Analysis for McEliece and Niederreiter Encryptions: An Alternative to Public Key Encryption

By

Chaw Lian Fong

A project report submitted in partial fulfilment of the

requirements for the award of MASTER OF MATHEMATICS - MEMAC170501

LEE KONG CHIAN FACULTY OF ENGINEERING AND SCIENCE

Universiti Tunku Abdul Rahman

JANUARY 2018

DECLARATION OF ORIGINALITY

I hereby declare that this project report entitled "**Analysis for McEliece and Niederreiter Encryptions: An Alternative to Public Key Encryption**" is my own work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : _____

Name : _____

ID No. : _____

Date : _____

APPROVAL FOR SUBMISSION

I certify that this project report entitled "**Analysis for McEliece and Niederreiter Encryptions: An Alternative to Public Key Encryption**" was prepared by **Chaw Lian Fong** has met the required standard for submission in partial fulfilment of the requirements for the award of MASTER OF MATHEMATICS - MEMAC170501 at Universiti Tunku Abdul Rahman.

Approved by,

Signature : _____

Supervisor : _____

Date : _____

ACKNOWLEDGEMENTS

I want to thank my parents and friends for their support when I encountered problems in completing my project.

Chaw Lian Fong

Analysis for McEliece and Niederreiter Encryptions: An Alternative to Public Key Encryption

Chaw Lian Fong

ABSTRACT

Cryptography is essential for the security of transmission of information. Not only that, it also contribute to the security of communication channel and the safety of device. However, this may not be the case with the presence of post quantum devices. In this project, we study one of the post quantum cryptography which is known as the code based cryptography. It acts as an alternative to public key cryptosystem and is believed to be secured over post quantum attacks. We start by investigating various properties of the two main code based cryptosystems namely, McEliece and Niederreiter encryption schemes. The properties being investigate include the hardness problem, public and private key used, complexity and etc. Furthermore, we construct two variants of McEliece encryption schemes and also a variant of Niederreiter encryption scheme based on the closely relation between the generator matrix and the parity check matrix. Finally, the security of the variants of McEliece scheme are proven by reduction to Sendrier's work. Also, a security analysis for the variant of Niederreiter similar to the variant of McEliece encryption scheme is provided.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

In this modern era, the public tend to use the cryptographic primitives in daily life. The usage of cryptography includes the encrypted messages sent when log into any virtual account or when the public log into a secured wireless connection with a password and etc. Encryption and decryption have now become very common among the people and it is very important to ensure the security behind the system. In the modern cryptography context, most encryption schemes can be classified as perfectly secrecy encryption or computationally secure encryption. For perfectly secrecy, the system requires absolutely no information about the message is leaked regardless of the underlying information that the eavesdropper has. However, a computationally secure encryption is defined as the encryption scheme that leakes only a very small amount of information (about 2^{-60} which is a very small number) with bounded computational power[23]. In real life, both of the perfectly secrecy and also the computationally secure system are considered secure and practical to use in real life. However, the scenario may change with the presence of quantum computer since many well established hardness problems become tractable.

For classical and also modern cryptography, the system uses the digital computing that encodes the data into binary digits called 'bits'. However, the quantum computing differs from binary digital computing as it uses the quantum bits called 'qubits'. Instead of using the binary digits of $\{0, 1\}$, qubits also include the field called the superposition state. As a result, the cryptanalysis ability of the quantum computer is highly enhanced and the security of the current cryptosystem may be broken easily. With this, the cryptography system that used in real life is considered to be no longer safe as they are vulnerable under the attack of quantum computation. Although quantum computation is still in developing state, a secure cryptography system that can endure the attack of quantum computer is very crucial for the modern digital life. Unlike other number theory based cryptography that are vulnerable under post quantum attack, code based cryptography is believed to be safe to resist the post quantum attack. There are two well-known code based cryptosystems, namely, the McEliece and Niederreiter versions shown in Algorithms 1 and 2, respectively in Chapter 3.

McEliece encryption scheme is the first scheme that uses randomization in encryption process which is proven to be post quantum resistant. The scheme is constructed on the hardness problem of decoding of any arbitrary general code. On the other hand, Niederreiter encryption scheme is another simplified version of McEliece encryption scheme that uses the concept of parity check matrix. The scheme is called the simplified version of McEliece encryption scheme as it uses the matrices with smaller dimension for key generation and encrption process. Nevertheless, Niederreiter encryption scheme [6] is about 10 times faster as compared to McEliece encryption scheme.

In this research, the important properties of the McEliece encryption scheme and Niederreiter encryption schemes are compare and contrast. Besides, we propose a new variant for both McEliece and Niederreither encryption schemes with improvised of the security of public key. Last but not least, the research aims to prove the security of both variants of McEliece and Niederreiter encryption schemes by using reduction.

1-1 Objectives

1. To compare and contrast McEliece encryption scheme with Niederreiter encryption scheme by investigating the underlying hardness problems, key sizes and computational complexity.

2. To propose new variant of the McEliece encryption schemes which improve the security of the public key.

3. Since McEliece and Niederreiter encryption schemes are computationally equivalence, we proposed a new secure variant of Niederreiter scheme and the security is proven similar to the variant of McEliece encryption scheme.

1-2 Problem Statement

Current encryptions schemes that higly based on number theory concept is vulnerable under the post-quantum attack. The encrytion schemes that resist the post quantum attack consists of some defects and considered to be relatively not effective as compared to the cryptosystems that we used today. However, it is very essential to study the encryption schemes that resist the post quantum attack and thus ensure the security for the future. As a result, one of the encryption scheme namely the code-based encryption scheme is being studied in this project.

1-3 Research Questions

1. What is the similarity and differences between McEliece encryption scheme with Niederreiter encryption scheme in terms of hardness problems, key sizes and computational complexity?

2. How to improve the security of the public key in McEliece encryption scheme?

3. How to construct a variant for Niederreiter scheme similar to the variant of McEliece encryption scheme?

1-4 Literature Review

In 1971, Goppa [8] introduced the concept of fast decoding algorithm which is very crucial in coding process . In general, he introduced the concept of generalized linear code. With the continuation of Goppa's contribution, McEliece [11] succeeded in applying Goppa's code in public-key cryptography in 1978. In 1986, Niederreiter came out with a simplified version of McEliece's cryptographic scheme with a smaller public key size [12].

As compared to the well-known RSA system that used today, McEliece encryption scheme is considered relatively not famous because of the relatively large public key size. However, the system does has several very important advantages as compared to other encryption schemes. First of all, because of the relatively low complexity of the encoding and decoding, the scheme is considered to be fast. Next, the system also considerably secure under several attacks such as the exhaustive search attack and the information set decoding attack. For more information, refer to [3, 5].

For Niederreiter encryption scheme, the simplified version of McEliece encryption scheme can even run faster than it's predecessor. This allowed the scheme to be derived to signature scheme [7] that is infeasible for McEliece's encryption scheme. The derivation of the signature scheme can be done by hashing the message and the decrypted message is now added with the signature for enhanced security purposes.

With the same Goppa code used, McEliece encryption scheme and Niederreiter encryption scheme are consider to have the same degree of security [6]. Therefore, to achieve the CCA2 secured variant, conversion need to be done for both McEliece encryption scheme and Niederreiter encryption scheme. The popular Kobara and Imai's conversion in 2001 with McEliece encryption scheme is said to be as secure as the original scheme but with a relatively smaller public key size [9].

As the cryptosystem that applied number theory concept is said to be weak and vulnerable against the Shor's attack [16], and thus vulnerable under post quantum attack. Therefore, it is very essential to look into some of the cryptosystems that resist the post quantum attack such as the McEliece encryption scheme. According to Dowsley et al.(2009) [13], a CCA2 secure scheme can be constructed in the standard model with

the k-repetition CPA secure McEliece scheme. Furthermore, it has also been shown in [13, 14] with the existence of IND-CCA2 secure McEliece encryption scheme.

According to Rosen and Segev(2009), k-repetition McEliece scheme is IND-CCA2 secure but the key size may very large [18]. The large key size may give the result of impractical usage in real life although the scheme is secure. Therefore, we propose two new variants of McEliece encryption scheme.

The security assumption for code based cryptography depends on the hardness of decoding in a random linear code [2]. The very crucial problem to the hardness of coding theory is that the complexity increases exponentially with regards to the decryption process [1]. Next, there are also some security assumptions for McEliece's encryption scheme and Niederreiter encryption scheme such as the indistinguishability of Goppa codes [19]. The hardness problem related to coding is reliable and valid although is not proven because there is still no any valid break down of the system by efficient adversary known today.

A summary of some important literature reviews are provided in the Table 1.1.

Authors and Years	**Title**	**Contribution**	**Remarks**
Robert J. McEliece 1978	A Public-Key Cryptosystem Based On Algebraic Coding Theory	A new variant of public key encryption scheme (PKE) using error correcting codes.	The scheme is fast with low complexity and secured without any successful attack known until today. The public key size is large.
H. Niederreiter 1986	Knapsack-type cryptosystems and algebraic coding theory	A variant of McEliece scheme that is faster.	Smaller key size and signature scheme can be applied. Some special cases such as the usage of Reed Solomon code is broken.

Kobara, K. and Imai, H 2001	Semantically secure McEliece public-key cryptosystems-conversions for McEliece PKC	A CCA2 variant of McEliece scheme	The variant is said to be as secured as the original scheme.
McEliece, R. 2002	The Theory of Information and Coding	Revised version of McEliece Encryption scheme	The big key size is assumed to be solvable with technology advancement. However, the relatively big key size problem still retained as the dimension of the matrices used for key generation and encryption still remained the same.
Li, Y.X., Deng, R.H., Wang, X.M 1994	On the Equivalence of McEliece's and Niederreiter's Public-Key Cryptosystems	McEliece and Niederreiter scheme are considered to have the same degree of security.	Niederreiter scheme is safe.
Bernstein, Daniel J.; Lange, Tanja; Peters, Christiane 2008	Attacking and defending the McEliece cryptosystem	McEliece scheme is secured.	A CCA2 variant of McEliece scheme can be constructed

Jean-Charles Faugère , Ayoub Otmani , Ludovic Perret , and Jean-Pierre Tillich 2010	Algebraic Cryptanalysis of Compact McEliece's Variants – Toward a Complexity Analysis	A new algebraic approach to investigate the security of the McEliece cryptosystem	An efficient key-recovery attack against two compact variants.
Edoardo Persichetti 2012	Improving the efficiency of code based Cryptography	Reduced the public key size of McEliece scheme	A better flexibility, and improved resistance to all the known attacks variant.
Wen Wang , Jakub Szefer , and Ruben Niederhagen 2017	FPGA-based Key Generator for the Niederreiter Cryptosystem using Binary Goppa Codes	A secured, efficient, and tunable FPGA implementation of the key-generation algorithm for the Niederreiter cryptosystem using binary Goppa codes	Parameters with equivalent or exceeding the recommended security level.

