

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着公钥过大的问题。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性和性能。其中一个减小公钥的大小；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

第一章 绪 论 . . . . .	1
1.1 研究背景及意义 . . . . .	1
1.2 国内外研究现状 . . . . .	2
1.3 本文工作与主要贡献 . . . . .	4
1.4 研究重点与组织结构 . . . . .	5
第二章 预备知识与概念 . . . . .	6
2.1 有限域基础理论 . . . . .	6
2.2 编码基础知识 . . . . .	7
2.3 常见码的性质 . . . . .	8
2.3.1 GRS 码 . . . . .	8
2.3.2 Goppa 码 . . . . .	9
2.4 本章小结 . . . . .	9
第三章 经典方案介绍 . . . . .	10
3.1 K 近邻问题 . . . . .	10
3.2 密码学基础 . . . . .	11
3.2.1 参与者与攻击者 . . . . .	12
3.2.2 攻击类型 . . . . .	12
3.3 语义安全 . . . . .	13
3.4 同态加密 . . . . .	15
3.5 Paillier 公钥加密 . . . . .	18
3.6 本章小结 . . . . .	19



第四章 KNN 中的隐私保护问题研究 . . . . .	20
4.1 研究动机. . . . .	20
4.2 问题定义. . . . .	21
4.3 隐私需求. . . . .	22
4.4 隐私保护原语 . . . . .	22
4.5 证明 SMIN 安全性. . . . .	32
4.6 本章小结. . . . .	33
第五章 PPkNNONED 方案 . . . . .	34
5.1 K 近邻安全检索 . . . . .	37
5.2 多数类的安全计算. . . . .	39
5.3 复杂性分析 . . . . .	40
5.4 本章小结. . . . .	40
第六章 总结与展望 . . . . .	41
参考文献 . . . . .	42
致谢 . . . . .	49
发表论文和科研情况 . . . . .	51

# 插图

3.1 K 近邻示意图 . . . . .	10
-----------------------	----

# 表格

4.1	SMIN 协议中间结果 . . . . .	28
-----	-----------------------	----

# List of Algorithms

1	$SMIN(u', v') \rightarrow [min(u, v)], E_{pk}(s_{min(u,v)})$ (Part 1) . . . . .	25
2	$SMIN_n([d_1], E_{pk}(s_{d_1})), \dots, ([d_n], E_{pk}(s_{d_n})) \rightarrow ([d_{min}], E_{pk}(s_{d_{min}}))$ . .	30
3	$SF(\Lambda, \Lambda') \rightarrow \langle E_{pk}(c_1), \dots, E_{pk}(f(c_w)) \rangle$ . . . . .	31
4	$PPKNNONED(D', q) \rightarrow c_q$ . . . . .	36
5	$SCMC_k(E_{pk}(c'_1), \dots, E_{pk}(c'_k)) \rightarrow c_q$ . . . . .	39

# 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-Hellman 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. Shannon 《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 Niederreiter 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。

Niederreiter 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。Niederreiter 方案与 McEliece 方案的安全性是等价的，设计 Niederreiter 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 码更合适的方案。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 Niederreiter）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  $\mathbf{F}$  上定义了两个二元运算：加法 “+” 和乘法 “ $\bullet$ ”，如果  $(\mathbf{F}, +)$  是交换群， $\mathbf{F}$  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  $\mathbf{F}$  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  $a^d = 1 \pmod{q}$  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  $F[x] = \{f(x) | f(x) = a_n x^n + \dots + a_1 x + a_0, a_i \in \mathbb{F}, a_n \neq 0, n \geq 0\}$ ， $F[x]$  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  $\mathbf{F}$  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  $\alpha$  为根，首项系数为 1，并且次数最低的多项式称为  $\alpha$  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。 $\alpha$  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  $\mathbf{x}$  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:  $wt(\mathbf{x})$ 。汉明距离指的是两个长度相同的码字  $\mathbf{x}, \mathbf{y}$  之间, 位不同的总数, 一般记作:  $dist(\mathbf{x}, \mathbf{y})$ 。可以看出,  $dist(\mathbf{x}, \mathbf{y}) = wt(\mathbf{x} - \mathbf{y})$ 。

**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:  $C$ 。另外, 线性码  $C$  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  $C$ , 且最大纠错能力为  $\lfloor (d-1)/2 \rfloor$ 。

**定义 2.2.3 (生成矩阵与校验矩阵)** 线性码  $C$  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  $C$  的基。也就是说由矩阵  $G$  的行可以线性组合成线性码  $C$  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  $C$  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  $C$  时, 都有一个相应的解码算法  $D_C$ , 完成纠正加入差错向量的码字的工作, 也就是说, 对任意的  $\mathbf{e} \in \mathbb{F}^n$ ,  $wt(\mathbf{e}) < d/2$ ,  $\mathbf{x} \in C$ , 都有

$$D_C(\mathbf{x} + \mathbf{e}) = \mathbf{x}。$$

## 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  $C$ ，如果线性码  $C$  的任意一个码字的循环移位还是一个码字，即当  $a_0a_1\cdots a_{n-1} \in C$  时， $a_{n-1}a_0a_1\cdots a_{n-2} \in C$ ，则称  $C$  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  $1, \alpha, \alpha^2, \dots, \alpha^{r-1}$  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  $1, \alpha, \alpha^2, \dots, \alpha^{r-1}$  的一个线性组合。令

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

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

我们称  $B_q(n, \delta, \alpha)$  是码长为  $n$  并且设计距离为  $\delta$  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  $\alpha$  替换成  $\alpha^b$ , 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  $\delta$  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  $\alpha$  是  $\mathbb{F}_q$  的一个  $q - 1$  阶元素, 即  $\alpha$  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

## 2.4 本章小结

## 第三章 经典方案介绍

本节将介绍涉及到的一些基础知识，包括  $k$ -近邻问题、密码学基础、语义安全、同态加密、Pailler 公钥加密。

### 3.1 K 近邻问题

K 近邻算法 (K-Nearest Neighbor algorithm)，简称 KNN 算法，属于分类算法，也是数据挖掘算法常用算法之一。K 近邻算法即给定一个训练数据集，一个新的输入实例，在训练数据集中找出与该输入实例最邻近的 K 个实例，若这 K 个实例中多数属于某个类别，则该实例就属于这个类。

