Loong-2：基于秩距离码的 IND-CPA 安全的密钥封装

本问介绍 Loong-2 方案, Loong 中文意思为龙，包含一个基于编码的 IND-CPA 安全的密钥封装方案 Loong-2.CPAKEM，参加中国密码学会的后量子公钥部分竞赛，该方案基于秩距离支撑学习困难问题 (rank support learning problem), 该问题与秩距离校验子译码 (rank syndrome decoding) 困难问题同样困难，我们分别提供了 128、192 和 256 比特安全参数下的不同的参数设置。

第一设计者： 王丽萍
单位： 中国科学院信息工程研究所
电话： 18611628400
邮箱： wangliping@iie.ac.cn

其他设计者： 彭力强
单位： 中国科学院信息工程研究所
邮箱： pengliqiang@iie.ac.cn

其他设计者： 蔡家豪
单位： 中国科学院信息工程研究所
邮箱： caijiahao@iie.ac.cn

其他设计者： 戚艳红
单位： 中国科学院信息工程研究所
邮箱： qiyanhong@iie.ac.cn

联系人： 王丽萍
单位： 中国科学院信息工程研究所
电话： 18611628400
邮箱： wangliping@iie.ac.cn
通信地址： 北京市闵庄路 89 号

目录

1 设计方案

第一个基于编码的公钥密码是 1978 年 McEliece 提出的著名的 McEliece 公钥方案 [15]，该方案在合适的参数选取下至今仍然是安全的。最近，基于编码的公钥密码由于其能抵抗量子攻击和 NIST 全球征集后量子密码方案而受到越来越多的关注 [17]。

一般归纳来说，基于编码的公钥方案可以分为三类，McEliece 型和 Niederreiter 型密码方案是两类经典的加密方案，分别使用了码的生成矩阵和校验矩阵 [16]，常用的码类是二元 Goppa 码，也存在一些替换的码类，如 QC-MDPC (Quasi-cyclic Medium Density Parity Check) 码, QC-LRPC (Quasi-Cyclic Low Rank Parity Check) 码。科研人员们也尝试用 LDPC (Low Density Parity Check) 码、卷积码、Gabidulin 码、Reed-Muller 码、广义 Reed-Solomon 码来替换上述公钥框架中的 Goppa 码，但不幸的是均已被证明是不安全的 [3], [10], [19], [24], [25]。

第三类基于编码的方案的思想受格的启发，特别是格中的 ring-LWE (Learning with Errors) 问题，主要原因是基于该问题构造的公钥方案的公钥规模大大减少。因此，NIST 基于编码的候选方案中如 HQC, RQC, Ouroboros-R 等都是基于拟循环码的译码问题，由于拟循环码具有明显的代数结构，因而该问题类似于基于 ring-LWE 问题一样，并不能证明其是 NP-困难问题。在基于格的公钥方案中还有直接基于 LWE 困难问题的，如 FrodoKEM。

然而，在基于编码的方案中没有类似于直接基于秩距离校验子困难问题的方案，该问题已经被证明是 NP-困难问题 [8]，另外值得一提的是为什么我们关注的是秩距离，主要是由于秩距离有很好的特性，诸如 NIST 方案 LAKE, LOCKER, McNie, RQC 和 Ouroboros-R 等都使用了秩距离。

本文我们给出了名为 Loong-2 的一个新的基于编码的 IND-CPA 安全密钥封装 Loong-2.CPAKEM, 参加中国密码学会的后量子公钥部分竞赛，这个方案和 Loong-1 一样，都是基于秩距离支撑学习困难问题 (rank support learning problem), 该问题的困难性与秩距离校验子译码 (rank syndrome decoding) 困难问题，我们分别提供了 128、192 和 256 比特安全参数下的不同的参数设置。

本节安排如下：首先 1.1 节给出需要的关于编码的一些基本概念的结论，在 1.2 节展示我们的具体方案，在第 1.3 节给出三种参数设置。

1.1 预备知识

1.1.1 秩距离码

我们用黑体小写字母表示向量，大写字母表示矩阵，所有的向量都假定为行向量。令 $\mathbb{F}_{q^m}^n$ 为有限域 $\mathbb{F}_{q^m}$ 上的 n 维向量空间，其中 q 为素数幂，n, m 为正整数。本文仅考虑

