2020 届研究生硕士学位论文

分类号：_____ 学校代码：10269

密　级：_____ 学　　号：51174500168

華東師範大學

East China Normal University

硕 士 学 位 论 文

MASTER'S DISSERTATION

论文题目：基于编码的抗量子加密算法关键技术研究

院　　系：	计算机科学与软件工程学院
专 业 名 称：	软件工程
研 究 方 向：	密码与网络安全
指 导 教 师：	曾鹏 副教授
学位申请人：	周玉壮

2019 年 11 月

Dissertation for master degree in 2020 University Code: 10269
Student ID: 51174500168

EAST CHINA NORMAL UNIVERSITY

THE RESEARCH KEY TECHNOLOGY OF CODE-BASED ENCRYPTION SCHEME

Department:	School of Computer Science and Software Engineering
Major:	Software Engineering
Research direction:	Cryptography and Network Security
Supervisor:	Associate Professor Peng Zeng
Candidate:	Yuzhuang Zhou

2019.01

华东师范大学学位论文原创性声明

郑重声明：本人呈交的学位论文《基于位置服务的隐私保护关键技术研究》，是在华东师范大学攻读硕士/博士（请勾选）学位期间，在导师的指导下进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

作者签名:_____ 日期: 年 月 日

华东师范大学学位论文著作权使用声明

《基于编码的抗量子加密算法关键技术研究》系本人在华东师范大学攻读学位期间在导师指导下完成的硕士/博士（请勾选）学位论文，本论文的研究成果归华东师范大学所有。本人同意华东师范大学根据相关规定保留和使用此学位论文，并向主管部门和相关机构如国家图书馆、中信所和“知网”送交学位论文的印刷版和电子版；允许学位论文进入华东师范大学图书馆及数据库被查阅、借阅；同意学校将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于（请勾选）

（）1. 经华东师范大学相关部门审查核定的“内部”或“涉密”学位论文 *，于年月日解密，解密后适用上述授权。

（）2. 不保密，适用上述授权。

导师签名:_____ 本人签名:_____

年 月 日

* “涉密”学位论文应是已经华东师范大学学位评定委员会办公室或保密委员会审定过的学位论文（需附获批的《华东师范大学研究生申请学位论文“涉密”审批表》方为有效），未经上述部门审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权）。

周玉壮 硕士学位论文答辩委员会成员名单

姓名	职称	单位	备注
曹珍富	教授	华东师范大学	主席
张磊	研究员	华东师范大学	
朱浩瑾	副教授	上海交通大学	

摘　要

量子计算机的出现，深刻影响着计算机各个领域的发展。量子计算机的重要优越性就是量子并行计算，这使得量子计算机可以达到经典计算机无法达到的算力水平。于是，很多经典计算机无法解决的问题，量子计算机就能很快解决。这无疑是巨大的进步，但是从另一方面来看，超级计算能力对如今计算机构建的信息时代安全本身也是一种威胁。具体来看，在密码学领域，密码算法的安全性是基于当前计算机的计算能力的，在允许的破解成本下，计算机所拥有的算力无法破解密码，就说密码算法是安全的。研究表明，在理想情况下，现有的量子计算机，已经使经典的密码算法处于严重威胁之中。为了抵抗未来量子计算机带来的攻击，有多种抗量子加密算法已经处于研究之中。比如：基于 Hash 函数的公钥密码体制；基于格问题的公钥密码体制；基于多变量问题的公钥密码体制；基于编码问题的公钥密码体制。

如上的密码体制又叫做后量子密码，迄今为止，还没有量子计算机算法对其能进行有效的攻击。本文主要讨论的是基于编码的公钥加密方案。基于编码的公钥加密方案的是在 1978 年由 McEliece 首次提出的，该方案加密过程是将明文当作合法码字并加入可以纠正的错误向量，解密时根据码结构先进行有效译码纠错，再恢复明文。虽然目前基于编码的公钥加密方案公认是的可以抵抗量子攻击，且具有较高的安全性和实现效率，但是该密码技术仍无法大规模广泛应用。这主要是因为基于编码的公钥加密方法都存在这公钥过大的问题。后续重点的研究方向就是缩小算法的公钥尺寸。

在基于编码理论的加密方案中，有很多 McEliece 方案的变体。一般来说，这些变体的尝试总是利用两种基本方式来增强密码系统的安全和性能。其中一个是减小公钥的大小；另一个是提高解码算法的效率和纠错能力。与此同时，安全级别是一直追求的目标。在本文中，我们按照 BBCRS 方案的思想提出了一种新的公钥构造方式。这种改进增强了编码的纠错能力，并且可以更好地保护密钥结构。我们还在 BBCRS 方案中详细讨论了一些已知的攻击，结果表明我们的新方案在当前已知的攻击手段下是安全的。

关键词: 抗量子密码学，基于编码的加密学，McEliece 加密方案，公钥加密

ABSTRACT

The emergence of quantum computers has profoundly affected the development of various fields of computers. The important advantage of quantum computers is quantum parallel computing, which allows quantum computers to reach the level of computing power that classical computers cannot. As a result, many problems that classic computers cannot solve can be solved quickly by quantum computers. This is undoubtedly a huge improvement, but on the other hand, supercomputing power is also a threat to the security of the information age built by computers today. Specifically, in the field of cryptography, the security of the cryptographic algorithm is based on the computing power of the current computer. Under the allowed cracking cost, the computing power possessed by the computer cannot attack effectively the encryption scheme, and the cryptographic scheme is safe. Researches have shown that, under ideal conditions, existing quantum computers have placed classic cryptographic algorithms at serious risk. In order to resist the attacks brought by quantum computers in the future, a variety of anti-quantum encryption algorithms are under study. For example: Hash-based public key cryptosystem; lattice-based public key cryptosystem; Multivariate Public Key Cryptosystems; code-based public key cryptosystem.

The above cryptosystem is also called post-quantum cryptography. So far, no quantum computer algorithm has been able to effectively attack it. This article focuses on code-based public key encryption schemes. The code-based public key encryption scheme was first proposed by McEliece in 1978. The encryption process of the scheme is to treat the plaintext as a legal codeword and add a correctable error vector. When decrypting, the code structure is firstly decoded and corrected, and then resume the plaintext. Although the code-based public key encryption scheme is recognized to be resistant to quantum at-

tacks, and has high security and implementation efficiency, the cryptographic technology cannot be widely applied on a large scale. This is mainly because the public key encryption has the problem that the public key is too large. The next focus of research is to reduce the public key size of the algorithm.

There are many variations of the McEliece scheme in encryption schemes based on coding theory. In general, attempts at these variants always take advantage of two basic ways to enhance the security and performance of the cryptosystem. One is to reduce the size of the public key; the other is to improve the efficiency and error correction ability of the decoding algorithm. At the same time, high security levels are always important. In this paper, we propose a new public key construction method according to the idea of the BBCRS scheme. This improvement enhances the error correction capability of the code and can better protect the secret key structure. We also discussed some known attacks in detail in the BBCRS program, and the results show that our new solution is safe under the currently known attack methods.

Keywords: *Post-quantum cryptography; code-based cryptography; McEliece cryptosystem; public key encryption.*

目录

插图

表格

List of Algorithms

第一章　绪 论

1.1　研究背景及意义

随着信息时代的发展，信息安全，网络安全，系统安全在社会中的作用日益重要起来。互联网作为一个自由开放，虚拟交互的全球平台，可以使人们更便利的获取，发布信息，但与此同时，互联网与个人息息相关的的资源也受到不同程度的威胁，于是安全技术研究也蓬勃发展起来。密码学技术，是安全技术研究的基石，其实从古至今都不乏密码学的研究，自密码学从外交情报和军事领域走向公开后，社会信息流通的方式，也深刻影响着密码学的特点。传统计算机的出现，使得古典密码的破解变得容易，计算机网络的数据传输需要更安全的密码算法，于是产生了一些经典的加密算法，如：DES，AES 等对称加密算法，RSA，ECC（椭圆曲线加密算法）等非对称加密算法。密码设计者与密码分析者相互竞技，共同促进密码学平衡的发展。但是，量子计算机的问世，让密码学领域的格局发生了巨大的变化。

量子计算机作为第六类计算机，使用的计算方式和平常使用的普通计算机非常不同。量子计算机使用量子位进行计算，可以将普通计算机需要执行几十年的任务在几秒钟之内完成。目前出现的一些量子算法，如 Shor 算法【引用】和 Grover 算法【引用】已经对互联网中应用广泛的 RSA 算法、ElGamal 算法、ECC 公钥密码算法和 Diffie-Hellamn 密钥协商协议进行有效的密码攻击。如此以来，经典加密算法将受到严重的威胁，虽然在短时间内量子计算机的硬件成本和理论模型实际运作的难度不会让量子计算机真正的破解已经在商用的密码算法，但是为了防范未来量子计算机的攻击，许多种防御量子计算的加密算法也在研究之中。

比如：基于 Hash 函数的公钥密码体制、基于格问题的公钥密码体制、基于多变量问题的公钥密码体制、基于编码问题的公钥密码体制。本文主要讨论的是基于编码问题的公钥密码体制。作为抵御量子计算机攻击的算法，基于编码的加密算法的理论基础却不是来源于量子物理，它的理论来源是信息论，编码理论，代数理论等数学知识。

基于编码的加密算法，就不会受到 Shor 算法或者 Grover 算法的影响，从而保证网络通信的安全。50 年代，随着 C.E Shanon《通信的数学理论》的发表，信道编码定理给出了提高多类信道上传输消息的效率，采用性质良好的纠错码的指导。60 年代纠错码的研究进入快速发展期，期间有广泛应用的汉明码、Reed-Muller 码、BCH 码、Goppa 码等等。纠错码具备的检查错误或纠正错误的能力，被很好的应用到了公钥密码体制中。1978 年，McEliece 首次提出基于编码的公钥加密方案，采用可以快速译码的 Goppa 码，安全性依赖于一般线性码译码问题（NP-完全问题）。在一些已知的攻击算法中，其工作因子都是在 2^{70} 以上，具备较高的安全性。其变形方案 Niderreiter 公钥密码体制在公钥私钥设置中有所不同，但在安全性上被证明是等价的。

在之后的研究中，基于编码的加密方案在码的选择和公钥的构造方式上做了很多尝试，目的就是为了减小公钥大小和提升算法效率，使得在实际中快速落地，获得更好的发展，以做好应对几年之后量子攻击的准备。综上所述，研究基于编码的加密方案具有十分重要的意义。

1.2 国内外研究现状

公钥密码技术经历近五十年的发展，已经被广泛的应用到计算机系统领域，互联网通信领域中。工业界应用公钥加密的案例数不胜数。网络安全访问的实现就是集成了 RSA 的加密与解密功能，极大的解决了网络中的密钥分发与密钥管理的问题。相对于对称密码在数字签名领域的局限性，公钥密码技术可以很好的实现唯一性，私有性等数字签名的要求，通过私钥签名由对应的公钥去验证，其他

人无法冒名顶替，无法篡改签名内容，从而提供网络中的数字签名服务。在身份鉴别领域，通过公钥密码技术实现的方案在执行起来也比对称密码简单得多。公钥密码技术的安全性，基于两大数论难题：① 大整数分解的难解性问题，② 离散对数问题的难解性。后续在经典数论的基础上，RSA，ECC，Rabin 等算法都在不断的改进中，以适应越来越有挑战的密码攻击。但是随着 Shor，Grover 等量子算法针对性的出现，建立在经典数论的公钥密码技术处于集体沦陷的趋势，此时抗量子密码体制逐渐成为公钥密码研究的方向之一，目的就是为了防范量子计算机算法带来的安全威胁。

基于编码的加密算法，之所以能够抵御量子计算机的攻击，是因为量子攻击算法只能攻破上述两大难题，而攻不破 NP-完全问题。McEliece 在 1978 年首次提出基于编码的加密方案，它的安全性假设是二元随机码的译码问题以及 Goppa 码的随机码的区分问题，这两个问题在相关专家学者的研究中被证明是 NP-完全问题，原始的 McEliece 方案经过三十多年的密码分析，被认为是目前为止最安全的公钥密码方案之一。McEliece 方案采用一个随机二元不可约 Goppa 码作为私钥，也就是说保密 Goppa 码的生成矩阵，公钥是对生成矩阵进行混淆和交换后公开的随机生成矩阵。加密过程中，就是将明文加密成合法码字，并加入尽可能多可纠错的差错向量发送给消息接收者。消息接收者就可以使用私钥，消除密文中冗余的错误信息，从而正确译码。相对于 RSA，McEliece 方案因为 Goppa 码拥有快速译码的算法，在加解密的运算中都有高效的优势。但是公钥尺寸过大的问题，一直是 McEliece 方案无法真正取代经典公钥密码算法的主要原因之一。

Niderreiter 方案是 McEliece 方案的一种对偶变形，是一种背包型的密码体制。与 McEliece 方案不同，Niederreiter 方案首先运用函数将消息编码成一个错误向量，用错误向量代表消息，私钥同样是一个随机 GRS 码，公钥是码的校验矩阵混淆和交换后的矩阵，加密过程是计算公钥矩阵和错误向量一个伴随式，解密的时候利用 GRS 码的伴随式译码算法，恢复明文消息。Niderreiter 方案与 McEliece 方案的安全性是等价的，设计 Niderreiter 的初衷是为了缩小公钥的规模，在存储公

