Application of Polar Code-Based Scheme in Cloud Secure Storage

Zhe Li, Yiliang Han(✉), and Yu Li

College of Cryptographic Engineering, Engineering University of PAP, Xi'an 710086, Shaanxi, China
hanyil@163.com

Abstract. In view of the fact that the quantum computer attack is not considered in the cloud storage environment, this paper selects the code-based public key encryption scheme as the security protection measure in the cloud storage. Based on random linear code encryption scheme, it employs the structure of the RLCE scheme and Polar code polarization properties, using the Polar code as underlying encoding scheme, through the method of RLCEspad, putting forward a kind of improved public key encryption scheme which considers semantic security and is resistant to adaptively chosen ciphertext attacks. The improved scheme is applied to cloud storage to ensure that the storage environment will not be attacked by quantum computer while ensuring the confidentiality, availability and reliability.

Keywords: Post-quantum cryptography · Polar code · Random linear code encryption · Code-based scheme · Cloud storage

1 Introduction

At the beginning of the New Year in 2020, COVID-19 outbreak broke out in Wuhan. China is experiencing a major public health emergency with the fastest transmission rate, the widest infection range and the greatest difficulty in prevention and control since the founding of new China. While the whole country is doing its best to fight the epidemic, foreign hostile forces and domestic criminals have not stopped cyber-attacks and sabotage. According to relevant reports, some overseas hacker groups have taken advantage of the situation to carry out network penetration attacks on key areas of China with pneumonia and other relevant information [1]. We will integrate cyber security with political security, military security, economic security and biological security, systematically plan the development of a cybersecurity risk prevention and control and governance system, and comprehensively improve the country's ability to comprehensively manage cyber society. Cryptography will play a crucial role.

Cloud storage system should ensure availability (for, any legal customers can access the uploaded data from some networking equipment), reliability (outsource to the cloud user data) and efficient retrieval (outsource to the cloud user data), data sharing (between authorized users), security (including confidentiality and integrity), and other functional requirements/regulations [2].

J. Zeng et al. (Eds.): ICPCSEE 2020, CCIS 1257, pp. 676–689, 2020.
https://doi.org/10.1007/978-981-15-7981-3_50

The rapid development of quantum computer technology has a profound impact on the present and future society. At present, 5G communication, transportation, cloud storage, public opinion control and other fields are closely related to cryptography. Key cryptography techniques used in various fields are still relatively safe until the arrival of quantum computers. The proposed Shor algorithm and Grover [3] algorithm pose severe challenges to the security of classical cryptography schemes.

In order to deal with the threat of quantum computer to cryptography schemes based on classical mathematical difficulties, countries all over the world are seeking new cryptography schemes that can resist the attack of quantum computer, namely Post Quantum Cryptography [4]. The National Institute of Standards and Technology (NIST) of the United States launched the standardization project of post quantum algorithm in 2012, and the post quantum cryptography was recruited globally in 2016, and the second round of selection was completed in 2019 [5]. In the time when countries formulate post quantum cryptography algorithm standards, China has also promoted post quantum cryptography design competition and formulated post quantum cryptography standards. So far, post quantum cryptography standardization has entered the second round of selection [6]. Post quantum cryptography schemes mainly include 5 kinds [7], each of which has its unique application range. At present, code- based scheme and lattice - based scheme [8] have become the focus of research. Compared with other post quantum cryptography schemes, the code-based cryptography scheme is more suitable for the construction of encryption schemes. Among the post quantum cryptography, the code-based encryption scheme has a good research prospect.

The code-based cryptography scheme can be traced back to the McEliece scheme in 1978 [9]. Up to now, the scheme is still safe under the appropriate parameter selection, and it has entered the second round of PQC collection algorithm. Because the security of the Goppa code-based McEliece scheme is based on the decoding difficulty of general linear code, it can be specified to the NPC problem and has the characteristics of resisting the attack of quantum computer, many experts began to focus on this scheme. In 1986, Niederreiter [10] constructed the Niederreiter scheme using the dual form of the McEliece scheme. The security of the two schemes is equivalent. Both the McEliece scheme and Niederreiter scheme have the shortcoming of too large key size, which is difficult to be applied in the actual scenario. In view of the characteristics of poor utility of McEliece schemes, experts used other codes with more compact structures as the underlying codes to reduce the key size, and the constructed deformation schemes had security problems and were vulnerable to structural attacks [11].