$n \leq m$ 的情形。令 $\beta = \{\beta_1, \ldots, \beta_m\}$ 是 $\mathbb{F}_{q^m}$ 在 $\mathbb{F}_q$ 上的一组基，$\mathcal{F}_i$ 是 $\mathbb{F}_{q^m}$ 到 $\mathbb{F}_q$ 的映射，其中 $\mathcal{F}_i(u)$ 是 $u \in \mathbb{F}_{q^m}$ 在 β 基表示中的第 i 个坐标。对于任意的 $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{F}_{q^m}^n$, 相对应的矩阵 $\bar{\mathbf{u}} \in \mathcal{M}_{m,n}(\mathbb{F}_q)$, 其中 $\bar{u}_{i,j} = \mathcal{F}_i(u_j)$, $\mathcal{M}_{m,n}(\mathbb{F}_q)$ 表示所有的在 $\mathbb{F}_q$ 上的 $m \times n$ 矩阵, 因此一个向量 $\mathbf{u}$ 的秩重量定义为其相伴矩阵 $\bar{\mathbf{u}}$ 的秩, 记为 $w_R(\mathbf{u})$，想了解秩距离码，详见 [13]。

令整数 $k, 1 \leq k \leq n$, $\mathbb{F}_{q^m}$ 上长度为 n，维数为 k 的线性秩码 C 定义为 $\mathbb{F}_{q^m}^n$ 的一个具有秩测度的 k 维子空间，一个 $k \times n$ 的矩阵称为码 C 的生成矩阵如果它的行向量能张成这个码，码 C 的对偶 $C^\perp$ 是指 $\mathbb{F}_{q^m}^n$ 的向量子空间 C 的正交补空间，因而 $[n, k]$ 线性码 C 的一个校验矩阵 H 就是 $C^\perp$ 的一个 $(n-k) \times n$ 的生成矩阵。线性码 C 的最小距离是指 C 中非零码字的最小秩重量，记为 $d_R(C)$，因此可以唯一译 $\frac{d_R(C)-1}{2}$ 个秩错误，通常将线性码 C 记为 $[n, k, d_R(C)]$。

$\mathbb{F}_{q^m}^n$ 的任意一个向量 $\mathbf{x}$，我们定义 $\mathbf{x}$ 的支撑是指由 $\mathbf{x}$ 的坐标张成的 $\mathbb{F}_{q^m}$ 的一个 $\mathbb{F}_q$ 上的向量子空间，记为 $\mathrm{Supp}(\mathbf{x})$，即 $\mathrm{Supp}(\mathbf{x}) =< x_1, \ldots, x_n >_{\mathbb{F}_q}$，因此有 $w_R(\mathbf{x}) = \dim(\mathrm{Supp}(\mathbf{x}))$.

因此我们可以将一个向量的秩重量推广定义到一个矩阵的秩重量。

定义 1.1 令 $X = (x_{ij})_{1 \leq i \leq l, 1 \leq j \leq n}$ 为 $\mathbb{F}_{q^m}$ 上的 $l \times n$ 的矩阵，X 的支撑定义为由 X 的每一个分量生成的 $\mathbb{F}_{q^m}$ 的一个 $\mathbb{F}_q$ 的向量子空间，记为 $Supp(X)$, 即 $Supp(X) = \langle x_{11}, \ldots, x_{1,n}, \ldots, x_{l1}, \ldots, x_{ln} \rangle_q$，$X$ 的秩重量定义为 X 的支撑的维数，也记为 $w_R(X)$。

注意到 $\mathbb{F}_{q^m}$ 上的矩阵的秩重量并不是该矩阵的秩。

另外还常用到下面的一个结论：

$\mathbb{F}_{q^m}$ 上的维数为 w 的支撑的个数等于 $\mathbb{F}_{q^m}$ 上的维数为 w 的 $\mathbb{F}_q$ 线性子空间的个数：$\begin{bmatrix} m \\ w \end{bmatrix} = \prod_{i=0}^{w-1} \frac{q^m - q^i}{q^w - q^i} = \theta(q^{w(m-w)})$。

1.1.2 LRPC 码及其秩支撑恢复算法

低秩校验码 (Low Rank Parity Check, 简记为 LRPC) 是在论文 [6] 中引入的，LRPC 码由于其弱的代数结构和高效的译码算法已经广泛应用到基于编码的密码体制中。

定义 1.2 一个秩为 d, 码长为 n，维数为 k 的 $LRPC$ 码是指一个在 $\mathbb{F}_{q^m}$ 上的 $[n, k]$ 线性码，其校验矩阵 $H = (h_{ij})_{1 \leq i \leq n-k, 1 \leq j \leq n}$，由所有的 h_{ij} 张成的 $\mathbb{F}_q$ 上的向量子空间的维数为 d。