钥的时候只需要存储公钥的冗余部分。但是在将明文映射到差错向量时，加解密的速度较慢。

后续的研究中，相关专家学者在码的选择上做了充分的分析。因为 GRS 码密钥的紧凑性，基于 GRS 码的加密方案一经提出就被认为是比 Goppa 码更适合的方案。但根据【引用】的密码分析结果，基于 GRS 码的加密方案已经是不够安全的。在【引用】中，Gabidulin 编码也已经应用到 McEliece 加密方案中，Gabidulin 码在度量的选择上与其它方案不同，前者是基于秩度量，后者都是基于汉明度量。实际上因为 Gabidulin 码在 Frobenius 自同构下包含巨大的向量空间不变性导致结构上的弱点明显，基于秩度量的方案在安全性上也不可靠。1994 年，SideInikov 在他的研究中使用 Reed-Muller 码字构造了公钥加密算法，该算法具有非常高效的解码算法。后续还有很多其它的基于编码的密码方案。到目前为止，安全参数下的 McEliece 方案仍未被打破。另外，在【引用】中提出了 McEliece（或 Niderreiter）方案的数字签名方案。在【引用】中，低密度奇偶校验码 LDPC 表示前向纠错技术，并允许接近香农理论极限的现有好性质的码。它可以减小密钥大小，提高传输效率。但是在【引用】的分析中，它所特有的稀疏公钥矩阵可能会暴露私有代码结构属性。目前，有人建议使用中等密度奇偶校验码 MDPC 来设计公钥密码系统，希望在所提出方案的密钥安全性和公钥大小之间找到良好的平衡。尽管这些基于编码的加密或签名方案没有被完全攻破，但是这些方案因为安全参数下公钥尺寸过大，所以并不适应作为标准的加密程序。因此，我们迫切需要设计基于编码的紧凑公钥和私钥结构混淆良好的加密方案。

1.3 本文工作与主要贡献

2016 年，Baldi 等人提出了 McEliece 密码系统的新变体（简称 BBCRS 方案）。其中，McEliece 中置换矩阵 P 由低秩矩阵 R 和广义置换矩阵 T 构造的矩阵 Q 替代。这使得 BBCRS 方案能够实现私钥不再与公钥存在置换相等的特性，从而允许在 BBCRS 方案中采用一些高性能的码（例如 GRS 码）来减小公钥大小。不幸

的是，Baldi 等人提出的方案的生成密钥的两个案例都是不安全的。另一方面，我们注意到解密过程中需要猜测满足条件的向量以消除在解密过程期间添加的差错向量的影响，这将会是很大的工作量，尤其是当有限域的元素很多时。

受 BBCRS 方案的影响，我们提出了一种新的公钥生成方法，避免了 BBCRS 方案在解密过程中需要的大量试错操作。此外，我们的新方案中的解密方式能够最大化码字的纠错能力来处理所接收的密文，从而增加消息的传输率。因此，我们可以考虑减小消息长度以获得适当大小的密钥。此外，我们提出了构造矩阵 R 的新方式，其在效率和存储方面与构建在 BBCRS 方案中建议的第一种情况相当。这种新方式的主要优点是它可以更好地隐藏密钥结构。这使得在不降低安全级别的情况下，能够在方案中采用强结构的码族。

1.4 研究重点与组织结构

第一章为绪论，主要介绍了抗量子密码的研究背景及意义，也讨论了国内外的研究现状，接着对文章的主要工作做了基本概括，最后对文章的章节做了介绍。

第二章是预备知识与概念的梳理，包括编码理论基础知识、解码算法、常见的码的性质的简要介绍。

第三章主要是对经典的基于编码的加密方案，进行系统概述。第四章着重介绍本文实现的改进方案，原理及分析等。

第五章讨论了解码算法对加密方案的影响。

第六章总结全文内容，并展望未来研究工作的方向。

第二章　预备知识与概念

本章介绍基于编码的加密方案涉及到的有限域知识，编码知识等等。有限域理论是编码理论的一个重要数学基础，有限域上的元素，多项式更好的描述了码字在加解密中的计算过程，编码知识则是加密算法的理论依据，能够选择好性质的码是保证加密算法安全和高效的前提。

2.1　有限域基础理论

定义 2.1.1 (有限域) 在有限集合 **F** 上定义了两个二元运算：加法“$+$”和乘法“$\bullet$”，如果 $(\mathbf{F}, +)$ 是交换群，**F** 的非零元素对乘法构成交换群，而且乘法对加法满足分配律，则称 $(\mathbf{F}, +, \bullet)$ 是有限域，如果 **F** 集合元素的个数为 q，则记作 $\mathbb{F}_q$ 或者 $GF(q)$。简单起见，我们将二元域记作 $\mathbb{F}$。

定义 2.1.2 (本原元) 一个数域 $GF(q)$，具有最大阶的域元素为本原元，即本原元为 a，则 $a^d = 1(mod\, q)$ 成立，其中 $d = \psi(q)$，$\psi(q)$ 是欧拉函数。

定义 2.1.3 (不可约多项式) 有限域 $\mathbb{F}$ 上定义的多项式集合 $F[x] = \{f(x)|f(x) = a_n x^n + ... + a_1 x + a_0, a_i \in \mathbb{F}, a_n \neq 0, n \geq 0\}$，$F[x]$ 中次数大于 *1* 的多项式 $f(x)$ 不能写成两个低次多项式的乘积，称 $f[x]$ 是 **F** 上不可约多项式。

定义 2.1.4 (极小多项式) 设 $\mathbb{F}_q$ 是一个含有 q 个元素的有限域，$\mathbb{F}_p$ 是 $\mathbb{F}_q$ 的一个含有 p 个元素的子域，$\alpha \in \mathbb{F}_q$。$\mathbb{F}_p$ 上的以 α 为根，首项系数为 1，并且次数最低的多项式称为 α 在 $\mathbb{F}_p$ 上的极小多项式。这里 1 是 $\mathbb{F}_p$ 的单位元。

定义 2.1.5 (本原多项式) 设 $\mathbb{F}_(q^n)$ 是一个含有 q^n 个元素的有限域，$\mathbb{F}_q$ 是 $\mathbb{F}_(q^n)$ 的一个含有 q 个元素的子域。设 $\alpha \in \mathbb{F}_(q^n)$ 为 $\mathbb{F}_(q^n)$ 的一个本原元。α 在 $\mathbb{F}_q$ 上的极小多项式称为 $\mathbb{F}_q$ 上的一个本原多项式。

2.2 编码基础知识

基于编码的加密方案的设计，需要考虑编码理论的相关定理，码的性质，纠错能力影响因素等，本小节做基本的概念介绍。在本文中介绍编码理论我们集中在二元域 $\mathbb{F} = GF(2)$ 上。

定义 2.2.1 (汉明重量与汉明距离) 一个码字 $\boldsymbol{x}$ 的汉明距离定义为码字本身含有的非零位的个数，并表示为：$wt(\boldsymbol{x})$。汉明距离指的是两个长度相同的码字 $\boldsymbol{x},\boldsymbol{y}$ 之间，位不同的总数，一般记作：$dist(\boldsymbol{x},\boldsymbol{y})$。可以看出，$dist(\boldsymbol{x},\boldsymbol{y}) = wt(\boldsymbol{x}-\boldsymbol{y})$。

定义 2.2.2 (线性码) 向量空间 $\mathbb{F}^n$ 上的一个 k 维子空间定义为 $[n,k]$ 线性码，记作：$\mathcal{C}$。另外，线性码 $\mathcal{C}$ 的最小距离指的是，线性码中任意两个不同的码字之间的最小汉明距离，记作 d。我们可以用 $[n,k,d]$ 来表示最小距离为 d 的线性码 $\mathcal{C}$，且最大纠错能力为 $\lfloor (d-1) \rfloor /2$。

定义 2.2.3 (生成矩阵与校验矩阵) 线性码 $\mathcal{C}$ 的生成矩阵是指一个 $k \times n$ 的矩阵 G，G 的行可以构成线性码 $\mathcal{C}$ 的基。也就是说由矩阵 G 的行可以线性组合成线性码 $\mathcal{C}$ 中所有的码字。矩阵 G 的系统形式，就是通过矩阵变换的前 k 列组成单位矩阵。线性码 $\mathcal{C}$ 校验矩阵 H 是指生成矩阵的对偶形式，形状即 $(n-k) \times n$。

定义 2.2.4 (解码算法) 在基于编码的加密方案中，当我们选定一种 $[n,k,d]$ 线性码 $\mathcal{C}$ 时，都有一个相应的解码算法 $D_{\mathcal{C}}$，完成纠正加入差错向量的码字的工作，也就是说，对任意的 $\boldsymbol{e} \in \mathbb{F}^n$, $wt(\boldsymbol{e}) < d/2$, $\boldsymbol{x} \in \mathcal{C}$，都有

$$D_{\mathcal{C}}(\boldsymbol{x}+\boldsymbol{e}) = \boldsymbol{x}。$$

2.3 常见码的构造

编码理论指导我们寻找一些好码，使得信源信息经过编码后的，通过信道传输，在信道接收端可以实现自动纠错和检错。良好的纠错检错能力对基于编码的加密算法的作用是很关键的，应用表现好的码，就能减小方案中的码长，从而缩减公钥尺寸，极大的促进基于编码的加密方案的实行。本小节我们介绍几种常见的码，便于从中发现提升码性质的技术和方向。

2.3.1 GRS 码

GRS 码，即广义 Reed-Solomon 码，因为其码结构紧凑的优势，在早期作为 Goppa 码的竞争者应用在基于编码的加密方案中。这是一种特殊的 BCH 码，BCH 码是一种性质良好的循环码，首先循环码的定义是：

定义 2.3.1 (循环码) 设线性码 $\mathcal{C}$，如果线性码 $\mathcal{C}$ 的任意一个码字的循环移位还是一个码字，即当 $a_0a_1\cdots a_{n-1}\in\mathcal{C}$ 时，$a_{n-1}a_0a_1\cdots a_{n-2}\in\mathcal{C}$，则称 $\mathcal{C}$ 是一个循环码。

BCH 码是由三位学者 (R. C. Bose, D. K. Ray-Chaudhuri, A. Hocquenghem) 分别独立提出的，当码长不是很长时，纠错性能非常接近于理论值。BCH 码构造方便，且编码和译码过程容易，非常具有研究价值。

定义 2.3.2 (BCH 码) 设 $\mathbb{F}_q$ 上的一个 r 维向量空间为 $\mathbb{F}_{q^r}$，并且 $1,\alpha,\alpha^2,...,\alpha^{r-1}$ 是 $\mathbb{F}_{q^r}$ 在 $\mathbb{F}_q$ 上的一组基，则 $\mathbb{F}_{q^r}$ 中的任意一个元素都可以唯一地表示为 $1,\alpha,\alpha^2,...,\alpha^{r-1}$ 的一个线性组合。令

$$B_q(n,\delta,\alpha)=\{\mathbf{c}=c_0c_1c_2\cdots c_{n-1}|\mathbf{c}H^{\mathrm{T}}\}，\text{其中 } 1<\delta<n，\text{且}$$

$$H=\begin{pmatrix}1 & \alpha & \alpha^2 & \cdots & \alpha^{n-1}\\ 1 & \alpha^2 & (\alpha^2)^2 & \cdots & (\alpha^2)^{n-1}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & \alpha^{\delta-1} & (\alpha^{\delta-1})^2 & \cdots & (\alpha^{\delta-1})^{n-1}\end{pmatrix}. \tag{2.1}$$

我们称 $B_q(n,\delta,\alpha)$ 是码长为 n 并且设计距离为 δ 的 q 元 *BCH* 码。我们对 *BCH* 码做一点推广，假设 $b \geq 0$ 是一个非负整数，将矩阵 H 第一行中 α 替换成 α^b，其它按照规律生成，则称 $B_q(n,\delta,\alpha,b)$ 为广义 *BCH* 码。

定义 2.3.3 (RS 码) 设 $q \geq 3$ 是一个素数的幂次方，码长为 $q-1$ 并且设计距离为 δ 的 q 元 *BCH* 码 $B_q(q-1,\delta,\alpha)$ 称为 q 元 *Reed-Solomon* 码，简称 q 元 *RS* 码，记作 $S(q-1,\delta,\alpha)$，其中 α 是 $\mathbb{F}_q$ 的一个 $q-1$ 阶元素，即 α 是 $\mathbb{F}_q$ 的一个本原元。在广义 *BCH* 码的基础上，广义 *RS* 码也就记作 $S(q-1,\delta,\alpha,b)$。

广义的 Reed-Solomon 码与广义的 BCH 码的关系，如同 RS 码与 BCH 码的关系。

2.3.2 Goppa 码

Goppa 码是 McEliece 方案的首选线性码，而且到目前为止，没有有效的攻击方法可以对其产生威胁。1982 年有学者在【引用】证明 Goppa 码可以达到 Gilbert-Varshamov 界，良好的代数几何结构满足了快速解码的要求。1992 年，在【引用】中讲到一个长度为 n，最小距离为 d 的 Goppa 码，可以在 $O(n^3)$ 的时间复杂度内对含有不超过 $d/2$ 个错误的码字进行解码，在特殊的构造下，时间复杂度可以降到 $O(n^{7/3})$。