In 2016, Wang et al. [12] constructed a random linear code encryption scheme, namely RLCE public key encryption scheme, by inserting random columns in each column of the generated matrix, which was selected by the first round of PQC collection algorithm. Unlike other cryptography schemes that make use of compact codes, this scheme does not depend on the structure of any underlying code. The RLCE scheme achieves randomness by inserting random columns, and the security of the scheme depends on the NPC problem of linear random code decoding, which avoids

the structural attack introduced by the underlying coding structure. In 2017, Wang [13] further studied the padding method of RLCE scheme, aiming at the potential attack of RLCE scheme. The improved RLCE scheme reduced the key size, improved encryption and decryption performance, and the IND-CCA2 (adaptive selective ciphertext attack) security of the padding scheme was achieved. Matthews [14] at the 2019 CBC (code-based Cryptography) conference, it was proposed to use Hermitian code as the underlying code of RLCE scheme. At the 2019 A2C (Algebra, Codes and Cryptology) conference, Liu et al. [15] proposed to use Polar code as the underlying code of RLCE scheme. By taking advantage of the polarization nature of Polar code, the encryption and decryption complexity was reduced. The scheme proposed by Liu in literature [15] does not consider semantic security and is vulnerable to IND-CCA2.

Based on RLCE scheme and PolarRLCE scheme, this paper proposes an IND-CCA2 secure RLCE Public key encryption scheme based on Polar code, which is based on Polar code polarization property and improves key storage method. And the improved scheme will be applied to the cloud storage.

2 Basic Knowledge

2.1 Relevant Definitions

Definition 1. A Binary Input Discrete Memoryless Channel [16].
It can be expressed as W: $X \to Y$, X is the set of input symbols, Y is the set of output symbols, and the transition probability is $W(Y|X), x \in X, y \in Y$. For channel W, the channel after N times polarization can be expressed as W^N, then the transition probability of channel W^N is: $X^N \to Y^N$

$$W^N\left(y_1^N|x_1^N\right) = \prod_{i=1}^{N} W(y|x) \tag{1}$$

For a B-DMC, there are two important channel capacity parameters:
Symmetric Capacity:

$$I(W) \triangleq \sum_{y\in Y}\sum_{x\in X}\frac{1}{2}W(y|x)\log\frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)} \tag{2}$$

Bhattacharyya Parameter:

$$Z(W) \triangleq \sum_{y\in Y}\sqrt{W(y|0)W(y|1)} \tag{3}$$

I(W) is a measure of channel rate and *Z(W)* is a measure of channel reliability. Under the condition of equal probability input, the maximum rate of channel W in reliable transmission is *I(W)*. In the case that channel W only transmits 0 or 1, the upper limit of maximum likelihood judgment error probability is *Z(W)*. The value range of *I (W)* and *Z(W)* is both [0, 1]. *I(W)* and *Z(W)* satisfy the following relationship: if and only if $Z(W) \approx 0, I(W) \approx 1$; If and only if $Z(W) \approx 1, I(W) \approx 0$.

Definition 2. Channel Polarization.

Channel polarization is divided into two stages: Channel combination and Channel Splitting.

(1) Channel Combination

Combine N independent channels of B-DMC W and generate a vector channel W^N through recursion: $X^N \rightarrow Y^N$, where N is the power of 2, $N = 2^n, N \geq 0$. $u_1^N \rightarrow x_1^N$ is the mapping from the input of the complex channel $\mathrm{W_N}$ to the input of the original channel W^N. So, get $u_1^N = x_1^N \boldsymbol{G}_N . u_1^N$ is the original bit sequence, x_1^N is the encoded bit sequence, $\mathrm{G^N}$ is a N dimension generated matrix, code length is $N = 2^n$.

The transition probability of channel W_N and W^N has the following relationship:

$$W_N\left(y_1^N | u_1^N\right) = W^N\left(y_1^N | u_1^N \mathbf{G}_N\right), y_1^N \in Y^N, u_1^N \in X^N. \tag{4}$$

(2) Channel Splitting

The complex channel W^N formed by the combination of channels splits into N coordinate channels of binary input. $W_N^{(i)} : X \rightarrow Y^N \times X^{i-1}$, $1 \leq i \leq N$,

$$W_N^{(i)}\left(y_1^N, u_1^{i-1} | u_i\right) \triangleq \sum_{u_{i+1}^N \in X^{N-i}} \frac{1}{2^{N-1}} W_N\left(y_1^N | u_1^N\right) \tag{5}$$

Definition 3. Polarization Coding Principle.