LRPC 码的秩校验子译码问题：给定该码的一个校验矩阵 $H \in \mathbb{F}_{q^m}^{(n-k) \times n}$，秩为 d，校验子 $\mathbf{s} \in \mathbb{F}_{q^m}^{n-k}$，求向量 $\mathbf{x} \in \mathbb{F}_{q^m}^n$ 满足 $w_R(\mathbf{x}) \leq r$ 且 $H\mathbf{x}^T = \mathbf{s}^T$。

事实上，在我们的算法 Loong-2.CPAKEM 中，我们的 LRPC 码的译码问题给出了多个例子，具体如下：

多实例的 LRPC 码的秩校验子译码问题：给定该码的一个校验矩阵 $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$，秩为 d，校验子 $\mathbf{s}_1, \ldots, \mathbf{s}_t \in \mathbb{F}_{q^m}^{n-k}$，求一个向量 $\mathbf{x}_1, \ldots, \mathbf{x}_t \in \mathbb{F}_{q^m}^{n}$ 满足 $\mathrm{Supp}(\mathbf{x}_1) = \ldots = \mathrm{Supp}(\mathbf{x}_t)$, $w_R(\mathbf{x}) \leq r$ 且 $H\mathbf{x}_i^T = \mathbf{s}_i^T$, $1 \leq i \leq t$。

本方案中只需要恢复出由 $\mathbf{x}_1, \ldots, \mathbf{x}_t$ 张成的向量子空间 E，而不是 $\mathbf{x}_i$, 这个问题称为秩支撑恢复 (rank support recovery) 问题。单实例的秩支撑恢复算法在 [18] 中给出的，该算法给出的是拟循环码的秩支撑集合恢复算法，使用了 LRPC 码的一般译码算法 [7] 和 [1] 中的改进算法。由于不需要计算出每一个 $\mathbf{x}_i$, 因此算法对多实例的问题也成立，下面我们详细给出多实例的一般的 LRPC 码的秩支撑恢复算法。

下面算法中，令 F 表示由 $H = (h_{ij})$ 的所有的 h_{ij} 生成的维数为 d 的向量空间, 且其的基为 $\{F_1, \ldots, F_d\}$, E 表示由所有向量 $\mathbf{x}_1, \ldots, \mathbf{x}_t$ 中的所有分量生成的维数为 r 的向量空间。再令 S 表示由所有校验子向量中的分量生成的向量子空间，S_i 定义为 $S_i = F_i^{-1}.S = \langle F_i^{-1}s_{1,1}, F_i^{-1}s_{1,2}, \cdots, F_d^{-1}s_{t,n-k}\rangle$, 其中 F_i 是 H 的基元素，$S_{ij} = S_i \cap S_j$。

RS-recover$(H, \mathbf{s}, r)$

输入: $H = \langle F_1, F_2, \ldots F_d\rangle$, $H\mathbf{x}_i^T = \mathbf{s}_i^T$, $1 \leq i \leq t$, r (E 的维数)
输出: 向量空间 E
// 第 1 部分: 计算 $E.F$
1 计算 $S = \langle s_{1,1}, \ldots, s_{t,n-k}\rangle$
2 预计算每一个 S_i for $i = 1$ to d
3 预计算每一个 $S_{i,i+1}$ for $i = 1$ to $d-1$
4 for i from 1 to $d-2$ do
5 　tmp $\leftarrow S + F.(S_{i,i+1} \oplus S_{i_1,i+2} \oplus S_{i,i+2})$
6 　if $\dim(tmp) \leq rd$ then
7 　　$S \leftarrow tmp$
8 　end
9 end
// 第 2 部分: 恢复子空间 E
10 $E \leftarrow F_1^{-1}.S \cap \ldots \cap F_d^{-1}.S$
11 return E

上述算法在某些情形下会失败，参考 Ouroboros-R [18]，我们给出给出译码失败的概率。

命题 1.1 上述算法译码失败率为 $\max(q^{(2-r)(d-2)} \times q^{-((n-k)t-rd+1)}, q^{-2((n-k)t-rd+2)})$, 其中 r 是错误向量的秩重量。

1.1.3 基于编码的公钥密码的困难问题

本节我们给出本文需要的一些基于编码的困难问题。

定义 1.3 (秩校验子译码 (简写为 RSD) 问题) 给定一个随机线性码的校验矩阵 $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$，以及 $\mathbf{y} \in \mathbb{F}_{q^m}^{n-k}$，本问题的目标是求一个向量 $\mathbf{x} \in \mathbb{F}_{q^m}^{n}$ 满足 $w_R(\mathbf{x}) \leq w$ 且 $H\mathbf{x}^T = \mathbf{y}^T$。