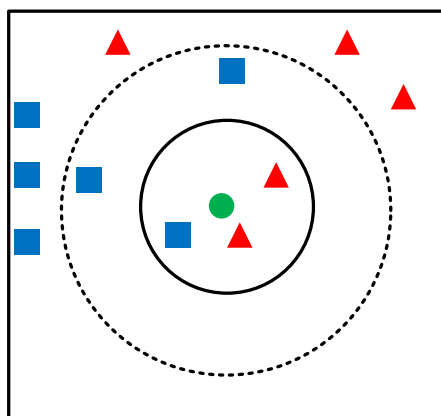


图 3.1: K 近邻示意图

如上图3.1所示，图中有两种不同的样本数据，分别用蓝色的正方形和红色的三角形表示，途中绿色的圆圈表示待分类的实例。现在假设我们并不知道途中绿色圆圈属于哪一个类别 (即不知道属于蓝色小正方形还是红色三角形)。接下来依次取不同的 K 值，观察绿色圆圈所属的类别。

- $K = 3$ , 则距离绿色圆圈最近的 3 个邻居分别为 2 个红色三角形和 1 个蓝色正方形, 因此根据统计学的方法, 基本可以判定绿色小圆圈属于红色三角形的类别。
- $K = 5$ , 观察可以发现, 距离绿色圆点最近的 5 个邻居分别是 2 个红色三角形和 3 个蓝色正方形, 基于统计方法, 基本可以判定绿色小圆圈属于蓝色正方形。

通过上述观察我们可以发现, 当无法判定当前待分类点从属于已知类别中的哪一类的时候, 可以根据统计学方法观察待分类点所处位置的特征, 衡量它周围邻居的权重, 从而将待分类点归为权重更大的那一类。

KNN 的算法描述大体可以分为以下三步骤:

- 依公式计算待分类点 Item 与已知分类点  $D_1, D_2, \dots, D_j$  之相似度。得到  $\text{Sim}(\text{Item}, D_1)$ 、 $\text{Sim}(\text{Item}, D_2)$ ……、 $\text{Sim}(\text{Item}, D_j)$ 。
- 对  $\text{Sim}(\text{Item}, D_1)$ 、 $\text{Sim}(\text{Item}, D_2)$ ……、 $\text{Sim}(\text{Item}, D_j)$  排序, 若超过相似度阈值  $t$  则放入邻居案例集合 NN。
- 取出邻居案例集合 NN 中的前  $k$  个, 求出这  $K$  个所属的类别, 依多数决, 得到 Item 可能类别。

## 3.2 密码学基础

密码体制分为对称密码体制和非对称密码体制。对称密码体制也称为私钥密码体制, 其加密方和解密方使用同一把密钥, 一个私钥密码体制一般由三部分组成, 一个密钥生成算法  $G()$ , 一个加密算法  $E()$ , 一个解密算法  $D()$  组成, 可以表示为  $(G, E, D)$ , 常用的对称加密算法有 DES, AES 等。非对称密码体制又叫做公钥密码体制, 加密方和解密方采用不同的密钥, 用来加密的密钥称为公钥, 用来解密的密钥称为私钥。常用的公钥加密算法有 RSA, ElGamal 等。

### 3.2.1 参与者与攻击者

**参与者** 参与者的行为通常决定了计算协议设计的难易程度，根据参与者在协议中的行为，可以将参与者分为三种类型。

**诚实的参与者** 诚实参与者是指在协议过程中参与者完全按照协议的要求完成各个步骤，并且对自己的输入以及中间结果进行保密。诚实参与者可以根据自己的输入以及输出信息推导出其他参与者的信息。

**半可信的参与者** 半可信的参与者是指在协议过程中，参与者完全按照协议的要求完成协议的各个步骤，并且可以将自己的输入以及输出信息透露给攻击者。诚实的参与者与半可信参与者的唯一区别就在于半可信的参与者可能会被攻击者妥协。

**恶意的参与者** 恶意的参与者是指在协议过程中，参与者完全按照协议的要求完成协议的各个步骤，但参与者不会将自己的输入以及输出信息透露给参与者，并且可以根据攻击者的意图对输入以及输出信息进行改变，甚至对协议进行终止。

**攻击者** 外部攻击者对恶意参与者的控制程度大致可以分为以下几种情形：

**被动攻击者** 被动攻击者有可以称为监听者，他不会去控制恶意参与者的行为，其主要职责就是负责监听恶意参与者的输入、输出以及一些中间结果，

**主动攻击者** 此类攻击者除了会对恶意参与者的输入、输出以及中间结果进行监听外还会对恶意参与者的行为进行控制。

### 3.2.2 攻击类型

根据密码分析者破译时已具备的前提条件，人们通常将攻击类型分为五种：唯密文攻击 (ciphertext-only attack)、已知明文攻击 (known plaintext attack)、选择



明文攻击 (chosen plaintext attack)、选择密文攻击 (chosen ciphertext attack)、选择文本攻击 (chosen text attack)。

**唯密文攻击** 密码分析者除了拥有截获的密文外，没有任何其他可以利用的信息，分析者通过对这些截获的密文进行分析得出明文或密钥信息。

**已知明文攻击** 密码分析者除了掌握相当数量的密文信息，还掌握了一些明文和用同一个密钥加密这些明文所对应的密文。

**选择明文攻击** 密码分析者不仅可以获得一定数量的明-密文对，还可以选择任何明文并在使用同一未知密钥的情况下能得到相应的密文。

**选择密文攻击** 密码分析者可以获得被加密密文所对应的明文，密码分析者的任务是推断出密钥以及其他密文信息所对应的明文。

**选择文本攻击** 选择文本攻击是选择明文和选择密文攻击的结合。

### 3.3 语义安全

加密算法通常需要有三个必备的部分，分别是：原始消息  $m$ ，加密密钥  $k$  以及加密算法。加密通常是为了对消息进行保密，最终目的就是为了在别人没有密钥的情况下，不能够从加密结果中恢复出原始的消息。

证明语义安全通常需要两种角色：攻击者 (Adversary) 和挑战者 (Challenger)。攻击者是指攻击加密算法的人，而挑战者是指接收攻击者挑战的人。Challenger 持有密钥  $k$ ，Adversary 的目的就是观察 Challenger 运行加密算法后，输出的加密结果是否“看起来一样”。下面定义两个实验来具体讲解 Adversary 和 Challenger：

