of correlation magnitude.

1.2 Overview of Dissertation

This dissertation is organized as follows. In Chapter 2, basic concept of RM code-based post quantum cryptosystems and pseudo random sequences are presented as preliminaries for understanding the whole of the dissertation. The definition of RM code and McEliece cryptosystems are introduced. Then, various attacks on RM code-based McEliece cryptosystems are described. In the part of code-based signature scheme, the CFS signature scheme and modification techniques are introduced. Lastly, in the part of pseudo random sequence, necessary definitions and notions for p-ary sequences and their cross-correlations are given. Also, some previous results for the bound of additive and multiplicative characters are introduced. In Chapter 3, the modified McEliece cryptosystems based on the RM code with sophisticated column puncturing of the generator matrix with or without inserting random columns are proposed. Also the security of the proposed McEliece cryptosystems in the various known attacks are verified. In Chapter 4, the modified CFS signature scheme based on the punctured RM code with random insertion is proposed. Then, the security of the proposed signature scheme is presented. Further, the implementation of the proposed signature scheme based on the punctured RM code with random insertion are given. In Chapter 5, the new families of p-ary sequences of period $p^n - 1/2$ with low maximum correlation magnitude is constructed. The decimation factors of the proposed sequences are 4 and $(p^n + 1)/2$. Finally, the concluding remarks are given in Chapter 6.

Chapter 2

Preliminaries

2.1 RM Codes

The RM code RM(r, m) is a linear code defined by Boolean functions of m variables and degree less than or equal to r for any integers m and r with $0 \leq r \leq m$. A Boolean function of m variables is evaluated on 2^m different positions, which corresponds to a codeword of length 2^m in RM(r, m). RM(r, m) is the set of codewords obtained by evaluating all the Boolean functions of m variables and degree less than or equal to r. The set of Boolean functions in the variables $v_1, \cdots, v_m$ of degree less than or equal to r is denoted by $B(r, \{v_1, \cdots, v_m\})$.

Definition 1 (Indicator vector)**.** *Let S be the set of all m-dimensional binary vectors*

$$S = F_2^m = \{x_1, \cdots, x_n\}.$$

Then the n-dimensional indicator vector I_H on subset $H \subset S$ is defined as

$$(I_H)_i = \begin{cases} 1, \text{if } \ x_i \in H \\ 0, \text{otherwise.} \end{cases}$$

The variables $v_1, v_2, \cdots, v_m$ are defined as the following vectors with length $n = 2^m$ in n-dimensional space F_2^n

$$v_0 = (1, 1, \cdots, 1)$$

and

$$v_i = I_{H_i}$$

where $1 \leq i \leq m$ and H_i is a hyperplane in F_2^m with dimension $m-1$ as

$$H_i = \{y \in F_2^m | y_i = 0\}.$$

The operation $\cdot$ of variables is defined as componentwise product.

Then, the generator matrix of the RM code can be constructed by using vectors and their products up to r times as

$$\{v_0, v_1, \cdots, v_m, \cdots, (v_{i_1} \cdots v_{i_2}), \cdots, (v_{i_1} \cdot v_{i_2} \cdots v_{i_r})\}$$

where $1 \leq i_k \leq m$. Then vectors corresponds to the rows of the generator matrix of RM code RM(r, m).

Example 1. *The RM(2,3) code is generated by the set*

$$\{v_0, v_1, v_2, v_3, v_1 \cdot v_2, v_1 \cdot v_3, v_2 \cdot v_3\}$$

and each v_1, v_2, v_3 are defined as

$$v_1 = (10101010)$$
$$v_2 = (11001100)$$
$$v_3 = (11110000).$$

Then, the generator matrix of RM code is given as

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The code parameters are given as $k = 7, n = 8$, and $d = 2$.

The fact that all Boolean functions generating codewords in $B(r-1, \{v_1, \cdots, v_m\})$ are also in $B(r, \{v_1, \cdots, v_m\})$ implies the following proposition.

Proposition 2. *For any integer m, we have*

$$\mathrm{RM}(0, m) \subset \mathrm{RM}(1, m) \subset \cdots \subset \mathrm{RM}(m, m).$$

The code length n, dimension k, and the minimum distance d of RM(r, m) are given as

$$n = 2^m, k = \sum_{i=0}^{r} \binom{m}{i}, d = 2^{m-r}.$$

We need the following definitions for the proposed RM code-based McEliece cryptosystem.

Definition 3. *The support of a codeword $c \in$ RM(r, m) is defined as the set of indices i such that $c_i \neq 0$, which is denoted by supp(c).*

Definition 4. *Let c be a codeword of C and L be an index set. Then, $\mathrm{proj}_L(c)$ is a sub-codeword which is composed of the components with indices in L from c. Also for a linear code C, we define $\mathrm{proj}_L(C) = \{\mathrm{proj}_L(c) | c \in C\}$.*

Example 2. *Let $c = (11011001)$ and $L = \{1, 2, 3, 7\}$. Then, $\mathrm{proj}_L(c) = (1100)$, which is composed of the 1st, 2nd, 3rd, and 7th components of c.*

Proposition 5 (Minder and Shokrollahi, 2007 [5])**.** *Let x be a codeword with the minimum weight in RM(r, m). Then, there exist $x_1, x_2, \cdots, x_r \in$ RM$(1, m)$ such that*

$$x = x_1 \cdot x_2 \cdot \cdots \cdot x_r$$

where x_i is a codeword with the minimum weight in RM$(1, m)$ and $x_i \cdot x_j$ denotes the componentwise multiplication.

Propositions 2 and 5 are used as the main tools for the Minder-Shokrollahi's attack, which will be explained in the following subsection.

2.2 Conventional Code-Based Cryptosystems

2.2.1 McEliece Cryptosystem

McEliece introduced a public key cryptosystem based on the difficulty of decoding of random linear codes, which consists of three algorithms such as key generation, encryption, and decryption as follows [1].

Key Generation: Let G be a $k \times n$ generator matrix of the (n, k) linear code. Let S be a $k \times k$ scrambling matrix and P an $n \times n$ permutation matrix. Bob generates the public key by calculating $G' = SGP$, where S, P, and G are the private keys of Bob. The error correction capability t of the linear code with generator matrix G is also disclosed.

Encryption: Alice generates a codeword corresponding to a message $m \in \{0, 1\}^k$ using Bob's public key (G', t). She chooses a random error vector $e \in \{0, 1\}^n$ with Hamming weight at most t and sends the ciphertext $c = mG' + e$ to Bob.

Decryption: When Bob receives the ciphertext c, he first multiplies P^{-1} to the right hand side of the ciphertext as $cP^{-1} = mSG + eP^{-1}$. Using decoding algorithm, Bob finds mS and then by multiplying S^{-1}, he can recover the original message m.

Originally, the generator matrix G of the McEliece cryptosystem is a generator matrix of the Goppa code and later there have been many suggestions of the generator matrices of the different linear codes including RM codes.

In this dissertation, we focus on the generator matrix of the RM codes. RM code-based cryptosystem is firstly proposed by Sidelnikov [4] and thus, it is also called as 'Sidelnikov cryptosystem'. In this cryptosystem, Sidelnikov introduced a large number of errors greater than the error correction capability t (for the case of RM code, $2^{m-r-1}-1$), e.g., for $(n, k, r) = (1024, 176, 3)$ or $(2048, 232, 3)$, the number of errors is greater than 200 or 400, respectively. Even with the excessive errors, it is known that the legitimate receiver can successfully remove them with high probability by using an efficient decoding algorithm of the RM code proposed by Sidelnikov [11]. This means

that an attacker should correct a larger number of errors than the error correctability t in the ciphertext using decoding of random linear code, which imposes more difficulty on the attacker. However, it turned out that the mathematical structure of RM code in the public key (a randomized generator matrix) reveals the secret information, the random permutation matrix, and private key using sophisticated attacking algorithms [5], [6]. It should be noted that the McEliece cryptosystem based on RM codes was broken by three known attacks.

In order to avoid the Minder-Shokrollahi's attack and the Chizhov-Borodin's attack, a modification scheme for generator matrices by inserting random columns into the random positions of the generator matrices was proposed [7]–[9]. In the subsequent discussion, we will use the following notations for the insertion of random columns into the generator matrix. Let L_I be a set of inserted column indices of the $k \times (n + |L_I|)$ generator matrix G_I. Then the permutation matrix P_I should be an $(n + |L_I|) \times (n + |L_I|)$ matrix. The encryption procedure is exactly the same as that of the original McEliece cryptosystems. In the decryption procedure, after multiplying P_I^{-1} to the received ciphertext, Bob can delete the elements with indices in L_I from $mSG_I + eP_I^{-1}$ and then the remaining decryption procedure is the same as that of the McEliece cryptosystems. It was known that when the insertion is solely used for their modification, it is not secure by the square code attack.

However, by using the random insertion after the sophisticated puncturing of generator matrix of the RM codes, it can be shown that the modified generator matrices do not reveal the secret to attackers anymore. We will show that the RM code-based cryptosystem can be resurrected by using the proposed modification of McEliece cryptosystem based on the punctured RM codes with random insertion

2.3 Attacks on McEliece Cryptosystems

In this subsection, we briefly describe the main ideas of decisional cryptanalyses for the McEliece cryptosystems based on RM codes.

2.3.1 Minder-Shokrollahi's Attack

One of the major objects of the attack on the McEliece cryptosystems is to find the permutation matrix P. Let $C = \mathrm{RM}(r,m)^\sigma$ be the permuted code of $\mathrm{RM}(r,m)$ for some unknown permutation σ. In the Minder-Shokrollahi's attack [4], the attack procedure to find σ is composed of three steps as follows:

1. Find codewords in C, which belong to $\mathrm{RM}(r-1,m)^\sigma$. It is required to find enough number of such codewords to build a basis of $\mathrm{RM}(r-1,m)^\sigma$.
2. Iterate the previous step until obtaining $\mathrm{RM}(1,m)^\sigma$.
3. Determine a permutation τ such that $\mathrm{RM}(1,m)^{\tau\cdot\sigma} = \mathrm{RM}(1,m)$. Then we have $\mathrm{RM}(r,m)^{\tau\cdot\sigma} = \mathrm{RM}(r,m)$. Then τ becomes P^{-1}.

As you can see, the first step is crucial for the success of this attack. Let $x \in C$ be a minimum weight codeword. Then, we define $C_{\mathrm{supp}(x)}$ as the shortened code of C on $\mathrm{supp}(x)$, that is, find only codewords which are zero on $\mathrm{supp}(x)$ in C and then puncture their components with indices in $\mathrm{supp}(x)$. For example, for $\mathrm{supp}(x) = \{1,4\}$, we have $c' = (c_2, c_3, c_5, \cdots, c_n) \in C_{\mathrm{supp}(x)}$. Clearly, the length of codewords in $C_{\mathrm{supp}(x)}$ is $n - |\mathrm{supp}(x)|$. Then, it is known that $C_{\mathrm{supp}(x)}$ is a concatenated code defined as

$$C_{\mathrm{supp}(x)} \subseteq \overbrace{\mathrm{RM}(r-1,m-r) \times \mathrm{RM}(r-1,m-r) \times \cdots \times \mathrm{RM}(r-1,m-r)}^{2^r-1} \tag{2.1}$$

where $\times$ denotes the direct product defined in [5].

Since the permutation is unknown, the position of $\mathrm{RM}(r-1,m-r)$ in $C_{\mathrm{supp}(x)}$ is supposed to be also unknown. However, the algorithm to find the position of $\mathrm{RM}(r-$

$1, m-r)$ is proposed by the Minder-Shokrollahi and thus a codeword in RM$(r-1, m)^{\sigma}$ can be determined, which corresponds to the first step in the above attack.

2.3.2 Chizhov-Borodin's Attack

Chizhov and Borodin [6] improves the computational complexity of Minder-Shokrollahi attack. Chizhov and From an RM code RM(r, m), RM$(2r, m)$ can be constructed with low polynomial-time complexity. Similarly, RM(kr, m) can easily be constructed. Moreover, RM$(m - r - 1, m)$, a dual code of RM(r, m), can also be constructed in low polynomial-time complexity. Thus, RM$(kr + l(m - 1), m)$ can be obtained and finally we have RM$(\gcd(r, m - 1), m)$. If $\gcd(r, m - 1) = 1$, then RM$(1, m)$ is directly found. Otherwise, RM$(r - 1, m)$ can be obtained by the Minder-Shokrollahi's attack. By iterating this procedure until we have $\gcd(r - k, m - 1) = 1$, RM$(1, m)$ can be found. Then it is straightforward to find the permutation τ, that is, P^{-1}.

2.3.3 Square Code Attack

According to the square code attack [9], applying insertion of random columns to the McEliece cryptosystem based on RM codes is insecure from the square code attack, which uses a property of the product of the random-column inserted RM codes. The product of codes is defined as follows.