定义 2.3.4 (Goppa 码) 假设 $\mathbf{g}(z)$ 是在有限域 $\mathbb{F}_{q^m}$ 上的一个首一多项式，$\mathcal{L} = \{\gamma_0,\gamma_2,...,\gamma_{n-1}\}$ 是 $\mathbb{F}_{q^m}$ 上含有 n 个元素的集合，当 $0 \leq i < n$ 时，满足 $\mathbf{g}(\gamma_i) \neq 0$，经典的 *Goppa* 码 $\Gamma(\mathcal{L},\mathbf{g})$ 指的是线性空间 $\mathbb{F}_q^n$ 中的所有 $(c_0,c_1,...,c_{n-1})$ 满足：

$$\sum_{i=0}^{n-1}(z-\gamma_i)^{-1}c_i = 0. \tag{2.2}$$

的码字集合。其中 $(z-\gamma_i)^{-1}$ 也可以写成 $-\mathbf{g}(z)^{-1}f(z)$，而 $(z-\gamma)f(z) \equiv 1(mod\,\mathbf{g}(z))$，且 $deg(f) < def(\mathbf{g})$。

2.4 常见的攻击模型

不同的攻击模型，攻击成功的难度就有所不同，如果一个密码系统在攻击门槛较低的情况下还能保证数据的完整，可靠与机密性，就说密码系统达到了该攻击模型下的安全等级。首先，我们的安全性是建立在敌手知道我们所使用的密码体制的，其它根据敌手掌握的信息不同，大致分为如下几个攻击模型：

2.4.1 唯密文攻击

这是一种最难的攻击模型，敌手想要攻击我们的密码系统，但是只能获得一串密文，其它的信息一概不知。由此可知，这对密码系统是最基本的要求，也是最容易防范的攻击。但是很多实际情况下，敌手总是能从大量的密文中，或者其它手段获取一定数量的明文，也就说该种攻击模型很少真正被敌手应用。

2.4.2 已知明文攻击

敌手拥有一定的明文串和对应的密文串，目的是发现密码系统的密钥。敌手会最大限度的利用明文串和密文串的对应关系，从中发现规律，以寻求对密钥的破解，一旦密钥被找到，则该密码系统则是完全被攻破。

2.4.3 选择明文攻击

选择明文攻击，允许敌手获取加密机的临时访问权限，敌手可以选定一些明文，并获得经过加密的密文。通常敌手会让一些明文的差别及其微小，以便比较加密后的密文，分析密钥的作用，从而获取加密的钥匙，密钥被找到，则敌手可完全攻破系统。

2.4.4 选择密文攻击

允许敌手可以获得解密机的临时访问权限，敌手就可以选定一些密文，并进行解密，获得相应的明文。在这种模型下，加入密钥没有及时更新，敌手可以获得其想要的密文对应的明文信息，虽然敌手没有完全攻破密码系统，本身完全攻

破的难度都是非常大的，但是敌手也可以达到自己的目的，就是获取自己想要的秘密信息。

2.4.5 自适应性选择密文攻击

在选择密文攻击的基础上，而又与选择密文攻击不同的是，敌手每次都可以根据之前密文解密的结果，决定接下来要解密的密文，也就是第 i 次解密所选定的密文依赖于前 $i-1$ 次密文解密的结果，而不是一次性直接获取一大段密文对应的明文信息。

2.4.6 不可区分性的自适应性选择密文攻击

假设现在有两个明文 $\mathbf{m_0}, \mathbf{m_1}$，分别对应密文 $\mathbf{c_0}, \mathbf{c_1}$，敌手在自适应性选择密文攻击的条件下，对密码系统进行攻击，但是不能直接解密 $\mathbf{c_0}, \mathbf{c_1}$，最后如果能区分出两个明文与两个密文的对应关系，则攻击成功。如果系统能抵御这种攻击，就说是在自适应选择密文攻击下具有不可区分加密。

2.5 本章小结

第三章　经典方案介绍

自基于编码的加密方案提出以来，相关专家和学者不断地改进这一极有希望在后量子时代代替基于传统数论的加密算法的候选者。从 McEliece 方案大致确定了基于编码的密码系统的加解密流程，后续的研究在公钥的设计上做了很多工作，我们需要在不影响私钥的秘密性的同时，又能保证对码结构的混淆不会扩散差错向量的影响，保留最大限度的纠检错能力，以使加解密的效率更高而成本更低。以下先探讨几个经典的基于编码的方案。

3.1　McEliece 方案

本小节要讨论的原始 McEliece 方案，到目前为止仍未被攻破，而且在现在，当时 McEliece 提出的推荐参数在安全和加解密方面同样也适用。但是有所缺憾的是，方案设计的公钥需要大的存储空间，相对于常用的 RSA 加密算法，公钥大小的局限十分明显。参数为 $n = 1024, k = 524, d = 101$ 的 McEliece 加密方案，公钥尺寸大概在 $1024 - bit$ 的 RSA 算法的 260 倍左右。降低公钥尺寸成为了学者研究的热点，但是后续方案在降低尺寸的同时，在安全性上都有所折扣，效率也不尽人意。原始 McEliece 公钥密码是采用 Goppa，先回顾一下 McEliece 方案的算法。

Algorithm 1 McEliece 密钥生成算法

Input: 系统安全参数：$n, t \in N$，其中 $t \ll n$。

Output: 公钥 G^{pub}，私钥 $(S, D_{\mathcal{C}}, P)$。

1: 密钥生成：对于给定的参数 n 和 t，产生下列矩阵。

- 矩阵 G：在有限域 $\mathbb{F}$ 上的信息位数为 k，最小距离为 $d \geq 2t + 1$ 的 Goppa 码 $\mathcal{C}$ 的 $k \times n$ 阶生成矩阵。
- 矩阵 S：$k \times k$ 阶的二元随机非奇异矩阵。

- 矩阵 P：$k \times n$ 阶的二元随机置换矩阵。

2: 然后计算方案的公钥 $G^{pub} = SGP$。

- 公钥：(G^{pub}, t)。
- 私钥：$(S, D_{\mathcal{C}}, P)$，有效译码算法 $D_{\mathcal{C}}$ 就是所用纠错码方案的陷门。

McEliece 在公钥尺寸的表现不够好，但是加密过程十分简单，位操作的复杂度较 RSA 减少很多，加密算法如下：

Algorithm 2 McEliece 加密算法

Input: 公钥 (G^{pub}, t)，长度为 k 的明文 $\mathbf{m}$。

Output: 密文 $\mathbf{c}$。

1: 随机选择一个汉明重量为 t 的随机向量 $\mathbf{e} \in \mathbb{F}^n$.

2: 加密，产生密文：

$$\mathbf{c} = \mathbf{m}G^{pub} + \mathbf{e}.$$

解密过程要利用到编码的纠检错机制，消除加入的差错向量的影响，进而进行解密操作。

Algorithm 3 McEliece 解密算法

Input: 密文 $\mathbf{c}$，私钥 $(S, D_{\mathcal{C}}, P)$。

Output: 明文 $\mathbf{m}$。

1: 解密密文 $\mathbf{c}$ 之前，首先计算:

$$\mathbf{c}P^{-1} = \mathbf{m}SG \oplus \mathbf{e}P^{-1}.$$

2: 然后对其进行译码，因为上一步的计算结果可以看成为是码的一个含有 t 个错误的码字，所以经过译码可以得到：

$$\mathbf{m}SG = D_{\mathcal{C}}(\mathbf{c}P^{-1}).$$

3: 最后，令集合 $J \subseteq \{0, 1, 2, ..., n\}$，需要通过矩阵变换使 G_j^{pub} 可逆，则进行如下计算可以得到明文。

$$\mathbf{m} = (\mathbf{m}SG)_j (G_j)^{-1} S^{-1}.$$

McEliece 方案在安全性上的表现是有优势的，可以达到 INA-CCA 安全。在已知的密码分析方法中，比如区分攻击、信息集攻击等，针对 McEliece 的攻击的工作因子大多在 2^{70} 以上。

3.2 Niederreiter 方案

1986 年，Niderreiter 对 McEliece 公钥密码方案做出改进，提出了一种 Niederreiter 公钥密码方案。该密码方案基于的困难问题也是随机线性码的译码困难问题，只是利用的角度有所不同。Niderreiter 密码方案隐藏了 Goppa 码的校验矩阵，于是在公钥尺寸上有所减少，但是还是没有达到实用的要求。

Algorithm 4 Niederreiter 方案

系统安全参数：$n, t \in N$，其中 $t \ll n$。

密钥生成阶段：

- 矩阵 H：在有限域 $\mathbb{F}$ 上的信息位数为 k，最小距离为 $d \geq 2t+1$ 的 Goppa 码 $\mathcal{C}$ 的 $(n-k) \times n$ 阶校验矩阵。
- 矩阵 A：随机选取的 $(n-k) \times (n-k)$ 可逆矩阵.
- 矩阵 P：随机选取的 $n \times n$ 阶的置换矩阵。

然后计算方案的公钥：$H^{pub} = AHP$，于是公钥为 (H^{pub}, t)，私钥为 $(A, D_{\mathcal{C}}, P)$，$D_{\mathcal{C}}$ 是码的伴随式译码算法。

加密阶段：

Niederreiter 首先将明文字符串映射成一个错误向量 $\mathbf{e}$，重量为 t。加密过程即计算一个伴随式，如下：

$$s = H^{pub}\mathbf{e}^{\mathrm{T}}.$$

加密者向解密者发送密文 s。

解密阶段：

为了解密密文，首先计算：

$$A^{-1}s = HP\mathbf{e}^{\mathrm{T}}.$$

利用码的伴随式译码算法 $D_{\mathcal{C}}$ 恢复出 $P\mathbf{e}^{\mathrm{T}}$，于是得到明文 $\mathbf{e}^{\mathrm{T}} = P^{-1}P\mathbf{e}^{\mathrm{T}}$。

Niederreiter 密码方案公钥尺寸更小，而安全性上被认为与 McEliece 密钥方案一致。但是在将明文映射称错误向量的操作上，会影响加解密的效率。Niederreiter 密码方案为了抵抗经典的信息集译码攻击，往往将参数设置的很大，故还不太适合实际应用。

后续也有大量研究 Niederreiter 的变形方案，其中 Maurich 等人提出的混合加密可以达到 IND-CCA 安全，而且选择 QC-LDPC 码可以在存储公钥的时候大大减少公钥的尺寸。

3.3 BBCRS 方案

2014 年，Marco Baldi、Marco Bianchi、France Chiaraluce、Joachim Rosenthal、Davide Schipani 等人设计了一种提升和加强 McEliece 安全性的方案，简称“BBCRS 方案”。之前及后续的 McEliece 方案变体的核心工作，就是更好的隐藏密钥结构，但是其纠错检错能力又不会大打折扣，如此就可以更好的避免被攻击成功的概率。BBCRS 方案，从原始 McEliece 方案的置换矩阵入手，考虑到置换矩阵暴露的安全性问题，采用一种由两个矩阵相加的形式代替置换矩阵，让多种经典的攻击手法找不到合适的入口，简言之，就是使得私钥结构与公开信息之间不具备置换转化的关系。BBCRS 方案的设计主要贡献在于密钥的生成，私钥结构隐藏更好就可以换取结构性更强的码，以带来公钥的紧凑，尺寸减小的优势。BBCRS 增加了一些系统参数，可以根据实际情况调整，以适应安全性，加解密效率，存储效率的平衡。

BBCRS 的密钥生成与原始的 McEliece 有很大的区别，且组成过程些许复杂。假设两个包含 w 个矩阵元素的集合 $\mathcal{A}$ 和 $\mathcal{B}$，内部元素都是随机选择并且保密的 $z \times n, z \leq n$ 矩阵，而矩阵元素都在有限域 $\mathbb{F}_q$ 上。

$$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_w\},\ \ \mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_w\}.$$

令矩阵 $R = \mathbf{a}_1^{\mathrm{T}}\mathbf{b}_1 + \mathbf{a}_2^{\mathrm{T}}\mathbf{b}_2 + \cdots + \mathbf{a}_w^{\mathrm{T}}\mathbf{b}_w$，则矩阵 R 就是一个密集的 $n \times n$ 矩阵。另一方面，定义一个在有限域 $\mathbb{F}_q$ 上的 $n \times n$ 稀疏矩阵 T，矩阵 T 的平均行重和列重都是一个常数 m, $m \ll n$。当 m 是一个整数时，又叫重量为 m 的广义置换矩阵，m 也可以不是整数，在 BBCRS 方案中，m 的值是在 $[1, 2]$ 范围中的。

最后经过矩阵 R 和 T 的加法计算，就是最终应用到公钥设计的矩阵 Q。即 $Q = R + T$，在原始 McEliece 方案中替换掉置换矩阵 P。在该种构造方式下，BBCRS 方案提出了两套参数设定，在 $w = 2$ 前提下，① $\mathbf{a}_2 = \mathbf{0}$；②$\mathbf{b}_2 = \mathbf{1} + \mathbf{b}_1$。其中 $\mathbf{0}$ 和 $\mathbf{1}$ 分别指的是全零矩阵和全 1 矩阵。在此设定下，保证矩阵 R 的秩为 z。

在进行加解密之前，关于加密过程中随机产生的差错向量还要满足一些性质，才能被正确的解码。对于差错向量 **e**，预定 $wt(\mathbf{e}) \leq \lfloor \frac{t}{m} \rfloor$。其中 m 是矩阵 T 的平均行重或列重，t 是纠错码的最大纠错个数。而且差错向量 **e** 还有如下限制：

$$(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_w) \cdot \mathbf{e}^{\mathrm{T}} = \mathbf{0}. \tag{3.1}$$

BBCRS 的详细算法过程如下：

Algorithm 5 BBCRS 公钥密码体制 McEliece 版本

系统安全参数：$n \in N, t \in N, m, z \in N, w \in N$，其中 $t, m \ll n$。

密钥生成阶段：

- 矩阵 G：在有限域 $\mathbb{F}$ 上的信息位数为 k，最小距离为 $d \geq 2t+1$ 的 Goppa 码 $\mathcal{C}$ 的 $k \times n$ 阶生成矩阵。
- 矩阵 S：随机选取的 $k \times k$ 可逆矩阵。
- 矩阵 Q：如上文所述，由 $R+T$ 计算得出。