实验 0：

- Adversary 任意选择两个消息  $m_0, m_1$ 。

- Adversary 将两个消息  $m_0, m_1$  发送给 Challenger。
- Challenger 运行加密算法，加密消息  $m_0$ ，并将加密结果发送给 Adversary。

实验 1:

- Adversary 任意选择两个消息  $m_0, m_1$ 。
- Adversary 将两个消息  $m_0, m_1$  发送给 Challenger。
- Challenger 运行加密算法，加密消息  $m_1$ ，并将加密结果发送给 Adversary。

Adversary 的目的就是通过观察加密结果，判断加密结果是属于  $m_0$  还是  $m_1$  的加密结果值。上述两个实验的唯一区别就在于挑战者返回的是  $m_0$  还是  $m_1$  的加密结果。上述两个实验可以合并为一个实验，即：

实验 b:

- Adversary 任意选择两个消息  $m_0, m_1$ 。
- Adversary 将两个消息  $m_0, m_1$  发送给 Challenger。
- Challenger 运行加密算法，加密消息  $m_b$ ，并将加密结果发送给 Adversary。

如果 Challenger 能够以  $1/2$  的概率执行实验 0，以  $1/2$  的概率执行实验 1。那么，如果 Adversary 只有  $1/2$  的概率判断正确的加密结果，即有：

$$Pr[c \leftarrow E(k, m_0)] = \frac{1}{2}$$

$$Pr[c \leftarrow E(k, m_1)] = \frac{1}{2}$$

这时候引入 Adversary 成功的优势的概念，成功优势是指如果 Challenger 以  $1/2$  的概率执行实验 0，以  $1/2$  的概率执行实验 1，那么 Adversary 是否能够有比  $1/2$  更大的概率猜测正确。因为 Adversary 能够利用的只有加密结果值。因此如果有更大的概率，那么概率的增加只能是从加密结果中得来的。我们定义：

$$Adv_{ss}[A, E] = |Pr[c \leftarrow E(k, m_0)] - Pr[c \leftarrow E(k, m_1)]|$$

若攻击者的一点优势也没有，那么这个优势则为 0，根据上面的定义有：

$$Adv_{ss}[A, E] = |Pr[c \leftarrow E(k, m_0)] - Pr[c \leftarrow E(k, m_1)]| = |\frac{1}{2} - \frac{1}{2}| = 0$$

若加密结果能够反映出任何有关加密结果的消息，那么 Adversary 成功的几率就会增加，即 Adversary 能够以更大的概率判断加密结果属于  $m_0$  还是  $m_1$ ，这个时候有：

$$Pr[c \leftarrow E(k, m_0)] = \frac{1}{2} \pm \frac{1}{2}\varepsilon, pr[c \leftarrow E(k, m_1)] = \frac{1}{2} \mp \frac{1}{2}\varepsilon$$

其中  $\varepsilon$  属于 Adversary 增加的的概率，这时候 Adversary 的攻击优势为：

$$Adv_{ss}[A, E] = |Pr[c \leftarrow E(k, m_0)] - Pr[c \leftarrow E(k, m_1)]| = |\frac{1}{2} \pm \varepsilon - \frac{1}{2} \mp \varepsilon| = \varepsilon$$

由于 Adversary 拿到了加密结果，因此成功的优势一定会有所增加。密码学中定义，这个优势在多项式复杂度内可忽略就可以了，即这个优势小到无法用一个以位数为基本单位 ude 多项式来描述。比如，对于一个 128 位安全常数的加密算法，安全常数通常可以达到  $1/2^{128}$ ：位数是 128 位，那么成功的优势为  $2^{-128}$ ，可以发现，这不是一个有关位数的多项式，而是一个指数式。

综上可以发现，语言安全性的定义是一种基于游戏的安全定义。

### 3.4 同态加密

同态加密 (Homomorphic Encryption) 最早由 Ron Rivest, Leonard Adleman 就以银行为背景提出的。Craig Gentry 第一次构造出了全同态的加密方案。同态加密方案关注的是数据处理安全，它提供了一种对加密数据进行处理的功能，即别人可以对数据进行相应的处理，但处理过程成不会泄露任何关于原始内容的信息，同时拥有密钥的用户对处理后的数据进行解密，得到的结果与对明文进行相应处理的结果一样。例如，Alice 想让金匠加工一把金锁，但是金匠有可能会偷取部分金子，因此 Alice 想出了一种方法能够避免工人偷取金子。

- Alice 首先将金子锁在一个密封住的盒子里面，并且在盒子上安装了一个手套。
- 工人加工金子的时候需要戴上手套，但是盒子是锁住的，所以工人是拿不到金块，就算是金块叫加工掉下来的碎片也得不到。
- 待工人加工完后，Alice 将盒子收回来，并用钥匙将盒子打开，得到加工后的产品。

这里的盒子对应着加密算法、盒子上的锁对应着用户持有的密钥、将金块放入盒子并且锁上对应着用同态加密方案对数据进行加密、工人加工的过程即阐明了同态的特性，即在无法取得数据的条件下对加密数据进行相应的处理、最后的开锁对应着对处理后的加密数据进行解密，得到处理后的结果。

同态加密在云计算中有很好的应用场景，几乎就是为云量身打造的，我们可以考虑如下的应用场景：假如一个用户想要处理某个数据，但是由于自身计算机硬件设备较差，不能对数据进行处理。这时候他可以使用云计算的概念，通过云来进行相应的处理从而得到结果。但是用户如果将数据不做任何处理直接交给云的话，可能会造成隐私泄露以及其他的一些安全问题。这时候，用户可以通过同态加密，然后通过云来对加密数据进行直接处理，并将处理后的结果返回给用户。这样的话用户向云服务提供商付款得到了处理后的结果，云服务提供商在不知道用户原始数据的情况下处理数据。

云计算应用场景下同态加密处理数据的整个过程如下：

- Alice 加密数据，并将加密后的数据外包给云服务提供商。
- Alice 向云提交数据的处理方法 (函数)，这里用  $f$  表示。
- 云服务提供商根据函数  $f$  对数据进行相应的处理，并且将处理后的结果发送给 Alice。