Definition 6 (Product of codes)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be linear codes of length n. Then the product code denoted by $\mathcal{A} * \mathcal{B}$ is the vector space spanned by all the componentwise product $\mathbf{a} \cdot \mathbf{b} = (a_1 b_1, \cdots, a_n b_n)$, where $\mathbf{a} \in \mathcal{A}$ and $\mathbf{b} \in \mathcal{B}$. When $\mathcal{A} = \mathcal{B}$, $\mathcal{A} * \mathcal{A}$ is called the square code of $\mathcal{A}$, denoted by $\mathcal{A}^2$.*

Let G_L be a $k \times (n + |L_I|)$ matrix obtained by inserting $|L_I|$ random columns into the generator matrix of an RM code RM(r, m) and $\mathcal{C}$ be the code spanned by the rows of G_L. The index set $L_I \subset \{1, \cdots, n + |L_I|\}$ is the set of indices that defines inserted locations of random columns. And let $\mathcal{C}_i$ be the code generated by the matrix

$G_{L,i}$ obtained by deleting the ith column of G_L. Then the following two cases occur with high probability.

$$\dim \mathcal{C}_i^2 = \begin{cases} \dim \mathcal{C}^2 - 1, & \text{if } i \in L_I \\ \dim \mathcal{C}^2, & \text{if } i \notin L_I. \end{cases} \tag{2.2}$$

With this argument, an attacker can discover the set L_I using public key of the McEliece cryptosystem based on RM codes in the polynomial time.

2.4 Conventional Code-Based Signature Scheme

In this section, we introduce the conventional code-based signature scheme (i.e., CFS signature scheme [17]), the RM codes, and its puncturing method with random insertion. The CFS signature scheme is transformed from the Niederreiter cryptosystem [14].

2.4.1 Niederreiter Cryptosystem

Niederreiter cryptosystem [14] is proposed by Niederreiter in 1986 which is codebased cryptosystem. Niederreiter cryptosystem is also based on the nature of the syndrome decoding problem known as NP-complete problem. The difference of the Niederreiter cryptosystem from McEliece cryptosystem is that the former uses parity check matrix in public/private key rather than the latter uses generator matrix. However, McEliece cryptosystem and Niederreiter cryptosystem are proven to be equivalent [15]. The key generation, encryption, and decryption procedures of Niederrieter cryptosystem are given as follows.

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.

2.4.2 CFS Signature Scheme

Courtois, Finiasz, and Sendrier proposed the first practical code-based signature scheme, called CFS signature scheme. It is based on the Niederreiter cryptosystem and consists of three stages, namely, key generation, signing, and verification, as follows.

Key Generation: Choose a parity check matrix H of an (n, k) binary t-error correcting Goppa code with a decoding algorithm γ. The decoding algorithm γ will produce an error vector e with Hamming weight of less than or equal to the error correctability t if it exists, or output $\perp$, otherwise. Let S be an $(n-k)\times(n-k)$ scrambling matrix and Q be an $n \times n$ permutation matrix. Construct the public key $H' = SHQ$, where S, Q, and H are the private keys.

Signing: To sign a message M,

1) $i \leftarrow i + 1$

2) $e' = \gamma(S^{-1}h(h(M)|i))$

3) if e' is $\perp$, go to Step 1).

Output $(M, e = e'(Q^{-1})^T, i)$

Verification:

Compute $\hat{s} = H'e^T$ and $s = h(h(M)|i)$.

The signature is valid if s and $\hat{s}$ are equal.

The security of the CFS signature scheme is based on the hardness of solving a syndrome decoding problem, which is known as an NP-complete problem [16].

Definition 7 (Syndrome decoding problem)**.** *Given an $r \times n$ parity check matrix H, a syndrome $s \in \{0,1\}^r$, and an error correctability $t > 0$, find an error vector e in $He^T = s$ with Hamming weight of less than or equal to t.*

In the signing process of the CFS signature scheme, the hashed message $h(h(M)|i)$ with counter i is treated as a syndrome. However, it is known that the ratio of successfully decodable syndromes is only $1/t!$ for the case of the CFS signature scheme with binary t-error correcting Goppa code [17]. Therefore, to obtain a valid signature, we need to search a valid error vector by carrying out $t!$ trial decodings on average, and thus t should be small.

In the proposed signature scheme, we consider another decoding method, called a complete decoding problem, which is finding a nearest codeword to the received vector in the vector space.

Definition 8 (Complete decoding problem [17])**.** *Given an $r \times n$ parity check matrix H and a syndrome $s \in \{0,1\}^r$, find an error vector e with the minimum Hamming weight in $\{e|He^T = s\}$.*

Complete decoding problem is known as the most difficult computational problem in decoding [27]. Complete decoding makes it possible to find an error vector with Hamming weight of greater than t for the given syndrome at the cost of large computational complexity [24], [25]. However, when we apply the complete decoding to the

signing in the CFS signature scheme with binary Goppa code, there is a limitation in that the value of δ in $w = t + \delta$ cannot be sufficiently large.

In addition, there are some security drawbacks to the CFS signature scheme: (i) the parity check matrix of high rate Goppa code can be distinguished from a random matrix, and thus the CFS signature scheme is insecure under the EUF-CMA, and (ii) it has poor scaling of the parameters based on the security as in the following description. The error correcting parameter t needs to be small because the number of operations required for the generation of valid signature is significantly dependent on t, that is, $t!t^2m^3$, where $n - k = tm$. The public key size of the CFS scheme is $(n - k)n = tm2^m$, and it is known that decoding attacks require $A = 2^{tm/2}$ operations. Thus the decoding attack complexity A is only a polynomial function of the key size with small power, that is, $A \approx \text{keysize}^{t/2}$. Therefore, because t should be kept as a relatively small value of up to 12 to reduce successful signing time, we need to significantly increase the key size itself for higher security.

2.4.3 RM Code and Its Modification

In this paper, we proposed the CFS signature scheme using the modified RM codes. Because the RM code and its modified one can be decoded through complete decoding, they can improve the security drawback of the CFS signature scheme with binary Goppa code by extending the error correctability t to the error weight parameter w.

1) Complete Decoding of RM Code

Because the recursive decoding of the RM code can find the coset leader of the received vector, this decoding can be considered as closest coset decoding [25], [24]. In fact, closest coset decoding is the same as complete decoding. Therefore, the RM code-based CFS signature scheme is worth considering. However, simply modifying the CFS signature scheme by replacing it with an RM code is easily broken by the well-known Minder-Shokrollahi and Chizhov-Borodin attacks, which are used in the RM

code-based McEliece cryptosystem [26]. With these attacks, the private keys S, H, and Q of the cryptosystem can be obtained from the public key $H' = SHQ$. Therefore, we need to modify the RM code structure to achieve security under known attacks, while maintaining the complete decodable property of the RM code [26].

2) Puncturing RM Code with Random Insertion

In [26], the punctured RM code with random insertion is introduced to construct a secure RM code-based public key cryptosystem. Similarly, the puncturing and random insertion methods can be applied to an RM code-based signature scheme. In fact, the puncturing of a generator matrix G is equivalent to row and column deletions of a parity check matrix H [13]. Because a signature scheme uses a parity check matrix H, the modified H with row and column deletions and random row insertion will be used for the proposed signature scheme. The modification method of the (n, k) RM code is given as follows.

a) Row and Column Deletions of Parity Check Matrix

The systematic forms of generator and parity check matrices are given as

$$G = [I_k|P], H = [P^T|I_{n-k}] \tag{2.3}$$

where P is $k \times (n-k)$ matrix given as

$$P = [p_1 p_2 \cdots p_{n-k}] \tag{2.4}$$

with column vectors p_i of size k. The generator matrix G can be punctured by deleting columns of matrix P. Let P' be a p column deleted matrix from P. Then, the generator matrix and parity check matrix of the punctured RM code are given as

$$G_p = [I_k|P'], H_p = [P'^T|I_{n-k-p}], \tag{2.5}$$

respectively. It can then be easily checked that

$$G_p H_p^T = 0_{k\times(n-k-p)}. \tag{2.6}$$

b) Modification of Parity Check Matrix with Random Row Insertion

The punctured generator matrix G_p can be modified by inserting random columns into P'. Then, the punctured generator matrix with random column insertion is denoted as $G_m = [I_k|P'']$, where p random columns are inserted in P''. Then, the generator matrix and its parity check matrix are given as

$$G_m = [I_k|P''], H_m = [P''^T|I_{n-k}]. \tag{2.7}$$

Clearly, we have $G_m H_m^T = 0_{k\times(n-k)}$.

2.5 Sequences

Several notations and definitions are given as follows.

1. Let p be an odd prime such that $p \equiv 3 \bmod 4$ and n be an odd positive integer, where $q = p^n$.

2. Let $\mathbb{F}_q$ be the finite field with q elements and α be a primitive element of $\mathbb{F}_q$.

3. The trace function from $\mathbb{F}_q$ to $\mathbb{F}_p$ is defined as

$$\mathrm{Tr}_1^n(x) = \sum_{i=0}^{n-1} x^{p^i}.$$

4. $\omega = e^{\frac{2\pi i}{p}}$ is a primitive complex pth root of unity, where $i = \sqrt{-1}$.

5. For some $\beta \in \mathbb{F}_q^*$, a p-ary m-sequence of period $q-1$ is defined as

$$m(t) = \mathrm{Tr}_1^n(\beta\alpha^t).$$

6. Let $a(t)$ and $b(t)$ be p-ary sequences of period N. A cross-correlation between $a(t)$ and $b(t)$ is defined as

$$C_{a,b}(\tau) = \sum_{t=0}^{N-1} \omega^{a(t)-b(t+\tau)}.$$

If $a = b$, then the cross-correlation function becomes the autocorrelation function, denoted by $C_a(\tau)$. Let S be a family of sequences of period N. Then the maximum magnitude of correlation values of the sequences in S is defined as

$$C_{\max} = \max\left\{|C_{a,b}(\tau)| : a, b \in S, 0 \leq \tau \leq N-1, \tau \neq 0 \text{ if } a = b\right\}.$$

Chapter 3

Punctured Reed–Muller Code-Based McEliece Cryptosystems

3.1 Modifications of RM Code-Based Cryptosystem

In this section, we propose modifications of McEliece cryptosystems based on the RM codes. The proposed cryptosystems start from the modification of generator matrix of an RM code RM(r, m), where $n = 2^m$ and $k = \sum_{i=0}^{r} \binom{m}{i}$. Here, we will consider two modifications: (i) modification by puncturing at the minimum number of the specified locations of the generator matrix and (ii) modification by both the sophisticated puncturing and randomly inserting columns of the generator matrix.

3.1.1 Modification by Puncturing

The proposed modification of McEliece cryptosystem can be presented by the following three algorithms.

1) Key Generation

1-1) Puncturing: Let G be a $k \times n$ generator matrix of RM code, $C = \text{RM}(r, m)$. Then find a minimum weight codeword x in C and $\text{supp}(x)$ and find a minimum weight codeword y in $\text{proj}_{\text{supp}(x)}(C)$. Then, we have the set of indices L_D corresponding to

$\mathrm{supp}(y)$ in the original code indices and finally delete columns with indices in L_D from G, which is denoted by G_D.

1-2) Generating S and P: Let S be a $k \times k$ scrambling matrix and P be an $(n - |L_D|) \times (n - |L_D|)$ permutation matrix. The public key is generated by calculating $G'_D = SG_DP$. The number of arbitrary random errors $t' = \lfloor t - |L_D|/2 \rfloor$ of G_D is known to others together with G'_D as the public keys. Note that the parameter t is determined according to the decoding algorithm by Sidelnikov which can correct almost all errors up to some limit greater than the original error correction capability of RM codes [11]. The private keys are P, S, G, and L_D.

2) Encryption

Alice encrypts a message $m \in \{0,1\}^k$ using Bob's public key (G'_D, t'). She chooses a random error vector $e \in \{0,1\}^{n-|L_D|}$ with Hamming weight at most t' and sends the ciphertext $c = mG'_D + e$ to Bob.

3) Decryption

When Bob receives the ciphertext c, he first multiplies P^{-1} to the right hand side of the ciphertext as $cP^{-1} = mSG_D + eP^{-1}$ and then inserts the erasure mark '?' in the jth positions, $j \in L_D$ for an erasure decoding. Alternatively, Bob can randomly insert '0' or '1' instead of the erasure mark '?' and then he can apply the conventional decoding algorithm since the erasures can be treated as errors. After decoding, multiplying S^{-1} to mS, he can recover the original message m.

In the modified McEliece cryptosystems, we can exchange the order of encryption procedures, that is, puncturing step and key generation step. It is not difficult to check that in the proposed modification of McEliece cryptosystem, deleting the columns of G' after calculating $G' = SG_DP$ and deleting the columns of G before calculating $G'_D = SGP$ are equivalent.

Algorithm 1 Puncturing procedure

Input: $k \times n$ generator matrix G of RM code

Output: $k \times (n-p)$ punctured generator matrix G_D

1. Randomly pick a minimum Hamming weight code x from C.

2. Randomly pick a minimum weight codeword y from $\mathrm{proj}_{\mathrm{supp}(x)}(C)$.

3. Choose p, such that $\mathrm{wt}(y) \leq p \leq 2\mathrm{wt}(y)$.

4. Randomly choose the set of indices L_D such that $\mathrm{supp}(y) \subseteq L_D$ and $|L_D| = p$.

5. Delete the columns with indices in L_D from G, which is denote by G_D.

3.1.2 Modification With Puncturing and Insertion

Although the modification of RM code-based McEliece cryptosystem by insertion to increase security level was proposed [8], this cryptosystem was broken by the square code attack [9]. However, it can be shown that the proposed cryptosystem which simultaneously uses the sophisticated puncturing and random insertion can prevent the square code attack as well as the other known attacks.