The basic idea of polarization coding is to send data bits only on the coordinate channel $W_N^{(i)}$, where $Z\left(W_N^{(i)}\right)$ approaches 0 ($I\left(W_N^{(i)}\right)$ approaches 1).

2.2 RLCE Encryption Scheme

Table 1, 2 and 3.

Table 1. RLCE key generation

RLCE Key generation
Input: (n,k,d,t,w), $n,k,d,t>0$,$w\in\{1,2\cdots n\}$, $k+1\geq d\geq 2t+1$.
Output: G^{pub}, (S,G_1,P,A).
1. G: $k\times n$ order generator matrix of code C whose dimension on domain F is k.
2. Generate w random column vectors $r_1,r_2,\cdots,r_w$, and by inserting w random $k\times 1$ column vectors into generator matrix G, obtained the $k\times(n+w)$ matrix G_1, $G_1=(g_1,\cdots,g_{n-w},g_{n-w+1},r_1,\cdots g_n,r_w)$.
3. To mix the columns, choose w random non-singular binary 2×2 non-singular matrices $A_1,A_2,\cdots,A_W$.
4. Denote $A=[1,\cdots 1,A_1,A_2,\cdots,A_W]$ the $(n+w)\times(n+w)$ nonsingular invertible matrix.
5. Let S be a randomly chosen $k\times k$ non-singular matrix, P be the $(n+w)\times(n+w)$ permutation matrix.
6. Output $k\times(n+w)$ public key $G^{pub}=SG_1AP$, Private key (S,G_1,A,P).

Table 2. RLCE encryption

RLCE encryption
Input: Public key G^{pub}, message $m\in F_2^k$, error vector $e\in F_2^{n+w}$.
Output: Ciphertext $c\in F_2^{n+w}$.
1. $c=mG^{pub}\oplus e$, $w_H(e)\leq t$.

Table 3. RLCE decryption

RLCE decryption
Input: Ciphertext $c\in F_2^{n+w}$, Private key (S,G_1,A,P).
Output: Message $m\in F_2^k$, or decoding error identification $\perp$.
1. $cP^{-1}A^{-1}=mSG_1\oplus eP^{-1}A^{-1}=(c'_1,c'_2,\cdots c'_{n+w})$.
2. From length of $n+w$ vector $cP^{-1}A^{-1}$ delete w column vectors, get length of n $c'=(c'_1,c'_2,\cdots,c'_{n-w+1},c'_{n-w+3},c'_{n-w+5},\cdots,c'_{n+w-1})$.
3. $c'=mSG_1\oplus e'$, $e'\in F_2^n$, $w_H(e')\leq t$.
4. Using a decoding algorithm, to calculate $m'=mS$, $m=m'S^{-1}$.
5. Calculate the hamming weight $w=wt(c-mG_1)$, if $w\leq t$, output message m; else, output $\perp$.

2.3 Message Encoding

In order to resist the known attacks of code-based schemes, the security of the code-based schemes can be guaranteed by means of message encoding (message padding).In general, there are three ways to encode messages in ciphertext [13]:

(1) basicEncoding: Encode information within the vector $m\in GF_2^k$ and the ciphertext is $c=mG+e$.In this case, we can encode $mLen=mk$ bits information within each ciphertext.

(2) mediumEncoding: In addition to basicEncoding, further information is encoded in the non-zero. In this case, we can encode $mLen = m(k+t)$ bits information within each ciphertext.

(3) advancedEncoding: In addition to mediumEncoding, further information are encoded within the choice of non-zero entries within e. Since there are $\binom{n+w}{t}$ candidates for the choice of non-zero entries within e, we can encode $mLen = m(k+t) + \left\lfloor \log 2 \binom{n+w}{t} \right\rfloor$ bits information within each ciphertext.