- Alice 收到函数  $f$  处理后的数据后，使用自己的密钥对数据进行解密，得到结果。

我们可以很直观的得到同态加密方案应该包含的一些函数：

- 密钥生成函数 (KeyGen): KeyGen 函数由 Alice 运行，产生用于加密原始数据 (Data) 的密钥 (key)，这期间可能会包含一些公开常数 (PP)。
- 加密函数 (Encrypt): Encrypt 由 Alice 运行，用生成的密钥 Key 对用户数据 Data 进行加密，得到密文 (CT)。
- 评估函数 (Evaluate): Evaluate 函数由云服务提供商进行运行，根据用户给定的数据处理函数  $f$ ，对密文进行相应的处理，得到结果相当于用密钥对  $f(\text{Data})$  进行操作。
- 解密函数 (Decrypt): Decrypt 由 Alice 运行，用于得到云服务提供商处理后的结果  $f(\text{Data})$ 。

根据数据处理函数  $f$ ，实际上同态加密方案可以分为两类：

- 全同态加密方案 (Fully Homomorphic Encryption, FHE): FHE 意味着可以支持任意给定的处理函数  $f$ ，只要函数  $f$  可以通过算法描述，能够计算机实现即可。虽然 FHE 方案是一个很好的方案，也非常强大，但由于计算开销很大，通常无法在实际生产中使用。
- 半同态加密方案 (Somewhat Homomorphic Encryption, SWHE): SWHE 意味着数据处理函数只支持一些特定的函数  $f$ 。SWHE 方案相对 FHE 来说比较弱，但开销也相应的会变小不少，容易实现，因此在实际的项目中可以实现。

同态加密方案的最基本安全性是语义安全性，即密文不会泄露任何关于明文的信息，语义安全的具体介绍可以见上一节的描述。

有时候可能需要保护数据机密的同时数据处理函数  $f$  也需要保密，直观的说就是，云服务提供商不仅得不到任何关于原始数据的信息，连数据是如何处理的

也不知道，只能按照给定的算法执行，然后返回用户所需要的结果。这种特性很厉害，不过现在还没有 FHE，甚至连 SWHE 也没有。

### 3.5 Paillier 公钥加密

设  $n=pq$ ，其中  $p, q$  为大素数， $\varphi(n)=(p-1)(q-1)$ ， $\varphi(n)$  为欧拉函数。Caemichael 函数  $\lambda(n)=lcm(p-1)(q-1)$  ( $lcm$  代表最小公倍数)。 $|Z_{n^2}^*|=\varphi(n^2)=n\varphi(n)$ ， $\forall \omega \in Z_{n^2}^*$ ，为了方便表示，接下来所有的  $\lambda(n)$  直接用  $\lambda$  进行表示。根据 Caemichael 的性质，有  $\omega^\lambda=1 \bmod n$ ， $\omega^{n\lambda}=1 \bmod n^2$ 。

Paillier 公钥密码体制是基于困难问题假设的，设  $n=pq$ ，其中  $p, q$  为大素数， $g \in Z_{n^2}^*$ ，整数值的函数  $\varepsilon_g$  定义如下： $Z_n \times Z_n^* \rightarrow Z_{n^2}^*$ ， $(x, y) \rightarrow g^x \cdot y^n \bmod n^2$ 。如果  $g$  在  $Z_{n^2}^*$  中的阶为  $n$  的倍数，则  $\varepsilon_g$  是一一映射。这样的话对于给定的  $\omega \in Z_{n^2}^*$ ， $x \in Z_n$ ，则  $\exists y \in Z_n^*$ ，使得  $\varepsilon_g(x, y) = \omega$ 。这样的  $\varepsilon_g(x, y)$  称为  $\omega$  的  $n$ -residuosity class，用  $[[\omega]]_g$  表示。目前认为，对于给定的  $n, g, \omega$ ，计算  $[[\omega]]_g$  是困难问题，即所谓的 Composite Residuosity Assumption(CRA)。但是根据  $p, q$  的内容，即依据  $\lambda$ ，可以计算出任意的  $[[\omega]]_g$ 。事实上，设  $S_n = \{u < n^2 | u = 1 \bmod n\}$ ， $\forall u \in S_n$ ， $L(u) = \frac{u-1}{n}$ ，则有如下的结果： $[[\omega]]_g = \frac{L(\omega^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n$ 。

Paillier 的加解密过程分为三个部分：

- 初始化阶段：设  $n=pq$ ，其中  $p, q$  为大素数，选取  $g \in Z_{n^2}^*$ ，使得  $\gcd(L(g^\lambda \bmod n^2), n) = 1$ ， $\gcd$  代表最大公约数，其中公钥  $pk$  为  $(n, g)$ ，私钥  $sk$  为  $(p, q)$  等价于  $\lambda(\lambda = lcm(p-1, q-1))$ 。
- 加密阶段：设有明文  $m \in Z_n$ ，且  $m < n$ 。选择一个随机数  $r < n$ ，则密文为  $c = g^m \cdot r^n \bmod n^2$ 。
- 解密阶段：密文  $c < n^2$ ，明文  $m = \frac{L(c^\lambda \bmod n^2)}{L(g^\lambda \bmod n^2)} \bmod n$ 。

Paillier 公钥加密方案满足加法和乘法同态：

- 加法同态:  $D_{sk}(E_{pk}(a + b)) = D_{sk}(E_{pk}(a) * E_{pk}(b) \bmod N^2)$ 。其中  $E_{pk}()$  表示使用公钥加密,  $D_{sk}$  表示使用私钥进行解密。
- 乘法同态:  $D_{sk}(E_{pk}(a * b)) = D_{sk}(E_{pk}(a)^b \bmod N^2)$ 。其中  $E_{pk}()$  表示使用公钥加密,  $D_{sk}$  表示使用私钥进行解密。

### 3.6 本章小结

本章主要对隐私保护总的基础知识进行描述, 包括  $k$ -近邻问题、密码学基础知识、语义安全、同态加密以及 Paillier 公钥加密。

$k$ -近邻问题介绍了基本概念,  $k$ -近邻算法属于分类算法, 是数据挖掘中算法中常用算法之一。介绍中, 通过案例对  $k$ -近邻问题进行了形象的描述, 并在后续给出了  $k$ -近邻算法涉及到的三个重要步骤。

密码学基础中首先介绍了对称与非对称密码体制, 对参与者与攻击者进行了介绍, 其实中参与者含有诚实参与者、半可信参与者以及参与者, 攻击者分为被动攻击者以及主动攻击者。攻击类型中, 人们通常将其分为 5 种: 唯密文攻击、已知明文攻击、选择明文攻击、选择密文攻击以及选择文本攻击。