然后计算方案的公钥：$G^{pub} = S^{-1}GQ^{-1}$，于是公钥为 (G^{pub}, t, m)，私钥为 $(S, D_{\mathcal{C}}, R, T)$，$D_{\mathcal{C}}$ 是码的译码算法。

加密阶段：

首先随机选取一个满足 $wt(\mathbf{e}) \leq \lfloor \frac{t}{m} \rfloor$ 和式 3.1 的差错向量 **e**，对明文 $\mathbf{m} \in \mathbb{F}^k$ 计算：

$$\mathbf{c} = \mathbf{m}G^{pub} + \mathbf{e}.$$

加密者向解密者发送密文 **c**。

解密阶段：

为了解密密文，首先计算：

$$\mathbf{c}Q = \mathbf{m}S^{-1}G \oplus \mathbf{e}Q = \mathbf{m}S^{-1}G \oplus \mathbf{e}(R+T).$$

接下来的问题就是对 $\mathbf{c}Q$ 进行译码，重点就在于 $\mathbf{e}(R+T)$ 的差错位数是否在码的有效纠错能力之内。首先讨论 $\mathbf{e}R$，在上文提到两种设定下，根据式 3.1，$\mathbf{e}R$ 结果分别是：

$$\text{① } \mathbf{a}_2 = \mathbf{0},\ \mathbf{e}R = 0,$$
$$\text{② } \mathbf{b}_2 = \mathbf{1} + \mathbf{b}_1,\ \mathbf{e}R = \mathbf{e} \cdot \mathbf{a}^{\mathrm{T}} \cdot \mathbf{1}.$$

当 $\mathbf{e}R = 0$ 时，$\mathbf{e}Q = \mathbf{e}T$，由矩阵 T 的平均行重或列重以及差错向量 **e** 的重量范围可知，$\mathbf{e}T$ 可以被正确译码。而针对 $\mathbf{e}R \neq 0$ 的情况，BBCRS 决定在可能值的范围里穷举搜索，直到找到一个值 $\lambda = \mathbf{e}R$，令 $\mathbf{e}(R+T) - \lambda$，从而可以正确译码，并计算出 $\mathbf{m}S^{-1} = D_{\mathcal{C}}(\mathbf{c}Q)$，最后有私钥矩阵 S 右乘，还原明文 **m**。

观察 BBCRS 公钥密码体制的公钥构造可知，矩阵 Q 对私钥进行隐藏和混淆，使得公钥与私钥之间不存在原始 McEliece 方案中的置换关系，从而防止了利用公

私钥之间置换关系进行的区分攻击。BBCRS 公钥密码体制在安全性和加解密效率的表现与矩阵 Q 的选择，矩阵 T 的选择息息相关。具体来说就是，如何保证矩阵 Q 对差错向量 $\mathbf{e}$ 的计算作用不会扩散错误的个数，从而导致解码失败。这就涉及矩阵 R 的构造矩阵集合 $\mathcal{A}$ 和 $\mathcal{B}$ 中元素的关系设定，可以使得 $\mathbf{e}R = 0$ 或者在解码过程可以消除 $\mathbf{e}R$ 的影响。这其中还牵涉到对差错向量的一些限制，一方面是其最大重量减小，一方面要满足式 3.1。式 3.1 在 BBCRS 的文章中做了分析，这是存在子码攻击的威胁的，通过计算可知公钥的一部分子码在式 3.1 的作用下，矩阵 R 的混淆作用完全消失，也就是说式 3.1 可以让公钥的子码与私钥的子码回归到置换相等的关系，这就是方案的私钥秘密优势荡然无存。BBCRS 也给出了解决的办法，但是期间涉及一些穷举和猜测，虽然可以避免公私钥子码之间置换的关系，但是加解密的效率和信息利用率不会很高。

矩阵 T 是另一个影响解码成功率的因素，矩阵 T 可以是一个广义的置换矩阵，与此同时为了加密过程中错误向量的错误位数不超过允许的范围内，对随机选取的差错向量的错误个数做了降低。但是当 m 不是整数时，矩阵 T 的平均行重和列重就无法保证 $\mathbf{e}T$ 的错误个数不会超过解码能力。当差错向量的非零位在计算 $\mathbf{e}T$ 过程中集中作用在行重大于 1 的行，计算结果非零位的总和就可能超过码的纠错能力 t。于是，矩阵 T 的选择下，存在着解码失败的概率。

BBCRS 也提出了 Niederreiter 版本的公钥密码体制。具体流程如下：

Algorithm 6 BBCRS 公钥密码体制 Niederreiter 版本

系统安全参数：$n \in N, t \in N, m, z \in N, w \in N$，其中 $t, m \ll n$。

密钥生成阶段：

- 矩阵 H：在有限域 $\mathbb{F}$ 上的信息位数为 k，最小距离为 $d \geq 2t+1$ 的 Goppa 码 $\mathcal{C}$ 的 $(n-k) \times n$ 阶校验矩阵。
- 矩阵 S：随机选取的 $k \times k$ 可逆矩阵。
- 矩阵 Q：如上文所述，由 $R+T$ 计算得出。

然后计算方案的公钥：$H^{pub} = S^{-1}GQ^{\mathrm{T}}$，于是公钥为 (H^{pub}, t, m)，私钥为 $(S, D_{\mathcal{C}}, R, T)$，$D_{\mathcal{C}}$ 是码的译码算法。

加密阶段：

首先将明文字符串映射成一个错误向量 $\mathbf{e}$，满足 $wt(\mathbf{e}) \leq \lfloor \frac{t}{m} \rfloor$ 和式 3.1，加密过程即计算一个伴随式，如下：

$$\mathbf{c} = H^{pub}\mathbf{e}^{\mathrm{T}}.$$

加密者向解密者发送密文 $\mathbf{c}$。

解密阶段：

为了解密密文，首先利用私钥矩阵 S 计算：

$$\mathbf{c}' = S \cdot \mathbf{c} = H \cdot Q^{\mathrm{T}} \cdot \mathbf{e}^{\mathrm{T}} = H \cdot (\mathbf{e} \cdot Q)^{\mathrm{T}}.$$

前文提到的矩阵 Q 的特殊参数设定，可以让 $\mathbf{e} \cdot Q$ 简化到 $\mathbf{e} \cdot T$。此时解密者获得 $H \cdot T^{\mathrm{T}} \cdot \mathbf{e}^{\mathrm{T}}$，因为 $T^{\mathrm{T}} \cdot \mathbf{e}^{\mathrm{T}}$ 的重量小于等于码 $\mathcal{C}$ 的纠错能力 t，接着进行伴随式解码，并利用私钥矩阵 T 还原由明文映射成的错误向量 $\mathbf{e}$，再次反映射，得到明文 m。

该版本的 BBCRS 密码体制，安全性与 McEliece 版本等同，在公钥尺寸上可以是公钥矩阵转换成系统形式，从而在真正存储公钥的时候，只存储其中的一部分。具体可以参考 BBCRS 的描述。

3.4 本章小结

第四章　基于 BBCRS 公钥密码体制的改进

基于编码的加密方案，后续的改进工作主要集中在缩小公钥尺寸，提升解码效率两大方面。BBCRS 公钥密码体制在安全性提升的同时，其 Niederreiter 版本保证了公钥尺寸与原始 Niederreiter 等同；在安全等级相同的情况下，加解密效率高于基于传统数论的 RSA。为了缩小公钥尺寸，相关学者对各种能应用到 McEliece 加密方案的码进行了大量研究，比如低密度奇偶校验码，但是经过密码分析者的分析工作，这些码应用到 McEliece 加密方案中是不安全的，特别是改进的信息集解码攻击极易对这么设计产生威胁。但是 BBCRS 公钥密码体制的公钥设计，在一定程度上可以对码作出妥协，比如可以选择结构更紧凑的 GRS 码，在安全性上依然与应用 Goppa 的 McEliece 加密方案基本等同。

BBCRS 公钥密码体制，在公钥的设计上是敢于突破的，不管是 McEliece 及其变体方案，还是 Niederreiter 及其变体方案，在公钥的设计上总是类似的，公钥与私钥之间总是保留着置换的相等关系。BBCRS 选择直接将置换矩阵替换为其它形式的矩阵，在基于编码的公钥构造上提供了一些思路。本章内容，就是根据这种思路，提供了另一种公钥构造方式，该种方式可以达到 BBCRS 公钥密码体制中描述的公钥与私钥之间不再是置换相等的关系，从而在安全性上表现突出，在加解密表现上避免了 BBCRS 公钥密码体制出现的情况，是一种可取的改进方式。

4.1 研究动机

Marco Baldi 等人在 BBCRS 公钥密码体制发表后，总结到方案采用 GRS 码既可以保证安全性的提升，又能带来公钥尺寸的减小。在同一种安全性级别，在加解密的操作复杂度显著低于 RSA 公钥密码体制。但是在 2015 年，有密码分析学者提出的密钥恢复攻击可以在多项式时间内攻破采用 GRS 码的 BBCRS 公钥密码体制。具体来说，密码分析学者提出的攻击方案是针对矩阵 T 平均行重或列重范围在 $[1,2]$ 的情况，利用矩阵 T 行列重量分布的特点，将 BBCRS 公钥密码体制的安全性降低到利用区分者攻击手法可恢复私钥的安全性。

如此以来，在 BBCRS 公钥密码体制的基础上，规避暴露出来的弱点，就是待解决的主要问题。首先就是关于码的选择，BBCRS 公钥密码体制推荐使用 GRS 码，以便更好的利用 GRS 码结构紧凑的特性。但是这也成为后续攻击手段的切入点，如果继而采用原始的 Goppa 码，就必须提升方案的信息利用率，不然在纠错能力 ($\lfloor \frac{t}{m} \rfloor$) 已经有所减少的情况下，对通信效率将是较大的影响。其次，方案在解密过程中，不仅要考虑中间结果超过码的最大纠错能力，还要考虑消除 $\mathbf{e}R \neq \mathbf{0}$ 带来的穷举搜索的重复操作。另外，在加密过程中，随机选取的差错向量，重量已经有所减少，而且要满足 $(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_w) \cdot \mathbf{e}^{\mathrm{T}} = \mathbf{0}.$，这也带来了威胁较大的子码漏洞问题，虽然 BBCRS 公钥密码体制提出了解决之道，但是如果能放开这样的限制条件，或者改为其它的限制，方案的设计会更合理。

BBCRS 公钥密码体制替换原始 McEliece 加密方案的置换矩阵，不仅消除了公钥与私钥之间置换相等的关系，而且矩阵 Q 的组成让我们有更多的细粒度的参数设定，比如 w 的值决定集合中矩阵的个数，z 的值决定矩阵 R 的秩，BBCRS 公钥密码体制 Niederreiter 版本要存储的公钥尺寸，和矩阵 R 的秩是息息相关的。这样的设计可以让我们根据不同的码的特点，空间效率等因素选择最优的参数来实现加解密。本文在设计基于 BBCRS 的改进方案时，也是将置换矩阵替换成多个矩阵计算结果的形式。BBCRS 公钥密码体制在差错向量上的类似于陷门设定，在后续解密简化了步骤，这些都为新方案提供了思想指导。

4.2 改进方向

考虑到设计的可行性，本文主要对 McEliece 版本的 BBCRS 公钥密码体制进行公钥构造形式上的改进，参考以往 McEliece 方案的变体，随机非奇异矩阵 S 的作用基本不变，混淆生成矩阵为随机的矩阵，所以在新方案中，我们也保留这一左乘的形式。公钥的构造关键，就是原始 McEliece 方案的置换矩阵 P 这一部分，明文编码成含有错误码字，在进一步混淆私钥结构的过程中，就要考虑码字错误位数的扩散问题，置换矩阵因为每行每列都只有一个非零位，所以在计算 $\mathbf{e}P^{-1}$ 的时候不会出现错误位数大于 **e** 的汉明重量，即仍然可以正确的进行译码。回顾 BBCRS 公钥密码体制在这部分的考虑，其一是对差错向量有一定的约束，其二是对矩阵 T 在平均行重和列重做了规定，而且分析了解码失败的概率。于是在新公钥构造方式中，如何使约束成本最小化，混淆作用最大化就成了关键。

在 BBCRS 公钥密码体制的改进方向上，基本上确定矩阵 Q 是安全性和效率的关键部分，如何设计一个陷门，能把在加密过程中加入的混淆作用，在解密的时候消除掉是重中之重。首先我们想到可逆矩阵，可逆矩阵无疑是隐藏私钥矩阵的利器，但是我们不能忽略一个可逆矩阵与差错向量的计算结果，也就是说，如果采用可逆矩阵差错向量对汉明重量的要求，即码的纠错能力将毫无意义，因为可逆矩阵的扩散作用非常难以控制。

接下来，虽然可逆矩阵无法直接应用到公钥构造设计中，但是我们可以间接的加入可逆矩阵，也就是可逆矩阵作为参与者，承担一部分的作用。于是，关于替换置换矩阵的部分，就要像 BBCRS 那样，分为若干个部分。BBCRS 中公钥密码体制中，采用的是一个秩固定的密集型矩阵和一个稀疏的广义置换矩阵，通过矩阵相加计算矩阵 Q，在新公钥构造设计中，设想使一个特殊矩阵和可逆矩阵相加，来替换原始 McEliece 加密方案的置换矩阵。经过反复的推导，发现特殊矩阵与可逆矩阵相加这种方式，在解码的时候，无论特殊矩阵怎么努力，都无法吸收可逆矩阵在差错向量 **e** 的扩散作用。但是，在反复推导中，我们逐渐认识到可逆矩阵不仅可以隐藏私钥结构，也可以隐藏差错向量，或者说任意与之做乘法的向