最近 RSD 问题已经被证明是 NP 困难问题，因为可以概率约化到汉明距离下的校验子译码问题 [8], 正如我们已知的，汉明距离下校验子译码问题已被证明是 NP-难问题 [4]，上述 RSD 问题也称作 RSD 问题的搜索版，该问题记为 $\text{RSD}_{n,k,w}$。下面介绍 RSD 问题的判定版 (Dicisional RSD version)。

定义 1.4 (判定 RSD (简写为 DRSD) 问题) 给定 $(H, \mathbf{y}^T) \xleftarrow{\$} \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{(n-k)}$，判定 *RSD* 问题是问能否以不可忽略的优势判定 $(H, \mathbf{y}^T)$ 是由 *RSD* 问题产生的还是在 $\mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{(n-k)}$ 上的均匀分布。

事实上，我们的算法是基于下面的秩支撑学习 (rank support learning (RSL)) 问题 [9]。

定义 1.5 (RSL 问题) 给定一个随机线性码的校验矩阵 $H \in \mathbb{F}_{q^m}^{(n-k)\times n}$，和 $Y \in \mathbb{F}_{q^m}^{(n-k)\times l}$，其中 $l \leq n$，本问题目标在于找到一个矩阵 $X \in \mathbb{F}_{q^m}^{n\times l}$ 满足 $w_R(X) = w$ 且 $HX = Y$。

我们简记上述问题为 $\text{RSL}_{n,k,w,l}$ 问题，在论文 [9] 中，作者们已经证明了如下的结论。

命题 1.2 *$RSL_{n,k,w,l}$ 问题与 $RSD_{n,k,w}$ 问题一样困难.*

类似地，也可以定义判定 RSL 问题，记为 $\text{DRSL}_{n,k,w,l}$。由于对 $\text{DRSL}_{n,k,w,l}$ 问题已知最好的攻击算法就在于解 $\text{RSL}_{n,k,w,l}$ 问题的相同实例，因此我们可以假定 $\text{DRSL}_{n,k,w,l}$ 问题也是困难的。

1.1.4 安全性定义

回忆一下公钥密码中的标准的安全概念：选择明文攻击下的不可区分和选择密文攻击下的不可区分安全，简称为 IND-CPA 和 IND-CCA [21]。敌手 A 的 CPA 下的优势定义为

$$\mathrm{Adv}_{PKE}^{CPA}(A) = \left| Pr\left[b = b' : \begin{array}{l} (pk, sk) \leftarrow KeyGen() \\ (m_0, m_1) \leftarrow A(\text{Find: pk}) \\ b \leftarrow \{0,1\}, c^* \leftarrow Enc(pk, m_b) \\ b' \leftarrow A(Guess : c^*) \end{array} \right] - \frac{1}{2} \right|.$$

如果敌手 A 可以询问解密 oracle 除密文 c^* 之外的密文，则可以定义在 CCA 下的优势 $\mathrm{Adv}_{\mathrm{PKE}}^{\mathrm{CCA}}$ 。

密钥封装定义在选择明文攻击下和选择密文攻击的不可区分安全性。敌手 A 在选择明文攻击下的优势定义为

$$\mathrm{Adv}_{KEM}^{CPA}(A) = \left| Pr\left[b = b' : \begin{array}{l} (pk, sk) \leftarrow KeyGen() \\ b \leftarrow \{0,1\} \\ (c^*, K_0^*) \leftarrow Encaps(pk); K_1^* \leftarrow \mathcal{K} \\ b' \leftarrow A(Guess : c^*, K_b^*) \end{array} \right] - \frac{1}{2} \right|,$$

其中 $\mathcal{K}$ 为密钥空间。进一步如果敌手 A 可以询问解封装 oracle 除密文 c^* 之外的密文，我们可以定义在选择密文攻击下的优势 $\mathrm{adv}_{KEM}^{CCA}(A)$。

1.2 Loong-2.CPAKEM 算法描述

本小节我们提出了一个新的基于编码的 IND-CPA 安全的密钥封装，i.e., Loong-2.CPAKEM, 该方案类似于 Loong-1.CPAPKE, 但是不同之处在于 Loong-1.CPAPKE 中使用了 Gabidulin 码作为辅助码，但是 Loong-2.CPAKEM 方案中没有用明文，利用了 LRPC 码做为辅助码译码出噪声矩阵，再直接将噪声矩阵做了哈希函数，哈希值作为公共密钥。令 $\mathcal{H}$ 为 $\{0,1\}^* \to \{0,1\}^{512}$ 的哈希函数。