Table 1.1: Summary of Literature Review

1-5 Research Methodology

The first part for the research is to identify the properties of McEliece's encryption scheme and Niederreiter's encryption scheme. These important properties include hardness problem, key sizes and computational complexity. As both schemes are constructed based on the similar concept, both schemes are expected to have high degree of similarity and the differences between two schemes are the trade-off between efficiency and security.

Next, to improve the security of the public key in McEliece's scheme in terms of avoiding the public key to be permutation-equivalent to the secret code chosen, the permutation matrix is replaced to other codes to form a dense transformation matrix. After replacing the permutation matrix, the permutation equivalence between the public key and secret code is eliminated. For instance, the variant of McEliece encryption proposed in this research paper replaced the permutation matrix with just only the generator matrix to form the dense matrix to ensure the security and at the same time eliminate the permutation effect of the scheme.

The secure variant for McEliece encryption scheme is used as a model to construct a new secure variant for Niederreiter encryption scheme. For McEliece encryption scheme, the security is proven by reduction as shown in Sendrier's work[19]. In order to construct the secured variant for Niederreiter encryption scheme, a similar concept is used by applying the product construction concept to enhance the security of the variant and at the same time retain the original Neiderreiter encryption scheme as part of the new variant.

1-6 Expected Outcomes

1. The similarity and differences between McEliece's encryption scheme with Niederreiter encryption scheme in terms of hardness problem, key sizes, computational complexity, and computationally equivalence are identified and analysed.

2. An improved scheme that improves the security of the public key in McEliece scheme by avoiding the public key to be permutation-equivalent to the secret

code chosen is generated.

3. A secure variant for Niederreiter encryption scheme is constructed.

CHAPTER 2: INTRODUCTION TO CODING THEORY

All communication channels contain some degree of noises such as the electric impulses, amplitudes, location and etc. Communication become difficult with the present of noises. In order to have a successful communication, we can resend/retransmit the message so that the correct message is sent. This process is called redundancy in terms of data transmission. With some degree of redundancy, we can correct the received word to achieve a successful communication. Nowadays, error correcting codes can be used as construct some efficient encryption schemes which resist post quantum attack.

For an open channel error correcting code, the code will go through two important processes namely the **encoding** and **decoding**. The message m will first transmitted through encoding process to convert m into a codeword c. Then, c will transmitted through a noisy channel and reach the receiver. The codeword c that passed through the noisy channel will now consist of some degree of error. Next, c with error will now pass through decoding process and c will now be converted back to the message m.

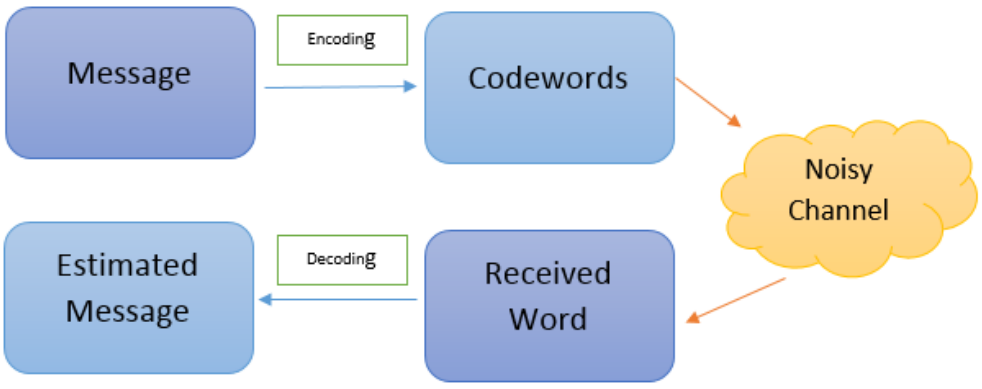

Figure 2.1: Encoding & Decoding

2-1 Linear Codes

Linear code is a well-known family of error correcting codes which apply many nice algebraic structures inherited from vector spaces [8, 11, 12].

Definition 2-1.1. [17] Suppose $\mathbf{F}_2^n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in \mathbf{F}_2\}$ is a n- dimensional vector space over the binary field $\mathbf{F}_2$. Then, C is a $[n, k]$- binary linear code if and only if C is k-dimensional subsapce of $\mathbf{F}_2^n$. The value n is known as the length of C and k is the dimensional of C.

Example 2-1.2. Let $\mathbf{F}_2^3$ be a 3-dimensional vector space over $\mathbf{F}_2 = \{0, 1\}$. Then, $\mathbf{F}_2^3 = \{000, 001, 010, 100, 101, 011, 110, 111\}$. Suppose $C =< 001, 010 >$ is a subspace of $\mathbf{F}_2^3$ generated by 001 and 010, Then, $C = \{000, 001, 010, 011\}$ is a 2-dimensional subspace of $\mathbf{F}_2^3$, Hence, C is a [3, 2]- binary liner code.

Next, we introduce another important parameter of a $[n, k]$- linear code C.

Definition 2-1.3. [17] Suppose $a = (a_1, a_2, \ldots, a_n), b = (b_1, b_2, \ldots, b_n) \in C$.

1. Hamming distance, denoted by $d(a, b)$ is the number of places where a_i differs from b_i, that is, $d(a, b) = |\{i \mid a_i \neq b_i\}|$.

2. Hamming weight denoted by, $wt(a)$ of the codeword a is the number of non-zero positions in a, that is, $wt(a) = |\{i \mid a_i \neq 0\}|$.

Definition 2-1.4. [17] Let C be an $[n, k]-$ linear code. The smallest Hamming distance of all the codewords is the minimum distance of C, denoted by $d(C)$.

$$d(C) = min\{d(a, b) \mid a, b \in C, a \neq b\}.$$

Example 2-1.5. Let $C = \{0000, 1010, 0101, 1111\}$.We compute the distance of any two distinct codewords.

d(0000,1010)=2	d(0000,0101)=2
d(1010,0101)=4	d(0101,1010)=4
d(1010,0000)=2	d(0101,0000)=2
d(0101,1111)=2	d(0000,1111)=4
d(1111,0101)=2	d(1111,0000)=4
d(1111,1010)=2	d(1010,1111)=2

As a result, the minimum distance $d(C) = 2$.

Next, for all $x, y \in C$, we see that $d(x, y) = d(x + y, \underline{0}) = wt(x + y)$. Because C is a linear code, then $\forall x, y \in C, x + y \in C$. Thus $d(x, y) = wt(x + y) = wt(v)$ where $v = x + y \in C$.Therefore, for the linear code $C = \{0000, 1010, 0101, 1111\}$, to compute $d(C)$ is equivalen to compute $wt(v)$ $\forall v \in C$.

wt(1010)=2,

wt(0101)=2 and

wt(1111)=4 .

Thus, the minimum weight $w(C) = 2$.

Example 2-1.5 can be generalized to the following theorem.

Theorem 2-1.6. *Let C be a linear code over F. Then, $d(C) = w(C)$.*

With the concept of minimum distance, we can say that a code is a linear three tuples $[n, k, d]$ - linear code. It is very important for us to investigate the distance of a code because it identify the error detecting and also the error correcting capabilities of a particular code.

Theorem 2-1.7. *[17] Let C be an $[n, k, d]$ - linear code over $\mathbf{F}_2$. Then C can detect any error patterns of weight less than or equal to $d - 1$, and C can correct any error patterns of weight less than or equal to $\frac{d-1}{2}$.*

Definition 2-1.8. [17] The dual code of C is defined as $C^{\perp} = \{u \in \mathbf{F}_2^n \mid u \cdot c = 0, \forall c \in C\}$.

Example 2-1.9. Let $C = \{0000, 1010, 0101, 1111\}$. The calculation to obtain $C^{\perp}$ from C is shown below.

First, we let $v = (x, y, z, w)$ and so $v \cdot 1010 = v \cdot 0101 = v \cdot 1111 = 0$. Then, we have

$$x + z = 0,$$

$$y + w = 0 \text{ and}$$

$$x + y + z + w = 0.$$

By solving this equation, we obtain $v = (-z, -w, z, w)$. Since z and w can be either 0 or 1. Therefore, we obtain $v = C^{\perp} = \{0000, 1010, 0101, 1111\}$.

The elements in Example 2-1.9 has a very nice structure and properties as all the elements are orthogonal to one another. The very special type of dual code as shown in Example 2-1.9 is called the self-dual code that defined as follow.

Definition 2-1.10. [17] C is self-dual if $C = C^{\perp}$.

Since linear code is a vector spaces, all elements of the code are generated by the bases. In coding theory, a basis is normally written in a matrix form. We called the basis matrix as generator matrix, G. Similarly, the basis matrix of the dual code is called the parity check matrix, H. Formally,

Definition 2-1.11. [17] Let C be an $[n, k, d]$ - linear code over $\mathbf{F}_2$. There exist a $k \times n$ generator matrix G and a $(n - k) \times n$ parity check matrix H.

(i) A generator matrix G for C is a matrix whose rows form a basis for C.

(ii) A parity check matrix H for C is a generator matrix for the dual code $C^{\perp}$.

The generator matrix and parity check matrix play a very important role in the encoding and decoding process, respectively.

In the following example, we use generator matrix G to encode the message such that for $m \in \mathbf{F}_2$, the respective codewords mG can be obtained.

Example 2-1.12. Let $G = \begin{pmatrix} 0101 \\ 1010 \end{pmatrix}$. Then the encoding process is as follows:

$$\mathbf{F}_2^2 \to C$$

$$m \to mG$$

$$00 \to 0000$$

$$01 \to 1010$$

$$10 \to 0101$$

$$11 \to 1111$$

As a result, the encoding yields the linear code, $C = \{0000, 1010, 0101, 1111\}$.