量或矩阵。那么我们是不是可以包装一下可逆矩阵，让其暴露出来一个乘法结果，该乘法结果与差错向量 **e** 满足一定的等式关系，从而在解密的时候可以利用这个等式关系，来达到控制差错向量错误位数的目的。

在改进方向上，我们反复斟酌，最终觉得这是可行的实现方法，下一节重点讨论方案的实现。

4.3 方案设计

在我们讨论的过程中，可逆矩阵基本确定要作为公钥构造设计的一部分，而另外一个特殊矩阵还需详细的分析。在公钥矩阵存储上，低秩矩阵的参与往往可以节约存储空间，但是在安全性上也有所下降，平衡好方案中各因素也是方案可行的关键。首先，密钥生成算法的流程如下：

Algorithm 7 基于 BBCRS 的改进方案密钥生成算法

Input: 系统安全参数：$n, t \in N$，其中 $t \ll n$。

Output: 公钥 $(G^{pub}, U^{-1}V)$，私钥 $(S, D_{\mathcal{C}}, U, V)$。

1: 密钥生成：对于给定的参数 n 和 t，产生下列矩阵。

- 矩阵 G：在有限域 $\mathbb{F}$ 上的信息位数为 k，最小距离为 $d \geq 2t+1$ 的 Goppa 码 $\mathcal{C}$ 的 $k \times n$ 阶生成矩阵。
- 矩阵 S：$k \times k$ 阶的二元随机非奇异矩阵。
- 矩阵 U：$n \times n$ 阶的二元随机可逆矩阵。
- 矩阵 V：$n \times n$ 阶的二元随机低秩矩阵，秩为 r。

2: 然后计算方案的公钥 $G^{pub} = S^{-1}G(U+V)^{-1}U$；接着计算另一公钥矩阵 $U^{-1}V$。

- 公钥：$(G^{pub}, U^{-1}V, t)$。
- 私钥：$(S, D_{\mathcal{C}}, U, V)$，有效译码算法 $D_{\mathcal{C}}$ 就是所用纠错码方案的陷门。

观察公钥的设计可知，可逆矩阵 U，与低秩矩阵 V 相加，使得满足可逆，外侧再右乘可逆矩阵 U，结果必然是一个可逆矩阵。于是我们就达到了以可逆矩阵替换原始 McEliece 加密方案的置换矩阵，相比较于 BBCRS 公钥密码体制的 $R+T$，一个可逆矩阵显然比其混淆性强，与此同时想要进行密钥恢复攻击几乎是不可能的。在接下来的加解密算法中，我们同 BBCRS 一样，需要对差错向量做一些预备限制。

Algorithm 8 基于 BBCRS 的改进方案加密算法

Input: 公钥 $(G^{pub}, U^{-1}V, t)$，长度为 k 的明文 $\mathbf{m}$。

Output: 密文 $\mathbf{c}$。

1: 随机选择一个汉明重量为 t 的随机向量 $\mathbf{e} \in \mathbb{F}^n$，但是要满足：

$$wt(\mathbf{e}) \leq t \quad \text{and} \quad \mathbf{e} \cdot U^{-1}V = \mathbf{0}. \tag{4.1}$$

2: 加密，产生密文：

$$\mathbf{c} = \mathbf{m}G^{pub} + \mathbf{e}.$$

为了确保解密过程含有错误的码字能被正确纠错，也就是码字含有错误的个数不能超过码的纠错能力，本方案对差错向量 $\mathbf{e}$ 做了式 4.1 的要求。式 4.1 本质上式规定了差错向量 $\mathbf{e}$ 的选择空间，之所以要求矩阵 V 是低秩矩阵，是因为低秩矩阵可以保证 $U^{-1}V$ 的解空间足够大，足以保证差错向量 $\mathbf{e}$ 的随机性，避免暴力攻击。

本方案的解密算法具体的实现如下：

Algorithm 9 基于 BBCRS 的改进方案解密算法

Input: 密文 $\mathbf{c}$，私钥 $(S, D_{\mathcal{C}}, U, V)$。

Output: 明文 $\mathbf{m}$。

1: 解密密文 $\mathbf{c}$ 之前，首先利用私钥矩阵 U 计算:

$$\mathbf{c} \cdot U^{-1} = \mathbf{m}S^{-1}G(U+V)^{-1} \oplus \mathbf{e} \cdot U^{-1}.$$

2: 接着利用私钥矩阵 U 与 V 的和 $(U+V)$ 右乘上述结果：

$$\begin{aligned} \mathbf{c} \cdot U^{-1}(U+V) &= \mathbf{m}S^{-1}G \oplus \mathbf{e} \cdot U^{-1}(U+V) \\ &= \mathbf{m}S^{-1}G \oplus \mathbf{e} \oplus \mathbf{e} \cdot U^{-1}V \\ &= \mathbf{m}S^{-1}G \oplus \mathbf{e}. \end{aligned} \tag{4.2}$$

3: 观察上一步的结果，对其进行译码，因为上一步的计算结果可以看成为是码的一个含有 t 个错误的码字，而且码字含有的错误必然不会超过码的纠错能力，所以经过译码可以得到：

$$\mathbf{m}S^{-1}G = D_{\mathcal{C}}(\mathbf{c} \cdot U^{-1}(U+V)).$$

4: 最后，令集合 $J \subseteq \{0, 1, 2, ..., n\}$，需要通过矩阵变换使 G_j^{pub} 可逆，则进行如下计算可以得到明文。

$$\mathbf{m} = (\mathbf{m}S^{-1}G)_j(G_j)^{-1}S.$$

综上，加密的计算方式同原始的 McEliece 加密方案，明文通过编码并加入一些错误处理成合理的码字，解密过程同样要利用到编码的纠检错机制，消除加入的差错向量的影响，才能进而正确进行解密操作。在加密过程中，我们同样对差错向量 **e** 作出了一定的约束，但是 BBCRS 公钥密码体制因此而导致的子码漏洞问题在本方案中是不存在的。我们保证差错向量的选择空间足够大，而且使其纠错能力上升到码的实际纠错能力，而 BBCRS 公钥密码体制在纠错能力上是有所折扣的。

解密工作类似于加密中混淆作用的反过程，但是要注意的是，加密过程中，差错向量是直接追加上去的，并没有参与混淆计算，也就是差错向量的重量和影响一直是差错向量本身带来的影响，而解密过程就大不相同，因为解密过程中，差错向量要跟解密时用到的私钥矩阵相计算，如此以来，一定要找到控制差错向量错误位数的方法，不然解密工作就会走向失败。由式 4.2 我们可以得出结论，本方案是可以正确进行加解密的基于编码的密码方案。

从方案的设计我们可以看到，原始的 McEliece 加密方案的置换矩阵，细化这部分的构造，将有很多的改进空间。如 BBCRS 公钥密码体制在广义置换矩阵 T 的基础上，再加上了一个密集矩阵 R，而矩阵 R 实际上是两个元素都是随机矩阵的集合通过计算而得。这两个矩阵集合的元素，就成为公钥构造的关键因素这也为本方案更细粒度的控制可逆矩阵和低秩矩阵提供了进一步的努力方向，此外，也提供了对于差错向量的控制的指导思路。

4.4 安全性分析

自原始 McEliece 加密方案在 1978 年提出以来，针对其进行的密码分析工作和相关变形方案一直在研究之中。基于编码的公钥加密方案需要满足最基本的安全等级就是被动攻击下的单向安全性，毫无疑问，BBCRS 公钥密码体制以及本文的改进方案都可以达到这样的要求。但是在接近实际的应用中，这样的安全等级是不可靠的，我们必须满足更好的安全要求，比如自适应选择密文攻击下的不可

区分性。BBCRS 在文中已经讨论了要达到自适应选择密文攻击下的不可区分性这种安全级别的方案实现，对安全级别起重要作用的就是替换原 McEliece 加密方案置换矩阵的矩阵 Q 所含有的私钥部分，而本文提出的是更为有效的可逆矩阵 $(U+V)^{-1}U$ 来代替置换矩阵。其次，差错向量的约束，以及后续的解密工作对比 BBCRS 公钥密码体制和本文方案，本文都有安全性和解密效率的提升。接下来我们从常见的对基于编码的加密方案的攻击来阐述本文方案的安全性。

4.4.1 解码攻击

经典的解码攻击是在不需要了解任何码的结构信息情况下，对一个随机线性码进行正确解码。也就是说，敌手不需要知道任何私钥信息，但是可以从密文 **c** 中恢复明文 **m**。首先我们先解释一下传统的伴随式译码攻击。

1. 敌手获得公钥矩阵 G^{pub}，然后根据 $G^{pub}\cdot(H^{pub})^{\mathrm{T}}=\mathbf{0}$ 计算出公开的校验矩阵 H^{pub}；

2. 接下来对密文 **c** 计算其伴随式，过程如下：

$$\begin{aligned}\mathbf{c}(H^{pub})^{\mathrm{T}} &= (\mathbf{m}G^{pub}\oplus\mathbf{e})(H^{pub})^{\mathrm{T}}\\ &= \mathbf{m}G^{pub}(H^{pub})^{\mathrm{T}}+\mathbf{e}(H^{pub})^{\mathrm{T}}\\ &= \mathbf{e}(H^{pub})^{\mathrm{T}}.\end{aligned}\tag{4.3}$$

3. 通过式 4.3 的验证，敌手会努力寻找正确的差错向量 **e**，于是就可以从 $\mathbf{c}=\mathbf{m}G^{pub}+\mathbf{e}$ 式中提取 $\mathbf{m}G^{pub}$，通过高斯消除法就可以进一步获取明文 **m**。

在这个过程中，敌手需要产生所有重量不超过码的纠错能力 t 的差错向量 **e**，就是采用暴力破解法，在敌手尝试攻击我们的系统时，需要尝试验证的工作量可以通过如下计算得到：

$$\sum_{i=1}^{t}\binom{n}{i}\tag{4.4}$$

前文已经提到，BBCRS 公钥密码体制的纠错能力已经不再是码的纠错能力 t，而是 $wt(\mathbf{e}) \leq \lfloor t/2m \rfloor$，于是 BBCRS 公钥密码体制在抵御经典解码攻击时的工作量是：

$$\sum_{i=1}^{\frac{t}{2m}} \binom{n}{i} \tag{4.5}$$

在这里 n 指得是码字的长度，一般情况下，只要 n, t 是正常的参数级别，要想从式 4.4 来暴力破解几乎式不可能的，但是对于 BBCRS 公钥密码体制，式 4.5 显示暴力搜索的空间已经减小很多，而且解码攻击已经加入了很多新的技术来加速攻击的效果。我们的方案在实际执行的时候，差错向量 **e** 的选择空间也不再是式 4.4 所描述的完全空间，而是受式 4.1 的约束和私钥矩阵 V 的秩 r 的影响，暴力搜索差错向量 **e** 需要尝试 2^{n-r} 次，在 McEliece 公钥加密方案推荐的参数下，这将是非常大的数字，所以说在这一点我们改进的方案会更安全。

仅仅在经典的伴随式译码攻击下保证安全性式远远不够的，在基于编码的加密算法的解码攻击中，最常见，也是发展最迅速的就是信息集解码攻击。其主要思想是通过确定 k 个坐标位置，使得生成矩阵对应的 k 列构成满秩矩阵，并且接收到的码字对应的位置都是没有错误的，由此获得传递中的明文信息。对于确定这样 k 个坐标位置的算法很多，比如 Stern 算法，MMT 算法等。但是无论如何，信息集解码攻击都是依赖差错向量非错误位的个数，也就是零位的个数，这样才能使 k 的选取成功概率更大。相较于 BBCRS 公钥密码体制，我们扩展了加密方案差错向量的最大错误位数，对于信息集解码攻击意味着更高的复杂度，从而保证了改进方案的安全性。

一些论文中，提出了加速解码攻击的方案，可以快速的找到正确的差错向量 **e**，以恢复明文信息。其基本的思路是将差错向量 **e** 分成若干个部分，将汉明重量分配到每一部分中，根据各部分汉明重量总和跟差错向量本身的汉明重量相等，从而提高暴力搜索差错向量 **e** 的速度。相关学者已经指出，该种解码攻击手段极易在 q 元域上攻击失败，我们的改进方案是很容易应用到 q 元域上的。不仅如此，

我们可以适当的减小码的长度，以使 t/n 的比值增大，如此以来，信息集攻击想要在没有错误的坐标位置还原明文信息将变得更加困难，所以说我们的方案是可以抵御各类解码攻击威胁的。

4.4.2 密钥恢复攻击

密钥恢复攻击，意味着敌手已经获得了加解密设计的密钥信息，敌手可以自由加密或解密信息，也就是密码系统被完全攻破。大部分情况下，这是很难做到的，而且敌手的野心一般也不需要这么大，敌手有时候只想获得特定的密文信息，而不是耗费大量攻击成本去还原密钥。关于 BBCRS 公钥密码体制在密钥恢复攻击的攻击情况，相关学者在【引用】中就提出可以在多项式时间内攻破采用 GRS 码的 BBCRS 公钥密码体制，该篇文章的主要工作主要使 BBCRS 公钥密码体制的安全性下降，其中的私钥矩阵 T 退化成一个普通的置换矩阵，然后根据已经研究出的攻击手段进行后续攻击。该种攻击针对的情况是很特殊的，其对 BBCRS 公钥密码体制中的参数 m 要求是在 $[1,2]$ 范围内。BBCRS 公钥密码体制就可以通过调整采用的码来抵御这种攻击，或者对 m 的值进行一定程度的优化。同样，我们的新方案也可以直接规避开 GRS 码，其实也可以保留 GRS 码，接下来我们重点讨论一下改进方案在密钥恢复攻击的下的安全层次。