需要注意的是，我们译 LRPC 码使用的算法是多实例 LRPC 码的译码算法。

Loong-2.CPAKEM:

- **KeyGen**(): 取正整数 n, n_1, n_2, m，素数幂 q, 公钥种子 $\mathrm{seed}_H \xleftarrow{\$} \{0,1\}^{256}$, 公钥矩阵 $H \in \mathbb{F}_{q^m}^{n \times n} := \text{Shake-256}(\mathrm{seed}_H)$。私钥种子 $\sigma \xleftarrow{\$} \{0,1\}^{320}$，噪声矩阵 $(X, Y) \in \mathbb{F}_{q^m}^{n \times n_1} \times \mathbb{F}_{q^m}^{n \times n_1} := \text{AES}(\sigma)$, 且 $w_R(X) = w_R(Y) = w$。令 $Q := HX + Y$，则公钥 $pk := (seed_H, Q)$，私钥 $sk := (X, Y)$。
- **Encaps**(pk): 噪声种子为 $\tau \xleftarrow{\$} \{0,1\}^{320}$, 噪声矩阵 $(R, E, E_2) \in \mathbb{F}_{q^m}^{n_2 \times n} \times \mathbb{F}_{q^m}^{n_2 \times n} \times \mathbb{F}_{q^m}^{n_2 \times n_1} := \text{AES}(\tau)$ 使得 $w_R(E_1) = w_R(R) = w_R(E_2) = w_e$。令 $F := Supp(R, E_1, E_2)$，计算 $K := \mathcal{H}(F)$, 且返回密文对 $(C_1, C_2) \in \mathbb{F}_{q^m}^{n_2 \times n} \times \mathbb{F}_{q^m}^{n_2 \times n_1}$, 其中

$$C_1 := RH + E_1, C_2 := RQ + E_2.$$

- **Decaps**(sk): 计算 $C_2 - C_1X := RY + E_2 - E_1X$，令 $S := Supp(X, Y)$，$F :=$ RS-Recover$(S, C_2 - C_1X, w_e)$，恢复出 $K := \mathcal{H}(F)$。

正确性： 方案 Loong-2.CPAKEM 的正确性依赖于对 LRPC 码的之支撑恢复算法的译码能力，具体来说，如果 LRPC 码能为 $C_2 - C_1X$ 正确译码，我们就得到

$$\mathrm{Dec}(sk, \mathrm{Enc}(R, E_1, E_2, w_e, pk)) = Supp(R, E_1, E_2).$$

1.3 Loong-2.CPAKEM 方案参数设置

Loong-2.CPAKEM 的实际安全性依赖于 DRSD 困难问题的实例，即满足 $Q = HX + Y$，小秩重量矩阵 X 和 Y 且支撑向量子空间含单位元 1，$w_H(X) = w_H(Y) = w$，需要译 $\mathbb{F}_{q^m}$ 上的 $[2n, n]$ 线性随机码。由于秘密密钥需要考虑 $1 \in \mathrm{Supp}(X, Y)$，需要求秘密密钥矩阵的含单位元的小重量为 w 的支撑子空间问题，该问题比求秘密密钥矩阵的秩重量为 $w-1$ 的问题困难，因此实际上我们的安全性参数需要满足 $\mathrm{DRSL}_{2n,n,w-1,n_1}$ 这一困难问题。

另外，本方案的 IND-CPA 安全性同样可以约化为译 $[2n + n_1, n]$ 的随机线性码的 DRSD 问题，小秩重量的矩阵 (R, E_1, E_2)，且 $w_R(R) = w_R(E_1) = w_R(E_2) = w_e$。因此，参数的选择根据 [1], [7] 中给出的目前最好的组合译码攻击计算复杂度，该计算复杂度在第 5 节中给出.