In general, the generator matrix is said to be in standard form if it has the form of$(I_k \quad A)$ where I_k is a $k \times k$ identity matrix.

Theorem 2-1.13. *[17] If $G = (I_k \quad A)$ is a generator matrix for a code C, then $H = (-A^T \quad I_{(n-k)})$ is a parity check matrix of C.*

Proof Consider the ith row of G, that has the form

$$v_i = (0, ..., 1, ..., 0, a_{i,1}, ..., a_{i,n-k})$$

where the 1 is in the $i$thposition. This is the vector of code C. The jth column of H^T is the vector

$$(-a_{1,j}, ..., -a_{n-k,j}, 0, ..., 1, ..., 0)$$

where the 1 is in the $(n-k+j)$th position. To obtain the jth element of $v_i H^T$, take the dot product of v_i and H^T,

$$v_i \cdot H^T = 1 \cdot (-a_{1,j}) + a_{i,j} \cdot 1 = 0$$

Therefore, H^T annihilates every row v_i of G. Since every element of C is a sum of rows of G, we find that $vH^T = 0$ for all $v \in C$.

From the fact of the left null space of an $m \times n$ matrix of rank r has dimension $n-r$ from linear algebra, H^T has rank $n-k$ since it has I_{n-k} as a submatrix. Therefore, the left null space has dimension k. Since C is contained in this null space, and C has dimension k, it must equal the null space, that prove what the theorem said.

Q.E.D.

This theorem shows the conversion between the generator matrix and also the parity check matrix in the standard form.

With the proof, we know that

$$\forall v \in \mathbf{F}, v \in C \leftrightarrow vH = 0.$$

This property gives an important role in decoding a code using the nearest neighbor decoding method and syndrome decoding method with coset. Before going into the decoding problem, we introduce the concept of coset.

Definition 2-1.14. [17] Let C be an $[n, k, d]$ - linear code over $\mathbf{F}_2$, and let u be any vector of length n over $\mathbf{F}_2$. The coset of C is defined to be the set $u + C = \{v + u : v \in C\}$. Note that $u \in u + C$.

Theorem 2-1.15. *[17] Suppose C is an $[n, k, d]$ - linear code over $\mathbf{F}_2$. Then,*

(i) Every coset contains exactly $|C| = 2^k$ vectors;

(ii) Any two cosets are either equivalent or disjoint.

Proof To prove for (i), we follow the definition as $u+C$ has at most $|C| = 2^k$ elements. With this, the two elements $u + c$ and $u + c'$ of $u + C$ are equal if and only if $c = c'$. Thus, $|u + C| = |C| = 2^k$.

To prove for (ii), we consider two cosets $u + C$ and $v + C$ and suppose $\alpha \in (u + C) \cap (v + C)$. Since $\alpha \in (u + C)$, $u + C = \alpha + C$ as $(u + C)$ is the subset of $\alpha + C$. Hence, $u + C = v + C$. **Q.E.D.**

Definition 2-1.16. [17] A coset leader is the vector having the minimum weight. With more than one minimal weight vector, choose either one to be the coset leader.

With coset, now we introduce decoding as the process of guessing the original message being sent from the codeword received. A simple yet efficient decoding method called nearest decoding method is stated as follow[17]:

Algorithm 1 Nearest neighbor decoding

1: Error pattern e is transmitted along with the received codeword c. Let m be the message codeword, we have $c = m + e$.
2: Error pattern e and received codeword c are within the same coset.
3: Choose another word e with least weight from the coset $c + C$, where C is the linear code.
4: We have $c = m + e$.

The nearest neighbor decoding algorithm is used to solve the decoding problem stated as the input and output.

Input: The received words $m = c + e$ where m is message sent and e is the error.

Output: The codeword m.

With nearest neighbor decoding, another more efficient method called the syndrome decoding is stated as follow :

Algorithm 2 Syndrome decoding for linear code.

1: Compute $s = cH$, where s is the syndrome.
2: List down all cosets and identify coset leader u.
3: With u, $m = c - u$, else return $\perp$.

Similarly, Algorithm 2 is used to solve the decoding problem of the input and desire output.

Input: H and c.

Output : m.

The algorithm is valid with $wt(e) \leq t$, where $t \leq \lfloor \frac{(d-1)}{2} \rfloor$, where d is the minimum distance of the code. As a result, we obtain

$$cH = (c + e)H = 0 + eH = eH.$$

Since the weight is within t, we can determine the error pattern from coset leader..

2-2 Families of Codes and Parameters

In this section, we list the parameters of some well-known families of codes.

Codes	n	k	d	**Remark**
Binary Hamming Codes	$2^r - 1$	$2^r - 1 - r$	3	$r \geq 2$
q-ary Hamming Codes	$\frac{q^r-1}{q-1}$	$\frac{q^r-1}{q-1} - r$	3	$r \geq 2$
Golay Codes G_{24}	24	12	8	-
Binary Golay Codes G_{23}	23	12	7	Obtained from G_{24} by deleting the last coordinate of every codeword.
Reed -Muller Codes	2^m	$\binom{m}{0} + \binom{m}{1} + ... + \binom{m}{r}$	2^{m-r}	$0 \leq r \leq m$.
Binary Bose, Chaudhuri & Hocquenghem (BCH) Codes	$2^m - 1$	$g(x) := lcm[M^{(a)}(x), M^{(a+1)}(x), ..., M^{(a+(2t-1)-2)}(x)]$	$2t - 1$	$3 \leq m \leq 6$
Reed-Solomon Codes	$q - 1$	$q - \delta$	δ	$q - 1 \geq \delta \geq 2$.

Table 2.1: Codes and Parameters

2-3 Families Of Linear Codes And Bounds

Next, we introduce some of the famous codes and the corresponding bounds. Given a linear code, the optimum condition for the code is when having large bandwidth (transfer large number of codewords) and large minimum distance (high error correcting capability). However, to have both large condition is not quite possible in real life practice as the increase of bandwidth of the codeword will result in the drastic decrease of the minimum distance and vice versa. To achieve a balance tradeoff between the two condition, several important bounds are introduced over the past years to achive the largest possible value of bandwidth with the reasonably minimum distance.

2-3-1 Hamming Codes

Hamming code is considered as the classic code for which the encoding and decoding process can be carried out easily. Besides that, the code is also has the ability to correct up to one error. The parameter is shown in Table 2.1.

The code is bounded by the Hamming bound that described as follows. For a proof, refer to [17].

Theorem 2-3.1. *For any integer $q \geq 1$ and integers n, d such that $1 \leq d \leq n$, there is*

$$A_q(n,d) \leq \frac{q^n}{\sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} (q-1)^i}.$$

A q-ary code is called a perfect code if it attains the Hamming bound.

2-3-2 Rank Codes

Rank codes or Gabidulin codes are non binary linear error correcting codes that apply rank metric concept instead of Hamming concept. It applies the concept of singleton bound which stated formerly as follows.

Theorem 2-3.2. *With any integer $q > 1$, $1 \leq d \leq n$, where $n > 0$. We obtain*

$$A_q(n,d) \leq q^{(n-d+1)}$$

If q is prime, the parameter of linear code satisfy

$$d \leq n-k+1$$

A linear code is defined as maximum distance separable (MDS) if the parameter

$$d = n-k+1$$

2-3-3 Reed Solomon Codes

Reed Solomon code is non-binary cyclic error-correcting code which is MDS that based on univariate polynomials over finite fields that can correct up to multiple errors.

Theorem 2-3.3. *Suppose C is an $[n,k,d]$ - linear code over F.The generator matrix and parity check matrix of C denoted by G and H, respectively. We have,*

(i) C is maximum distance separable if every columns of parity check matrix and generator matrix is linearly independent.

(ii) Dual code of C denoted by $C^{\perp}$ is also maximum distance separable if C is maximum distance separable.

2-3-4 Cyclic Codes

Another famous family of linear code is the cyclic code which is the most commonly used linear code in daily life due to some of the special properties.

Definition 2-3.4. A subset of S of $\mathbf{F}_q^n$ is cyclic if $x' = (x_{n-1}, x_0, x_1, ..., x_{n-2}) \in S$ whenever $x = (x_0, x_1, ..., x_{n-1}) \in S$ (x' is cyclic shift of x).

Example 2-3.5. The binary linear code

$C = \{000000, 100100, 010010, 110110, 101101, 011011, 111111\}$ is a cyclic code since $x = (x_0, x_1, ..., x_6, x_7) \in C$, $x' = (x_7, x_0, x_1, ..., x_6)$also $\in C$.

2-4 Hardness Problems

The major goal for a cryptography scheme is to create some cryptographic primitives that provably secured. The combination of cryptography and coding theory results in the code based cryptography scheme that consists of the problem that is hard to be proven secure. However, the problems are 'assumed' to be difficult in real life practice.

Definition 2-4.1. [23] Provable security refers to the inability of the adversary attacking model to solve the underlying hard problem in the implementation of a scheme relative to the definition of the scheme and the assumptions being used.

There are a few commonly well-known hard problems that based on number theory that are believed to be intractable and are applied in daily practical real-life usage. Although it is impractical to prove the hardness problems, it is believed that the complexity to solve the hardness is very high and is impossible to solve it efficiently within a short period of time. As a result, the encryption scheme with hardness problem is hard to break and is secured.

2-4-1 Integer Factorization Problem

Integer factorization in term of number theory refers as the decomposition of composite number into the product of smaller prime factors. In a cryptographic scheme, the factor is normally restricted to two large prime factors which is hard to decompose. Simply speaking, the integers are the product of two large prime factors that are hard to solve, $n = pq$ where p an q are the large prime factors. This hard problem is easy to apply and provably secure in the encryption scheme that used in real life such as the RSA encryption scheme[25].

2-4-2 Discrete Longarithm Problem

Another well-known hard problem based on number theory is the discrete longarithm problem. To explain this problem, let fix a prime p. Then, we let α and β be non-zero integers mod p. Suppose

$$\beta \equiv \alpha^x \pmod{p}$$

Given β and α, the problem of finding x is the discrete longarithm problem. The exhaustic search through all possible component is only possible for small p and not feasible for large p and therefore considered hard to solve in practical. One of the example of encryption scheme that practise discrete longarithm problem in real parctical life is the ElGamal cryptosystem [25].