对 BBCRS 公钥密码体制多项式时间内的密钥恢复攻击，目的是为了恢复私钥矩阵 R，T。具体来说，首先敌手会通过算法定位私钥矩阵 T 的行重或者列重大于 1 的位置；然后通过线性变换将其转换为置换矩阵，接下来的攻击工作会利用 GRS 码的结构性质及其平方码的构造，这部分工作详细介绍可以参考【引用】。对我们的改进方案来说，我们的私钥 $(S, D_{\mathcal{C}}, U, V)$ 不含有置换矩阵的成分，非奇异矩阵 U 和低秩矩阵 V 对密钥恢复攻击的工作是一项极大的挑战，特别需要关注的是 $(U+V)^{-1}U$ 的部分，这是一个非奇异矩阵，将使得密钥恢复攻击无处着手，密钥恢复攻击很多时候会用到区分者攻击的手段，下一小节我们详细讨论区分者的攻击情况。

在这里我们讨论相关学者提出的一个对 McEliece 加密方案变体进行高效密钥恢复攻击的方法。该方案变体采用的是附带卷积码信息的 Goppa 码，作者指出，对私钥附带的信息带有很大的随机性，所以其密钥结构已经不在矩阵敌手可以利用的代数结构，私钥的隐藏效果显著，于是就可以采用代数结构较强的码。但是这并不完全有效，在作者的分析中，可以利用公钥寻找一些汉明重量较低的码字，用其还原附带的卷积码结构。实际上，这种高效的密钥恢复攻击对我们的方案也是无效的。上述不安全的方案的私钥生成矩阵经过置换矩阵 P 的作用，紧接着穿插进卷积部分的列，置换矩阵 P 经由我们的私钥矩阵 $(U+V)^{-1}U$ 替换，将会杜绝这种风险。还有一种针对准循环码的密钥恢复攻击，也是利用了置换矩阵参与公钥构造的特点，对于我们的方案公钥的构造方式，此种攻击则不会构成威胁。

4.4.3 区分者攻击

正如我们上文提到的，区分随机线性码与 Goppa 码是 NP 困难问题，但是当我们使用结构特殊或者结构性较强的码时，就有可能遭受区分者攻击，区分者攻击技术在基于编码的加密方案的密码分析中用得很多，目的是为了获取我们的私钥结构信息。抵御区分者攻击的关键就是对私钥进行有效的混淆，在我们的改进方案中，对公钥的构造使得公钥看起来更随机，在其它方案中，置换矩阵的存在总是敌手利用相关技术获得信息的入口，消除置换矩阵的混淆作用时敌手进行攻击的前提。而本文的新方案采用右乘一个可逆矩阵 $(U+V)^{-1}U$ 的做法，这是很有效的，解决根源威胁的方式，我们接下来还是要讨论几种不同的区分者攻击方式，大部分都是针对 GRS 码的分析，我们也推荐采用除 GRS 码以外的码。但是为了公钥尺寸的问题，我们也可以调整参数，使得采用 GRS 码同样保证是安全的。

我们的新方案，推荐采用码率较高的码，以适应码长对公钥尺寸的影响。无独有偶，在【引用】中，作者提出了一种多项式时间内针对高码率的区分者攻击。事实上，该文章攻击的思想是利用与码息息相关的多项式在线性系统中的解空间问题，期间涉及到密钥恢复的一些手法，最后可以由推导的公式判断码的种

类。作者提供了通用随机码，替换代码和 Goppa 代码的显式公式。可以发现这项工作与私有代码的生成矩阵结构密切相关，并且该攻击方法不可避免的利用了 McEliece 加密方案的置换矩阵 P，在不包含置换矩阵的新公钥密码体制下，这将是无效的，我们还可以根据攻击情况调整码的种类选择。调整参数也是我们的预防手段之一。

假设我们采用 GRS 代码，即便我们能够安全地隐藏其结构，我们也必须考虑为它设计的区分攻击。如【引用】中所述，GRS 代码的平方具有异常小的维度，在使用 GRS 代码的一些密码系统中，可以通过计算公钥的平方码的维度来检测公钥的生成矩阵中的随机列信息。在某些情况下，计算各种子码以及公钥的各种穿插版本的平方码维度仍然可以从公钥中获取私钥的一部分信息。它们的共同弱点是密钥生成中的置换操作。不失一般性的情况下，我们的计划在这方面是具有明显的安全优势。

4.5 效率分析

在本节我们讨论新方案在存储效率和计算效率两方面的表现，为了清晰说明新方案的改进效果，我们会结合经典的 McEliece 变体方案及 BBCRS 公钥密码体制的效率分析。我们会利用具体的公式估计改进方案在操作复杂度上的表现。

相比较经典 McEliece 加密方案，BBCRS 公钥密码体制及本文的改进方案都在密钥生成阶段，解密阶段多出一些操作，但是，这些操作都是简单的位运算，而且改进方案是稳定，有序的进行计算的，在这一点上，BBCRS 公钥密码体制会有意外情况需要处理。由上文可知，解码过程有可能解码失败，而且在处理差错向量 **e** 的约束问题时，需要进行穷举操作，对译码的效率非常不利。而在我们的改进方案中，是不需要考虑这些操作的。在密钥生成阶段，我们较经典方案多出一次可逆矩阵的乘法计算，另外我们还计算了 $U^{-1}V$，为了对差错向量进行约束使其合法合理的被正确纠错。其实在产生私钥矩阵很 U, V 之后，上述两个计算可以并行执行，并不会影响正确的结果。

与此同时我们需要额外的 $n \times n$ 的存储空间，也就是存储 $U^{-1}V$。在存储方面较 BBCRS 公钥密码体制显示出了劣势。这对于推广基于编码的加密方案非常不利，因为公钥的尺寸问题正是阻碍方案的发展主要因素之一。幸运的是，我们可以进行其他调整以避免公钥尺寸的膨胀。矩阵 $U^{-1}V$ 的存储大小可以减小到 $n \times n - r \times r$。因为奇异矩阵可以转换为仅包含单位矩阵的部分。我们的改进方案还原了原始 McEliece 加密方案的纠错能力，这意味着每个码字都可以携带更多的信息位，这也使得我们能够适当地减少明文消息的长度，同时也保证通信的信息效率要求。实际上，当我们需要获得一个满足条件 $\mathbf{e}U^{-1}V = \mathbf{0}$ 的随机差错向量 $\mathbf{e}$ 时，我们只是从齐次线性方程的解空间中随机选取一个解。所以说解空间的大小至关重要，我们必须保证差错向量的足够随机化以预防暴力破解。

关于解密过程，我们首先计算 $\mathbf{c}U^{-1}$，然后我们进行矩阵乘法 $\mathbf{c}U^{-1}(U+V)$。根据经典的安全参数，与 McEliece 公钥加密方案的变体相比，本方案需要额外 n 长度向量和 $n \times n$ 矩阵之间的基本乘法。我们需要再次考虑适当的 n 值以加速我们的密码系统。接下来，解码工作将按顺序执行。让我们在这里回顾 BBCRS 公钥密码体制的解决方案，它是在矩阵乘法 $\mathbf{c}Q$ 之后猜测满足 $\mathbf{e}R = \lambda$ 的正确值 λ。随着 z 值的变化，例如 $z = 1$ 增加到 $z = n$ 的过程中，猜测的操作次数将呈指数级增加。假设 q 为有限域 $\mathbb{F}_q$ 中的元素数，每次选择一个可能值 λ 时，解密过程都将进行迭代计算，判断能否解码结果，直到解码操作成功为止。显然，迭代计算阶段涉及解码算法和其他操作，这极大地影响了密码体制的速度，接下来我们会统计实验数据来说明方案之间在各方面的差异。

4.6 本章小结

第五章　PPkNNONED 方案

这一章节将介绍新的 $k-$NN 密文上的隐私保护方案，在接下来的篇幅中将以 PPKNNONED 表示 $k-$NN 密文上的隐私保护方案。在第三章中粗略的对 PPKNNONED 进行了介绍，本章将在第三章介绍的基础上进一步的进行深入扩展。正如之前所提到的，我们假设 Alice's 的数据库包含 n 条记录 $D = \langle t_1, ..., t_n \rangle$ 以及 m+1 个属性，其中 $t_{i,j}$ 表示记录 $t_{i,j}$ 的第 j 个属性值。初始化阶段，Alice 对他的数据库的属性加密，即计算 $E_{pk}(t_{i,j}), 1 \leq i \leq n, 1 \leq j \leq m+1$，其中第 (m+1) 列包含类标签，用 D' 表示加密数据库。假设在将来的分类处理中 Alice's 也将数据库 D' 外包给云。不失一般性，可以假设所有的属性值以及他们的欧几里得距离处于区间 $[0, 2^l)$ 中。另外，设 w 表示在记录集合 D 中唯一的类标签数量。

在问题设定当中，我们假设存在两个非合谋的半可信云服务提供商，分别用 C_1 和 C_2 表示，他们一起形成一个联合云。在这个设定下，Alice 将他加密的数据库 D' 外包给 C_1，并将私钥 sk 外包给 C_2。这里数据库拥有着 Alice 有可能用他自己的私有服务器去取代 C_2。但是，如果 Alice 拥有一个私有服务器，我们可以认为 Alice 就没有必要将数据外包出去。使用 C_2 的主要目的由以下两个原因。(A) 收到计算资源以及专业技术的限制，Alice 的最佳选择就是将数据的管理以及操作任务外包给云来做。例如，Alice 也许想访问他的数据并且使用智能手机或其他任何受计算资源限制的终端设备分析出结果。(B) 假设 Bob 想从 Alice 那里维持他的输入查询以及私人访问模式。在这个例子当中，如果 Alice 使用一个私有服务器，那么他就不得不通过 C_2 执行计算假设，在这得目的就是否定将加密数据外包给 C_1。

通常 Alice 使用私有服务器还是云服务提供商，实际上依赖于他的资源限制。特别的，在问题设定中，我们宁愿使用 C_2，避免出现上述提及的问题。在我们的方案中，Alice 将加密数据外包给云后，将来不需要参加任何计算。

PPKNNONED 方案的主要目标是根据数据库 D' 在隐私保护的前提下对用户的查询记录进行分类。考虑到一个授权用户 Bob 想在 C_1 中基于 D' 对他的查询记录 $q = \langle q_1, ..., q_m \rangle$ 进行分类。PPKNNONED 方案主要包含以下两个步骤：

- 阶段 1——K 近邻安全检索 (SRKNN). 在这一过程中，Bob 首先以密文的形式发送他的查询请求给 C_1。之后，C_1 和 C_2 参与一系列的子协议对输入查询 q 安全检索类标签相对应的 K 近邻。在这一步的最后，K 近邻的加密标签仅 C_1 可知。

- 阶段 2——多数类的安全计算 (SCMC). 就步骤一，C_1 和 C_2 计算在查询 q 的 k 近邻中类标签的多数。在这一步的结尾，仅 Bob 知道和他查询记录 q 所对应的类标签。

Algorithm 10 $PPKNNONED(D', q) \rightarrow c_q$

Require:C_1 有 D' 以及 π;C_2 有 sk;Bob 有 q
Bob:
(a). 计算 $E_{pk}(q_j), 1 \leq j \leq m$
(b). 将 $E_{pk}(q) = \langle E_{pk}(q_1), ..., E_{pk}(q_m) \rangle$ 发送给 C_1
C_1 和 C_2:
(a). C_1 从 Bob 端接收 $E_p(q)$
(b). **for** $i = 1$ **to** n **do**:
    •$E_{pk}(d_i) \leftarrow SSED(E_{pk}(q), E_{pk}(t_i))$
    •$[d_i] \leftarrow SBD(E_{pk}(c'))$
for $s = 1$ **to** k **do**:
    (a). C_1 和 C_2:
        •$([d_{min}], E_{pk}(I), E_{pk}(c')) \leftarrow SMIN_n(\theta_1, ...\theta_n)$ 其中
            $\theta_i = ([d_i], E_{pk}(I_{t_i}), E_{pk}(t_{i.m+1}))$
        •$E_{pk}(c'_s) \leftarrow E_{pk}(c')$
    (b). C_1 :
        •$\triangle \leftarrow E_{pk}(I)^{N-1}$
        • **for**$i = 1$ to n **do**:　　$\tau_i \leftarrow E_{pk}(i) * \triangle$
            $\tau'_i \leftarrow \tau_i^{r_i}, r_i \epsilon_R \mathbb{Z}_N$
        •$\beta \leftarrow \pi(\tau')$; 将 β 发送给 C_2
    (c). C_2 :
        •$\beta'_i \leftarrow D_{sk}(\beta_i), 1 \leq i \leq n$
        • 计算 $U', 1 \leq i \leq n$:
            if $\beta'_i = 0$ **then** $U'_i = E_{pk}(0)$
            else $U'_i = E_{pk}(0)$
        将 U' 发送给 C_1
    (d). $C_1 : V \leftarrow \pi_{U'}^{-1}$
    (e). C_1 和 C_2:
        •$E_{pk}(d_{i,j}) \leftarrow SBOR(V_i, E_{pk}(d_{i,\gamma}))$
$SCMC_k(E_{pk}(c'_1), ..., E_{pk}(c'_k))$