Pointcheval padding [17]:

$$c = Enc(G, r_1, H_1(m\|r_2))\|(H_2(r_1) \oplus (m\|r_2) \tag{6}$$

Fujisak-Okamato padding [18]:

$$c = Enc(G, r_1, H_1(m\|r_1))\|(H_2(r_1) \oplus m) \tag{7}$$

Kobara-Imai's α-padding [19]:

$$c = Enc(G, y_1, H_1(m\|r_1))\|y_2(H_2(r_1) \oplus m),\ y_1\|y_2\ = H_2(H_1(m\|r_1) \oplus (m\|r_1) \tag{8}$$

Kobara-Imai's β-padding:

$$c = y_1\|Enc(G, y_2, H_1(r_1)),\ y_1\|y_2\ = (r \oplus H_1(H_2(r) \oplus m)) \oplus (H_2(r) \oplus m) \tag{9}$$

Kobara-Imai's γ-padding:

$$c = y_3\|Enc(G, y_1, y_2) \tag{10}$$

$$y_3\|y_2\|y_1 = (r \oplus H_1(H_2(r) \oplus (m\|const)))\|(H_2(r) \oplus (m\|const)) \tag{11}$$

Let H_1; H_2 be random oracles (they could be pseudo-random-bits generators or hash functions) that output random strings of appropriate lengths and let r_1; r_2 be randomly selected strings with appropriate length.

2.4 Pre-computation for Private Key

Table 4.

Table 4. Pre-computation for private key

Pre-computation for Private key [20]
Input: 1. $(n+w) \times (n+w)$ permutation matrix P. 2. $k \times (n+w)$ generator matrix G. 3. integer u_0 $(u_0 < k)$. Output: $k \times (u_0+1)$ matrix X. 1. $0 \le I_1 \le I_2 \le \cdots I_u \le k$ be a list of all integers in the interval $[0, k-1)$ with $P[I_i] \ge n-w$ for all $1 \le i \le u$. If $u > u_1$ return an error. 2. $0 \le J_1 < J_2 < \cdots J_{k-u} < k$ be a list of all integers in the interval $[0, k-1)$ with $P[J_i] < n-w$ for all $1 \le i \le k-u$. 3. $k \le T_1 < T_2 < \cdots T_u < n+w$ be the first u integers such that $P[T_i] < n-w$ for all $1 \le i \le u$. 4. W be a $u \times u$ matrix such that $W[i][j] = G[i][j]$ for all $1 \le i, j \le u$. 5. V be a $(k-u) \times u$ matrix such that $V[i][j] = G[i][j]$ for all $1 \le i \le k-u$ and $1 \le j \le u$. 6. $U = (P_2[T_0], \cdots, P_2[T_u])$ be a $1 \times u$ matrix.

3 Improved RLCE Public Key Encryption Scheme Based on Polar Code

3.1 Improved RLCE Public Key Encryption Scheme

The improved RLCE Public key encryption scheme based on Polar code proposed in this paper is based on the RLCE encryption scheme proposed by Wang and the RLCE encryption scheme based on Polar code proposed by Liu. Aiming at the shortcomings of the Polar RLCE encryption scheme proposed by Liu, which is larger in key size and has no IND-CCA2 security, an improved scheme is proposed. In this paper, Polar code is used as the underlying code of RLCE encryption scheme by virtue of its polarization property and low decoding complexity. The improved scheme still adopts the structure of Wang's RLCE scheme, without changing the form of RLCE encryption scheme. The IND-CCA2 formal security proof of Wang's RLCE encryption scheme was applied to the encryption scheme based on Polar code, the Polar RLCE encryption scheme was padded with messages, each ciphertext was semantically secure mediumEncoding, and the public key matrix was converted into the system matrix, part of the private keys were estimated to reduce the key storage space. This scheme is similar to the RLCE encryption scheme, including key generation (polarRLCE.keysetup), encryption (polarRLCE.enc), and decryption (polarRLCE.dec).

3.1.1 PolarRLCE.KeySetup

(1) Parameter selection: (n,k,d,t,w), $n,k,d,t>0$, $w\in\{1,2\cdots n\}$, $k+1\geq d\geq 2t+1$.
(2) G: $k\times n$ order generating matrix of Polar code C whose dimension on domain F is k and whose minimum distance d $\geq$ 2t + 1.
(3) G_1: Generate w random column vectors $r_1,r_2,\cdots,r_w$, and by inserting w random $k\times 1$ column vectors into generator matrix G, obtained the $k\times(n+w)$ matrix G_1, $G_1=(g_1,\cdots,g_{n-w},g_{n-w+1},r_1,\cdots g_n,r_w)$.
(4) A: $(n+w)\times(n+w)$ nonsingular invertible matrix,

$$A=\begin{pmatrix} I_{n-w} & \cdots & 0 \\ \vdots & A_1 & \vdots \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_w \end{pmatrix},$$

$A_1=\begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix},\cdots A_w=\begin{pmatrix} a_{w,11} & a_{w,12} \\ a_{w,21} & a_{w,22} \end{pmatrix}\in F_2^{2\times 2}$, $A_1,A_2,\cdots,A_W$ be 2×2 nonsingular invertible matrix.
(5) $S:k\times k$ nonsingular invertible matrix.
(6) $P:(n+w)\times(n+w)$ permutation matrix.
(7) $G^{pub}:k\times(n+w)$, Public key $G^{pub}=SG_1AP$.