1) Key Generation

1-1) Puncturing: The random puncturing procedure is the same as that in the previous subsection by L_D, whose code is called a punctured code.

1-2) Insertion: Let $L_I = \{l_1, \cdots, l_I\}$ be a set of randomly chosen $|L_I|$ column indices, where $1 \leq l_i \leq n - |L_D| + |L_I|, 1 \leq i \leq |L_I|$. Then, insert random columns into G_D, denoted by G_{DI}, where columns with indices in L_I are inserted columns.

1-3) Generating S and P: Let S be a $k \times k$ scrambling matrix and P be an $(n-|L_D|+|L_I|) \times (n-|L_D|+|L_I|)$ permutation matrix. The public key is generated by calculating $G'_{DI} = SG_{DI}P$. The number of random errors, $t' = \lfloor t - |L_D|/2 \rfloor$ of G'_{DI} is known to others together with G'_{DI} as the public keys. The private keys are P, S, G, L_D, and L_I.

2) Encryption:

Alice encrypts a message $m \in \{0,1\}^k$ using Bob's public key (G'_{DI}, t'). She chooses a random error vector $e \in \{0,1\}^{n-|L_D|+|L_I|}$ with Hamming weight at most t' and sends the ciphertext $c = mG'_{DI} + e$ to Bob.

3) Decryption:

When Bob receives the ciphertext c, he first multiplies P^{-1} to the right hand side of the ciphertext as $cP^{-1} = mSG_{DI} + eP^{-1}$. He deletes the inserted elements with indices in L_I of cP^{-1} and then he inserts the erasure mark '?' in the jth positions, $j \in L_D$ for an erasure decoding or randomly inserts '0' or '1' for normal error decoding of RM codes. After decoding procedure, multiplying S^{-1} to mS, he can recover the original message m.

3.2 Security of the Proposed Cryptosystems

In this section, we will discuss the security of the proposed cryptosystems from the known attacks such as Minder-Shokrollahi's attack, Chizhov-Borodin's attack, square code attack, and information set decoding.

3.2.1 Secure Against Minder-Shokrollahi's Attack

Let C be a permutated RM code, i.e., $C = \mathrm{RM}(r,m)^\sigma$ by the permutation σ, that is, the permutation matrix P and x be the minimum weight codeword in C. We would like to find the minimum $|L_D|$, where the Minder-Shokrollahi's attack does not work. Remember that all punctured positions are included in $\mathrm{supp}(x)$ in the proposed puncturing method. Let C' be the punctured code of C by the index set L_D. The first step of the Minder-Shokrollahi's attack is to find the minimum weight codewords. Let x' be a minimum weight codeword of C', which is a punctured codeword of $x \in C$. Then we can find $C'_{\mathrm{supp}(x')}$. The support set of the punctured codeword is denoted as

$$\mathrm{supp}(x') = \{a_1, a_2, \cdots, a_{2^{m-r}-|L_D|}\}$$

with $|\mathrm{supp}(x')| = 2^{m-r} - |L_D|$.

Clearly, the code lengths of $C_{\mathrm{supp}(x)}$ and $C'_{\mathrm{supp}(x')}$ are the same and $C_{\mathrm{supp}(x)} \subseteq C'_{\mathrm{supp}(x')}$ since all deleted positions are included in $\mathrm{supp}(x)$. Now, we are interested in the case of $C_{\mathrm{supp}(x)} \neq C'_{\mathrm{supp}(x')}$, that is, $C_{\mathrm{supp}(x)} \subsetneq C'_{\mathrm{supp}(x')}$, for which the Minder-Shokrollahi's attack does not work. In the following theorem, we can determine the minimum number of punctured positions in order to prevent the Minder-Shokrollahi's attack.

Theorem 9. *For an RM code* $\mathrm{RM}(r,m)$*, at least* $|L_D| = 2^{m-2r}$ *is required for* $C_{\mathrm{supp}(x)} \subsetneq C'_{\mathrm{supp}(x')}$*. And the support set of the minimum weight codeword in* $\mathrm{proj}_{\mathrm{supp}(x)}(C)$ *is the essential punctured locations, where* x *is the minimum weight codeword of* C*.*

Proof. It is not difficult to check that $\mathrm{RM}(r, m-r) = \mathrm{proj}_{\mathrm{supp}(x)}(C)$. Thus, the minimum weight codeword of $\mathrm{proj}_{\mathrm{supp}(x)}(C)$ is 2^{m-2r}. Let y be the minimum weight codeword in $\mathrm{proj}_{\mathrm{supp}(x)}(C)$. Then, there exists $z \in C$ such that

$$y = \mathrm{proj}_{\mathrm{supp}(x)}(z). \tag{3.1}$$

Then, $\mathrm{proj}_{N\setminus\mathrm{supp}(x)}(z)$ clearly belongs to $C'_{\mathrm{supp}(x')}$ but not to $C_{\mathrm{supp}(x)}$, where $N = \{1, 2, \cdots, n\}$. Thus, $C_{\mathrm{supp}(x)} \subsetneq C'_{\mathrm{supp}(x')}$. □

Example 3. *Consider an RM code* $RM(2,5)$*. Suppose that the permutation matrix and scrambling matrix are identity matrices for simplicity. Clearly, one of the minimum weight codewords is* $x = (1111111100\cdots00) \in RM(2,5)$ *and* $\mathrm{proj}_{\mathrm{supp}(x)}(C) = \mathrm{RM}(2,3)$*. Then, one of the minimum weight codewords in* $\mathrm{proj}_{\mathrm{supp}(x)}(C)$ *is* $y = (10001000)$*. Also, we set* $L_D = \{1,5\}$ *and the punctured codeword of* x *is* $x' = (?111?11100\cdots00) = (11111100\cdots00)$*. And* z *in (3.1) is*

$$z = (10001000|10001000|10001000|10001000). \tag{3.2}$$

Since $C_{\mathrm{supp}(x)}$ *forms* $\mathrm{RM}(1,3) \times \mathrm{RM}(1,3) \times \mathrm{RM}(1,3)$*,* $\mathrm{proj}_{N\setminus\mathrm{supp}(x)}(z)$ *does not belong to* $C_{\mathrm{supp}(x)}$*. However,* $\mathrm{proj}_{N\setminus\mathrm{supp}(x)}(z)$ *belongs to* $C'_{\mathrm{supp}(x')}$ *by definition. And* $C'_{\mathrm{supp}(x')} \subseteq (\mathrm{RM}(1,3) + \{\mathbf{0}, (10001000)\}) \times (\mathrm{RM}(1,3) + \{\mathbf{0}, (10001000)\}) \times (\mathrm{RM}(1,3) + \{\mathbf{0}, (10001000)\})$.

Example 4. *Consider an RM code, RM*$(1,4)$*. Suppose that the permutation matrix and scrambling matrix are identity matrices for simplicity. Then, one of the minimun weight codewords is* $x = (1111111100000000)$*. Then,* $\mathrm{proj}_{\mathrm{supp}(x)}(C) = \mathrm{RM}(1,3)$ *and minimum weight of* $\mathrm{proj}_{\mathrm{supp}(x)}(C)$ *is 4. Thus, to prevent the Minder-Shokrollahi's attack at least 4 components should be punctured. If we puncture less than 4 components, the attack cannot be prevented. This can be described as follows. Let* $L_D = \{1,2,3\}$*, where* $|L_D| = 2^{m-2r} - 1 = 3$*. Let* C' *be a punctured code and* $x' = (111110\cdots0)$*. Then the generator matrix of* $C'_{\mathrm{supp}(x')}$ *is*

$$C'_{\mathrm{supp}(x')} = (0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,1). \tag{3.3}$$

The $C'_{\mathrm{supp}(x')}$ *is equal to* $C_{\mathrm{supp}(x)}$ *and attacker can proceed to the next step of the Minder-Shokrollahi's attack. But, if* $L_D = \{1,2,3,4\}$*, the generator matrix of* $C'_{\mathrm{supp}(x')}$ *is given as*

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{3.4}$$

Then, we have $C_{\mathrm{supp}(x)} \subsetneq C'_{\mathrm{supp}(x')}$.

Similarly, it is not difficult to check that

$$C'_{\mathrm{supp}(x')} \subseteq \overbrace{C'_i \times C'_i \times \cdots \times C'_i}^{2^r-1} = (C'_i)^{2^r-1} \tag{3.5}$$

where

$$C'_i = \mathrm{RM}(r-1, m-r) + \{\mathbf{0}, \text{an element in the basis of } \mathrm{RM}(r, m-r)\}. \tag{3.6}$$

It is difficult to correctly decompose component codes C'_i from $C'_{\mathrm{supp}(x')}$ because many random component codes of $C'_{\mathrm{supp}(x')}$ can form

$$\mathrm{RM}(r-1, m-r) + \{\mathbf{0}, \text{an element in the basis of } \mathrm{RM}(r, m-r)\}$$

.

3.2.2 Secure Against Chizhov-Borodin's Attack

The Chizhov-Borodin's attack uses the property that the dual code of the RM code is also the RM code. For the punctured RM codes, their dual codes are the shortened RM codes [13]. It is not possible to recover the RM codes from the shortened RM codes, because some rows and columns are deleted from the generator matrix. Therefore, the Chizhov-Borodin's attack cannot be applied to the McEliece cryptosystem based on the proposed punctured RM codes.

Example 5. *Let G be a generator matrix of RM$(1,3)$ and G_D denote a generator matrix of a punctured RM code of RM$(1,3)$, where the first and the second columns of the generator matrix are deleted. Then we have*

$$G_D = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \tag{3.7}$$

and the generator matrix of its dual code is given as

$$G_D{}^{\perp} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{3.8}$$

The dual code of RM$(1,3)$ is also RM$(1,3)$ but the dual code of the punctured RM code is the shortened code of the RM code. That is, $G_D^{\perp}$ is the generator matrix of the shortened RM code of RM$(1,3)$ but the deleted rows by shortening cannot be recovered.

3.2.3 Secure Against Square Code Attack

For $m = 2r$, let us consider an RM code $\mathrm{RM}(r, m)$, with insertion of $|L_I|$ random columns, where the code length is $2^m + |L_I|$. Then, its square code becomes $\mathcal{C}^2 = \mathrm{RM}(m, m)$ with $|L_I|$ random column insertions. The minimum weight codeword of

Table 3.1: Required complexity for operations of the square code attack on RM$(r, 2r)$ code

	Number of square code attack cases	Required operations
RM$(5, 10)$	$2^{17.5}$	$2^{67.5}$
RM$(6, 12)$	$2^{21.3}$	$2^{81.3}$

RM(m, m) is 1 and there are m minimum weight codewords in RM(m, m). Let x be one of the minimum weight codewords in RM(m, m). Assume that x has 1 in jth position. If we delete the jth column of G, then $\dim \mathcal{C}_j^2 = \dim \mathcal{C}^2 - 1$ with high probability. Thus there are $|L_I| + m$ indices of reducing dimension of C by 1. Thus, the probability to find randomly inserted columns is $1/\binom{|L_I|+m}{|L_I|}$.

If $m = 10$ and $|L_I| = 10$, $1/\binom{|L_I|+m}{|L_I|} = 1/\binom{20}{10} \approx 2^{-17.5}$. In addition, it requires $O(n^5) \approx 2^{50}$ operations for each case [9]. Thus, inserting 10 random columns requires $2^{67.5}$ more operations to succeed in the square code attack. Since the attacker has to apply the Minder-Shokrollahi's attack after success of the square code attack, it takes more time. The required complexity for operations of the square code attack on RM$(r, 2r)$ for $r = 5, 6$ is given in Table 3.1. Therefore, it is difficult to apply the square code attack on RM$(r, 2r)$ with insertion of random columns.

For $m > 2r$, using the puncturing of the RM codes, we can also prevent the square code attack because (2.2) does not hold anymore. In the case of $m = 2r$, we can find that if there is codeword with weight one in C^2, the dimension of C^2 can be reduced by deleting a column. Although the square code of $\mathrm{RM}(2r, m)$ does not contain codeword with weight 1, we can control the weight by puncturing method. It is known that the square code is RM$(2r, m)$ and the minimum weight of the square code is 2^{m-2r}. Since $2^{m-2r} \geq 2$, deleting one column does not reduce the dimension [9]. However, puncturing more than $2^{m-2r} - 1$ columns, (2.2) does not hold similarly to the case of RM$(r, 2r)$. Thus the square code attack cannot be applied to the proposed

McEliece cryptosystems. The minimum weight of punctured code in RM$(r,m)^2$ is reduced when we puncture the codeword in the support set of the minimum weight codeword in RM(r,m). Also, the square code attack is effectively prevented when $m-2r=1,2$, because the number of required puncturing bits is small.

3.2.4 Information Set Decoding Attack

The information set decoding attack is based on finding k error free bits c_k of ciphertext randomly. An adversary can choose k columns of G'_{DI} with error free indices of the ciphertext, which is denoted by $G'^{(k)}_{DI}$. Then, $c_k = m \cdot G'^{(k)}_{DI} + e_k$ with $e_k = 0$ and the decryption is done by $m = c_k \cdot G'^{(k)-1}_{DI}$. Lee and Brickell [12] generalized the information set decoding attack for $e_k \neq 0$, where the weight of e_k can be less than or equal to a given integer j. The complexity of the attack is given as

$$W_j = T_j(k^3 + N_j k) \tag{3.9}$$

where $T_j^{-1} = \sum_{i=0}^{j} \binom{t}{i}\binom{n-t}{k}/\binom{n}{k}$ and $N_j = \sum_{i=0}^{j}\binom{k}{i}$. For an RM code RM$(r,m)$, the dimension is $k = \sum_{i=0}^{r}\binom{m}{i}$ and let t be the bit error correctability. The minimum number of the punctured bits for the proposed cryptosystem is equal to $p = 2^{m-2r}$. Then the number of correctable bit errors after puncturing is given as $t' = t - 2^{m-2r-1}$ and the number of columns of the generator matrix reduces to $n' = 2^m - 2^{m-2r}$. Since the term $k^3 + N_j k$ is independent of n and t, it is reasonable to compare the term T_j for complexity of the information set decoding attack.