另外，我们在 Loong-2.CPAKEM 参数设置中根据命题 1 还给出了译码失败率 (DFR)，主要是 LRPC 码的译码存在译码失败的情形。

表 1 给出了 Loong-2.CPAKEM 方案在 128 比特、192 比特和 256 比特安全级别下三种安全级别下的参数选取。

表 1: Loong-2.CPAKEM 的参数设置

Instance	q	m	n	n_1	n_2	w	w_e	Security
Loong-2.CPAKEM-I	2	71	44	7	8	6	5	128
Loong-2.CPAKEM-II	2	83	50	8	8	7	6	192
Loong-2.CPAKEM-III	2	107	64	8	8	7	7	256

表 2 给出了 Loong-2.CPAKEM 方案在 128 比特、192 比特和 256 比特安全级别下公钥、私钥、密文、会话密钥的理论规模参数值，字节为单位，另外给出了经典复杂度和量子复杂度，单位是比特。其中，公钥 pk 由 (ρ, Q) 组成，其规模大小为 $nmn_1 + 256$ 比特，即 $\frac{nmn_1}{8} + 32$, 秘密密钥 sk 由 X 和 Y 组成，其规模大小为 $2nn_1m$ 比特，但是实

际上它可以由 40 字节的种子生成，密文由 (C_1, C_2) 生成，其规模大小为 $nn_2m + n_2n_1m$ 比特，即 $\frac{mn_2(n+n_1)}{8}$ 字节，会话密钥 512 比特，即 64 字节。

表 2: Loong-2.CPAKEM 方案下的参数的理论规模

Instance	pk size	sk size	ct size	ss size	DFR	Security	Q-security
Loong-2.CPAKEM-I	2766	40	3621	64	2^{-39}	128	87
Loong-2.CPAKEM-II	4182	40	4814	64	2^{-43}	192	117
Loong-2.CPAKEM-III	6880	40	7704	64	2^{-41}	256	152

本节最后，表 3 展示了 Loong-1.CCAKEM 方案、Loong-2.CPAKEM 方案与 NIST 某些第二轮候选方案的公钥规模比较，我们的方案与 FrodoKEM、Classic McEliece 和 NTS-kem 方案相比，公钥规模小的多，而安全性相当，因此我们的方案可以作为长期安全的后量子候选方案。

表 3: Comparison on sizes of public keys (in bytes)

Instance	128 bits	192 bits	256 bits
Classic McEliece	368,282		1,046,737
NTS-kem	319,488	929,760	1,419,704
Loong-1.CCAKEM	3156	5489	10840
Loong-2.CPAKEM	2766	4182	6880
RQC	786	1411	1795
HQC	2819	5115	7417
LEDAKem	3,480	7,200	12,384
BIKE-I	2541	5474	8181
BIKE-II	1271	2737	4094
BIKE-III	2757	5421	9033
Ouroboros-R	676	807	1112
LOCKER	737	1048	1191
Frodo	9,616	15,632	

2 性能分析

我们编了程序来实现本算法，但是不知道什么原因没有跑出结果，因此我们没有给出 Loong-2.CPAKEM 的性能分析。

3 Loong-2. CPAKEM 的安全性证明

本节我们证明 Loong-2.CPAKEM 在 DRSD 困难假设下和在 random-oracle 模型下 IND-CPA 安全的。我们假设方案中用到的哈希函数 $\mathcal{H}$ 是一个 random oracle 模型。

定理 3.1 对于任意一个敌手 A，存在一个敌手 B 使得

$Adv^{CPA}_{Loong\text{-}2.CPAKEM} \leq Adv^{DRSD}_{2n,n,w,n_1}(B) + Adv^{DRSD}_{2n+n_1,n,w_e,n_2}(B).$

证明: 令 A 是执行 IND-CPA 安全实验的敌手，称该实验为 G_1，即

$$\mathrm{Adv}^{\mathrm{CPA}}_{\text{Loong-1.CPAPKE}}(A) = |Pr[b = b' \text{ in game } G_1] - 1/2|,$$

实验 G_2 中, 密钥生成部分 KeyGen() 产生的 $Q = HX + Y$ 由随机生成的矩阵替换，因此可以验证存在一个敌手 B 以与实验 G_1 相同的运行时间使得

$$|Pr[b = b' \text{ in game } G_1] - Pr[b = b' \text{ in game } G_2]| \leq \mathrm{Adv}^{\mathrm{DRSL}}_{[2n,n,w-1,n_1]}(B),$$

由于 $(I\ H)\begin{pmatrix} X \\ Y \end{pmatrix} = Q$, 其中 $(I\,H)$ 是在 $\mathbb{F}_{q^m}$ 上的随机线性码 $[2n, n]$ 的校验矩阵，X 和 Y 是秩重量为 w 的随机选取的矩阵。

在实验 G_3，挑战密文生成的 $C_1' = RH + E_1$ 和 $C_2' = RQ + E_2$ 由随机均匀选取替换的，因而存在一个敌手 B 以与 A 相同的运行时间使得

$$|Pr[b = b' \text{ in game } G_2] - Pr[b = b' \text{ in game } G_3]| \leq \mathrm{Adv}^{\mathrm{DRSL}}_{[2n+n_1,n,w_e,n_2]}(B),$$