2-4-3 Hardness Problem In Coding Theory

Unlike the practical usage of the number theory problem, there are two classical hardness problems in coding that are different from number theory. The first problem is called the maximum likelihood decoding problem (MLD).

Given a linear code C over $\mathbf{F}$ and a vector $v \in \mathbf{F}^n$, the objective is to find $m \in C$ such that $d(v, m)$ is minimal. The problem arise when the adversary try to break the scheme by correcting certain errors in the codewords m where the received word, $w = m + e$ such that the e refers to the errors. A unique solution present if $wt(e) \leq t$, where $t \leq (d-1)/2$ and d is the minimum distance of the code. The problem is believed to be hard in general.

Next, another problem called the syndrome decoding problem (SDP) that applied the concept of parity check matrix to find the syndrome. Given an $(n-k) \times n$ parity check matrix H for an $[n, k]$-linear code C over F, a syndrome $s \in \mathbf{F}^{(n-k)}$ can be generated. The goal is to find $e \in \mathbf{F}^n$ such that $s = He^T$. Similar to MLD, the hardness of syndrome decoding problem is not proven, but is believed to secure in practical usage.

Algorithm 3 Codeword Finding Problem (CFP)

Instance: The set of all possible generator matrices of C, an integer $w > 0$ and a matrix G_C.

Find: A codeword v of weight $\leq w$ in C with generator matrix G_C.

Algorithm 4 Syndrome Decoding Problem (SDP)

Instance: The set of all possible generator matrices of C, an integer $t > 0$ and a matrix H_C and $s \in \mathbf{F}_2^{n-k}$.

Find: A vector $e \in \mathbf{F}_2^n$ of $wt(e) \leq t$ such that $eH_C = s$.

CHAPTER 3: MODERN CRYPTOGRAPHY

3-1 Introduction to Modern Cryptography

Modern cryptography refers to the post-1980's cryptography system that differs from the classical cryptography on definition, the precise assumption and the rigorous proof of security. Modern cryptography is defined as the study of mathematical techniques for securing digital information, systems, and distributed computations against adversarial attacks [23].

With a formal definition of the modern cryptography, it is a very essential element for a cryptosystem to have a very well defined formal definition of security. As to design any cryptographic primitives or protocols, it is very important for the formal definition to act as the guideline for what the system do and ensure the security purpose of the system. Next, the precise assumption is very important to modern cryptography. Although not proven, a well-defined and stated assumption can provide some rigorous proof for the security of a system.

Public key encryption is one type of cryptography system that use the concept of double keys that are the public key and also private or secret key. The public key may be distributed widely in public and can be get easily from the encryption scheme while the private or secret key must be kept in secret to ensure the security of the system. The public key encryption can rely on several important mathematical algorithms such as integer factorization and also discrete logarithm that are believed to be hard so that the system is secure.

In a public key encryption scheme, any message can be encrypted by using the easily accessible public key. However, the decryption process can only be carried out using one or several important secret key(s) that are only known to the owner. The secret key must be kept in secret and must be unknown to the adversary. Besides, the secret key must also attain some degree of computational complexity so that the adversary is hard to recover the secret ket and reveal the underlying message.

The RSA cryptosystem is proposed by Rivest, Shamir, and Adleman in 1977 and

named from the founder of the scheme. The scheme is an example of public key cryptography that practice the integer factorization problem. The problem of factorizing the two large prime numbers is hard and thus the scheme is secure and is one of the mostly used cryptosystem in daily practical life due to the hardness problem that is safe. As we have infinitely number of primes in mathematics system and the primes that used in each scheme are kept in secret to ensure the security of the system.

ElGamal crypotosystem is another well-known public key encryption that is applied in daily practical life. Unlike RSA system, ElGamal encryption relies on the hardness problem of discrete logarithm problem. The scheme may be referred to ineffective as the probabilistic encryption may create many possible ciphertexts.

3-1-1 McEliece Encryption Scheme

Next, we discuss two very important code based public key encryptions that are the McEliece encryption scheme and Niederreiter encryption scheme. Both schemes rely on syndrome decoding problem.

Definition 3-1.1. Public key encryption is defined as 3-tuples(KeyGen, Enc, Dec) probabilistic polynomial time algorithm such that

(i) Gen : Take the input of security parameter to generate the secret key(sk) and public key (pk).

(ii) Enc : Encryption process that take in message m and public key to generate cipher text c.

(iii) Dec : Decryption process that decrypt the cipher text with secret key to obtain the message.

Algorithm 5 McEliece Encryption Scheme [11]

1: The parameter is defined over $n, t \in N$ where $t \leq n$.

2: **Gen :**

(i) Compute G that is a $k \times n$ generator matrix of code C over $\mathbf{F}$ with dimension k and minimum distance of $d \geq 2t + 1$.

(ii) Compute S that is a $k \times k$ random binary non-singular matrix.

(iii) Compute P that is a $n \times n$ random permutation matrix.

(iv) Compute $k \times n$ matrix $G^{pub} = SGP$.

(v) Public key is (G^{pub}, t), while private key is (S, D_G, P) whereD_G is the efficient decoding algorithm for G.

3: **Encryption :**

(i) Encrypt a plaintext $m \in \mathbf{F}^k$ by choosing a vector $z \in \mathbf{F}^n$ of weight t. Then, by randomly, compute the ciphertext, $c = mG^{pub} + z$.

4: **Decryption :**

(i) With ciphertext , first, compute the inverse of $P = P^{-1}$. Then calculate the c by using the formula : $cP^{-1} = (mS)G + zP^{-1}$.

(ii) Apply the decoding algorithm for G, D_G and obtain $mSG = D_G(cP^{-1})$.

(iii) Let $J \subset \{1, \ldots, n\}$ be a set such that G_J^{pub} is invertible and compute $m = (mSG_J)(G_J)^{-1}S^{-1}$.

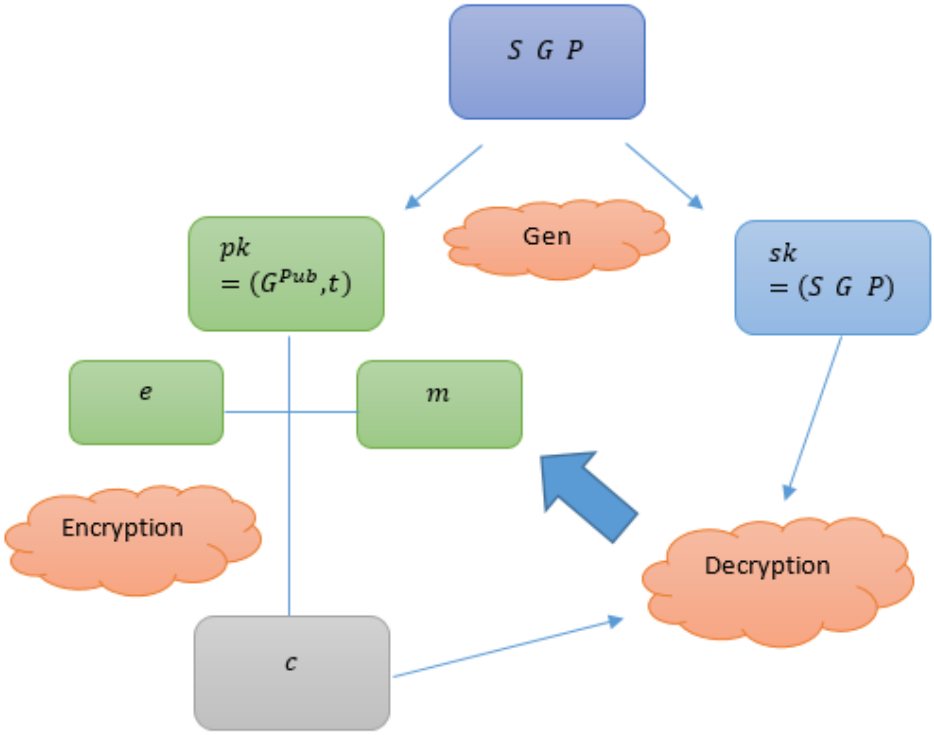

Figure 3.1: McEliece Encryption Scheme

3-1-2 Niederreiter Encryption Scheme

Algorithm 6 Niederreiter Encryption Scheme [12]

1: The parameter is defined over $n, t \in N$ where $t \ll n$.

2: **Gen :**

(i) Compute H that is a $(n-k) \times n$ check matrix of code G which can correct up to t errors.

(ii) Compute P that is a $n \times n$ random permutation matrix.

(iii) Compute $(n-k) \times n$ matrix $H^{pub} = MHP$ whose columns span the column space of HP.

(iv) Public key is (H^{pub}, t) while private key is (P, D_G, M)whereD_G is the efficient decoding algorithm for G.

Encryption :

(i) Encrypt a plaintext $m \in F^k$ by choosing a vector $e \in (0,1)^n$ of weight t. Then, by randomly, compute the ciphertext, $c = H^{pub}e^T$.

Decryption :

(i) With ciphertext , first, compute $M^{-1}c = HPe^T$.

(ii) Apply the decoding algorithm for G, D_G in order to recover e^T.

(iii) Obtain the plaintext $e^T = P^{-1}Pe^T$.