For example, consider the case of RM$(3,10)$. Using the decoding algorithm in [11], the number of correctable bit errors is $t = 200$. For the case of $(n,k,t) = (1024,176,200)$ without puncturing, the approximate value of T_j is 2^{61}. For the case of $(n',k,t') = (1008,176,192)$ with puncturing, we have $T_j \approx 2^{60}$. In Table 3.2, for the cases of RM$(3,10)$ and RM$(3,11)$ with or without puncturing, the corresponding values of T_j are compared.

Table 3.2 tells us that the computational complexity of the information set decoding is slightly reduced after puncturing columns. Thus, the effect of puncturing

Table 3.2: Comparison of the proposed cryptosystems with original cryptosystems in terms of information set decoding

		(n, k, t)	Number of punctured bits	T_j
RM$(3, 10)$	without puncturing	(1024,176,200)	0	2^{61}
	with puncturing	(1008,176,192)	16	2^{60}
RM$(3, 11)$	without puncturing	(2048,232,420)	0	2^{82}
	with puncturing	(2016,232,404)	32	2^{80}

columns of generator matrix for the proposed cryptosystems should be minimized. However, the proposed cryptosystem use the punctured RM codes with random column insertion and thus the security level of the proposed cryptosystems is maintained.

Table 3.3 compares the public key size and security level of the proposed cryptosystems to those of the conventional McEliece cryptosystems. Comparing McEliece64 to RM$(3, 10)$ and McEliece80 to RM$(3, 10)$, it can be verified that the public key size is reduced under the same security level.

Table 3.3: Comparison of the proposed cryptosystems with original McEliece cryptosystems in terms of public key size and security level

Cryptosystems	(n, k, t)	Public key size	Security level
McEliece64	(1024,524,50)	65.5 kB	2^{64}
McEliece80	(2048,1751,27)	437.75 kB	2^{80}
RM$(3, 10)$ with puncturing	(1008,176,200)	21.67 kB	2^{64}
RM$(3, 11)$ with puncturing	(2016,232,404)	57.1 kB	2^{80}

Chapter 4

A New Signature Scheme Based on Punctured Reed–Muller Code With Random Insertion

4.1 New Signature Scheme Using Punctured RM Code With Random Insertion

In this section, we propose a new code-based signature scheme, which is a modified version of the CFS signature scheme based on punctured RM codes with random insertion. The proposed signature scheme can improve the probability of successful signing and guarantee EUF-CMA security, which is composed of three stages, namely, key generation, signing, and verification as follows.

4.1.1 Proposed Signature Scheme

1) Key Generation

1-1) Puncturing with random insertion: Let G be a $k \times n$ generator matrix of the RM code, RM(r, m). In this dissertation, we assume the systematic form of the RM code and L_D is a set of indices of puncturing positions in the parity part P of the systematic form of the generator matrix, which was described for the nonsystematic

Algorithm 2 Puncturing procedure of generator matrix [26]

Input: $k \times n$ generator matrix G of RM code

Output: index set L_D

1. Randomly pick a minimum Hamming weight codeword x from C.
2. Randomly pick a minimum weight codeword y from $\mathrm{proj}_{\mathrm{supp}(x)}(C)$.
3. Choose p such that $\mathrm{wt}(y) \leq p \leq 2\mathrm{wt}(y)$.
4. Randomly choose the set L_D of indices such that $\mathrm{supp}(y) \subseteq L_D$ and $|L_D| = p$.

RM code in Algorithm 1 [26]. The procedure for puncturing generator matrix and determining the set L_D is described in Algorithm 1 and two important notations for Algorithm 1 are defined as follows.

Definition 10 ([26]). *The support of a codeword $c \in RM(r, m)$ is defined as the set of indices i such that $c_i \neq 0$, denoted as supp(c).*

Definition 11 ([26]). *Let c be a codeword of C and L be an index set. Then, $\mathrm{proj}_L(c)$ is a sub-codeword composed of the components with indices in L from c. In addition, for a linear code C, we define $\mathrm{proj}_L(C) = \{\mathrm{proj}_L(c)|c \in C\}$.*

Because the puncturing procedure of the generator matrix of RM code is given in Algorithm 1, the parity check matrix H corresponding to G can be modified. Some of elements in L_D of Algorithm 1 may be in the information part I of the generator matrix $G = [I|P]$, but we modify the generator matrix into the systematic form such that all elements of L_D should be in the parity part P. Using L_D in Algorithm 1, a modification algorithm of the parity check matrix corresponding to the punctured generator matrix is proposed in Algorithm 2, where the systematic form of the generator matrix $G = [I|P]$ and the parity check matrix $H = [P^T|I]$ are used.

Further, the generator and parity check matrices are row-scrambled and column-permuted to generate the public key in the signature scheme. Thus, without a loss of

Algorithm 3 Modification of parity check matrix of punctured RM code in systematic form

Input: $k \times n$ generator matrix $G = [I|P]$ of the systematic form of RM code.

Output: modified parity check matrix H_m.

1. Let $L_D = \{n-k-p+1, n-k-p+2, \cdots, n-k\}$ be an row index set in the systematic form of parity check matrix $H = [P^T|I]$ using p in Algorithm 1.

2. Replace the last p rows of the parity check matrix H by the binary random vectors r_i, $n-k-p+1 \leq i \leq n-k$ denoted as H_m, where

$$r_{ij} = \begin{cases} 1, & \text{for } j = i+k \\ 0, & \text{for } n-p+1 \leq j \leq n, j \neq i+k \\ \text{random selection of binary bits, otherwise.} \end{cases}$$

generality, we can assume that the last p columns of P in G are punctured, and thus we have $L_D = \{n-k-p+1, n-k-p+2, \cdots, n-k\}$. Therefore, the modification of the parity check matrix is described in Algorithm 2.

Then, the modified parity check matrix H_m can be described as in Fig. 4.1, where R is a $p \times (n-p)$ binary random sub-matrix with row vectors $r_i = (r_{ij})$, $n-k-p+1 \leq i \leq n-k$, $1 \leq j \leq n-p$, and P' is the last p column deleted version of P.

The deleted and inserted rows are not necessarily the same number as well as the same position but here, we assume that they are the same.

1-2) Generation of $S, Q,$ and H_m: Let S be an $(n-k) \times (n-k)$ scrambling matrix and Q be an $n \times n$ permutation matrix. The public key is generated by calculating $H' = SH_mQ$, and the private keys are S, H_m, and Q.

2) Signing

For a message M, counter i, and hash function h, define the syndrome as $s = h(h(M)|i)$, which is the same as that of the CFS signature scheme.

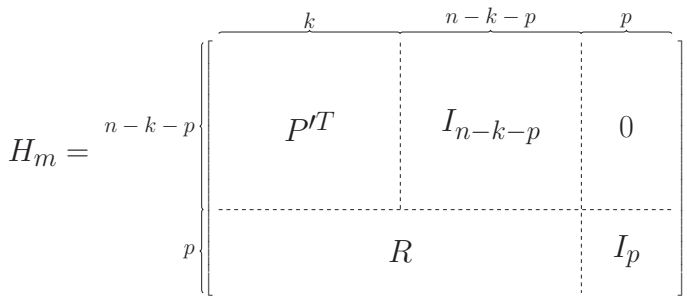

Figure 4.1: Modified parity check matrix of the proposed signature scheme.

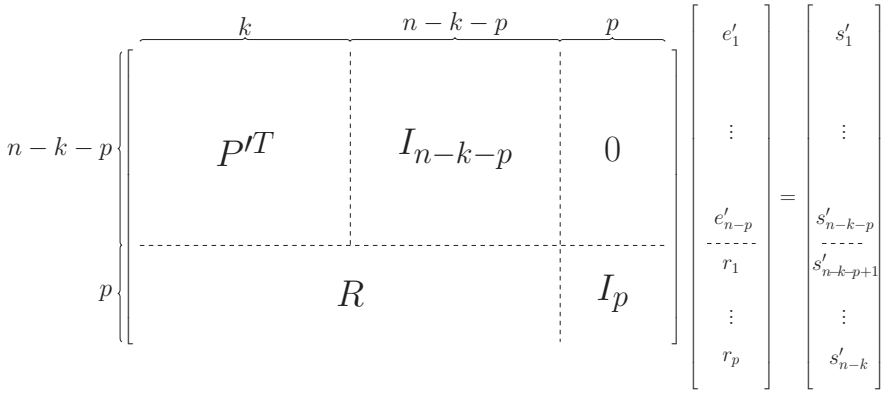

Figure 4.2: Signing process of the proposed signature scheme.

2-1) Find the closest coset: Find the error vector e such that $SHQe^T = s$. Let $e'^T = Qe^T$ and $s' = S^{-1}s$. Then, $He'^T = s'$. Decode the error vector e' by the closest coset decoding.

2-2) Find the punctured part of the error vector: Because the parity check matrix H is random row-deleted and inserted as H_m, we have to replace the last p elements of e' by $e'_p = [r_1, r_2, \cdots, r_p]$, denoted as $e' = [e'_{n-p}|e'_p]$, such that $H_m e'^T = s'$. Let $s' = [s'^T_{n-k-p}|s'^T_p]^T$, where s'_{n-k-p} and s'_p denote the first $n-k-p$ and last p

elements of s', respectively. Then, $H_m e'^T = s'$ can be rewritten as

$$\begin{bmatrix} [P'^T | I_{n-k-p}] e'^T_{n-p} \\ R e'^T_{n-p} + e'^T_p \end{bmatrix} = \begin{bmatrix} s'_{n-k-p} \\ s'_p \end{bmatrix}.$$

Thus, we have

$$e'^T_p = s'_p + R e'^T_{n-p}$$

and thus $e' = [e'_{n-p} | s'_p + R e'_{n-p}]$.

If the Hamming weight of e' is larger than the error weight parameter w, then we increase the counter i and apply the signing process again, where w is larger than the error correctability t. The maximum number of iterations of the counter i is given as N, which will be discussed in the next subsection.

If $\mathrm{wt}(e') \leq w$, compute $e^T = Q^{-1} e'^T$ and the signature σ is then given as $\sigma = (M, e, i)$.

3) Verification

Check $\mathrm{wt}(e) \leq w$ and $H' e^T = h(h(M)|i)$. If TRUE, then return ACCEPT; else, return REJECT.

4.1.2 Preprocessing for Error Weight Parameter

In the proposed signature scheme, choosing the error weight parameter w is significant for balancing security level and time for successful signing, where w is larger than the error correctability t. To determine what is the appropriate value of w in the proposed signature scheme, we perform simulations for random syndromes. For N random syndromes s, we find the minimum Hamming weight error vector e satisfying $H' e^T = s$ by carrying out complete decoding. The required number N of counters i, the corresponding error weight parameter w, and probability of successful signing in the signing stage are listed in Table 4.1.

Assume that N is the maximum number of signing trials for the successful signing in the signing stage. The signing is successful if the complete decoded error weight is

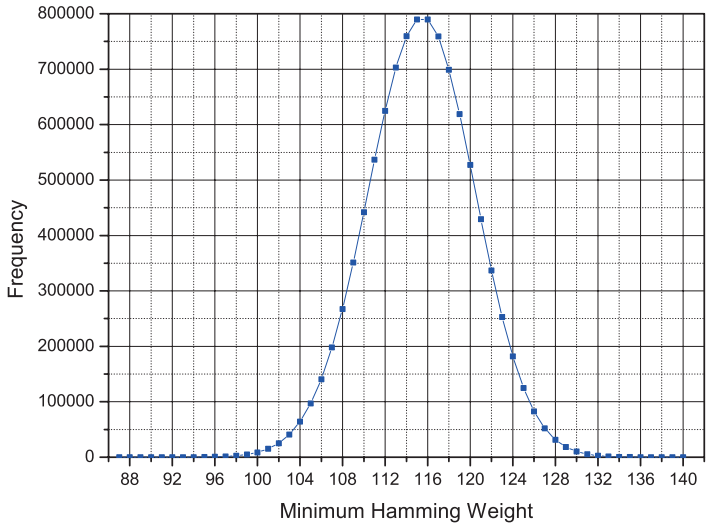

Figure 4.3: Distribution of Hamming weights of coset leaders among 10^7 in RM$(5, 10)$.

less than or equal to w for the hashed message with counter i, $h(h(M)|i)$. Let X_i be the Hamming weight of error vector by the complete decoding for counter i. Then the probability of successful signing is given as

$$\begin{aligned} \text{prob}\left\{\min_{i\leq N}(X_i) \leq w\right\} &= 1 - \text{prob}\left\{X_1 > w, X_2 > w, \cdots, X_N > w\right\} \\ &= 1 - (\text{prob}\{X_1 > w\})^N \end{aligned} \tag{4.1}$$

where each X_i is assumed to be i.i.d. The probability $\text{prob}\{X_1 > w\}$ can be numerically obtained by the distribution of Hamming weights of coset leaders in the complete decoding. Thus, N and w can be selected for successful signing for the given RM code using (6). For RM$(5, 10)$, the distribution of Hamming weights of coset leaders is numerically obtained from Fig. 4.3, where the minimum Hamming weight of error vectors among 10^7 random syndromes is 87. In Table 4.1, the probability of successful signing for RM(5, 10) is listed for parameters N and w using (4.1).