语义安全部分以简单易懂的方式进行了讲解, 语义安全实际上就是一种 Adversary 和 Challenger 之间进行的一种游戏。同态加密部分同意以案例的方式进行描述, 同态加密在云计算中有很好的应用场景, 几乎就是为云量身打造的, 试想, 如果有哪家存储公司能够将同态加密很好的应用起来, 那么必将是一家独大, 达到技术的垄断, 可由于效率问题, 目前为止还没有哪个全同态加密方案能够应用于实际的项目中。在 Paillier 中重点介绍了 Paillier 加解密过程中的三个步骤: 初始化阶段, 加密阶段以及解密阶段。Paillier 公钥加密方案属于半同态加密方案, 满足加法和乘法同态。

## 第四章 KNN 中的隐私保护问题研究

### 4.1 研究动机

云计算正在颠覆着传统企业对数据存储，访问以及处理的操作方法。作为一个新兴的计算模式，在成本效率、灵活性以及管理开销都优于之前进行本地处理的一些传统方法，因此云计算受到了众多企业的关注与追捧，亚马逊，微软等这些互联网大亨，也把云计算作为企业的一项核心业务。大部分情况下，企业更擅于将计算操作而不是数据本身委托给云，尽管云计算提供了巨大的优势，但由于隐私和安全问题，很多企业对云计算还是持慎重使用的态度，或者对具有高度敏感性的数据进行加密后再外包至云端。但是，当数据被加密，无论底层属于哪种加密方案，在数据未解密的情况下进行数据挖掘都是异常繁琐且效率低下的。除了上述所叙说的，还有其他的隐私安全问题被关注，例如：

- 假设保险公司将他们加密过的顾客数据及相关的数据挖掘任务外包给云。当公司代理想确定一个潜在客户的风险级别的时候，代理可以使用分类的方法去确定客户的风险级别。首先，代理需要生成一条包含顾客特定个人信息的查询请求数据  $q$ ，例如，身份证号，婚姻状况，年龄，信用额度等信息。这些查询请求数据会发送至云端，云将会计算相应查询数据  $q$  的所属类标签。然而，由于查询请求  $q$  包含了顾客的一些敏感个人信息，为了保护顾客的个人信息的，查询请求  $q$  在发送至云端的时候也应该进行加密。

上述案例说明了在云端进行密文上的数据挖掘（DMED）的时候，当涉及到用户的敏感信息的时候，这些敏感信息也应该被加密。此外，云通过观察数据的访问模式，也可能获得实际数据项中一些有用或敏感信息，即使实际数据项可能



是被加密过的 [49][50]。因此，云中密文上数据挖掘的隐私安全问题需要考虑到如下三点：

- 加密数据的保密性
- 用户查询记录的保密性
- 隐藏数据的访问模式

现存的数据挖掘隐私保护解决方案（数据微扰方案或者安全多方计算）并不能很好的解决 DMED 问题。基于数据扰乱的方法不具有语义安全性，所以数据扰乱技术不能用于加密高度敏感的数据，另外，由于数据被扰乱了，因此不能获得精准的数据挖掘的结果。基于安全多方计算的方法假设数据是分布式的，不能对每个参与方进行加密。另外，许多中间计算需要依赖于未加密的数据。在本章，将会提出一个新的方法能够有效的解决云中密文上数据挖掘的问题。分类问题是数据挖掘中比较常见的一种任务，因此具体将会在分类问题上进行研究。由于每种分类问题都有各自的优势与弊端，在这里将着重选择 k-近邻分类算法用于处理云计算环境中密文数据的分类。

## 4.2 问题定义

假设 Alice 拥有数据库  $D$ ， $D$  由  $m+1$  个属性以及  $n$  条记录（元组） $t_1, t_2, \dots, t_n$  组成。设  $t_{i,j}$  表示元组  $t_i$  的第  $j$  个属性值。首先，Alice 对属性值进行加密，即计算  $E_{pk}(t_{i,j})$ ，其中  $1 \leq i \leq n$ ， $1 \leq j \leq m+1$ ，其中  $m+1$  列为类标签。这里假设底层的加密方案是语义安全的，用  $D'$  表示加密数据库。假设将来 Alice 也会将  $D'$  的分类过程外包给云端。

设 Bob 为得到授权的用户，Bob 利用 K-近邻分类算法，基于加密数据库  $D'$  对输入记录  $q = \langle q_1, q_2, \dots, q_m \rangle$  进行分类。我们称这一过程为云中密文数据上 K-近邻分类的隐私保护问题（PPKNNONED）。关于 PPKNNONED 的形式化定义如下：

$$\text{PPKNNONED}(D', q) \rightarrow c_q$$

其中  $c_q$  表示以加密数据库  $D'$  和查询请求  $q$  作为输入，应用 K-近邻分类算法，得到  $q$  所属的分类标签。

### 4.3 隐私需求

PPKNNONED 协议是一个能够满足密文数据上语义安全的 K-近邻分类器。在协议中，加密后的数据外包给云端，Alice 无需参与计算，因此，没有信息透露给 Alice。另外，协议满足下面的隐私需求：

- 云不能获取原始数据库（明文） $D$  中的内容以及任何中间结果。
- 云不能获取 Bob 关于查询请求  $q$  的任何信息。
- Bob 除了分类标签  $c_q$  外不能获取其他任何信息。
- 数据访问模式，例如，查询  $q$  对应的  $k$ -近邻记录不应该被 Bob 以及云获取，防止任何的推理攻击。

在协议中，假设云能够获取到的中间结果要么是新生成的随机加密，要么就是随机数。因此，查询请求  $q$  对应的  $k$  个近邻以及所属的标签对云来说是保密的。另外，向云发送加密查询请求后，Bob 本地无需涉及任何计算。因此，数据访问模式进一步的得到保护。

### 4.4 隐私保护原语

本节将介绍用于构造 KNN 协议的一些子协议且所有子协议都是构建在两方半可信的模型下。假设存在两个半可信方  $P_1$  和  $P_2$ ，其中仅有  $P_2$  拥有 Pailler 私钥  $sk$ ，公钥  $pk$  是公开的。子协议如下：