这是由于 $\begin{pmatrix} I_n & & H^T \\ & I_{n_1} & Q^T \end{pmatrix}\begin{pmatrix} E_1^T \\ E_2^T \\ R^T \end{pmatrix} = \begin{pmatrix} C_1^T \\ C_2^T \end{pmatrix}$, 其中 $\begin{pmatrix} I_n & & H^T \\ & I_{n_1} & Q^T \end{pmatrix}$ 是 $[2n+n_1, n]$-随机线性码的系统化校验矩阵，H, Q 随机均匀选取，E_1, E_2, R 随机选取，且秩重量为 w_e。

注意到在实验 G_3, 由挑战密文得到的 C_2 与 b 是独立的，因此 $Pr[b = b' \text{ in game } G_3] = \frac{1}{2}$，结论得证。 □

4 现有攻击

目前对 RSD 问题存在两种通用攻击，一种是组合译码攻击，一种是利用 Gröbner 基的代数攻击。

最新的组合译码攻击见论文 [1] 和 [7]，主要的结论如下，该结论也是我们选择参数的主要依据。

给定 $\mathbb{F}_{q^m}$ 上的 $[n, k]$ 秩码 $\mathcal{C}$，译一个秩距离为 w 的码字，目前最好的组合译码攻击的计算复杂度为

$$O((nm)^3 q^{w\lceil \frac{m(k+1)}{n} \rceil - m}). \tag{1}$$

应用 Grover 搜索，我们可以给出其的组合译码的量子复杂度为

$$O((nm)^3 q^{\frac{1}{2}(w\lceil\frac{m(k+1)}{n}\rceil-m)}). \tag{2}$$

至于代数攻击，当 $q=2$ 时，对于上述问题所用的时间复杂度公式如下 $q^{w\lceil\frac{w(k+1)-(n+1)}{w}\rceil}$ [11]。

5 优势与不足

本文主要提出了 IND-CPA 安全的公钥加密方案 Loong1.CPAPKE 和 IND-CCA 安全的密钥封装 Loong1.CCAKEM，和一个新的 IND-CPA 安全的密钥封装方案 Loong2.CPAKEM，这两个方案均基于秩支撑学习的困难问题，该问题等价于已被证明是 NP-hard 的秩距离校验子译码问题，因此这些方案的安全性非常有保障，我们有理由相信我们的方案可以作为长期安全的后量子公钥方案的好的候选方案。具体的优势如下：

安全性. 我们方案设计的理念在于安全性第一，效率其次，我们的方案基于一个 NP-困难问题，安全性证明也很简单，而现有的基于编码的各种方案中，要么是基于混淆 Goppa 码，要么利用拟循环码，都是具有一定的代数结构，因而不能排除未来会出现致命的攻击，而我们使用的是无任何代数结构的随机线性码，因此具有很高的安全性。

公钥规模. 我们 Loong-1.CCAKEM 和 Loong-2.CPAKEM 两个方案与已经进入第二轮的一些候选方案相比，比如与目前经历 40 年考验的 Classic McEliece 方案、直接基于 LWE 困难问题的 FrodoKEM、基于 McEliece 方案的 NTS-KEM 相比，我们的安全性同样很高但是公钥规模小的太多。即便与基于汉明距离的拟循环结构的进入第二轮的一些候选方案如 HQC, BIKE, LEDAkem 等相比公钥规模相差不多，而我们的安全性基于无结构的随机码的译码问题。因此，我们的方案可以成为长期安全的后量子公钥方案的有力的候选方案。

译码失败率. 在 Loong-1.CPAPKE 和 Loong-1.CCAKEM 中由于我们使用的是 Gabidulin 码的译码算法，因此不存在译码失败率；而 Loong-2.CPAKEM 方案中由于使用的是 LRPC 码的译码，因而存在译码失败率，但是我们已经给出了其的译码失败率。

不足之处：

效率低. 由于我们直接使用的基于困难等价的秩距离校验子译码困难问题，使用的是随即线性码的无任何代数结构的校验矩阵，因此性能实现上会比使用代数结构如拟循环码的要慢。

Loong-2 方案的无 IND-CCA 的转换. 由于 Loong2.CPAKEM 方案中使用了 LRPC

码的译码技术，不能获得一个可忽略的译码失败率，因而转换成更高的安全性，比如 IND-CCA 安全的方案就很困难。

参考文献

[1] N. Aragon, P. Gaborit, A. Hautevile, and and J.-P. Tillich. Improvement of generic attacks on the rank synd rome decoding problem. 2017. Pre-print.available at https://www.unilim. fr/pages-perso/philipe.gaborit/newGRS.pdf.