From Table 4.1, using the parameter $w = 99$, the probability of successful signing is almost 1 for $N = 10000$ in RM(5, 10). If we simulate for more syndromes,

Table 4.1: The probability of successful signing for parameters N and w in RM$(5, 10)$

$w \setminus N$	10,000	20,000	40,000
90	0.01	0.02	0.04
93	0.1	0.18	0.33
96	0.41	0.83	0.97
97	0.83	0.97	1
98	0.97	1	1
99	1	1	1

smaller weight of e can be obtained at the cost of longer signing time. That is, if we choose the smaller error weight parameter w and the larger N in the proposed signature scheme, the level of security becomes higher, but the time for successful signing is increased. In Table 4.2, the error weight parameter w that can be successfully signed with $N = 10,000$ trials and its signing time in a straightforward implementation on Intel i7 Processor of 3.0 GHz for each RM code are described.

4.1.3 Additional Modification of the Algorithm

The replaced (punctured and inserted) part of the signature is generated as

$$e_p'^T = s_p' + Re_{n-p}^T.$$

Hence the probability of occurrence of ones in the punctured/inserted part of the signatures is about to 1/2 and the probability of occurrence of ones in the unpunctured part is smaller than w/n. Precisely, since we choose e's having Hamming weight smaller than or equal to w as signature, e_p having larger Hamming weight is likely to be discarded. Hence, the probability of the occurrence of ones in the punctured/inserted part is slightly lower than 1/2. In fact, it is about 45% in RM(4, 12).

By using this difference of the probabilities, an attacker can figure out the punc-

Table 4.2: An error weight parameter w and their signing time for 10,000 random syndromes

	$(n, k, d_{\min})$	Error weight parameter w for successful signing	Successful signing time
RM(4, 10)	(1024, 386, 64)	192	2.26 sec
RM(5, 10)	(1024, 638, 32)	97	2.21 sec
RM(5, 11)	(2048, 1024, 64)	306	4.45 sec
RM(5, 12)	(4096, 1586, 128)	855	9.50 sec
RM(6, 12)	(4096, 2510, 64)	458	8.94 sec

tured/inserted elements in the signature. However, in order to avoid the possible threats, the algorithm can be slightly modified and the parameters such that the probabilities of ones in the unpunctured and punctured/inserted parts are the same.

Let $e_p'^T = s_p' + Re_{n-p}'^T$ and $e' = [e_{n-p}'|s_p'^T + e_{n-p}'R^T]$.

If e' does not satisfies the following two conditions

i) $\mathrm{wt}(e') \leq w$

ii) $\{\mathrm{wt}(e_p') < w_p\}$ or

$\{\mathrm{wt}(e_p') = w_p$ and $rand < q\}$, where $rand$ is a randomly generated number in $[0, 1)$.

then increase the counter i and apply the signing process again.

The proposed signature scheme is summarized as in Algorithm 4.

Algorithm 4 The proposed signature scheme

Preprocessing:

For a given modified (n, k) RM code and the security level larger than 128 bits, derive (N, w) for successful signing as in Table 4.2.

Key Generation:

Generate random matrices S, Q, and R.

Generate H_m as in Algorithm 2.

Compute $H' = SH_mQ$.

Signing:

Do

Choose random number i_r.

Find syndrome $s = h(h(M)|i_r)|h(h(h(M)|i_r))|\cdots|h^{(l)}(h(M)|i_r)$ and compute $s' = S^{-1}s$.

Set $e'_{v[i]} = s'_i$ for $1 \leq i \leq n-k$. Then, $He'^T = s'$ holds.

Perform complete decoding and find new e' such that $He'^T = s'$.

Find $e'^T_p = s'_p + Re^T_{n-p}$ and thus $e' = [e'_{n-p}|e'_p]$.

Until satisfying the following two conditions:

i) $\mathrm{wt}(e') \leq w$

ii) $\{\mathrm{wt}(e'_p) < w_p\}$ or
$\{\mathrm{wt}(e'_p) = w_p$ and $rand < q\}$, where $rand$ is a randomly generated number in $[0, 1)$

Compute $e^T = Q^{-1}e'^T$ and thus signature is $\sigma = (M, e, i)$.

Verification:

Check $\mathrm{wt}(e') \leq w$ and $H'e^T = h(h(M)|i)$.

If True, then return ACCEPT, else return REJECT.

4.2 Implementation of the Proposed Code-Based Signature Scheme

4.2.1 List of Parameter Sets

Parameter set RM(4, 12)

Use RM code RM(4,12) with $w = 1295, w_p = 7, q = 59/256$, and $p = 16$.

Parameter set RM(5, 11)

Use RM code RM(5,11) with $w = 311, w_p = 2, q = 220/256$, and $p = 8$.

Parameter set RM(6, 12)

Use RM code RM(6,12) with $w = 472, w_p = 2, q = 1$, and $p = 8$.

Parameter set RM(6, 13)

Use RM code RM(6,13) with $w = 1441, w_p = 4, q = 23/256$, and $p = 16$.

4.2.2 Description of Platform

The following measurements were collected using a desk-top computer with CPU an Intel —Intel(R) Xeon(R) CPU E5-2698 v4 2.20GHz— running at 2.2 GHz. This computer has 128GB of RAM and runs Ubuntu 16.04 LTS. Benchmarks have ran on one core of the CPU. The source code can be compiled by `make` in the directory `/pqsigrm412`, `/pqsigrm612`, `/pqsigrm613` with proper source code file containing `main()` function. Since the signing algorithm is a probabilistic algorithm, number of iterations at signing varies. The following result is the average of 100 experiments. For the detailed descriptions about the success probability of the signing, see 2.B.1.

Table 4.3: CPU cycles of pqsigRM with —Intel(R) Xeon(R) CPU E5-2698 v4 2.20GHz—

	Security	Key generation	Singing	Verification
`/pqsigRM-4-12`	128	14639777783	3971208456	139814898
`/pqsigRM-6-12`	196	6395769782	3275234719	198607502
`/pqsigRM-6-13`	256	72162115384	1087667252	956410761

NIST says that the "NIST PQC Reference Platform" is "an Intel x64 running Windows or Linux and supporting the GCC compiler." Our system is an Intel x64 running Linux and supporting the GCC compiler. Beware, however, that different Intel CPUs can output different results.

4.2.3 Time

The following measurements are CPU cycles for run `/pqsigrm412`, `/pqsigrm612`, `/pqsigrm613` at —Intel(R) Xeon(R) CPU E5-2698 v4 2.20GHz—. The measurements are given in Table 4.3.

4.2.4 Space

Sizes are straightforwardly calculated from parameters (and confirmed in various experiments). Specifically, the size of public keys is 328KB for `pqsigRM-4-12`, 489KB for `pqsigRM-6-12`, 2056KB for `pqsigRM-6-13`. Signatures are $528, 528$, and 1040 bytes longer than message. The size of keys is given in Table 4.4.

4.2.5 How Parameters Affect Performance

The performance can be controlled by parameters N and w. We can control the singing time and the difficulty of the attack by changing w, the maximum weight of the e in signature. Decreasing w, the probability that a desired error will be obtained is

Table 4.4: Public key and secret key size of pqsigRM (byte)

	Public key	Secret key
/pqsigRM-4-12	336804	1382118
/pqsigRM-6-12	501176	334006
/pqsigRM-6-13	2105344	2144166

lowered. Therefore, N should increase inversely with w. This will increase the signing time. However, increasing w, we can expect the better security. Current system uses fixed values, but by changing w, we can create a digital signature system with flexible security. Thus, w, the secrecy, N, and signing time are trade-offs. We set the value of N as 10,000.

4.3 Security Analysis of the Proposed Code-Based Signature Scheme

4.3.1 EUF-CMA

In this subsection, we prove that the proposed signature scheme is secure under EUF-CMA. The proposed signature scheme can be simplified as follows. We can consider finding e satisfying $H'e^T = h(h(M)|i)$ as a signing process, where the parity check matrix of a linear code is H_m and γ is the decoding algorithm in the signature scheme in the previous section. Then, the proposed signature scheme can be considered to be the same as the original CFS scheme for the EUF-CMA security check.

1) Key Generation:

Let H_m be a parity check matrix of a modified RM code with the decoding algorithm γ. Then, the private keys are S, Q, and H_m, and the public keys are $H' = SH_mQ$ and w.

2) Signing:

To sign a message M,

For $i = 1$ to N

$e' = \gamma(S^{-1}h(h(M)|i))$

if $\mathrm{wt}(e') \leq w$, go to *.

end

* Output $(M, e = e'(Q^{-1})^T, i)$

3) Verification

Check $\mathrm{wt}(e) \leq w$ and $H'e^T = h(h(M)|i)$. If TRUE, then return ACCEPT; otherwise, return REJECT.

To prove the EUF-CMA security, we need the following assumption and proposition. The differences between the proposed signature scheme and the CFS signature scheme are; i) the use of a different code, namely, a modified RM code rather than a Goppa code, and ii) the use of complete decoding instead of syndrome decoding.

Assumption 12 (RM code distinguishability problem)**.** *There is no probabilistic polynomial time (PPT) distinguisher $\mathcal{D}$ that can distinguish $H' = SH_mP$ from a randomly generated parity check matrix H_R.*

To the best of our knowledge, there are no known algorithms for distinguishing a modified parity check matrix H' of an RM code from H_R up to now, and thus we set the following assumption. Specially, the random submatrix R in H_m is inserted, which strengthens the indistinguishability from H_R.

Proposition 13 (Hardness of complete decoding [16])**.** *The complete decoding problem for an (n, k, t) linear code is an NP-complete problem if the Hamming weight of e is less than $n/3$.*

With this assumption and proposition, the following theorem holds.

Theorem 14. *The proposed modified RM code-based signature scheme is EUF-CMA secure.*

Proof. The proof of this theorem is almost the same as the proof of the EUF-CMA security of the CFS signature scheme [28]. It was proved in [28] that a variant of a CFS scheme is EUF-CMA secure if certain assumptions are true. However, it was shown that the assumption that distinguishing Goppa codes from random codes is difficult is not true for the case of some parameters (small t) used in the CFS signature scheme. Thus, we can follow the logic of the proof in [28] because the adopted assumption for the proposed modified RM code-based signature scheme is still valid. We define the sequence of games $G_0, G_1, \cdots, G_5$ in the same way as in [28]. Let G_0 be the original security game, that is, the EUF-CMA game, and G_5 be solving the syndrome decoding problem.

The main differences between the proposed signature scheme and the CFS signature scheme are mostly in Games G_3 and G_5. In the proof of the CFS signature scheme, Game G_3 discusses the Goppa code distinguishing problem, but for a small t, it turns out to be distinguishable from a random code. In the case of the proposed signature scheme, we adopt the modified RM code distinguishing problem in Assumption 5 because there has been no way to prove the distinguishability up to now. Although Game G_5 is related to the syndrome decoding problem in the proof of the original CFS signature scheme, we will replace it using the complete decoding problem, which is known as an NP-complete problem. A full description of this proof is given as follows.

The challenger $\mathcal{C}$ plays a sequence of games $G_0, G_1, \cdots, G_5$. Here, G_0 corresponds to the standard EUF-CMA game as mentioned above. In G_0, the adversary $\mathcal{A}$ tries to forge a signature. If the adversary $\mathcal{A}$ successfully forges the signature, then $\mathcal{A}$ wins the game G_0. Successive games are given through slight modifications of the preceding games. Let $\Pr[G_i]$ be the winning probability of each game G_i. We then have to prove that the probability of the winning condition of these games is proved to be arbitrarily

small through all of the intermediate games.

G_0: The challenger $\mathcal{C}$ obtains the private and public keys using a key generation algorithm. The adversary $\mathcal{A}$ obtains the public key H', and can access a hash oracle $\mathcal{H}$ and signing oracle Σ. Let q_h and q_s be the maximum numbers of queries made by the adversary $\mathcal{A}$ to the hash oracle and the signing oracle, respectively. The procedure of G_0 is given in Algorithm 5. Then, the winning probability of G_0 is given as

$$\Pr[G_0] = \mathrm{succ}^{\mathrm{EUF-CMA}}(\mathcal{A}). \tag{4.2}$$

Algorithm 5 G_0 (EUF-CMA)

1. $(H', S, H, Q) \leftarrow \mathrm{keygen}(\mathcal{C})$
2. Set the oracles $\mathcal{H}$ and Σ
3. $(M^*, \sigma^*, i^*) \leftarrow \mathcal{A}^{\mathcal{H},\Sigma}(H')$
4. If $\mathcal{H}(M^*, i^*) = H'\sigma^{*T}, \mathrm{wt}(\sigma^*) \leq w$, and Σ did not provide σ^*, then

   $\mathcal{A}$ wins the game

   else

   $\mathcal{A}$ loses the game

   end

G_1: In this game, the challenger modifies the hash oracle $\mathcal{H}$ by $\mathcal{H}'$. In $\mathcal{H}'$, the challenger uses a list Λ that consists of counter values of $i = \Lambda(M)$ for message M such that $\mathcal{H}(M, i)$ is a decodable syndrome and another list $\Lambda_{\mathcal{H}}$ to store a valid syndrome-error pair that was already produced in the previous queries. If there is no element corresponding to the input, the output is $\perp$. The modified hash oracle $\mathcal{H}'$ produces syndromes according to Algorithm 6, and finally produces $q_h + q_s + 1$ syndromes. In

addition, it is known that the relation of $\Pr[G_0]$ and $\Pr[G_1]$ is given as

$$|\Pr[G_1] - \Pr[G_0]| \leq \epsilon_0 \tag{4.3}$$

where $\epsilon_0 = 1 - \left(1 - \frac{1}{2^{n-k}}\right)^{q_h+q_s+1}$ [28].