- 乘法安全 (Secure multiplication, SM). 协议考虑  $P_1$  以及输入  $(E_{pk}(a), E_{pk}(b))$ ，输出  $E_{pk}(a * b)$  至  $P_1$ ，其中  $a$  和  $b$  对  $P_1, P_2$  是保密的，处理过程中  $P_1$  和  $P_2$  无法获知关于  $a$  和  $b$  的任何信息。

- 欧几里得距离平方安全 (Secure squared euclidean distance ,SSED). 协议中  $P_1$  输入  $(E_{pk}(X), E_{pk}(Y))$ , 安全计算向量  $X$  和  $Y$  之间欧几里得距离的平方。这里  $X$  和  $Y$  是  $m$  维的向量, 其中  $E_{pk}(X) = \langle E_{pk}(x_1), \dots, E_{pk}(x_m) \rangle$ ,  $E_{pk}(Y) = \langle E_{pk}(y_1), \dots, E_{pk}(y_m) \rangle$ 。输出值  $E_{pk}(|X - Y|^2)$  仅  $P_1$  可见。
- 比特分解安全 (Secure bit-decomposition,SBD).  $P_1$  输入  $E_{pk}(z)$ ,  $P_2$  对  $z$  逐比特安全加密计算, 其中  $0 \leq z \leq 2^l$ 。输出值  $[z] = \langle E_{pk}(Z_1), \dots, E_{pk}(Z_l) \rangle$  仅  $P_1$  可知。这里  $z_1$  和  $z_l$  分别为整数  $z$  的最高和最低有效位。
- 最小值安全 (Secure minimum, SMIN). 在协议中,  $P_1$  持有私有输入  $(u', v')$ ,  $P_2$  持有  $sk$ , 其中  $u' = ([u], E_{pk}(s_u)), v' = ([v], E_{pk}(s_v))$ 。这里  $s_u$  表示和  $u$  相关的隐私信息, 同理  $s_v$  表示和  $v$  相关的隐私信息。最小值安全的目标是  $P_1$  和  $P_2$  共同对  $u$  和  $v$  之间最小的数进行逐比特加密计算, 即输出值是  $([min(u, v)])$ 。另外, 他们计算  $E_{pk}(s_{min}(u, v))$ 。在此协议期间, 值  $E_{pk}(s_{min}(u, v))$  仅  $P_1$  可知, 关于  $u, v, s_u$  以及  $s_v$  的任何信息,  $P_1$  和  $P_2$  都不可知。
- $n$  个数中的最小值安全 (Secure minimum out of  $n$  numbers). 协议中, 考虑  $P_1$  的  $n$  个加密向量  $([d_1], \dots, [d_n])$  以及他们各自的加密信息,  $P_2$  的私钥  $sk$ 。这里有  $[d_i] = \langle E_{pk}(d_{i,1}), \dots, E_{pk}(d_{i,l}) \rangle$ , 其中  $d_{i,1}$  和  $d_{i,l}$  分别为整数  $d_i, 1 \leq i \leq n$  的最高和最低有效位。  $d_i$  的保密信息由  $s_{d_i}$  计算得出。  $P_1$  和  $P_2$  共同计算  $[min(d_1, \dots, d_n)]$ 。以及  $E_{pk}(s_{min(d_1, \dots, d_n)})$ 。在协议最后时, 输出  $([min(d_1, \dots, d_n)], E_{pk}(s_{min(d_1, \dots, d_n)}))$  仅  $P_1$  可知。在协议期间, 关于  $d_i$  以及他们的保密信息没有泄露给  $P_1$  和  $P_2$ 。
- Bit-OR 安全 ( Secure Bit-OR,SBOR).  $P_1$  输入  $(E_{pk}(o_1), E_{pk}(o_2))$ 。  $P_2$  安全计算  $E_{pk}(o_1 \vee o_2)$ , 其中  $o_1$  和  $o_2$  为二进制比特串。输出值  $E_{pk}(o_1 \vee o_2)$  仅  $P_1$  可知。
- 频率安全 (Secure frequency). 协议中,  $P_1$  输入  $(\langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle, \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle)$ ,  $P_2$  在安全对集合  $\langle c'_1, \dots, c'_k \rangle$  中  $c_j$  出现的频率进行加密,

用  $f(c_j)$  表示, 其中  $1 \leq j \leq w$ 。在这里明确假设  $c_j$  是唯一的,  $c'_i \in \{c_1, \dots, c_w\}, 1 \leq i \leq k$ 。输出  $(\langle E_{pk}(f(c_1)), \dots, E_{pk}(f(c_w)) \rangle)$  仅有  $P_1$  可知。协议中,  $P_1$  和  $P_2$  无法获知关于  $c'_i, c_j$  以及  $f(c_j)$  的任何信息, 其中  $1 \leq i \leq k, 1 \leq j \leq w$ 。

这里我将会提出一个新的解决方案, 也可以说是一种比上述协议能够更有效实现的方案。主要会讨论“最小值安全”, “n 个数之外的最小安全”以及“频率安全”细节上的一些问题, 并针对他们中的每一个提出一个新的解决方案。

**最小值安全 SMIN** 此协议中, 假设  $P_1$  的输入为  $(u', v')$ ,  $P_2$  私钥为  $sk$ , 其中  $u' = ([u], E_{pk}(s_u))$ ,  $v' = ([v], E_{pk}(s_v))$ 。这里  $s_u$  和  $s_v$  分别表示  $u$  和  $v$  对应的秘密消息。SMIN 的主要目标是安全计算  $\min(u, v)$  中各个比特位的加密值, 用  $[\min(u, v)]$  表示。这里  $[u] = \langle E_{pk}(u_1), \dots, E_{pk}(u_l) \rangle$ ,  $[v] = \langle E_{pk}(v_1), \dots, E_{pk}(v_l) \rangle$ , 其中  $u_1$  和  $u_l$  分别表示  $u$  的最高和最低有效比特位, 同理  $v_1$  和  $v_l$  分别表示  $v$  的最高和最低有效比特位。另外,  $P_1$  和  $P_2$  取  $u$  和  $v$  两者的最小的一个, 对保密信息进行加密, 解密结果为  $E_{pk}(s_{\min(u, v)})$ 。在 SMIN 协议结束阶段, 输出值  $([\min(u, v), E_{pk}(s_{\min(u, v)})])$  仅  $P_1$  可知。