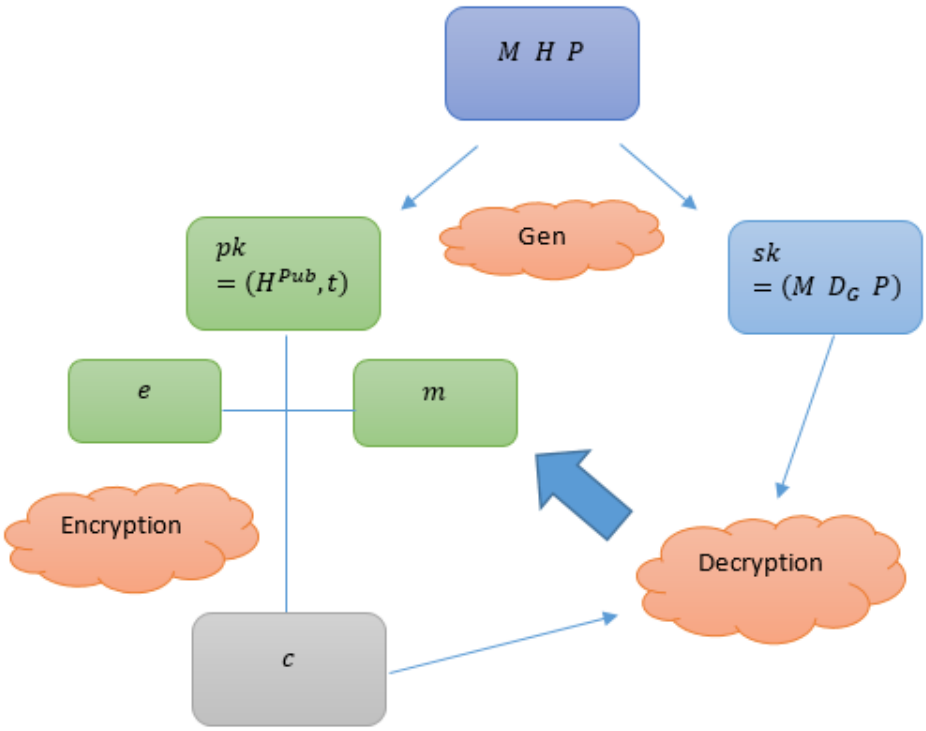

Figure 3.2: Niederreiter Encryption Scheme

3-2 Security Model For Encryption Scheme

For cryptographic scheme constructed, one can easily claim that the scheme is 'secure'. However, for the truth, the scheme may not as secure as it seemed to be. Therefore, a standard definition is stated to differentiate between the secured scheme and also the scheme that may consist of high risk against cryptanalysis.

Definition 3-2.1. Secured encryption scheme is defined as regardless of any information an adversary already has, a ciphertext should not leak any additional information about the underlying plaintext.

According to the definition, the scheme is not secure if some additional information is leaked although the information leakage may seemed unimportant to the plaintext message that send to the recipient. As a result, a few attacks can be used to prove the security of a particular scheme.

(i) Ciphertext only attack refers to the adversary that can only observe the ciphertext and try to get the underlying plaintext message from the observation.

(ii) Known-plaintext attack refers to the adversary that can try to deduce the underlying plaintext message from some plaintext/ciphertext generated using the same key.

(iii) Chosen-plaintext attack refers to the adversary that can obtain plaintext/ciphertext pairs for plaintexts of its choice.

(iv) Chosen-ciphertext attack refers to the adversary that deduce some underlying additional information from the decryption of the ciphertext of its choice.

To illustrate the security model, one of the possible attack namely the adaptive chosen ciphertext attack (CCA2). CCA2 is defined as the attack where the attacker sends some ciphertexts to be decrypted.

The decrypted results from CCA2 are used to select the subsequent ciphertext. The ciphertext is modified to have predictable property. The attack gradually reveal the underlying information of the key. This can only be done when the public key has

the property of ciphertext malleability. To prevent the attack, the malleability should be reduced to minimum. In complexity cryptography, indistinguishability CCA2 (IND-CCA2) is often used to show the non-malleability property.

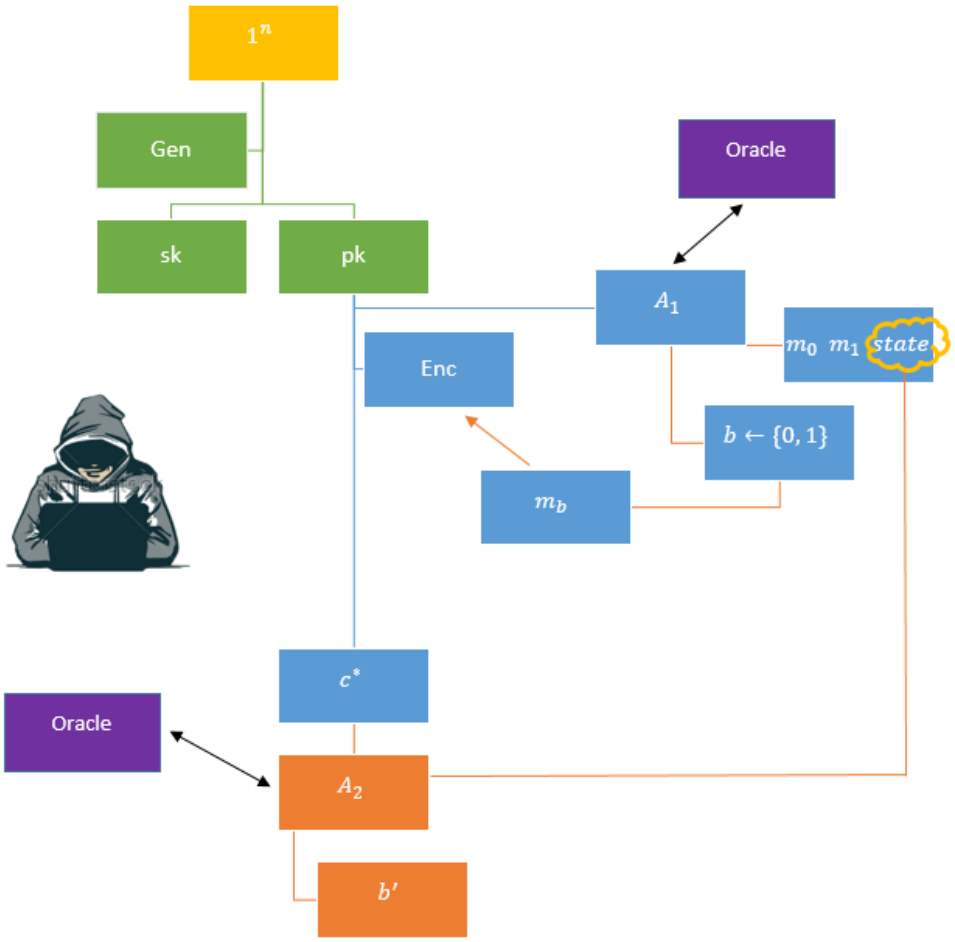

Figure 3.3: IND-CCA2 Experiment

CHAPTER 4: VARIANTS OF MCELIECE ENCRYPTION SCHEMES

4-1 Motivations

In the original McEliece scheme, the encryption scheme makes use of a generator matrix G_C. On the other hand, the Niederreiter scheme make use of a parity check matrix $G_{C^\perp} = H_C$. However, there is a closed relationship between the generator matrix G_C and the parity check matrix H_C, as shown in Diagram 4.1. The main ingredient in this relationship is a method known as t-method, which is illustrated as follow.

Suppose G_C is a generator matrix of a linear code C. We perform a sequence of elementary row operations to G_C, so that G_C can be written in the following form:

G' = $[I_k\ A]$, where A is a $k \times (n-k)$ matrix.

For some cases, we just need to perform successive columns permutation to G_C to obtain G'. Once we obtain G', we form $H' = \begin{bmatrix} A \\ I_{n-k} \end{bmatrix}$ and then perform the inverse columns permutation to obtain H_C, the parity check matrix of C.

Since C is a linear code, then we can construct the corresponding dual code $C^\perp$. Clearly, $C^\perp$ is also linear, so we can find a generator matrix and a parity check matrix for $C^\perp$, which are denoted as $G_{C^\perp}$ and $H_{C^\perp}$, respectively. The relationship between $G_{C^\perp}$ and $H_{C^\perp}$ is also shown in diagram 4.1.

As a result, the interrelationship between the matrices contribute to two new variants of McEliece schemes, which will be proposed in the next section.

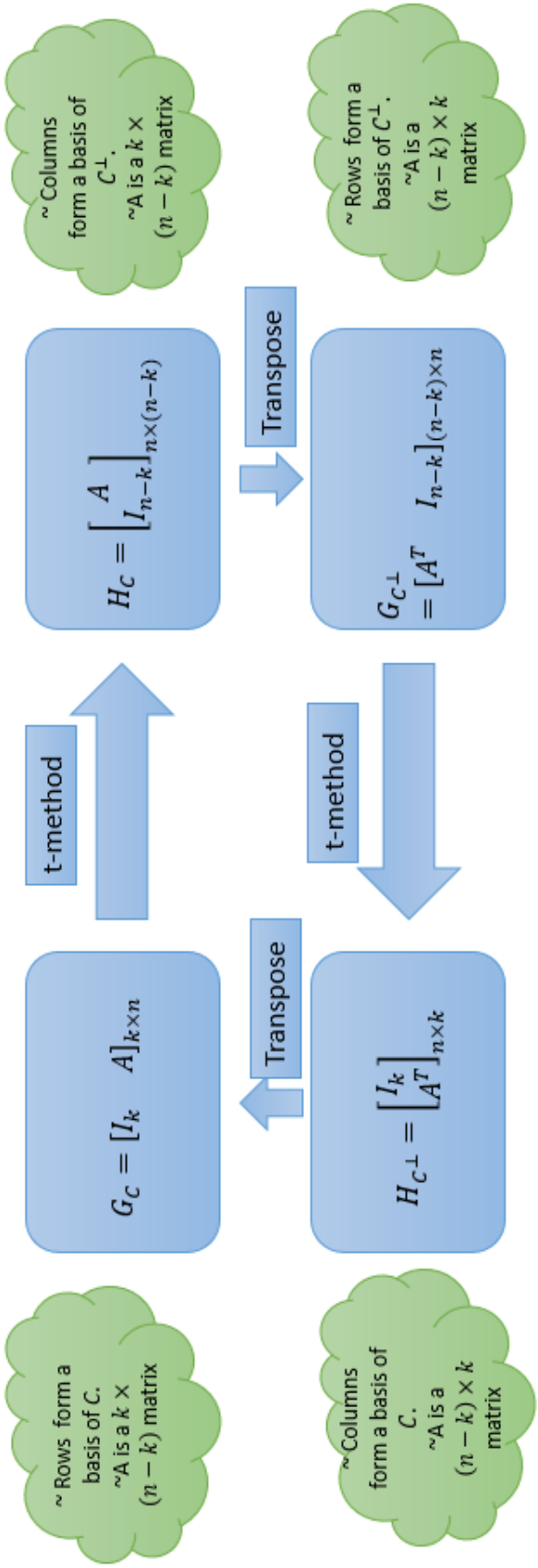

Figure 4.1: Motivation to New Variant.