G_2: In G_2, the challenger replaces the signing oracle Σ with Σ'. The modified signing oracle queries $\mathcal{H}'$ on $(M, \Lambda(M))$ according to Algorithm 7. In addition, the winning probability relation of G_1 and G_2 is derived as

$$|\Pr[G_2] - \Pr[G_1]| \leq \epsilon_1 \tag{4.4}$$

where $\epsilon_1 = 1 - \left(1 - \frac{q_s}{2^{n-k}}\right)^{q_h}$.

G_3: In G_3, the challenger replaces the key generation algorithm with the selection of a random binary parity check matrix. The selected parity check matrix is taken as the public key. Because neither the hash oracle nor the signature oracle uses the hash function and the private keys, the difference in the winning probabilities of G_2 and G_3 is the same as the distinguishing probability between the modified RM code and a random binary code, that is,

$$|\Pr[G_3] - \Pr[G_2]| \leq \epsilon_{\text{distinguish}}. \tag{4.5}$$

By Assumption 5, the value of $\epsilon_{\text{distinguish}}$ is negligible. The description of G_3 is given as Algorithm 8.

G_4: G_4 is conditioned by an adversary making a forgery on a particular hash query. The challenger first obtains a random $c \xleftarrow{R} \{1, \cdots, q_h + q_s + 1\}$. Adversary $\mathcal{A}$ wins the game if the cth query to $\mathcal{H}'$ is made on (M^*, i^*). Because c is randomly chosen from $q_h + q_s + 1$ possibilities, the winning probability of G_4 is given as

$$\Pr[G_4] = \frac{\Pr[G_3]}{q_h + q_s + 1}. \tag{4.6}$$

G_5: In this game, the challenger modifies the hash oracle to output a random syndrome s^* to the cth query. The winning probability of G_5 is the same as the winning

Algorithm 6 Game G_1 ($\mathcal{H}'$: simulation of $\mathcal{H}$)

Input: a pair (M, i)

Output: a syndrome s

1. If $\Lambda(M) = \perp$, then
   $\Lambda(M) \xleftarrow{R} \{1, \cdots, 2^{n-k}\}$
2. $(s, e) \leftarrow \Lambda_{\mathcal{H}}(M, i)$
3. If $i \neq \Lambda(M)$, then

   If $s = \perp$, then

   $s \xleftarrow{R} F_2^{n-k}$

   $\Lambda_{\mathcal{H}}(M, i) \leftarrow (s, \perp)$

   end

   return $\mathcal{H}(M, i) = s$

   else

   If $s = \perp$, then

   $e \xleftarrow{R} \{y \in F_2^n | \mathrm{wt}(y) \leq w\}$

   $s \leftarrow He^T$

   $\Lambda_{\mathcal{H}}(M, i) \leftarrow (s, e)$

   end

   return $\mathcal{H}'(M, i) = s$

   end

Algorithm 7 Game G_2 (Σ': simulation of Σ)

Input: a message M

Output: a signature (i, σ)

1. If $\Lambda(M) = \perp$, then

   $\Lambda(M) \xleftarrow{R} \{1, \cdots, 2^{n-k}\}$

   end

2. $\mathcal{H}'(M, \Lambda(M))$

3. $(s, x) \leftarrow \Lambda_{\mathcal{H}'}(M, \Lambda(M))$

4. $\Lambda(M) \leftarrow \perp$

5. Return $\Sigma(M) = (i, x)$

Algorithm 8 Game G_3

Input: a parity check matrix H

Output: a bit b

1. Given w, set the oracles $\mathcal{H}'$ and Σ'

2. $(M^*, \sigma^*, i^*) \leftarrow \mathcal{A}^{\mathcal{H}', \Sigma'}(H)$

3. If $\mathcal{H}'(M^*, i^*) = H\sigma^{*T}$, $\mathrm{wt}(\sigma^*) \leq w$, and Σ' did not provide σ^*, then

   $b = 1$

   else

   $b = 0$

   end

probability of G_4. The detailed procedure for G_5 is given in Algorithm 9. Note that this game is the same as solving the complete decoding problem. Then,

$$\Pr[G_5] = \Pr[G_4] \leq \epsilon_{\text{complete}}. \tag{4.7}$$

From Proposition 6, the value of $\epsilon_{\text{complete}}$ is negligible.

Algorithm 9 Game G_5

Input: an adversary $\mathcal{A}$

1. $c \xleftarrow{R} \{1, \cdots, q_{\mathcal{H}} + q_{\Sigma} + 1\}$
2. $H^* \xleftarrow{R}$ (n, k) binary code for given w
3. $s^* \xleftarrow{R} F_2^{n-k}$
4. Set the oracles $\mathcal{H}'$ and Σ'
5. $(M^*, \sigma^*, i^*) \leftarrow \mathcal{A}^{\mathcal{H}', \Sigma'}(H^*)$
6. If $\begin{cases} \mathcal{H}'(M^*, i^*) = H\sigma^{*T} \\ \text{wt}(\sigma^*) \leq w \end{cases}$ and $\begin{cases} \Sigma' \text{ did not provide } \sigma^* \\ c - \text{th query to } \mathcal{H}' \text{ was } (M^*, i^*), \end{cases}$ then

   $\mathcal{A}$ wins the game

   else

   $\mathcal{A}$ loses the game

   end

Combining all of the above equations, (4.2)–(4.7), we have

$$\text{succ}^{\text{EUF}-\text{CMA}}(\mathcal{A}) \leq (q_h + q_s + 1)\epsilon_{\text{complete}} + \epsilon_{\text{distinguish}} + \epsilon_0 + \epsilon_1. \tag{4.8}$$

Hence, the probability of a successful forgery is negligible if the punctured RM codes with random insertions are indistinguishable from random linear codes and the com-

Table 4.5: The security of the proposed signature scheme for $N = 10,000$

	$(n, k, d_{\min}, w)$	$\frac{\sum_{i=0}^{w} \binom{n-k}{i}}{2^{n-k}} \times C$
RM(5, 11)	(2048, 1024, 64, 312)	$\leq 2^{-128}$
RM(4, 12)	(4096, 794, 256, 1285)	$\leq 2^{-128}$
RM(6, 12)	(4096, 2510, 64, 472)	$\leq 2^{-192}$
RM(6, 13)	(8192, 4096, 128, 1367)	$\leq 2^{-256}$

plete decoding problem is intractable. Thus, the proposed signature scheme is EUF-CMA secure. □

4.3.2 Forgery Attack

The attacker tries to forge the signature with public key H' and hashed message $h(h(M)|i)$. Assume that the public key H' is systematic, where $H' = [H_0|I]$, I is an $(n-k) \times (n-k)$ identity matrix, and H_0 is an $(n-k) \times k$ matrix. Then, attacker computes z satisfying the following equation

$$H'z^T = [H_0|I][z_1|z_2]^T = s = h(h(M)|i) \tag{4.9}$$

where z_1 and z_2 are vectors with size k and $n-k$, respectively. The attacker can let z_1 be an all-zero vector and $z_2 = s$. If the Hamming weight of z_2 is less than or equal to w, then the forgery is successful. The probability of successful forgery is given as

$$\frac{\sum_{i=0}^{w} \binom{n-k}{i}}{2^{n-k}}. \tag{4.10}$$

Table 4.5 presents the security for each RM code and error weight parameters w of error vectors given in Table 4.2.

4.3.3 Information Set Decoding Attack

Information set decoding is brute force method that find the error vector e such that $He^T = s$ and $\mathrm{wt}(e) \leq w$. The algorithm for information set decoding is given in Algorithm 10. Stern [27] optimized the complexity of the information set decoding.

Algorithm 10 Information set decoding attack [27]

- Input: a $k \times n$ matrix H, integer w
- Output: a non-zero codeword of Hamming weight $\leq w$
  - Pick $n \times n$ permutation matrix P
  - Compute $H' = UHP = (I|R)$
  - Compute all the sum of p rows or less of H', if one of those sums has weight $\leq w$ then stop and return it.

The complexity is given as follows.

$$WF = Kl\binom{n}{w} \Big/ \binom{k/2}{g}\binom{n-k-l}{w-2g}$$

where $l = \log_2\binom{k/2}{g}$, g denotes the number of chosen components of error vector, and the hidden parameter K is considered as $\frac{\log_2 n}{2}$ for actual computation. However, in our signature scheme, there are many n-tuple error vector with Hamming weight less than or equal to w for each syndrome. The number of n-tuple error vectors with Hamming weight less than or equal to w is approximately

$$N = \binom{n}{w} \Big/ 2^{n-k}.$$

Dividing the complexity by N, the total computational complexity for forgery attack is given as

$$WF = Kl2^{n-k} \Big/ \binom{k/2}{g}\binom{n-k-l}{w-2g}$$

The value of WF for each RM code is given in Table 4.6.

Table 4.6: The WF for each RM code

	$(n, k, d_{\min}, w)$	g in WF	WF
RM(5, 11)	(2048, 1024, 64, 312)	13	$\geq 2^{133}$
RM(4, 12)	(4096, 794, 256, 1285)	11	$\geq 2^{133}$
RM(6, 12)	(4096, 2510, 64, 472)	22	$\geq 2^{180}$
RM(6, 13)	(8192, 4096, 128, 1367)	34	$\geq 2^{322}$

Chapter 5

New Families of p-ary Sequences of Period $\frac{p^n-1}{2}$ With Low Maximum Correlation Magnitude

5.1 Known Sequences With Low Correlations

Pseudo random sequences with low correlation are widely used in random number generation and wireless communications, that is, code division multiple access, spread spectrum, cryptography, and error correcting codes. Especially, in the post quantum cryptosystems, pseudo random sequences can be used to generate random matrices. That is, by adopting pseudo random sequences, the private key, such as S, P, R in cryptosystems, can be obtained rapidly.

Many papers on sequence families with good correlation properties have been published. Kasami [32], [33] proposed a binary sequence family with the optimal correlation property with respect to Welch's lower bound. Further, there are lots of research results for the nonbinary sequence families. Liu and Komo [38] generalized the Kasami sequence family to the p-ary case and Kumar and Moreno [36] constructed a p-ary sequence family with correlation magnitude upper bounded by $1 + \sqrt{p^n}$ using bent function. Jang, Kim, No and Helleseth [31] also proposed a p-ary sequence family with the optimal correlation property. Muller [39] also proposed two p-ary sequence

families, whose correlation magnitude is upper bounded by $1 + 2\sqrt{p^n}$ and $1 + \sqrt{p^n}$, respectively. Seo, Kim, No, and Shin [40] derived the cross-correlation distribution of p-ary sequences which have good correlation property. Choi, Lim, No, and Jung [29] also proposed a p-ary sequence family with correlation magnitude upper bound $\frac{p+1}{2}\sqrt{p^n}$ and family size $\sqrt{p^n}$.

Recently, p-ary sequence families with half period, that is, $N = \frac{p^n-1}{2}$ have been proposed. Generally, half period sequences can have larger family size. Kim, Choi, and No [35] constructed the first p-ary sequence family of half period using Kloosterman sum. This sequence family has large family size of $4N$ and their correlation magnitude is upper bounded by $2\sqrt{N + \frac{1}{2}}$ for an odd prime $p \equiv 3 \pmod 4$ and an odd integer n. This result is further generalized by Kim, Chae, and Song [34], that is, they generalized this sequence family to all odd prime p. Xia and Chen [43] constructed new sequence families having family size $4N$ and the correlation magnitude upper bounded by $\frac{p}{\sqrt{2}}\sqrt{N + \frac{1}{2}} + \frac{1}{2}$. In this dissertation, we propose new p-ary sequence families of half period, whose correlation property is almost the same as that by Kim, Choi, and No [35]. For comparison of those well known p-ary sequence families with good correlation properties, their parameters are listed in the Table 5.1.

Weil bound for exponential sums is often used to prove the upper bound on the magnitude of correlation values [42]. There are three types of Weil bounds. The first one is the sum of multiplicative character. The second one is sum of additive character, and the last one is the sum of multiplication of additive and multiplicative characters (hybrid type). Han and Yang [30] used multiplicative characters of Weil bound to derive the upper bound on the magnitude of correlation values. Wang and Gong [41] constructed polyphase sequence families whose correlation magnitude is derived from the Weil bound of exponential sums. They applied all three types of Weil bounds to the proof of the upper bounds.