PPKNNONED 方案涉及的主要步骤由算法10给出，详细介绍 PPKNNONED 方案中的每个阶段。

5.1 K 近邻安全检索

在这一步骤中，Bob 首先加密查询 q 属性，也就是说，计算 $E_{pk}(q) = \langle E_{pk}(q_1), ..., E_{pk}(q_m)\rangle$ 并将 $E_{pk}(q)$ 发送给 C_1。阶段 1 所涉及的主要步骤由算法10的步骤 1 到 3 所描述。接受 $E_{pk}(q)$ 时，C_1 私有输入 $(E_{pk}(q), E_{pk}(t_i))$，C_2 私钥 sk 共同参与到 SSED 协议。这里，$E_{pk}(t_i) = \langle E_{pk}(t_{i,1}), ..., E_{pk}(t_{i,m})\rangle, 1 \leq i \leq n$。输出为 q 和 t_i 之间欧几里得距离的平方，用 $E_{pk}(d_i)$ 表示，即 $d_i = |q - t_i|^2$。正如前面所提到的，$E_pk(d_i), 1 \leq i \leq n$ 仅 C_1 知道。我们强调加密向量间的精准欧几里得距离由于涉及到平方根，因此计算是很难完成的。但是，在我们的问题中，由于保留了之间的相对顺序，因此能够计算出欧几里得距离平方。之后使用 SBD 协议，C_1 输入 $E_{pk}(d_i)$ 以及 C_2 安全的计算出 d_i 的各个比特的加密值。在这里，输出值 $[d_i] = \langle E_{pk}(d_{i,1}), ..., E_{pk}(d_{i,1})\rangle$ 只有 C_1 可知，其中 $d_{i,1}$ 和 $d_{i,l}, 1 \leq i \leq n$ 分别为 d_i 的最低和最高有效位。

在这之后，C_1 和 C_2 以迭代的方式计算查询 q 相对应的 k 个近邻的类标签。具体的说，在第一轮迭代过程中，他们计算 $E_{pk}(c'_1)$，第二轮迭代时计算 $E_{pk}(c'_2)$，依此类推下去。这里 c'_s 表示查询 q 的第 s 个近邻的类标签，$1 \leq s \leq k$。在第 k 轮的迭代过程中，仅有 C_1 知道 $\langle E_{pk}(c'_1), ..., E_{pk}(c'_k)\rangle$。首先考虑第一次迭代。$C_1$ 和 C_2 共同计算 $d_1, ..., d_n$ 中最小值的个体比特位的加密值，并使用 $SMIN_N$ 协议加密位置以及 d_{min} 相对应的类标签，也就是 C_1 输入 $(\theta_1, ..., \theta_2)$ 同 C_2 的密钥 sk 计算 $([d_{min}], E_{pk}(I)), E_{pk}(c')$，其中 $\theta_i = ([d_i], E_{pk}(I_{t_i}), E_{pk}(t_{i,m+1})), 1 \leq i \leq n$。这里 d_{min} 表示 $d_1, ..., d_n$ 中最小的值；I_{t_i} 以及 $t_{i,m+1}$ 分别表示唯一标识符以及数据记录 t_i 相对应的类标签。具体的说，$(I_{t_i}, t_{i,m+1})$ 是和 t_i 相关的秘密信息。简单期间，假设 $I_{t_i} = i$。在输出阶段，I 和 c' 表示索引值以及 d_{min} 对应的类标签。输出值 $([d_{min}], E_{pk}(I), E_{pk}(c'))$ 只有 C_1 可知。C_1 在本地执行以下操作：

- 分配 $E_{pk}(c')$ 给 $E_{pk}(c'_1)$。根据 SMIN 协议，c' 等价于和 d_{min} 对应的数据记录的类标签。因此，q 的大部分近邻的类标签都是一样的。

- 计算 I 和 i 区别的加密值，其中 $1 \leq i \leq n$。即，C_1 计算 $\tau_i = E_{pk}(i) * E_{pk}(I)^{N-1} = E_{pk}(i-I), 1 \leq i \leq n$。

- 随机化 τ_i 得到 $\tau'_i = \tau_i^{r_i} = E_{pk}(r_i * (i-1))$，其中 r_i 为 $\mathbb{Z}_N$ 中的随机数。$\tau'_i, 1 \leq i \leq n$ 是 0 或者随机数的加密结果值。另外值得注意的是在 τ' 中恰好有一项是 0 的加密结果值 (当且仅当 $i = I$)，剩下的为随机数的加密结果值。使用随机置换函数 π(π 仅 C_1 可知) 排列 τ' 得到 $\beta = \pi(\tau')$ 并发送给 C_2。

在收到 β 后，C_2 解密其中的固定组件得到 $\beta'_i = D_{sk}(\beta_i), 1 \leq i \leq n$。之后，$C_2$ 计算一个加密向量 U' 的长度 n，如果 $\beta'_i = 0$ 则计算 $U'_i = E_{pk}(1)$，否则等于 $E_{pk}(0)$。由于在 τ' 中有一项为 0 的加密值，这进一步说明了，U' 中确切有一项为 1 的加密值，剩下的为 0 的加密结果值。需要特别注意的是，如果 $\beta'_k = 0$，则 $\pi^{-1}(k)$ 是 d_{min} 对应的数据记录的索引值。之后 C_2 将 U' 发送给 C_1，接收到 U' 后，C_1 执行逆置换，有 $V = \pi^{-1}(U')$。在 V 中确切有一项是 $E_{pk}(1)$。剩余的为 0 的加密结果值。另外，如果 $V_i = E_{pk}(1)$，则 t_i 是最接近于 q 的元组。但是，C_1 和 C_2 不知道 V 中的哪一项和 $E_{pk}(1)$ 相对应。

最终，C_1 根据以下原因更新距离向量：

- 值得注意的是，在进一步的计算中距查询 q 的第一个近元组应该会被排除在外。但是，由于 C_1 不知道记录对应的 $E_{pk}(c'_1)$，我们需要在下一轮迭代的时候消除再次选择这个记录的可能性。为此，C_1 放弃更新对应 $E_{pk}(c'_1)$ 距离到最小值，即 $2^l - 1$。更具体的，C_1 在 C_2 的帮助下使用 SBOR 协议更新距离向量，$E_{pk}(d_{i,\gamma}) = SBOR(V_i, E_{pk}(d_{i,\gamma}))$，$1 \leq i \leq n, 1 \leq \gamma \leq l$。当 $V_i = E_{pk}(1)$ 时，对应的距离向量 d_i 是设置的最小值。即在这里例子下，有 $[d_i] = \langle E_{pk}(1), ..., E_{pk}(1) \rangle$。另一方面，当 $V_i = E_{pk}(0)$ 时，OR 操作不会影响相应的加密距离向量。

Algorithm 11 $SCMC_k(E_{pk}(c'_1), ..., E_{pk}(c'_k)) \rightarrow c_q$

Require:$\langle E_{pk}(c_1), ..., E_{pk}(c_w)\rangle, \langle E_{pk}(c'_1), ..., E_{pk}(c'_k)\rangle$ 仅 C_1 可知
　　sk 仅 C_2 可知
C_1 和 C_2：
(a). $\langle E_{pk}(f(c_1)), ..., E_{pk}(f(c_w)) \leftarrow SF(\Lambda, \Lambda')\rangle$, 其中 $\Lambda = \langle E_{pk}(c_1), ..., E_{pk}(c_w)\rangle$
　　$\Lambda' = \langle E_{pk}(c'_1), ..., E_{pk}(c'_k)\rangle$
(b). **for** $i = 1$ to w **do**:
　　$[f(c_i)] \leftarrow SBD(E_{pk}(f(c_i)))$
　　end for
(c). $([f_{max}], E_{pk}(c_q) \leftarrow SMAX_w(\psi_1, ..., \psi_w))$, 其中
　　$\psi_i = ([f(c_i)], E_{pk}(c_i)), 1 \leq i \leq w$
C_1:
(a). $\gamma_q \leftarrow E_{pk}(c_q) * E_{pk}(r_q)$, 其中 $r_q \in_R \mathbb{Z}_N$
(b). 将 γ_q 发送给 C_2，r_q 发送给 Bob
C_2:
(a). 从 C_1 接收 γ_q
(b). $\gamma'_q \leftarrow D_{sk}(\gamma_q)$; 将 γ'_q 发送给 Bob
Bob:
(a). 从 C_1 接收 r_q,C_2 接收 γ'_q
(b). $c_q \leftarrow \gamma'_q - r_q mod N$

上述的过程将会进行 k 轮迭代，在每轮迭代 $[d_i]$ 对应的当前选择标签设置为最小值。但是，C_1 和 C_2 不知道哪个 $[d_i]$ 被更新了。在第 s 轮迭代，$E_{pk}(c'_s)$ 仅 C_1 可知。在第一阶段结束时，C_1 有查询 q 的 k 个近邻的加密标签列表 $E_{pk}(c'_1), ..., E_{pk}(c'_k)$。

5.2 多数类的安全计算

不失一般性，假设 Alice's 数据集 D 包含 w 个不重复的类标签 $c = \langle c_1, ..., c_w\rangle$。Alice 将加密后的类列表外包给 C_1。即在数据的外包阶段，Alice 将 $\langle E_{pk}(c_1), ..., E_{pk}(c_w)\rangle$ 以及加密数据库外包给 C_1。为了安全起见，Alice 也许会在列表中添加一些假的分类来保护一定数量的类标签，即 w 来自于 C_1 和 C_2。但是，为了简单起见，我们假设 Alice 不在列表中添加任何的假分类。

第二阶段期间，C_1 同私有输入 $\Lambda = \langle E_{pk}(c_1), ..., E_{pk}(c_w)\rangle$、$\Lambda' = \langle E_{pk}(c'_1), ..., E_{pk}(c'_k)\rangle$,

以及 C_2 的密钥 sk 安全计算 $E_{pk}(c_q)$。这里 c_q 表示 $c'_1, ..., c'_k$ 中的多数类标签。在第二阶段末尾，仅 Bob 知道类标签 c_q。

算法11描述了阶段 2 设计的步骤。首先，C_1 和 C_2 使用 k 个邻近集合作为输入共同对每个类标签出现的频率进行加密。即，他们使用 (Λ, Λ') 计算 $E_{pk}(f(c_i)), 1 \leq i \leq w$ 作为 SF 协议的输入。输出值 $\langle E_{pk}(f(c_1)), ..., E_{pk}(f(c_w))\rangle$ 仅 C_1 可知。之后 C_1 的 $E_{pk}(f(c_i))$ 以及 C_2 的 sk 参与安全比特分解协议，从而来计算 $[f(c_i)]$，也就是说，对 $f(c_i), 1 \leq i \leq w$ 的各个为进行矢量加密。之后，C_1 和 C_2 共同参与 $SMAX_w$ 协议。简略的，$SMAX_w$ 利用 SMAX 的子规则以迭代的方式最终计算 $([f_{max}], E_{pk}(c_q))$。这里 $[f_{max}] = [max(f(c_1), ..., f(c_w))]$,c_q 表示 Λ' 之外的多数类。结束时，输出值 $([f_{max}], E_{pk}(c_q))$ 仅 C_1 可知。这步之后，C_1 计算 $\gamma_q = E_{pk}(c_q + r_q)$，其中 r_q 是 $\mathbb{Z}_N$ 中的随机数，仅 C_1 可知。接着 C_1 将 γ_q 发送给 C_2，r_q 发送给 Bob。C_2 收到 γ_q 后，解密得到随机多数类标签 $\gamma'_q = D_{sk}(\gamma_q)$ 并发送给 Bob。最终，在从 C_1 接收到 r_q，从 C_2 收到 γ'_q 后，Bob 计算 q 对应的输出类标签 $c_q = \gamma'_q - r_q mod N$。

5.3 复杂性分析

阶段 1 的计算复杂度主要来自于加密和求幂，复杂度为 $O(n * (l + m + k * l * log_2 n))$。另一方面，阶段 2 的计算复杂度同样来自于加密和求幂，复杂度为 $O(w * (l + k + l * log_2 w))$。一般情况下，当 $w << n$，则阶段 1 的计算复杂度应该明显高于阶段 2。

5.4 本章小结

本章详细介绍了 PPKNNONED 隐私保护方案，对 PPKNNONED 方案进行了算法描述。对 K 近邻安全检索进行了详细描述。另外对多数类别的安全计算进行了描述，也给出了相关的算法描述，最后对复杂性进行了分析。

第六章　总结与展望

随着智能终端的普及，移动互联网进入了快速发展时期。GPS 定位技术以及无线通信的日以完善给人类的衣食住行提供了极大的便利，如今在交通、医疗、教育等这些和民众生活息息相关的服务行业无时无刻不依赖着智能终端、无线通信等技术的支持。基于位置的服务也作为新起之秀被终端民众所追捧。由于用户的位置也许会揭露出一些敏感个人信息，而用户的当前位置完全暴露在 LBS 服务提供商面前，因此对用户的位置隐私构成了威胁，如何保护用户的位置隐私，并能够在保护用户隐私的同时还能给用户提供相应的位置服务是当下研究的热点话题。

为了保护用户的隐私，在过去的几年中出现了各种各样的隐私保护分类技术。但是这些数据不适合将加密数据外包至第三方的场景。在本篇文章中，我们提出了一种新的，且适合云端加密数据的隐私保护 KNN 分类方案。方案能够保证数据以及用户查询输入的隐私性，并且能够隐私数据访问模式。

在 PPKNNONED 性能方面有所欠缺，组要涉及到方案中的第一步 $SMIN_n$ 的效率性。因此在接下来的工作中，将会研究各种关于 $SMIN_n$ 解决方案，并用来扩展到我们的分类算法中。

参考文献

[1] SWEENEY L. k-anonymity: a model for protecting privacy[J]. International Journal on Uncertainty,Fuzziness and Knowledge-based System, 2002, 10(5): 557 – 570.