Public key $G^{pub}=SG_1AP$, Private key (S,G_1,A,P).

3.1.2 PolarRLCE.Enc

(1) The sender selects the plaintext $m\in F_2^k$, randomly pick the error vector $e=[e_1,\cdots,e_{n+w}]\in F_2^{n+w}$, the hamming weight of e the error vector is at most t, $wt_H(e)\leq t$.
(2) The sender uses the public key G^{pub} of the receiver to encrypt the plaintext m, get the ciphertext $c\in F_2^{n+w}$.
(3) $c=mG^{pub}\oplus e$.

3.1.3 PolarRLCE. Dec

(1) The receivers receive the ciphertext c,use own private key, get

$$cP^{-1}A^{-1}=mSG_1\oplus eP^{-1}A^{-1}=(c'_1,c'_2,\cdots c'_{n+w}) \tag{12}$$

(2) From length of $n+w$ vector $(c'_1,c'_2,\cdots c'_{n+w})$ delete frozen bit column vectors, get length of n, $c'=(c'_1,c'_2,\cdots,c'_{n-w+1},c'_{n-w+3},c'_{n-w+5},\cdots,c'_{n+w-1})$, $c'=mSG_1\oplus e'$.

(3) The receiver uses the SC decoding algorithm of Polar code, decode c' to obtain plaintext m.
(4) The receiver calculates hamming weight of $wt = wt(c - mG_1)$, if $wt \leq t$, Output message m; else, output $\perp$.

The key storage space is reduced by transforming the public key matrix into the system matrix and no longer storing the private key S.

3.2 Message Padding

The PolarRLCE schemes proposed by Liu has no IND-CCA2 security. This paper adopts the method of message padding to make the improved scheme IND-CCA2 security. This paper adopts the RLCEspad message padding scheme designed by Wang, which is suitable for encrypting short length information. After the information is filled in, the information is converted into information bits or other information in the RLCE scheme. If mediumEncoding or advancedEncoding is used, the error vector e is also partially encoded. The general encoding used in this paper is that the populated information can be encoded in non-zero terms (Table 5).

RLCEspad $(mLen, k_1, k_2, k_3, t, m, r)$:

a. $k_1 + k_2 + k_3 = \left\lceil \frac{mLen}{8} \right\rceil$, $v = 8(k_1 + k_2 + k_3) - mLen$.
b. H_1, H_2, H_3 be a random oracle, takes any-length inputs and outputs k_2, $k_1 + k_2$, k_3 bytes Output bit string.
c. $m \in \{0,1\}^{8k_1}$ be a message to be encrypted, $r_1 \in \{0,1\}^{8k_3 - v}$, $r = r_1 \| 1^v$ be a randomly selected binary bit string.

RLCE padding process is as follows:

Random select $1 \leq l_1 < l_2 < \cdots l_t \leq n + w$, $e_1 = l_1 \| l_2 \cdots \| l_t \in \{0,1\}^{16t}$, By calculation r to get e_1.

$$y = ((m \| H_1(m, r, e_1)) \oplus H_2(r, e_1)) \| H_3(m \| H_1(m, r, e_1)) \oplus H_2(r, e_1))) \tag{13}$$

Convert y to $(y_1, e_1) \in GF_2^{k+t}$, $y_1 \in GF_2^k$, $e_1 \in GF_2^t \cdot e \in GF_2^{n+w}$, for $0 \leq i \leq t$, $e[l_i] = e_1[i]$; for $j \neq l_i$, $e[j] = 0$. output (y_1, e_1). Ciphertext $c = y_1 G + e$.