我们假设  $0 \leq u, v < 2^l$  并提出一种新的 SMIN 协议。准确的说, 新的 SMIN 协议的主要思想是  $P_1$  通过投掷硬币, 随机选择一个函数  $F$ ,  $F$  可以是  $u > v$  或  $u < v$ ,  $P_2$  不经意执行函数  $F$ 。由于函数  $F$  是随机选择的并且只有  $P_1$  可知, 因此函数  $F$  的结果对  $P_2$  来说是不经意的。根据  $F$  的选择以及结果的比较,  $P_1$  可以在本地使用同态的性质计算  $[\min(u, v)]$  以及  $E_{pk}(s_{\min(u, v)})$ 。

算法1列出了 SMIN 协议的大体步骤。一开始,  $P_1$  最初随机选择函数  $u > v$  或  $v > u$  作为  $F$ 。然后使用 SM 协议,  $P_1$  在  $P_2$  帮助下计算  $E_{pk}(u_i * v_i)$ , 其中  $1 \leq i \leq l$ 。之后,  $P_1$  本地执行协议的以下几个关键步骤。

- 一般情况下, 对于任意给定的两个比特  $o_1$  和  $o_2$ , 有性质  $o_1 \oplus o_2 = o_1 + o_2 - 2(o_1 * o_2)$ 。根据此性质, 对  $u_i$  和  $v_i$  逐比特异或。  $T_i = E_{pk}(u_i) * E_{pk}(v_i) *$

$$E_{pk}(u_i * v_i)^{N-2}。$$

- 初始化  $H_0 = E_{pk}(0)$ ，如果  $T$  中存在  $E_{pk}(1)$ ，保存第一次出现的  $E_{pk}(1)$ ，计算加密向量  $H$ 。 $H$  的其余部分通过  $H_i = H_{i-1}^{r_i} * T_i$  计算。注意，在  $H$  中出现的最多项是  $E_{pk}(1)$ ，剩余的项为 0 或随机数的加密值。
- 之后， $P_1$  计算  $\phi_i = E_{pk}(-1) * H_i$ ，在这里“-1”等价于  $\mathbb{Z}_N$  下的“N-1”。根据上述的讨论，由于  $H_i$  最多出现一次等于  $E_{pk}(1)$ ，因此，很清楚  $\phi_i = E_{pk}(0)$  最多出现一次。另外，如果  $\phi_j = E_{pk}(0)$ ，那么索引  $j$  位于  $u$  和  $v$  之间第一个比特位不同位置处。

---

**Algorithm 1**  $SMIN(u', v') \rightarrow [min(u, v)], E_{pk}(s_{min(u, v)})$  (Part 1)

---

```

1: Require:  $P_1$  有  $u' = ([u], E_{pk}(s_u))$  以及  $v' = ([v], E_{pk}(s_v))$ ,  $0 \leq u, v \leq 2^l$ ,  $P_2$  有  $sk$ 
2:  $P_1$  :
3: (a). 随机选取函数  $F$ 
4: (b). for  $i = 1$  to  $l$  do:
5:      $E_{pk}(u_i * v_i) \leftarrow SM(E_{pk}(u_i), E_{pk}(v_i))$ 
6:      $T_i \leftarrow E_{pk}(u_i \oplus v_i)$ 
7:      $H_i \leftarrow H_{i-1}^{r_i} * T_i; r_i \in_R \mathbb{Z}_N$  且  $H_0 = E_{pk}(0)$ 
8:      $\Phi_i \leftarrow E_{pk}(-1) * H_i$ 
9:     if  $F : u > v$  then:
10:         $W_i \leftarrow E_{pk}(u_i) * E_{pk}(u_i * v_i)^{N-1}$ 
11:         $\Gamma_i \leftarrow E_{pk}(v_i - u_i) * E_{pk}(\hat{r}_i); \hat{r}_i \in_R \mathbb{Z}_N$ 
12:     else
13:         $W_i \leftarrow E_{pk}(v_i) * E_{pk}(u_i * v_i)^{N-1}$ 
14:         $\Gamma_i \leftarrow E_{pk}(u_i - v_i) * E_{pk}(\hat{r}_i); \hat{r}_i \in_R \mathbb{Z}_N$ 
15:         $L_i \leftarrow W_i * \Phi_i^{r_i}; r_i' \in_R \mathbb{Z}_N$ 
16: (c). if  $F : u > v$ 
17:     $\delta \leftarrow E_{pk}(s_v - s_u) * E_{pk}(\bar{r})$ 
18:    else
19:     $\delta \leftarrow E_{pk}(s_u - s_v) * E_{pk}(\bar{r}), \bar{r} \in_R \mathbb{Z}_N$ 
20: (d).  $\Gamma' \leftarrow \pi_1(\Gamma), L' \leftarrow \pi_2(L)$ 
21: (e). 将  $\delta, \Gamma'$  以及  $L'$  发送给  $P_2$ 。

```

---

现在，根据  $F, P_1$  构造两个加密向量  $W$  和  $\Gamma$ ，其中  $1 \leq i \leq l$ :

---

 $SMIN(u', v') \rightarrow [min(u, v)], E_{pk}(s_{min(u, v)})$  (Part 2)

---

22:  $P_2$  :23: (a). 从  $P_1$  接收  $\delta, \Gamma'$  以及  $L'$ 24: (b). 解密:  $M_i \leftarrow D_{sk}(L'_i), 1 \leq i \leq l$ 25: (c). **if**  $\exists j$  使得  $M_j = 1$  **then**26:        $\alpha \leftarrow 1$ 27:       **else**28:        $\alpha \leftarrow 0$ 29: (d). **if**  $\alpha = 0$  **then**30:        $M'_i \leftarrow E_{pk}(0), 1 \leq i \leq l$ 31:        $\delta' \leftarrow E_{pk}(0)$ 32:       **else**33:        $M'_i \leftarrow \Gamma'_i * r^N$ , 其中  $r \in_R \mathbb{Z}_N, 1 \leq i \leq l$ 34:        $\delta' \leftarrow \delta * r_\delta^N$ , 其中  $r_\delta \in_R \mathbb{Z}_N$ 35:       将  $M', E_{pk}(\alpha)$  以及  $\delta'$  发送给  $P_1$ 36:  $P_1$ :37: (a). 从  $P_2$  接收  $M', E_{pk}(\alpha)$  以及  $\delta'$ 38: (b).  $\tilde{M} \leftarrow \pi_1^{-1}(M'), \theta \leftarrow \delta' * E_{pk}(\alpha)^{N-\bar{r}}$ 39: (c).  $\lambda_i \leftarrow \tilde{M}_i * E_{pk}(\alpha)^{N-\hat{r}_i}, 1 \leq i \leq l$ 40: (d). **if**  $F : u > v$  **then**:41:        $E_{pk}(s_{min(u, v)}) \leftarrow E_{pk}(s_u) * \theta$ 42:        $E_{pk}(min(u, v)_i) \leftarrow E_{pk}(u_i) * \lambda_i, 1 \leq i \leq l$ 43:       **else**44:        $E_{pk}(s_{min(u, v)}) \leftarrow E_{pk}(s_v) * \theta$ 45:        $E_{pk}(min(u, v)_i) \leftarrow E_{pk}(v_i) * \lambda_i, 1 \leq i \leq l$ 