Table 5.1: Comparison with some well-known sequence families

Family	Alphabet	n	Period N	Family size	$C_{\max}$
Liu and Komo[38]	odd p	even	p^n-1	$\sqrt{N+1}$	$\sqrt{N+1}+1$
Jang *et al.*[31]	odd p	even or odd	p^n-1	$N+1$	$\sqrt{N+1}+1$
Kumar and Moreno[36]	odd p	even or odd	p^n-1	$N+1$	$\sqrt{N+1}+1$
Seo *et al.*[40]	odd p	even	p^n-1	$\sqrt{N+1}$	$2\sqrt{N+1}+1$
Choi *et al.*[29]	odd p	even	p^n-1	$\sqrt{N+1}$	$\frac{p+1}{2}\sqrt{N+1}+1$
Kim *et al.*[35]	$p \equiv 3 \pmod 4$	odd	$\frac{p^n-1}{2}$	$4N$	$2\sqrt{N+\frac{1}{2}}$
Kim *et al.*[34]	all p	even or odd	$\frac{p^n-1}{e}$	e^2N	$2\sqrt{eN+1}$
Xia and Chen[43]	$p \equiv 1 \pmod 4$	even or odd	$\frac{p^n-1}{2}$	$4N$	$\frac{p}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}$
	$p \equiv 3 \pmod 4$	even	$\frac{p^n-1}{2}$	$4N$	$\frac{p}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}$
Proposed family S	$p \equiv 3 \pmod 4$	odd	$\frac{p^n-1}{2}$	$4N$	$\frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}$

5.2 Characters and Weil Bound

There are two types of characters, that is, additive characters and multiplicative characters as follows [37].

Definition 15 (Additive character)**.** *For $\beta \in \mathbb{F}_q$, an additive character of $\mathbb{F}_q$ is defined as*

$$\psi(x) = e^{\frac{2\pi i \mathrm{Tr}_1^n(\beta x)}{p}}, x \in \mathbb{F}_q$$

and ψ_0, $\psi(x)$ with $\beta = 0$, denotes the trivial additive character such that $\psi_0(x) = 1$ for all $x \in \mathbb{F}_q$.

Definition 16 (Multiplicative character)**.** *Let g be a fixed primitive element of $\mathbb{F}_q$. For each $j = 1, 2, \cdots, q-2$, a multiplicative character of $\mathbb{F}_q$ is defined as*

$$\chi(g^k) = e^{\frac{2\pi i j k}{q-1}}$$

where $\chi(0) = 0$ and χ_0, $\chi(g^k)$ with $j = 0$, denotes the trivial multiplicative character such that $\chi_0(x) = 1$ for all $x \in \mathbb{F}_q^*$.

We consider the quadratic character η in this paper, which is defined as

$$\eta(y) = \begin{cases} 1, & \text{if } y \text{ is nonzero square in } \mathbb{F}_q \\ -1, & \text{if } y \text{ is nonzero nonsquare in } \mathbb{F}_q \\ 0, & \text{if } y = 0. \end{cases}$$

Lemma 17 (Gaussian sum [37])**.** *Let ψ be an additive character of $\mathbb{F}_q$ and χ be a multiplicative character of $\mathbb{F}_q$. Then the Gaussian sum $G(\psi, \chi)$ is defined as*

$$G(\psi, \chi) = \sum_{x \in \mathbb{F}_q} \psi(x)\chi(x),$$

which satisfies

$$G(\psi,\chi)=\begin{cases} p^n-1 & \textit{for } \psi=\psi_0 \textit{ and } \chi=\chi_0 \\ 0 & \textit{for } \psi=\psi_0 \textit{ and } \chi\neq\chi_0 \\ -1 & \textit{for } \psi\neq\psi_0 \textit{ and } \chi=\chi_0 \end{cases}$$

and for $\psi\neq\psi_0$ *and* $\chi\neq\chi_0$,

$$|G(\psi,\chi)|=q^{1/2}.$$

The following Weil bounds are often used to prove the correlation property of the sequence.

Theorem 18 (Weil bound [42])**.** *Let* ψ *be a nontrivial additive character of* $\mathbb{F}_q$ *and* χ *be a nontrivial multiplicative character of* $\mathbb{F}_q$ *with order* M *and* $\chi(0)=0$*. Let* $f(x)\in\mathbb{F}_q[x]$ *with degree* e *and* $g(x)\in\mathbb{F}_q$ *with* s *distinct roots in* $\overline{\mathbb{F}}_q$*, where* $g(x)\neq c\cdot h^M(x)$ *for some* $c\in\mathbb{F}_q$ *and* $h(x)\in\mathbb{F}_q[x]$*, and* $\overline{\mathbb{F}}_q$ *denotes the algebraic closure of* $\mathbb{F}_q$*. Then*

$$\left|\sum_{x\in\mathbb{F}_q}\chi(g(x))\psi(f(x))\right|\leq(e+s-1)\sqrt{q}.$$

Theorem 19 (Additive type of Weil bound [37])**.** *Let* $f\in\mathbb{F}_q[x]$ *be of degree* $n\geq 1$ *with* $\gcd(n,q)=1$ *and let* ψ *be a nontrivial additive character of* $\mathbb{F}_q$*. Then*

$$\left|\sum_{x\in\mathbb{F}_q}\psi(f(x))\right|\leq(n-1)\sqrt{q}.$$

5.3 New Sequence Families and Their Correlation Bound

In this section, we will propose two new p-ary sequence families of period $N=\frac{p^n-1}{2}$ and family size $4N$ and derive their correlation bound. Let $m(t)$ be a p-ary m-sequence of period $q-1$. We consider the sequence $m(2t)$ and $m(dt)$, where $d=4$ and $N+1$. Since $q-1$ is even, the decimated sequence $m(2t)$ has the period N. Since $\gcd(q-1,d)=2$ for both cases, the period of $m(dt)$ is also N. Then we define the new p-ary sequence family of period N and family size $4N$ as

$$S = \{m(2t+i) + m(d(t+l)+j) | 0 \le i, j \le 1, 0 \le l \le N-1\}.$$

We will show that the magnitude of cross-correlation and nontrivial autocorrelation values of the sequences in the family S is upper bounded by $\frac{\sqrt{3}}{2}\sqrt{N+\frac{1}{2}} + \frac{1}{2}$. For the proof of the upper bound, we use Theorems 18 and 19.

The correlation function between two sequences in S, $m(2t+i_1)+m(d(t+l_1)+j_1)$ and $m(2t+i_2)+m(d(t+l_2)+j_2)$, except for the trivial autocorrelation ($\tau = 0, i_1 = i_2, j_1 = j_2, l_1 = l_2$), is given as

$$\begin{aligned} C(\tau) &= \sum_{t=0}^{N-1} \omega^{\mathrm{Tr}_1^n(\alpha^{2t+i_1})+\mathrm{Tr}_1^n(\alpha^{d(t+l_1)+j_1})-\mathrm{Tr}_1^n(\alpha^{2(t+\tau)+i_2})-\mathrm{Tr}_1^n(\alpha^{d(t+\tau+l_2)+j_2})} \\ &= \sum_{t=0}^{N-1} \omega^{\mathrm{Tr}_1^n(\alpha^{2t}(\alpha^{i_1}-\alpha^{2\tau+i_2})+\alpha^{dt}(\alpha^{dl_1+j_1}-\alpha^{d\tau+dl_2+j_2}))}. \end{aligned}$$

Let $a = \alpha^{i_1} - \alpha^{2\tau+i_2}$ and $b = \alpha^{dl_1+j_1} - \alpha^{d\tau+dl_2+j_2}$. Then

$$C(\tau) = \sum_{t=0}^{N-1} \omega^{\mathrm{Tr}_1^n(a\alpha^{2t}+b\alpha^{dt})}.$$

We will derive the upper bound of $C_{\max}$ for $d = 4$ and $N+1$ in the following two theorems.

Theorem 20. *For $d = 4$, we have*

$$C(\tau) = \sum_{t=0}^{N-1} \omega^{\mathrm{Tr}_1^n(a\alpha^{2t}+b\alpha^{4t})}.$$

Then, the maximum magnitude of $C(\tau)$ is given as

$$C_{\max} \le \frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}} + \frac{1}{2}.$$

Proof. Let $x = \alpha^{2t}$ and QR be the set of quadratic residues of $\mathbb{F}_q$. Then we have

$$\begin{aligned} C(\tau) &= \sum_{x \in \mathrm{QR}} \omega^{\mathrm{Tr}_1^n(ax+bx^2)} \\ &= \frac{1}{2}\left(\sum_{x \in \mathbb{F}_q^*} \omega^{\mathrm{Tr}_1^n(ax+bx^2)} + \sum_{x \in \mathbb{F}_q^*} \eta(x)\omega^{\mathrm{Tr}_1^n(ax+bx^2)} \right). \end{aligned} \tag{5.1}$$

Since the trivial autocorrelation case is excluded, it is easy to check that $a = b = 0$ should not be considered because $i_1, i_2, j_1, j_2 \in \{0, 1\}$.

(i) $b = 0$ and $a \neq 0$:

In this case, (5.1) can be rewritten as

$$\frac{1}{2}\left(\sum_{x\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ax)} + \sum_{x\in\mathbb{F}_q^*}\eta(x)\omega^{\mathrm{Tr}_1^n(ax)}\right). \tag{5.2}$$

The first term in (5.2) is given as

$$\sum_{x\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ax)} = -1. \tag{5.3}$$

Let $\chi = \eta, g(x) = x$, and $f(x) = ax$ in Theorem 18. Then the second term in (5.2) is computed as

$$\left|\sum_{x\in\mathbb{F}_q^*}\eta(x)\omega^{\mathrm{Tr}_1^n(ax)}\right| \leq \sqrt{q}. \tag{5.4}$$

From (5.3) and (5.4), (5.2) can be computed as

$$\begin{aligned}
|C(\tau)| &= \frac{1}{2}\left|\left(\sum_{x\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ax)} + \sum_{x\in\mathbb{F}_q^*}\eta(x)\omega^{\mathrm{Tr}_1^n(ax)}\right)\right| \\
&\leq \frac{\sqrt{q}+1}{2} \\
&= \frac{\sqrt{2N+1}}{2} + \frac{1}{2} \\
&= \frac{1}{\sqrt{2}}\sqrt{N+\frac{1}{2}} + \frac{1}{2}.
\end{aligned} \tag{5.5}$$

(ii) $b \neq 0$:

From Theorem 19 with $f(x) = ax + bx^2$, the first term in (5.1) can be derived as

$$\left|\sum_{x\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ax+bx^2)}\right| \leq \sqrt{q}+1. \tag{5.6}$$

Let $\chi = \eta, g(x) = x$, and $f(x) = ax + bx^2$ in Theorem 18. Then, the second term in (5.1) is computed as

$$\left|\sum_{x\in\mathbb{F}_q^*}\eta(x)\omega^{\mathrm{Tr}_1^n(ax+bx^2)}\right| \leq 2\sqrt{q}. \tag{5.7}$$

From (5.6) and (5.7), we have

$$\begin{aligned}\frac{1}{2}\left|\left(\sum_{x\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ax+bx^2)}+\sum_{x\in\mathbb{F}_q^*}\eta(x)\omega^{\mathrm{Tr}_1^n(ax+bx^2)}\right)\right| &\le \frac{3}{2}\sqrt{q}+\frac{1}{2}\\ &= \frac{3}{\sqrt{2}}\sqrt{\frac{q}{2}}+\frac{1}{2}\\ &= \frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}. \qquad (5.8)\end{aligned}$$

From (5.5) and (5.8), we prove the theorem. □

Theorem 21. *Let $d = N + 1$. $C(\tau)$ can be rewritten as*

$$C(\tau)=\sum_{t=0}^{N-1}\omega^{\mathrm{Tr}_1^n(a\alpha^{2t}+b\alpha^{(N+1)t})}. \qquad (5.9)$$

Then, the maximum magnitude of $C(\tau)$ can also be derived as

$$C_{\max}\le\frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}.$$

Proof. Let $x=\alpha^t$. It is easy to check that -1 is a nonsquare in $\mathbb{F}_{p^n}$ for an odd integer n and an odd prime $p\equiv 3 \pmod 4$. Let $x=y^2$ for a square x and $x=-y^2$ for a nonsquare x.

Since $N+1$ is even, we have the same form of

$$\mathrm{Tr}_1^n(ax^2+bx^{N+1})=\mathrm{Tr}_1^n(ay^4+by^2)$$

for both $x=y^2$ and $-y^2$. Then (5.9) can be rewritten as

$$\begin{aligned}C(\tau) &= \frac{1}{2}\sum_{y\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ay^4+by^2)}\\ &= \frac{1}{2}\left(\sum_{y\in\mathbb{F}_q^*}\omega^{\mathrm{Tr}_1^n(ay^2+by)}+\sum_{y\in\mathbb{F}_q^*}\eta(y)\omega^{\mathrm{Tr}_1^n(ay^2+by)}\right). \qquad (5.10)\end{aligned}$$

Since (5.10) is the same as (5.1) by swapping a and b, the proof is the same as that of Theorem 20. Thus the proof is done. □

Table 5.2: Simulation results of $C_{\max}$ and number of correlation values for some p and n

p	n	N	$\frac{C_{\max}}{\sqrt{N}}$	Number of distinct values
3	3	13	2.1650	5
	5	121	2.1259	6
	7	1093	2.1219	6
	9	9841	2.1214	6
7	3	171	2.0304	94
	5	8403	2.0951	852
11	3	665	2.0003	450

Theorem 22. *: The family size of S is $4N$.*

Proof. If there are two cyclically equivalent sequences in S, then their cross-correlation value is equal to N. From Theorems 18 and 19, the magnitude of the cross-correlation values of arbitrary two sequences are upper bounded by $\frac{3}{\sqrt{2}}\sqrt{N+\frac{1}{2}}+\frac{1}{2}$ and thus the sequences in S are cyclically inequivalent. □

Even though the maximum magnitude of correlation values of the proposed sequence families is upper bounded, the number of distinct correlation values increases as N becomes large. Table 5.2 shows the number of distinct correlation values and the normalized maximum magnitude of $C_{\max}$ by $\sqrt{N}$ for some p and n. In case of $p = 3$ and odd n, the number of distinct correlation values is less than 6 and the correlation distribution is studied in [43].