Table 5. RLCEspad

RLCEspad
Input: $mLen, t, m, r$, H_1, H_2, H_3.
Output: (y_1, e_1), c.
1. Calculate $v = 8(k_1 + k_2 + k_3) - mLen$.
2. Calculate $e_1 = l_1 \| l_2 \cdots \| l_t \in \{0,1\}^{16t}$, for $m \in \{0,1\}^{8k_1}$, $r_1 \in \{0,1\}^{8k_3 - v}$, $r = r_1 \| 1^v$.
3. (mediumEncoding): $y = ((m \| H_1(m, r, e_1)) \oplus H_2(r, e_1)) \| H_3(m \| H_1(m, r, e_1)) \oplus H_2(r, e_1)))$.
4. Convert y to $(y_1, e_1) \in GF_2^{k+t}$, output (y_1, e_1), get ciphertext $c = y_1 G + e$.

4 Performance Analysis

Public key size:

$$G^{pub} = SG_1AP = [I|Q] \tag{14}$$

The size of G^{pub} is $k \times (n+w)$, by using the system matrix, the improved G^{pub} size is $k \times (n+w-k)$.

Private key size:
Private key (V, W^{-1}, G_1, A, P), the size of V is $(k-u) \times u$, the size of W^{-1} is $u \times u$, the size of G_1 is $k \times (n+w)$, the size of A is $(n+w) \times (n+w)$, the size of P is $(n+w) \times (n+w)$. Pre-computation for private key W^{-1}, the improved scheme is no longer stored $k \times k$ matrix S, reduced private key storage space (Table 6).

Table 6. Comparison of public key size of codes-based schemes (Kbytes)

ISD	Our	PolarRLCE [15]	HermitianRLCE [14]	GRSRLCE [12]	GoppaMcEliece [9]
128	98	98	103	183	188
192	256	256	198	440	490
256	380	380	313	1203	900

From the table, get the analysis:

(1) At the same bit security level, the public key size of the proposed scheme is the same as that of the Liu scheme, but the proposed scheme has IND-CCA2 security, and part of the private keys are estimated in this paper, which can significantly reduce the storage space of the private keys.
(2) At the same bit security level, the public key size of the cryptography scheme based on Polar code is the smallest in this paper.

5 Security Analysis

5.1 Brute Force Attack

A brute force attack is a trial-and error method used to obtain the correct keys. By taking the form of a system matrix, the private key are (V, W^{-1}, G_1, A, P).In this paper, Polar code is used as the underlying code of the scheme, and its equivalence class code family of permutation matrix P is large. It is difficult to find the correct one in polynomial time through exhaustive attack, and it is difficult to operate the ciphertext. $cP^{-1}A^{-1} = mSG_1 \oplus eP^{-1}A^{-1} = (c'_1, c'_2, \cdots c'_{n+w})$, unable to recover the correct plaintext. In addition, according to the parameters selected by Liu scheme, it is not feasible for the attacker to find the other three types of keys by exhaustive method. Therefore, exhaustive attacks do not affect the security of the scheme in this paper.

5.2 Information Set Decoding Attack

Information set decoding attacks are by far the most effective attacks against McEliece encoding schemes. Among all known attacks, the attacks based on information set decoding have the lowest computational complexity. Stern [21] firstly proposed an information set decoding attack on the McEliece scheme. Since then, in order to improve the effectiveness of the attack on the code-based scheme, there have been many improved information set decoding attacks [22]. Information set decoding attack does not attack the cryptography scheme by using the underlying structure of codes, but by searching information set. The working factor of information set decoding attack increases with the increase of error vector. For RLCE encryption schemes, the information set decoding attack looks for the number of columns of the public key G^{pub}, not the number of columns of the private key G_1. A random selection of k bits from the $n+w$ bit ciphertext containing t errors constitutes, a bit composition of the corresponding position is selected from the error vector, and a matrix is formed by selecting the corresponding column. If the selected bits do not contain the errors bits, that is $e_k = 0$, the attacker can easily recover the plaintext m.

For randomly selected k columns from the public key of the RLCE encryption scheme, the probability that the ciphertext contains no errors at these locations is $\frac{\binom{n+w-t}{k}}{\binom{n+w}{k}}$. In this paper, the difficulty of finding the error-free position of the selected error vector is e_k increased by inserting a random w column. Given the appropriate parameters, it is extremely difficult for an attacker to decode the information set to obtain the plaintext m. Therefore, by inserting random columns, this paper guarantees that the cryptography scheme is not affected by the information set decoding attack.