[2] M. R. Albrecht, R. Player, S. Scott, On the concrete hardness of learning with errors, Journal of Mathematical Cryptology 9 (3) (2015)169-203.

[3] M. Baldi, QC-LDPC code-based cryptography, Ser. Springer Briefs in Electrical and Computer Engineering, Springer International Publishing, 2014

[4] E. Berlekamp, R. McEliece and H. Van Tilborg. On the inherent intractability of certain coding problems, IEEE Trans. Information Theory 24(3)(1978) 384-386.

[5] A. Fujisaki and T. Okamoto, Secure integration of asymmetric and symmetric encryption schemes, CRYPTO'99, LNCS Vol. 1666, Springer, Heidelberg, 537-554

[6] P. Gaborit, G. Murat, O. Ruatta, and G. Zémor. Low rank parity check codes and their application to cryptography. In Proceedings of the Workshop on Coding and Cryptography WCC' 2013, Bergen, Norway, 2013.

[7] P. Gaborit, O. Ruatta and J. Schrek. On the complexity of the rank syndrome decoding problem. IEEE Trans. Information Theory 62(2) (2016) 1006-1019.

[8] P. Gaborit and G. Zemor, On the hardness of the decoding and the minimum distance problem for rank codes, IEEE Trans. Information Theory 62 (12) (2016) 7245-7252.

[9] P. Gaborit, A. Hauteville, D. H. Phan, J.-P. Tillich: Identity-Based Encryption from Codes with Rank Metric. CRYPTO 2017: 194-224.

[10] G. Landais and J.-P. Tillich. An effcient attack of a McEliece cryptosystem variant based on convolutional codes, PQCRYPT 2013, 102-117.

[11] F. Levy-dit-Vehel, L. Perret, Algebraic decoding of rank metric codes. Proceedings of YACC06, 2006.

[12] R. Lindner and C. Peikert. Better key sizes (and attacks) for LWE-based encryption. CT-RSA 2011, LNCS 6558, 319-339, 2011.

[13] P. Loidreau, Properties of codes in rank metric, http://arxiv.org/abs/cs/ 0610057.

[14] P. Loidreau. A welch-berlekamp like algorithm for decoding gabidulin codes. In Coding and cryptography, 36-45. Springer, 2006.

[15] R. J. McEliece. A public key crytosystem based on algebraic coding theory, DSN progress report 44 (1978) 114-116.

[16] H. Niederreiter, Knapsack-type cryptosystems and algebraic coding theory, Problems of Control and Information Theory 15 (1986), 159-166.

[17] NIST. Post quantum crypto project. http://csrc.nist.gov/groups/ST /post-quantum-crypto, 2017. Available at https://csrc.nist.gov /Projects/Post-Quantum-for-Cryptography/Post-Quantum-Cryptography-Standardization/call-for-Proposalls.
List of First Round candidates available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-1-submissions.

[18] J.-C. Deneuville, P. Gaborit, and G. Zémor. Ouroboros: A simple, secure and efficient key exchange protocol based on coding theory. In Tanja Lange and Tsuyoshi Takagi, editors, Post-Quantum Cryptography 8th International Workshop, PQCrypto 2017, June 18-34, 2017.

[19] R. Overbeck. Structural attacks for public key cryptosystems based on Gabidulin codes, Journal of Cryptology 21(2008) 280-301.

[20] K. Pietrzak, Cryptography from learning parity with noise, SOFSEM 2012, LNCS 7147, 99-114, 2012.

[21] C. Rackoff and D. R. Simon. Non-Interactive Zero-Knowledge Proof of Knowledge and Chosen Ciphertext Attack. In CRYPTO'91 (LNCS), Joan Feigenbaum (Ed.), Vol. 576. Springer, Heidelberg, 433-444.

[22] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. STOC'05, 84-93, 2005.

[23] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Comput. 26 (5)(1997) 1484-1509.

[24] V. M. Sidelnikov. A public-key cryptosystem based on binary Reed-Muller codes, Discrete Mathematics and Applications 4 (1994) 191-207.

[25] V. M. Sidelnikov, S. O. Shestakov. On insecurity of cryptosystems based on generalized Reed-Solomon codes. Discrete Mathematics and Applications 2 (1992) 439-444.