4-2 "The" Variants

Algorithm 7 Variant of McEliece Encryption Scheme Using Linear Code C

1: **System Parameters:** Integer n, t with $t \leq n$. A family $\mathbf{A}$ of $2-$ary $[n, k, 2t+1]-$ linear codes.

2: Suppose G_C=[I_k A] is the generator matrix of $C \in \mathbf{A}$ written in standard form and $H_C = \begin{bmatrix} A \\ I_{n-k} \end{bmatrix}$ is the corresponding parity check matrix of C.

3: **Key Generation:**

G_C: $k \times n$ generator matrix written in standard form for a randomly chosen C from $\mathbf{A}$,

H_C: The $n \times (n-k)$ parity check matrix of C such that $G_C H_C = [0]$, where $[0]$ is the $k \times (n-k)$ zero matrix,

D_C: The syndrome decoding algorithm of C.

(I) Public key : G_C.

(II) Private key :H_C.

(III) Plaintext space: $\mathbf{F}_2^k$

(IV) Ciphertext space: $\mathbf{F}_2^n$

4: **Encryption:** $Enc : \mathbf{F}_2^k \rightarrow \mathbf{F}_2^n$. $Enc(m) = mG_c + e$, where e is the randomly chosen vector of length n with $wt(e) \leq t$.

5: **Decryption:** $Dec : \mathbf{F}_2^n \rightarrow \mathbf{F}_2^k$. $Dec(v) =$ the first k position of $D_c(vH_c)$.

The correctness of the scheme is shown as follows:

For any plaintext $m \in \mathbf{F}_2^k$,

$Dec(Enc(m)) = Dec(mG_C + e)$ = the first k position of $D_C((mG_C + e)H_C)$.

Next, we compute

$$\begin{aligned}(mG_C + e)H_C &= mG_CH_C + eH_C \\ &= m[0] + eH_C \\ &= \underline{0} + eH_C \\ &= eH_C\end{aligned}$$

$\therefore Dec(Enc(m)) =$ the first k position of $D_C(eH_C)$.

Since $wt(e) = t$ and $d(C) = d \leq 2t + 1$, then the error pattern e is correctable, and so

$$\begin{aligned}D_C(eH_C) &= mG_C \\ &= m[I_kA] \\ &= (m, mA).\end{aligned}$$

Therefore, $Dec(Enc(m)) =$ the first k position of $(m, mA) = m$.

Algorithm 8 Variant of McEliece Encryption Scheme Using Linear Dual Code $C^{\perp}$

1: **System Parameters:** Integer n, s with $s \leq n$. A family $\mathbf{B}$ of $2-$ary $[n, n-k, 2s+1]-$ linear codes.

2: Suppose $G_{C^{\perp}}$ =$[B^T\ I_{n-k}]$ is the generator matrix of $C^{\perp} \in \mathbf{B}$ written in standard form and $H_{c^{\perp}} = \begin{bmatrix} I_k \\ B \end{bmatrix}$ is the corresponding parity check matrix of $C^{\perp}$.

3: **Key Generation:** $G_{C^{\perp}}$: $(n-k) \times n$ generator matrix written in standard form for a randomly chosen $C^{\perp}$ from $\mathbf{B}$,

$H_{C^{\perp}}$: The $n \times k$ parity check matrix of $C^{\perp}$ such that $G_{C^{\perp}} H_{C^{\perp}} = [0]$, where $[0]$ is the $(n-k) \times k$ zero matrix,

$D_{C^{\perp}}$: The syndrome decoding algorithm of $C^{\perp}$.

(I) Public key : $G_{C^{\perp}}$ =$[B^T\ I_{n-k}]$.

(II) Private key : $H_{C^{\perp}} = \begin{bmatrix} I_k \\ B \end{bmatrix}$.

(III) Plaintext space: $\mathbf{F}_2^{n-k}$

(IV) ciphertext space: $\mathbf{F}_2^n$

4: **Encryption:** $Enc : \mathbf{F}_2^{n-k} \to \mathbf{F}_2^n$. $Enc(m) = mG_{C^{\perp}} + e$.

5: **Decryption:** $Dec : \mathbf{F}_2^n \to \mathbf{F}_2^{n-k}$. $Dec(v)$ = the last $n-k$ position of $D_{C^{\perp}}(vH_{C^{\perp}})$.

The correctness of the scheme is shown as follows:

For any plaintext $m \in \mathbf{F}_2^{n-k}$,

$Dec(Enc(m)) = Dec(mG_{C^{\perp}} + e)$ = the last $n-k$ position of $D_C((mG_{C^{\perp}} + e)H_{C^{\perp}})$.

Next, we compute

$$\begin{aligned}(mG_{C^{\perp}} + e)H_{C^{\perp}} &= mG_{C^{\perp}}H_{C^{\perp}} + eH_{C^{\perp}} \\ &= m[0] + eH_{C^{\perp}} \\ &= \underline{0} + eH_{C^{\perp}} \\ &= eH_{C^{\perp}}\end{aligned}$$

$\therefore Dec(Enc(m))$ = the last $n-k$ position of $D_C(eH_{C^{\perp}})$.

Since $wt(e) = s$ and $d(C) = d \leq 2s + 1$, then the error pattern e is correctable, and so

$$\begin{aligned} D_C(eH_{C^\perp}) &= mG_{C^\perp} \\ &= m[B^T I_{n-k}] \\ &= (mB^T, m). \end{aligned}$$

Therefore, $Dec(Enc(m))$ = the last $n - k$ position of $(mB^T, m) = m$.

4-3 Small Game Example

Variant of McEliece Encryption scheme using C

Given $G_C = \begin{bmatrix} 1000111 \\ 0100110 \\ 0010011 \\ 0001101 \end{bmatrix}$ as the public key. We can construct the private key

$$H_C = \begin{bmatrix} 111 \\ 110 \\ 011 \\ 101 \\ 100 \\ 010 \\ 001 \end{bmatrix}.$$

$$Enc : \mathbf{F}_2^4 \to \mathbf{F}_2^7.$$

With plaintext $m = 0101$, we have $v = mG_C + e$ where we let $e = 0001000$.

As a result, we get $v = 0101011 + 0001000 = 0100011$.

To decrypt v, we get $vH_c = 101$ that indicate the 4th position in H_C. Therefore, we can detect and correct the error, $e = 0001000$ and get back the plaintext $m = 0101$.

Next, we may look into another related example using the variant of $C^\perp$.

Variant of McEliece Encryption scheme using $C^{\perp}$

From the generator matrix and parity check matrix of C, we can generate generator matrix and parity check matrix for $C^{\perp}$, denoted by $G_{C^{\perp}}$ and $H_{C^{\perp}}$ respectively.

$$G_{C^{\perp}} = \begin{bmatrix} 1101100 \\ 1110010 \\ 1011001 \end{bmatrix} \text{ as the public key and}$$

$$H_{C^{\perp}} = \begin{bmatrix} 1000 \\ 0100 \\ 0010 \\ 0001 \\ 1101 \\ 1110 \\ 1011 \end{bmatrix} \text{ as the private key.}$$

With plaintext $m = 101$, we have $v = mG_{C^{\perp}} + f$ where we let $f = 1000000$.

we get

$$\begin{aligned} v &= 0110101 + 1000000 \\ &= 1110101. \end{aligned}$$

To decrypt v, we get $vH_{C^{\perp}} = 100$ that indicate the first position in $H_{C^{\perp}}$. Therefore, we can detect and correct the error, $f = 1000000$ and get back the plaintext $m = 101$.

4-4 Efficiency Analysis

Next, we perform an efficiency analysis.

Public key size: Since G_C=$[I_k \ A]_{k\times n}$, then A is of size $k \times (n-k)$. The key size only depends on A since we assume G_C is in standard form. Hence, the public key size is $k(n-k)$ bits.

Private key size: Since $H_C = \begin{bmatrix} A \\ I_{n-k} \end{bmatrix}$ and the size of A is also $k \times (n-k)$, then similar to public key size, the private key size is also $k(n-k)$ bits.

Encryption Time : Suppose we let w be the time needed to generate the error pattern e and t be the time needed to compute the multiplication of two elements from $\mathbf{F}_2$. Then, the total number of time needed to perform an encryption is $knt + w$.

Decryption Time : Suppose we let d be the time needed for the efficient decoder D_C to perform a single decoding algorithm. Then, the total number of time needed to perform a decryption is $dn(n-k)t$, where $n(n-k)t$ is the time needed to compute vH_C.

4-5 Security Analysis

Our scheme depends on the following two hardness problems which are stated in computational form.

Algorithm 9 Codeword Finding Problem (CFP)

Parameters : The set of all possible generator matrices of C and an integer $w > 0$.

Instance: A matrix G_C.

Problem : Find a codeword v of weight $\leq w$ in C with generator matrix G_C.

Algorithm 10 Syndrome Decoding Problem (SDP)

Parameters : The set of all possible generator matrices of C and an integer $t > 0$.

Instance: A matrix H_C and $s \in \mathbf{F}_2^{n-k}$.

Problem : Find a vector $e \in \mathbf{F}_2^n$ of $wt(e) \leq t$ such that $eH_C = s$.

Theorem 4-5.1. *CFP is polynomially equivalent to SDP (CFP=SDP).*

Proof Suppose we can solve CFP, that is, given a generator matrix G_C, we can find codeword v of weight less than or equal to w in C. We note that $v = mG_C$, where $m \in \mathbf{F}_2^k$ and G_C = $[I_k\ A]$. Therefore, $v = m[I_k\ A] = (m, mA)$, where mA is a vector of length $n-k$.

Next, let v' be another codeword in C. By using v and v', we construct an error pattern e of weight t which is a solution to SDP. Since $d(C) = 2t+1$, we may assume $d(v, v') = 2t+1$. Without loss of generality, we may assume that $w \leq 2t+1$. If v, v' do not satisfied the distance formula, we simply replace with another pair of codewords.