5.3 Reaction Attack

Reaction attack is that the attacker modifies a small amount of ciphertext c and sends the modified ciphertext c to the receiver to observe whether the receiver can correctly decode. For a given ciphertext c, the attacker randomly selects the location i, adds errors in the location i, sends the added wrong ciphertext c to the receiver, and observes whether the receiver can decode correctly. If the receiver can decode correctly, it means that the attacker has added an error in the position i.If the receiver decoding fails, the attacker did not add an error to the location i. Repeat the operation until the attacker gains an k error-free bit. This article resists response attacks by inserting a random w column and then padding it with a message. The Polar code-based RLCE encryption scheme proposed by Liu did not consider the response attack. This paper analyzed the possible response attacks and found that, unlike the scheme with the underlying code of hamming cyclic code, this scheme could resist the possible response attacks through the method of message padding.

5.4 Key Recovery Attack

The basic idea of key recovery attack is to recover the correct private key from the public key G^{pub}. A key recovery attack is a structural attack, usually against a specific code, for example QC-LDPC codes, GRS codes etc. The scheme proposed in this paper inserts random columns. At the same time, random nonsingular matrix A is used to mix the inserted random columns, the structure of the private key is scrambled. After mixing, the frozen bit column is randomly deleted from the ciphertext, since the attacker cannot know the polarization property of the Polar code, the complexity of the private key structure is further increased, and it is difficult for the attacker to recover the correct private key structure through the public key in polynomial time. Bardet et al. [23] proposed an attack to determine the minimum weight of Polar code, and solved the equivalence problem of polarized code relative to decreasing monomial code. The random columns inserted into the improved scheme in this paper disturb the original structure of the generated matrix, so the security of the scheme in this paper will not be affected by reference [23]. Since the structure of Polar code is different from that of hamming cyclic code, GRS code and RM code, the structural attack against the underlying codes of cyclic code, GRS code and RM code will not threaten the security of the scheme in this paper.

5.5 Adaptively Chosen Ciphertext Attacks

In this paper, the improved system matrix is used to encrypt the plaintext m, which may lead to the IND-CCA2 security reduction. Liu's Polar code-based RLCE encryption scheme does not have semantic security, and an example of a correlation attack is illustrated: two different ciphertexts can be obtained by encrypting the same plaintext twice, and two different ciphertexts can be compared and analyzed to find the original message.

$y = ((m\|H_1(m,r,e_1)) \oplus H_2(r,e_1))\|H_3(m\|H_1(m,r,e_1)) \oplus H_2(r,e_1)))$. According to the message padding method described above, this paper adopts the RLCEspad padding method to perform mediumEncoding for each ciphertext, so that some messages are encoded in the non-zero term of the error vector e. Before sending the plaintext m, the sender scrambles the information through three random oracle models H_1, H_2, H_3, encodes part of the message in the non-zero term of the error vector e, and then sends the encoded message to the sender. In order to obtain the plaintext, the difficulty of attacking the populated cryptography scheme is equivalent to that of the original McEliece scheme based on Goppa code. Therefore, the scheme after message padding by RLCEspad method in this paper has semantic security and can achieve IND-CCA2 security.

6 Application of Polar Code-Based Scheme in Cloud Secure Storage

We apply the improved code-based scheme to cloud storage. In our cloud storage system, we adopt public-key encryption and symmetric encryption based on encoding. Specifically including the message sender, message receiver, a number of cloud storage servers. Specific details of cloud storage can be referred to the literature [24].

7 Conclusion

Based on RLCE encryption scheme and PolarRLCE encryption scheme, this paper takes advantages of Polar with more equivalence classes and low complexity of encryption and decryption, proposing an IND-CCA2 secure RLCE public key encryption scheme. After security analysis, it can resist the known information set decoding and other attacks. Compared with the scheme based on hamming code, this scheme can resist the known structural attacks.

Polar has become a research hotspot due to its unique polarization and low decoding complexity. The application of Polar code to the code-based cryptography scheme has a good prospect in the future. In this paper, the public key encryption scheme based on Polar code is applied to cloud storage to ensure that the data stored in the cloud environment is safe enough under the attack of quantum computer. In the following research, the application prospect of public key encryption scheme based on Polar code can be further broadened, and the code-based scheme can be applied to more practical scenarios. Applying code-based cryptography schemes to blockchain is an interesting work.

Acknowledgment. This work was supported the National Natural Science Foundation of China (No.61572521); The Scientific Foundation of the Scientific Research and Innovation Team of Engineering University of PAP (No.KYTD201805).