Chapter 6

Conclusion

In this dissertation, the RM code-based cryptosystems, signature scheme, and pseudo random sequences are studied.

First, the secure modification methods for the McEliece cryptosystems based on the punctured RM codes are proposed. We find the exact number and locations of puncturing of the generator matrix of the original RM codes to prevent the various known attacks. While the previous McEliece cryptosystem based on RM codes is vulnerable to some known attacks such as the Minder-Shokrollahi's attack, the Chizhov-Borodin's attack, and square code attack, the proposed punctured RM code-based McEliece cryptosystems can resist these attacks. The security level of the proposed cryptosystems is maintained due to puncturing and random insertion and thus the proposed cryptosystems can be revived.

Next, a new signature scheme based on the punctured RM code with random insertion is proposed. The proposed signature scheme improves the Goppa code-based CFS signature scheme by increasing the probability of successful signing using the complete decoding method. In addition, the proposed signature scheme can avoid some known attacks for the RM code-based cryptosystem using the puncturing method with random insertion. The optimal parameters of the signing time and security were derived. It was also proved that the proposed signature scheme achieves EUF-CMA se-

curity.

Last, for an odd positive integer n and an odd prime p such that $p \equiv 3 \mod 4$, two new families of p-ary sequences with low maximum correlation magnitude are constructed where the period of sequences is $N = \frac{p^n-1}{2}$ and the family size $4N$. The sequences in the family are obtained using shift and additions of the decimated p-ary m-sequences $m(2t)$ and $m(dt)$, where $d = 4$ and $N + 1$. The upper bound for the magnitude of cross-correlation and nontrivial autocorrelation values of the sequences in the family S can be evaluated as $\frac{3}{\sqrt{2}}\sqrt{N + \frac{1}{2}} + \frac{1}{2}$ using the Weil bound and the family size is four times the period of sequences, $4N$.

Bibliography

[1] R. J. McEliece, "A public-key cryptosystem based on algebraic coding theory," DSN Progress Report, vol. 44, pp. 114–116, 1978.

[2] V. M. Sidelnikov and S.O. Shestakov, "On insecurity of cryptosystems based on generalized Reed-Solomon codes," *Discrete Mathematics and Applications,* vol. 1, no. 4, pp. 439–444, 1992.

[3] S. R. Shrestha and Y.-S. Kim, "New McEliece cryptosystem based on polar codes as a candidate for post-quantum cryptography," in *Proc. ISCIT 2014* Incheon, Korea, Sep 24-26, 2014, pp. 368–372.

[4] V. M. Sidelnikov, "A public-key cryptosystem based on binary Reed-Muller codes," *Discrete Mathematics and Applications*, vol. 4 no. 3, 1994.

[5] L. Minder and A. Shokrollahi, "Cryptanalysis of the Sidelnikov cryptosystem," in *Proc EUROCRYPT* 2007, LNCS, vol. 4515, 2007, pp. 347–360.

[6] I. V. Chizhov and M. A. Borodin, "The failure of McEliece PKC based on Reed-Muller codes," *IACR Cryptology ePrint Archive*, Report 2013/287 (2013).

[7] A. Couvreur, P. Gaborit, V. Gauthier-Umaña, A. Otmani, and J.-P. Tillich, "Distinguisher-based attacks on public-key cryptosystems using Reed-Solomon codes," *Des. Codes. Cryptogr.*, vol. 73, no. 2, pp. 641–666, Nov. 2014.

[8] C. T. Gueye and E. H. M. Mboup, "Secure cryptographic scheme based on modified Reed-Muller codes," *International Journal of Security and its Applications*, vol. 7, no. 3, May 2013.

[9] A. Otmani and H. T. Kalachi, "Square code attack on a modified Sidelnikov cryptosystem," *Codes, Cryptology, and Information Security*, vol. 9084, pp. 173–183.

[10] M. Esmaeili, M. Dakhilalian, and T. A. Gulliver, "New secure channel coding scheme based on randomly punctured quasi-cyclic low-density parity check codes," *IET Commun.*, vol. 8, no. 14, pp. 2556–2562, Sep. 2014.

[11] V. M. Sidelnikov and A. S. Pershakov, "Decoding of Reed-Muller codes with a large number of errors," *Problems of Information Transmission,* Vol. 28, no. 3, pp. 269–282, Jan. 1993. (A translation of *Problemy Peredachi Informatsii*.)

[12] P. Lee and E. Brickell, "An observation on the security of McEliece's public key cryptosystem," *In Advances in Cryptology-EUROCRYPT'88*, vol. 330, pp. 275-–280, Springer Verlag (1989).

[13] E. C. Boyle and R. J. McEliece, "Asymptotic weight enumerators of randomly punctured, expurgated, and shortened code ensembles," in Proc. *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, Sep. 2008, pp. 910–917.

[14] H. Niederreiter, "Knapsack-type cryptosystems and algebraic coding theory," *Prob. Contr. Inf. Theory*, vol. 15, pp. 159–166, 1986.

[15] Y. X. Li and R. H. Deng, and X. M. Wang, "On the equivalence of McEliece's and Niederreiter's public-key cryptosystems," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 271–273, 1994.

[16] E. Berlekamp, R. McEliece, and H. van Tilborg, "On the inherent intractability of certain coding problems," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 384–386, May 1978.

[17] N. Courtois, M. Finiasz, and N. Sendrier, "How to achieve a McEliece-based digital signature scheme," in *Proc. Asiacrypt*, vol. 2248, 2001, pp. 157–174.

[18] J.-C. Faugere, V. Gauthier-Umaña, A. Otmani, L. Perret, and J.-P. Tillich, "A distinguisher for high-rate McEliece cryptosystems," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6830–6844, Oct. 2013.

[19] M. Baldi, M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani, "Using LDGM codes and sparse syndromes to achieve digital signatures," in *Proc. PQC 2013*, vol. 7932, 2013, pp. 1–15.

[20] D. Gligoroski, S. Samardjiska, H. Jacobsen, and S. Bezzateev, "McEliece in the world of Escher," IACR Cryptology ePrint Archive, Report2014/360, 2014.

[21] A. Phesso and J.-P. Tillich, "An efficient attack on a code-based signature scheme," in *Proc. PQC*, 2016, vol. 9606, pp. 86–103.

[22] K. P. Mathew, S. Vasant, and C. P. Rangan, "A provably secure signature and signcryption scheme using the hardness assumptions in coding theory," in *Proc. ICISC*, vol. 8565, 2013, pp. 342–362.

[23] M. Finiasz and N. Sendrier, "Security bounds for the design of code-based cryptosystems," in *Proc. Asiacrypt*, 2009, LNCS, vol. 5912, pp. 88–105.

[24] F. Hemmati, "Closest coset decoding of $u|u+v|$ codes," *IEEE J. Sel. Areas Commun.*, vol. 7, pp. 982–988, Aug. 1989.

[25] I. Dumer, "Recursive decoding and its performance for low-rate Reed–Muller codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 811–823, May 2004.

[26] W. Lee, J.-S. No, and Y.-S. Kim, "Punctured Reed–Muller code-based McEliece cryptosystems," *IET Commun.*, vol. 11, no. 10, pp. 1543–1548, Jul. 2017.

[27] R. Overbeck and N. Sendrier, "Code-based cryptography," *Post-Quantum Cryptography,* pp. 95–146, Springer.

[28] L. Dallot, "Towards a concrete security proof of Courtois, Finiasz, and Sendrier signature scheme," in *Proc. WEWoRC*, vol. 4945, 2007, pp. 65–77.

[29] S. T. Choi, T. H. Lim, J. S. No, and H. B. Chung, "On the cross-correlation of a p-ary m-sequence of period $p^{2m}-1$ and its decimated sequences by $\frac{(p^m+1)^2}{2(p+1)}$," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1873–1879, Mar. 2012.

[30] Y. K. Han and K. Yang, "New M-ary sequence families with low correlation and large size," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, Apr. 2009.

[31] J. W. Jang, Y. S. Kim, J. S. No, and T. Helleseth, "New family of p-ary sequences with optimal correlation property and large linear span," *IEEE Trans. Inf. Thoery*, vol. 50, no. 8, pp. 1839–1844. Aug. 2004.

[32] T. Kasami, "Weight distribution formula for some class of cyclic codes," Coordinated Science Laboratory, Univ. of Illinois, Urbana, Tech. Rep. R-285 (AD 632574), Apr. 1966.

[33] T. Kasami, *Weight distribution of Bose-Chaudhuri-Hocquenghem codes*, in *Combinatorial Mathematics and Its Applications.* Chapel Hill, NC: Univ. of North Carolina Press, 1969.

[34] D. S. Kim, H. J. Chae, and H. Y. Song, "A generalization of the family of p-ary decimated sequences with low correlation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7614–7617, Nov. 2011.

[35] J. Y. Kim, S. T. Choi, and J. S. No, "A new family of p-ary sequences of period $(p^n - 1)/2$ with low correlation," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3825–3829, Jun. 2011.

[36] P. V. Kumar and O. Moreno, "Prime-phase sequences with periodic correlation properties better than binary sequences," *IEEE Trans. Inf. Theory*, vol. 37, pp. 603–616, May 1991.

[37] R. Lidl and H. Niederreliter, *Finite Fields,* vol. 20, *Encyclopedia of Mathematics and Its Applications.* Amsterdam, The Netherlands: Addison-Wesley, 1983.

[38] S. C. Liu and J. F. Komo, "Nonbinary Kasami sequences over GF(p)," *IEEE Trans. Inf. Theory*, vol. 38, no. 4, pp. 1409–1412, Jul. 1992.

[39] E. N. Muller, "On the cross-correlation of sequences over GF(p) with short periods," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 289–295, Jan 1999.

[40] E. Y. Seo, Y. S. Kim, J. S. No, and D. J. Shin, "Cross-correlation distribution of p-ary m-sequence of period $p^{4k} - 1$ and its decimated sequences by $\left(\frac{p^{2k}+1}{2}\right)^2$," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3140–3149, Jul. 2008.

[41] Z. Wang, G. Gong, and N. Y. Yu, "New polyphase sequence families with low correlation derived from the Weil bound of exponential sums," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3990–3998. Jun. 2013.

[42] A. Weil, "On some exponential sums," in *Proc. Natl. Acad. Sci.* USA, vol. 34, no. 5, pp. 204–207, 1948.

[43] Y. Xia and S. Chen, "A new family of p-ary sequences with low correlation constructed from decimated sequences," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 6037–6046, Sep. 2012.

초 록

본 논문에서는 포스트 양자 암호시스템과 p진 수열 군에 대한 연구를 수행하였다. 본 논문은 세 개의 주요 연구 결과로 구성되었다.

먼저 첫 번째 연구 결과로, 천공 Reed-Muller (RM) 부호에 기반한 McEliece 암호시스템을 제안하였다. 기존의 RM 부호 기반 McEliece 암호시스템에 대한 공격이 제안된 암호시스템에 대해서는 효과가 없음을 증명하였다. 또한 이러한 공격들을 막기 위한 최적의 천공 기법을 찾아 내었다.

두 번째로, 랜덤 행 추가를 한 천공 RM 부호 기반 전자서명을 제안하였다. 새로운 전자서명은 기존의 Goppa 부호 기반 전자서명인 CFS 전자서명을 개선하였다. 하지만 제안한 수정된 RM 부호 기반 전자서명은 완전 복호를 할 수 있는 RM 부호의 재귀 복호 기법을 활용하여, 서명에 성공할 확률을 높임으로써 서명시간을 줄일 수 있고 또한, EUF-CMA 보안도 달성할 수 있음을 증명하였다.

마지막으로, $p \equiv 3 \bmod 4$인 소수 p와 홀수 n에 대해서 각각 2와 d로 데시메이션된 두 개의 p진 수열 간의 상호상관도를 분석하였다. 이 때 d 값은 두 가지 경우를 가지는데, $d = 4$, $d = (p^n + 1)/2$이다. 각각의 d 값에 대해 상호상관도 값의 상한을 구하고, 우수한 상호상관도 특성을 가지는 주기 $N = (p^n - 1)/2$인 두 개의 새로운 p진 수열군을 제안하였다.

주요어: 공개키 암호시스템, 랜덤 추가, 부호 기반 암호시스템, 천공 기법, 포스트 양자 암호, CFS 전자서명, m-수열, McEliece 암호시스템, p진수열, Reed-Muller 부호, Weil 경계

학번: 2012-20839