[2] RONGXING L, XIAODONG L, TOM H L, et al. Pseudonym changing at social spots: An effective strategy for location privacy in vanets[J]. IEEE Transactions on Vehicular Technology, 2012, 61(1): 86 – 96.

[3] GENTRY C. Fully homomorphic encryption using ideal lattices[C] // STOC. 2009: 169 – 178.

[4] GOLDWASSER S. Multi party computations: past and present[C] // Proceedings of 16th Annual ACM Symposium on Principles of Distributed Computing. [S.l.]: ACM, 1997.

[5] DWORK C. Differential privacy[M]. [S.l.]: Automata, languages and programming. Springer Berlin Heidelberg, 2006: 1 – 12.

[6] SWEENEY L. Achieving k-anonymity privacy protection using generalization and suppression[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 571 – 588.

[7] MYLES G, FRIDAY A, DAVIES N. Preserving privacy in environments with location-based applications[J]. IEEE Pervasive Computing, 2003, 2(1): 56 – 64.

[8] YOUSSEF M, ATLURI V, ADAM R N. Preserving mobile customer privacy: An access control system for moving objects and customer profiles[C] // Proceedings of the 6th international conference on Mobile data management. [S.l.] : ACM, 2005 : 67 – 76.

[9] BERESFORD R A, STAJANO F. Location privacy in pervasive computing[J]. IEEE Pervasive computing, 2003, 2(1) : 46 – 55.

[10] BAMBA B, LIU L, PESTI P. Supporting anonymous location queries in mobile environments with privacygrid[C] // Proceedings of the 17th international conference on World Wide Web. [S.l.] : ACM, 2008 : 237 – 246.

[11] CHOW Y C, MOKBEL F M, LIU X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service[C] // Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems. [S.l.] : ACM, 2006 : 171 – 178.

[12] MOKBEL F M, CHOW Y C, AREF W G. The new Casper: query processing for location services without compromising privacy[C] // Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment. 2006 : 763 – 774.

[13] YIU L M, JENSEN C S, HUANG X. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services[C] // In Proc.ICDE. 2008 : 366 – 375.

[14] SHANKAR P, GANAPATHY V, IFTODE L. Privately querying location-based services with SybilQuery[C] // Proceedings of the 11th international conference on Ubiquitous computing. [S.l.] : ACM, 2009 : 31 – 40.

[15] HU H, XU J, REN C. Processing private queries over untrusted data cloud through

privacy homomorphism[C] // Data Engineering (ICDE). [S.l.] : IEEE, 2011 : 601 – 612.

[16] KHOSHGOZARAN A, SHAHABI C. Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy[R]. [S.l.] : Advances in Spatial and Temporal Databases. Springer Berlin Heidelberg, 2007 : 239 – 257.

[17] GHINITA G, KALNIS P, KHOSHGOZARAN A. Private queries in location based services: anonymizers are not necessary[C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data. [S.l.] : ACM, 2008 : 121 – 132.

[18] GHINITA G, KALNIS P, SKIADOPOULOS S. PRIVE: anonymous location-based queries in distributed mobile systems[C] // Proceedings of the 16th international conference on World Wide Web. [S.l.] : ACM, 2007 : 371 – 380.

[19] WILLIAMS P, SION R. Usable PIR[C] // NDSS. 2008.

[20] ZHANGHAO. Research on information privacy protection technology based on location services[M]. [S.l.] : University of Science Technology China, 2014.

[21] YINJIE W. Privacy Preserving Data Publishing:Models and Algorithms[M]. [S.l.] : Tsinghua University Press, 2015.

[22] CHENG R, ZHANG Y, BERTINO E. Preserving user location privacy in mobile data management infrastructures[C] // Privacy Enhancing Technologies. [S.l.] : Springer Berlin Heidelberg, 2006 : 393 – 412.

[23] GRUTESER M, GRUNWALD D. Anonymous usage of location-based services through spatial and temporal cloaking[C] // Proceedings of the 1st international conference on Mobile systems, applications and services. [S.l.] : ACM, 2003 : 31 – 42.

[24] BERESFORD R A, STAJANO F. Mix zones: User privacy in location-aware services[J]. IEEE Pervasive computing, 2004.

[25] CHEYANZHE. Research on Key Technologies of user location privacy protection based on location services[D]. [S.l.] : Data resource system, 2013.

[26] KIDO H, YANAGISAWA Y, SATOH T. An anonymous communication technique using dummies for location-based services[C] // ICPS. [S.l.] : IEEE, 2005 : 88 – 97.

[27] XU J, TANG X, HU H, et al. Privacy-conscious location-based queries in mobile environments[J]. Parallel and Distributed Systems, IEEE Transactions on, 2010, 21(3) : 313 – 326.

[28] XU T, CAI Y. Feeling-based location privacy protection for location-based services[C] // Proceedings of the 16th ACM conference on Computer and communications security. 2009 : 348 – 357.

[29] ARDAGNA C A, CREMONINI M, DAMIANI E, et al. Location privacy protection through obfuscation-based techniques[G] // Data and Applications Security XXI. [S.l.] : Springer, 2007 : 47 – 60.

[30] KALNIS P, GHINITA G, MOURATIDIS K, et al. Preventing location-based identity inference in anonymous spatial queries[J]. Knowledge and Data Engineering, IEEE Transactions on, 2007, 19(12) : 1719 – 1733.

[31] WANG T, LIU L. Privacy-aware mobile services over road networks[J]. Proceedings of the VLDB Endowment, 2009, 2(1) : 1042 – 1053.

[32] HOSSAIN A-A, HOSSAIN A, YOO H-K, et al. H-star: Hilbert-order based star network expansion cloaking algorithm in road networks[C] // Computational Science and Engineering (CSE), 2011 IEEE 14th International Conference on. 2011 : 81 – 88.

[33] WANG S, WANG X S. In-device spatial cloaking for mobile user privacy assisted by the cloud[C] // Mobile Data Management (MDM), 2010 Eleventh International Conference on. 2010 : 381 – 386.

[34] PAN X, XU J, MENG X. Protecting location privacy against location-dependent attacks in mobile services[J]. Knowledge and Data Engineering, IEEE Transactions on, 2012, 24(8) : 1506 – 1519.

[35] JUNCHENG P, HUIMIN D, YINGHUI S, et al. Potential Attacks against k-Anonymity on LBS and Solutions for Defending the Attacks[G] // Advances in Computer Science and its Applications. [S.l.] : Springer, 2014 : 877 – 883.

[36] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. l-diversity: Privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1) : 3.

[37] LI N, LI T, VENKATASUBRAMANIAN S. t-closeness: Privacy beyond k-anonymity and l-diversity[C] // Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on. 2007 : 106 – 115.

[38] YAO A C. Protocols for secure computations[C] // Foundations of Computer Science, 1982. SFCS'08. 23rd Annual Symposium on. 1982 : 160 – 164.

[39] CLIFTON C, KANTARCIOGLU M, VAIDYA J, et al. Tools for privacy preserving distributed data mining[J]. ACM Sigkdd Explorations Newsletter, 2002, 4(2) : 28 – 34.

[40] GOLDWASSER S. Multi party computations: past and present[C] // Proceedings of the sixteenth annual ACM symposium on Principles of distributed computing. 1997 : 1 – 6.

[41] DU W, ATALLAH M J. Secure multi-party computation problems and their applications: a review and open problems[C] // Proceedings of the 2001 workshop on New security paradigms. 2001 : 13 – 22.

[42] OLESHCHUK V A, ZADOROZHNY V. Secure multi-party computations and privacy preservation: Results and open problems[J]. Telektronikk, 2007, 103(2) : 20.

[43] SHI E, CHAN T H, RIEFFEL E, et al. Privacy-preserving aggregation of time-series data[C] // Proc. NDSS : Vol 2. 2011 : 1 – 17.

[44] JUNG T, MAO X, LI X-Y, et al. Privacy-preserving data aggregation without secure channel: Multivariate polynomial evaluation[C] // INFOCOM, 2013 Proceedings IEEE. 2013 : 2634 – 2642.

[45] GHINITA G, KALNIS P, KHOSHGOZARAN A, et al. Private queries in location based services: anonymizers are not necessary[C] // Proceedings of the 2008 ACM SIGMOD international conference on Management of data. 2008 : 121 – 132.

[46] KUSHILEVITZ E, OSTROVSKY R. Replication is not needed: Single database, computationally-private information retrieval[C] // focs. 1997 : 364.

[47] FLATH D E. Introduction to number theory[J], 1989.

[48] KHOSHGOZARAN A, SHAHABI C, SHIRANI-MEHR H. Location privacy: going beyond K-anonymity, cloaking and anonymizers[J]. Knowledge and Information Systems, 2011, 26(3) : 435 – 465.

[49] di VIMERCATI S D C, FORESTI S, SAMARATI P. Managing and accessing data in the cloud: Privacy risks and approaches[C] // 2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS). 2012 : 1 – 9.

[50] WILLIAMS P, SION R, CARBUNAR B. Building castles out of mud: practical access pattern privacy and correctness on untrusted storage[C] // Proceedings of the 15th ACM conference on Computer and communications security. 2008 : 139 – 148.

致　　谢

从2014年9月进入密码与网络安全系至今，这两年半的时光转眼即逝，在华东师范大学的研究生生活无疑是我人生中宝贵的财富之一，在这两年半的时间中我学会了如何去发现问题，独立思考问题，如何将所学知识活学活用，让我对自己有了更加深刻的了解，如何更好的规划未来的人生道路。在毕业论文即将完成之际，我要对每一位支持我，学习上鼓励我、教导我，生活上关心我的人表示衷心的感谢。

首先，我要感谢密码与网络安全系的所有老师们，老师们具有一流的科研水平以及极高科研热情，带领我们以当下最先进，最前沿的眼光来审视当下各种大数据安全，云计算安全等问题，如今网络安全被列为教育部一级学科，从斯诺登事件到如今的电信你诈骗案，处处都体现了当今密码与网络安全的重要性。曹珍富老师是密码与安全方面的专家，在学术界很有名气，我们密码与网络安全系2014年刚成立，很高兴在曹老师的带领下，队伍不断壮大，我们系也受到越来越多国内外学者的关注，引进了许多安全方向的专家学者，研究成果如雨后春笋，我感动由衷的欣慰，感谢老师们为了系部的发展做出的各种努力，感谢董老师在研究生期间对我的教导，董老师强大的专业背景让我认识到了密码学的魅力之处，总能从董老师的耐心讲解中找到正确的求解思路，感谢董老师在研究室期间对我的信任与包容，不管是在个人发展以及学术研究上，董老师总能给我最大的帮助，理解与关心。感谢周俊老师，周俊老师是我们实验室的小老师同时也是我们的大师兄，有着老师的严谨以及师兄的担当，周老师在学术研究方面有很深的造诣，每次例会都能在不懂之处给我们进行耐心的讲解与指导。感谢沈佳辰老师，沈老

师为实验室付出了很多，对实验室同学们的生活、学习都很关心。感谢何道敬老师、张磊老师、曾鹏老师、王高丽老师等等，非常感谢每一位老师在过去的两年多时间对我的培养。

其次感谢所有 TDT 实验室陪伴我走过研究生生涯的师兄弟们，在这里对我来说就像是一个大家庭，两年多的时间，在这里见证了大家的喜怒哀乐，感谢宁建廷博士、王海江博士、李冬梅博士，郭莹博士，巩俊卿博士，曹楠源博士在我刚来实验室时对我的学术指导，感谢王丹，陈冬冬和我一起三足鼎立坚守实验室，感谢邓尔冬，张华君、来思远，赵晓鹏，郑锦文，毋萌，宋春芝，张晓东，王乾，郭婉芬，丁诗瑶等在 TDT 实验室学习的师兄姐弟妹们，感谢密码与网络安全系所有的同学们。还要感谢这两年多来包容我，陪伴我度过最美好的研究生生活的“逗逼”舍友们。

此外，我想感谢所有软件学院的老师和朋友，感谢陪我走过两年多研究生生活的的软件专硕班，在这个班集体中我遇到了积极向上，执着努力热爱生活的一帮程序员么。

最后，感谢我的父母，他们是我背后最坚强的后盾。虽然离家不算远，可每年只能在家陪伴他们短短的天数，每次回家，爸妈都会为我准备丰富的饭菜，从来不会因为我没有在他们身边而有多抱怨。他们教会了我成为一个正直，勤奋，有担当的人，总是默默的支持我的任何想多的事情，无私的付出。希望在将来不会辜负他们对我的期望，对得起父母的养育之恩。

孙浩

二零一七一月

攻读硕士学位期间发表论文和科研情况

■ 已发表软件著作权

[1] 孙浩, 基于 MVC 模式的信息发布与管理系统 V1.0 登记号：2016SR211875.

■ 攻读学位期间参加的科研项目

[1] 面向大数据系统的安全计算，课题号 61632012.

[2] 从属性基加密到功能加密的扩展安全模型与新方法研究，课题号 61672239