Let $v = a_1a_2...a_n$ and $v' = b_1b_2....b_n$. and define $S = \{i : a_i \neq b_i\}$. Note that $|S| = 2t + 1$.Furthermore, $S = A \cup B, A \cap B = \phi$ with $|A| = t$ and $|B| = t + 1$. We define $e = \begin{cases} 0, & i \notin S \\ b_i - a_i, & i \in A \\ 0, & i \in B \end{cases}$.

Clearly, $wt(e) = t$. Since, $d(C) \leq 2t + 1$, then C can correct the error pattern e and so $eH_C = s \neq \underline{0}$.

Next, suppose we can solve SDP, that is, given a parity check matrix H_C and $s \in \mathbf{F}_2^{n-k}$, we can find $e \in \mathbf{F}_2^n$ with $wt(e) \leq t$ such that $eH_C = s$. We perform the following computations:

$$\begin{aligned} e = eH_C &= \underline{0} + eH_C \\ &= m[0] + eH_C, \text{where } m \in \mathbf{F}_2^k \\ &= mG_CH_C + eH_C \\ &= (mG_C + e)H_C. \end{aligned}$$

If $mG_C + e = \underline{0}$, then $\underline{0}H_C = \underline{0}$, so $mG_C + e = \underline{0} \in C$ with weight =0. Hence, we are done.

Thus, we may assume $mG_C + e = w'$, $w' \neq \underline{0}$. So $w' = mG_C + e$ is the received word with error pattern e and $v = mG_C \in C$ such that $w' = v + e$ which is equivalent to $v = w' + e$.

Let $wt(w') = r \in \mathbf{Z}^+$. Knowing that $w(e) \leq t$, then $wt(w' * e) \leq r - t$. Hence, we see that

$$\begin{aligned} wt(v) &= wt(w' + e) \\ &= wt(w') + wt(e) - 2wt(w' + e) \\ &\leq r + t - 2(r - t) \\ &= 3t - r \end{aligned}$$

Without loss of generality, we may take $w = 3t - r$ which is a solution to CFP.

Q.E.D.

Theorem 4-5.2. *Breaking the variant of McEliece encryption for$C$ is as hard as solving CFP.*

Proof Let π be our variant of McEliece encryption, K_π be the keyspace of π and K_{Pb} be the set of all possible public keys (generator matrices of C).

Let E be the set of all error patterns $e \in F_2^n$ with $wt(e) \leq t$ (which are all error patterns that can be corrected by C) and $\omega = \{(G, e) : G_C \in K_\pi, e \in E\}$ equipped with a uniform distribution.

Define

1. A (T, ϵ)- distinguisher D of K_{Pb} if it runs in time at most T and the advantage of D for K_π is

$$Adv(D, K_{Pb}) = |{}^{Pr}_{\omega}[D(G) = 1|G \in K_{Pb}] - {}^{pr}_{\omega}[D(G_C) = 1]| \geq \epsilon.$$

2. A (T, ϵ) -decoder ϕ for (K_{Pb}, t) if it runs in time at most T and its Success probability $Succ(\phi) = {}^{Pr}_{\omega}[\phi(H, eH) = e] \geq \epsilon$.

3. A (T, ϵ) - adversary A against π if it runs in time at most T, it's Success probability is $Succ(A, K_{Pb}) = {}^{Pr}_{\omega}[A(H, eH) = e|H \in K_\pi \setminus K_{Pb}] \geq \epsilon$.

Suppose A is a (T, ϵ)- adversary against π. We define a distinguisher D which input generator matrix $G \in K_\pi$. Hence, by using the t-method to obtain the corresponding private key $H \in K_\pi \setminus K_{Pb}$. Then D pick up $e \in E$ randomly and uniformly. Next, D check whether $A(H, eH) = e$. If yes return 1, else return 0.

Thus, we have

$$\begin{aligned} {}^{Pr}_{\omega}[D(G) = 1] &= {}^{Pr}_{\omega}[A(H, eH) = e] \\ &= Succ(A) \end{aligned}$$

and

$$\begin{aligned} {}^{Pr}_{\omega}[D(G) = 1|G \in K_{Pb}] &= {}^{Pr}_{\omega}[A(H, eH) = e|H \in K_\pi \setminus K_{Pb}] \\ &= Succ(A, K_{Pb}). \end{aligned}$$

Since

$$Adv(D, K_{Pb}) = |^{Pr}_{\omega}[D(G) = 1|G \in KPb] -^{Pr}_{\omega} [D(G_C) = 1]|$$
$$= |Succ(A, K_{Pb}) - Succ(A)|$$

and so,

$$Adv(D, K_{Pb}) > -(Succ(A, K_{Pb}) - Succ(A)) = Succ(A) - Succ(A, K_{Pb})$$

$$\therefore Adv(D, K_{Pb}) + Succ(A, K_{Pb}) \geq Succ(A).$$

Since A is a (T, ϵ)- adversary against π, then $Succ(A, K_{Pb}) \geq \epsilon$. Thus we have either $Adv(D, K_{Pb}) or Succ(A) \geq \frac{\epsilon}{2}$.

The running time of D is equal to the running time of A increased by the cost for picking e and compute eH, which cannot exceed $O(n^2)$. So either A is a $(T, \frac{\epsilon}{2})$-decoder for (K_{Pb}, t) or D is a $(T + O(n^2), \frac{\epsilon}{2})$ -distinguisher for K_{Pb}. Note that a $(T, \frac{\epsilon}{2})$- decoder for (K_{Pb}, t) is a solution to SDP.

Q.E.D.

Next, we look at a particular type of attack to our scheme, namely, the key distinguishing attack, that is, by producing one word of weight w in the dual code $C^{\perp}$ is enough to distinguish a public key from a random matrix.

Lemma 4-5.3. *Suppose there is a $v' \in C^{\perp}$ with $wt(v') = w$, then we can find a solution to CFP.*

Proof Without loss of generality, we may assume that v' is in the 1^{st} row of the generator matrix of $C^{\perp}$ (We may assume also that $G_{C^{\perp}}$ is in the standard form). Then by applying the t-method to $G_{C^{\perp}}$ to obtain $H_{C^{\perp}}$, the parity check matrix of $C^{\perp}$ with v'' is the 1^{st} column of $H_{C^{\perp}}$. Since $v' \cdot v'' = 0$, then $v'' \in (C^{\perp})^{\perp} = C$ and $wt(v') = wt(v'') = w$. Hence, v'' is a solution to CFP. In general, $(v' \neq v'')$

Q.E.D.

Theorem 4-5.4. *If there is a (T, ϵ)-adversary which is a (T, ϵ)-distinguisher for π, then we can find a solution for CFP in $C^{\perp}$.*

Proof Define the second decoder:

1. A (T', ϵ')- distinguisher D of K'_{Pb} if it runs in time at most T' and the advantage of D for $K'_{\pi'}$ is

$$Adv(D, K'_{Pb}) = |\overset{Pr}{\omega}[D(G) = 1 | G \in K'_{Pb}] - \overset{pr}{\omega}[D(G_{C^{\perp}}) = 1]| \geq \epsilon.$$

2. A (T', ϵ') -decoder ϕ' for (K'_{Pb}, t) if it runs in time at most T' and its Success probability $Succ(\phi') = \overset{Pr}{\omega}[\phi'(H', fH) = f] \geq \epsilon'$.

3. A (T', ϵ') - adversary A against π' if it runs in time at most T', it's Success probability is $Succ(A, K'_{Pb}) = \overset{Pr}{\omega}[A(H', fH') = f | H' \in K'_{\pi'} \setminus K'_{Pb}] \geq \epsilon'$.

By using the proof in Theorem 4-5.2, we can find an adversary A' which is a $(T', \frac{\epsilon'}{2})-$decoder for $(K_{Pb'}, s)$, and hence produce a solution for CFP in $C^{\perp}$.

Q.E.D.

Theorem 4-5.5. *Breaking the variant of McEliece encryption for $C^{\perp}$ is as hard as solving CFP in $C^{\perp}$.*

Proof Directly follow from Theorem 4-5.4.

Q.E.D.

CHAPTER 5: VARIANTS OF NIEDERREITER ENCRYPTION SCHEME

5-1 The Scheme

Niederreiter encryption scheme has the similar concept and structure as the McEliece scheme. Both schemes are constructed based on the decoding hardness problem. Unlike McEliece scheme that relies on the generator matrix, Niederreiter encryption scheme is considered to be the variant of McEliece scheme that focuses more on the parity-check matrix H.

5-2 Variant of Niederreiter Scheme

Setup : Randomly choose a $[n, k, d]$-binary linear code C with a $(n-k) \times n$ parity check matrix H and $d = 2t + 1$ for $t \in \mathbf{Z}^+$.

Randomly pick a permutation matrix P of size $n \times n$, a $l \times n$ binary matrix G' and a $l \times (n-k)$ binary matrix F with $F = G'P^{-1}H^T$.

Key Generation :

Public key : H, P

Private key : G', F

Encryption : $m \in \mathbf{F}^l$, $e \in \mathbf{F}^{(n-k)}$

$Enc : \mathbf{F}^l \to \mathbf{F}^n \times \mathbf{F}^{n-k}$

$Enc(m) = (c_1 = mG' + e, c_2 = mF)$

The encryption is a product construction using the variant of McEliece encryption schme using C and the original Niederreiter scheme.

Decryption : $Dec : \mathbf{F}^n \times \mathbf{F}^{n-k} \to \mathbf{F}^l$

Step 1 : Compute $s' = c_1 P^{-1} H^T - c_2$

Step 2 : Use an efficient decoding algorithm for C, $\Phi_H(s') = eP^{-1}$

Step 3 : Compute $\Phi_H(s')P = e$

Step 4 : Solve $mG' = c_1 - e$ to obtain m.

The correctness of the scheme is shown as follows:

For any plaintext $m \in \mathbf{F}^l$, $e \in \mathbf{F}^{(n-k)}$,

$$Dec(Enc(m)) = Dec((c_1 = mG' + e, c_2 = mF))$$

Next, we compute

$$\begin{aligned} s' &= c_1 P^{-1} H^T - c_2 \\ \Phi(s') &= eP^{-1} \\ \Phi(s')P &= e \end{aligned}$$

By obtaining e, we can now solve $mG' = c_1 - e$ to obtain back m.