References

1. Hongzhe, D., Yongfang, Z.: Fire prevention and extinguishing: generation and management of secondary public opinions in COVID 19 public crisis events. J. Univ. Electron. Sci. Technol. **22**(2), 1–7 (2020)
2. Zhang, L., Xiong, H., Huang, Q., et al.: Cryptographic solutions for cloud storage: challenges and research opportunities. IEEE Trans. Serv. Comput. (2019)
3. Jordan, S.P., Liu, Y.K.: Quantum cryptanalysis: Shor, Grover, and Beyond. IEEE Secur. Priv. **16**(5), 14–21 (2018)
4. Bernstein, D.J., Lange, T.: Post-quantum cryptography. Nature **549**(7671), 188–194 (2017)
5. Ding, J., Steinwandt, R. (eds.): PQCrypto 2019. LNCS, vol. 11505. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25510-7
6. Wu, W.L.: Preface of special issue on block cipher. J. Cryptologic Res. **6**(6), 687–689 (2019)

7. Alagic, G., Alagic, G., Alperin-Sheriff, J., et al.: Status report on the first round of the NIST post-quantum cryptography standardization process. US Department of Commerce, National Institute of Standards and Technology (2019)
8. Nejatollahi, H., Dutt, N., Ray, S., et al.: Post-quantum lattice-based cryptography implementations: a survey. ACM Comput. Surv. (CSUR) **51**(6), 1–41 (2019)
9. Mceliece, R.J.: A public-key cryptosystem based on algebraic coding theory. DSN Prog. Rep. **42**(44), 114–116 (1978)
10. Niederreiter, H.: Knapsack-type cryptosystems and algebraic coding theory. Probl. Control Inf. Theory **15**(2), 159–166 (1986)
11. Faugère, J.C., Otmani, A., Perret, L., et al.: Structural cryptanalysis of McEliece schemes with compact keys. Des. Codes Crypt. **79**(1), 87–112 (2016)
12. Wang, Y.: Quantum resistant random linear code based Public key encryption scheme RLCE. In: 2016 IEEE International Symposium on Information Theory (ISIT), pp. 2519–2523. IEEE (2016)
13. Wang, Y.: Revised quantum resistant public key encryption scheme RLCE and IND-CCA2 security for McEliece schemes. IACR Cryptology ePrint Arch. **2017**, 206 (2017)
14. Matthews, Gretchen L., Wang, Y.: Quantum resistant public key encryption scheme HermitianRLCE. In: Baldi, M., Persichetti, E., Santini, P. (eds.) CBC 2019. LNCS, vol. 11666, pp. 1–10. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25922-8_1
15. Liu, J., Wang, Y., Yi, Z., Pei, D.: Quantum resistant public key encryption scheme polarRLCE. In: Gueye, C.T., Persichetti, E., Cayrel, P.-L., Buchmann, J. (eds.) A2C 2019. CCIS, vol. 1133, pp. 114–128. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-36237-9_7
16. Arikan, E.: Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans. Inf. Theory **7**(55), 3051–3073 (2009)
17. Fouque, P.-A., Pointcheval, D.: Threshold cryptosystems secure against chosen-ciphertext attacks. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 351–368. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45682-1_21
18. Fujisaki, E., Okamoto, T.: Secure integration of asymmetric and symmetric encryption schemes. J. Cryptology **26**(1), 80–101 (2013)
19. Kobara, K., Imai, H.: Semantically secure McEliece public-key cryptosystems -conversions for McEliece PKC. In: Kim, K. (ed.) PKC 2001. LNCS, vol. 1992, pp. 19–35. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44586-2_2
20. Wang, Y.: RLCE Key Encapsulation Mechanism (RLCE-KEM) Specification (2019)
21. Stern, J.: A method for finding codewords of small weight. In: Cohen, G., Wolfmann, J. (eds.) Coding Theory 1988. LNCS, vol. 388, pp. 106–113. Springer, Heidelberg (1989). https://doi.org/10.1007/BFb0019850
22. Welch, Z.D.: An Analysis of Potential Standards for Post-Quantum Cryptosystems. Carleton University (2019)
23. Bardet, M., Chaulet, J., Dragoi, V., Otmani, A., Tillich, J.-P.: Cryptanalysis of the McEliece public key cryptosystem based on polar codes. In: Takagi, T. (ed.) PQCrypto 2016. LNCS, vol. 9606, pp. 118–143. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29360-8_9
24. Zeng, P., Chen, S., Choo, K.K.R.: An IND-CCA2 secure post-quantum encryption scheme and a secure cloud storage use case. Hum.-Centric Comput. Inf. Sci. **9**(1), 1–15 (2019)