$\therefore Dec(Enc(m)) = m.$

5-2-1 Efficient Analysis

Next, we perform an efficiency analysis on variant of Niederreiter Encryption scheme.

1. Public key size: Since $H_C = \begin{bmatrix} A \\ I_{n-k} \end{bmatrix}_{n \times n-k}$, then A is of size $n \times n$. Furthermore, the permutation matrix P is also a $n \times n$ matrix. Therefore, the key size only depends on A since we assume H_C in standard form and also P. Hence, the public key size is $n \times n$ bits.

2. Private key size: As $G_C = [I_k A]$ and the size of A is exactly $n \times n$. Since F is a $l \times (n-k)$ matrix, the private key size is $l \times n$ bits.

3. Suppose we let w be the time needed to generate the error pattern e and t be the time needed to compute the multiplication of two elements from $\mathbf{F}_2$. Then, the total number of time needed to perform an encryption is $lnt + w$.

4. Suppose we let d be the time needed for the efficient decoder Φ_H to perform a single decoding algorithm. Then, the total number of time needed to perform a decryption is $dlnt$, where lnt is the time needed to compute $[\Phi_H(s')]P$.

5-3 Security Analysis

Since the public code generated by G' is not related to the secret code generated by H, attacking G' does not in any way expose the private code.

Thus, finding a minimal weight codeword in the dual of G' is useless (Lemma 4-5.3) since G' is randomly chosen. An attacker may attack the public matrix F and try to obtain H. Since G' is publicly known, then the attacker can use some well-known decomposition method to obtain $F = G'H_0$, where H_0 is a $n \times (n-k)$ matrix. Thus, we have $2^{n(n-k)}$ possible solutions H_0.

For the message attack, if G' has a smaller error-correction capability compared to H, then decoding using G will fail. Also, since G' is randomly generated, this is the SDP problem. For this proposed scheme, good parameters can be selected in order to increase the decoding failure probability of using G' while keeping a low decoding failure probability using H.

An attacker may also use c_2 to recover $m$by solving $c_2 = mF$. However, $c_2 = mF$ resembles the ciphertext from the Niederreiter encryption scheme, notice that there is no restriction on the weight of the message m, so attacking the Niederreiter encryption scheme does not threaten the security of our proposed scheme.

By using Theorem 4-5.2 and Lemma 4-5.3, we have the following results.

Theorem 5-3.1. *Breaking the variants of Niederreiter encryption scheme is as hard as breaking*

1. *the variant of McEliece encryption for C and*

2. *the Niederreiter encryption scheme.*

5-4 Comparison to McEliece Encryption Scheme and Niederreiter Encryption Scheme

	McEliece Encryption Scheme	**Niederreiter Encryption Scheme**	**Variants of McEliece**	**Variant of Niederreiter**
Hardness Problem	Syndrome Decoding Problem & Codewords Finding Problem	Syndrome Decoding Problem & Codewords Finding Problem	Syndrome Decoding Problem & Codewords Finding Problem	Syndrome Decoding Problem & Codewords Finding Problem
Public Key	G	H	$G_C \& G_{C^\perp}$	G', F
Private Key	H	G	$H_C \& H_{C6\perp}$	H, P
Original Code	Binary Goppa code	Binary Goppa code	Binary Linear Code	Binary Linear Code
Complexity	Complexity of decryption grow exponentionally with increasing key size.	Complexity of decryption grow exponentionally with increasing key size.	Complexity of decryption grow exponentionally with increasing key size.	Complexity of decryption grow exponentionally with increasing key size.
Security	Secure over all classical attack	Special case is broken such as the Reed Solomon code.	Proof of security through reduction	Proof of security through reduction
Security Parameters	(n, k, d)	$(n, n-k, d)$	(n, k, d)	(n, k, d)
Suggested Parameters	$n = 1024, k \geq 644, d = 38$ [11]	$n = 255, k = 133, d = 4$[22]	-	-

Table 5.1: McEliece vs Niederreiter

CHAPTER 6: CONCLUSION

As overall, this project achieves

1. Comparison of McEliece encryption scheme and Niederreiter encryption scheme.
2. Variant of McEliece encryption scheme is constructed using linear code C.
3. Second variant of McEliece encryption scheme is constructed using the dual code $C^{\perp}$.
4. Variant of Niederreiter encryption scheme is constructed.

We start by investigating the properties of McEliece and Niederreiter encryption scheme and found out that both scheme are highly similar and related to one another with the relationship between the linear code C and its dual code $C^{\perp}$.

Next, based on the motivations on the relationship of C and its dual, we constructed 2 variants of McEliece encryption scheme. The biggest advantage of both the variants is the non-permutation equivalent properties. The absence of the permutation matrix in the variants gives to the advantage where no permutation equivalency between the public key and the secret code. As a result, the adversary can no longer exploit the equivalency and recover the secret code. Thus, the security of the scheme is highly enhanced. Not only that, with a smaller public generator matrix without the permutation matrix, this further reduce the size of the public key in the encryption scheme.

Furthermore, we constructed the variant of Niederreiter encryption scheme by product construction using the variant of McEliece encryption scheme using C and the original Niederreiter encryption scheme. By doing that, not only the security of the scheme can be highly enhanced, we can also remain the original Neiderreiter scheme as part of the new variant. All the security of the variants of McEliece and Neiderreiter encryption scheme is proven by reduction.

Due to time constraint, this project only proves the security of the variants by reduction. For future study, to improve the security, some CPA and CCA secured variant

can be constructed based on the relationship of the C and its dual. Next, the parameters of some famous codes are also provided in the project so that future investigation can be carried out to study the most suitable codes to be use in the particular encryption scheme.

REFERENCES

[1] A. Barg (1998), *Complexity issues in coding theory.* In V.S. Pless and W.C. Huffman, editors, Handbook of Coding theory, volume I, chapter 7, 649–754. North-Holland.

[2] E. Berlekamp, R. McEliece, and H. van Tilborg (1978) *on the inherent intractability of certain coding problems.* IEEE Transactions on Information Theory, 24(3):384–386.

[3] Daniel J. Bernstein, Lange, Tanja; Peters, Christiane (2008). *Attacking and defending the McEliece cryptosystem.* Proc. 2nd International Workshop on post quantum Cryptography. Lecture Notes In Computer Science. 5299: 31–46. doi:10.1007/978-3-540-88403-3-3.

[4] Daniel J. Bernstein, Johannes Buchmann, Erik Dahmen (2009). *post quantum Cryptography* ,Springer-Verlag Berlin Heidelberg.

[5] F. Chabaud (1995), *On the security of some cryptosystems based on error-correcting codes,* Advances in Cryptology – EU – Springer - Verlag, 131 – 139.

[6] Y.X. Li, R.H. Deng, X.M Wang,(1994), *On the Equivalence of McEliece's and Niederreiter's Public-Key Cryptosystems*IEEE Transactions on Information Theory 40, 271–273.

[7] N.T. Courtois, M. Finiasz, and N. Sendrier, (2001, December). How to achieve a McEliece-based digital signature scheme. In *International Conference on the Theory and Application of Cryptology and Information Security* (157-174). Springer Berlin Heidelberg.

[8] V.D.Goppa (1971), *Rational representation of codes and (L, g) codes*,Probl. Peredach. Inform. 7(3), 41-49.

[9] K. Kobara, and H. Imai (2001), *Semantically secure McEliece public-key cryptosystems- conversions for McEliece PKC.* In Practice and Theory in Public Key Cryptography - PKC '01 Proceedings. Springer Verlag.

[10] R. McEliece, (2002). *The theory of information and coding*. Cambridge University Press.

[11] Robert J. McEliece (1978). *A Public-Key Cryptosystem Based On Algebraic Coding Theory*(PDF). DSN Progress Report. 44: 114–116. Bibcode: 1978DSNPR..44..114M.

[12] H. Niederreiter (1986). *Knapsack-type cryptosystems and algebraic coding theory.* Problems of Control and Information Theory. Problemy Upravlenija i Teorii Informacii. 15: 159–166.

[13] Rafael Dowsley et al. (2009) *A CCA2 Secure Public Key Encryption Scheme Based On The McEliece Assumptions In The Standard Model* 240-251. Springer.

[14] David Mandell Freeman et al. (2010) *More construction of lossy and correlation secure trapdoor functions* Public Key Cryptography, volume 6056 of Lecture Notes in Computer Science, pages 279-295. Springer.

[15] E. Persichetti, (2012) *Improving the efficiency of code based cryptography*Doctoral dissertation , Department of Mathematics, University of Auckland.

[16] Peter W. Shor.(1994) *Polynomial Time Algorithms for Discrete Logarithms and Factoring On a Quantum Computer* In Leornard M. Adleman and Ming-Deh A. Huang, editors, ANTS, volume 877 of Lecture Notes in Computer Science, page 289. Springer.

[17] San Ling, Chaoping Xing (2004). *Coding Theory:A First Course.* Cambridge University Press.

[18] Alon Rosen & Gil Segev (2009) *Chosen-ciphertext security via correlated products*. In Omer Reingold, editor, TCC, volume 5444 of Lecture Notes in Computer Science, pages 419-436. Springer.

[19] N. Sendrier (2002), *On the security of the McEliece public-key cryptosystem.* In M. Blaum, P. Farrell, and H. van Tilborg, editors, Proceedings of Workshop honouring Prof. Bob McEliece on his 60th birthday, pages 141–163. Kluwer.

[20] J.H. Van Lint, (2012). *Introduction to Coding Theory* (Vol. 86). Springer Science & Business Media.

[21] C.E. Shannon(1948) *A Mathematical Theory of Communication* Bell Systems Technical Journal, 27, 379-423, 623-656.

[22] Rongxing Lu, et al. (2010) *An efficient and provably secure public key encryption scheme based on coding theory* Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.274

[23] Jonathan Katz, et.al. (2015) *Introduction To Modern Cryptography* $2^n d$ ed, CRC Press.

[24] Wade Trappe, Lawrence Washington (2006). *Introduction to Cryptography with coding theory,*Pearson Education,Inc.

[25] Douglas R.Stinson (2006). *Cryptography Theory and Practice*, Chapman & Hall/CRC.