---

- 当函数  $F$  为:  $u > v$  时, 计算

$$W_i = E_{pk}(u_i * (1 - v_i))$$

$$\Gamma_i = E_{pk}(v_i - u_i) * E_{pk}(\hat{r}_i) = E_{pk}(v_i - u_i + \hat{r}_i)$$

- 当函数  $F$  为:  $v > u$  时, 计算

$$W_i = E_{pk}(v_i * (1 - u_i))$$

$$\Gamma_i = E_{pk}(u_i - v_i) * E_{pk}(\hat{r}_i) = E_{pk}(u_i - v_i + \hat{r}_i)$$

$\hat{r}_i$  为  $\mathbb{Z}_N$  中的一个随机数。根据观察, 若函数  $F$  为  $u > v$ , 当且仅当  $u_i > v_i$  时有  $W_i = E_{pk}(1)$ , 反之则  $W_i = E_{pk}(0)$ 。同理, 当函数  $F$  为  $v > u$ , 当且仅当  $v_i > u_i$  时有  $W_i = E_{pk}(1)$ , 反之则  $W_i = E_{pk}(0)$ 。另外, 依据  $F$ , 随机选取  $u_i$  和  $v_i$  不同处, 进行加密并存储  $\Gamma_i$ , 用于后面的计算。

$P_1$  结合  $\phi$  和  $w$  计算  $L$ 。更准确的说,  $P_1$  计算  $L_i = W_i * \phi_i^{r'_i}$ ,  $r'_i$  为  $\mathbb{Z}_N$  中的随机数。观察可知, 如果存在索引  $j$ ,  $\phi_j = E_{pk}(0)$ , 表示  $u$  和  $v$  比特位中出现的第一位。  $W_j$  存储相应的所需信息, 即在加密形式中是  $u_j > v_j$  还是  $v_j > u_j$ 。另外, 根据  $F$ ,  $P_1$  计算  $s_u$  和  $s_v$  之间的随机差异的加密结果, 并记录为  $\delta$ 。特别的, 若函数  $F$  为  $u > v$ , 则  $\delta = E_{pk}(s_v - s_u + \bar{r})$ 。否则  $\delta = E_{pk}(s_u - s_v + \bar{r})$ , 其中  $\bar{r} \in_R \mathbb{Z}_N$ 。

$P_1$  使用两个随机的排列函数  $\pi_1$  和  $\pi_2$  变更加密向量  $\Gamma$  和  $L$ ,  $P_1$  计算  $\Gamma' = \pi_1(\Gamma)$ ,  $L' = \pi_2(L)$ , 并将结果连同  $\delta$  发送给  $P_2$ 。  $P_2$  接收到结果后, 解密  $L'$  分量, 得到  $M_i = D_{sk}(L'_i)$ ,  $1 \leq i \leq l$ , 同时检查索引  $j$ 。也就是说, 如果  $M_j = 1$ , 那么  $P_2$  设置  $\alpha$  为 1, 否则设置  $P_2$  为 0。另外,  $P_2$  根据  $\alpha$  的值计算一个新的加密向量  $M'$ 。准确的说, 如果  $\alpha = 0$ , 则  $M'_i = E_{pk}(0)$ ,  $1 \leq i \leq l$ 。这里对于每个  $i$ ,  $E_{pk}(0)$  是不同的。另一方面, 当  $\alpha = 1$ ,  $P_2$  设置  $M'_i$  为  $\Gamma'_i$  的一个随机值, 也就是说  $M'_i = \Gamma'_i * r^N$ , 其中  $r^N$  来自于双随机化, 对于  $r \in_R \mathbb{Z}_N$ , 每个  $i$  应该有不同的  $r$ 。此外, 若  $\alpha = 0$ , 则  $\delta' = E_{pk}(0)$ 。但是, 当  $\alpha = 1$  时,  $P_2$  设置  $\delta'$  为  $\delta * r_\delta^N$ , 其中  $r_\delta$  是  $\mathbb{Z}_N$  中的随机数。随后,  $P_2$  将  $M'$ ,  $E_{pk}(\alpha)$  以及  $\delta'$  发送给  $P_1$ 。  $P_1$  接收到  $M'$ ,  $E_{pk}(\alpha)$  以及  $\delta'$  后, 计算  $M'$  的逆置换  $\tilde{M}$ ,  $\tilde{M} = \pi^{-1}_1(M')$ 。接着  $P_1$  执行以下的同态操作计算  $\min(u, v)$  的第  $i$  比特位的加密结果, 即  $E_{pk}(\min(u, v)_i)$ ,  $1 \leq i \leq l$ :

- 通过计算  $\lambda_i = \tilde{M}_i * E_{pk}(\alpha)^{N - \hat{r}_i}$  去除  $\tilde{M}_i$  的随机性。
- 如果函数  $F : u > v$ , 计算  $E_{pk}(\min(u, v)_i) = E_{pk}(u_i) * \lambda_i = E_{pk}(u_i + \alpha * (v_i - u_i))$ 。否则, 计算  $E_{pk}(\min(u, v)_i) = E_{pk}(v_i) * \lambda_i = E_{pk}(v_i + \alpha * (u_i - v_i))$ 。

另外, 根据  $F, P_1$  也可以根据以下步骤计算  $E_{pk} = (s_{\min(u, v)})$ 。如果  $F : u > v$ ,  $P_1$  计算  $E_{pk}(s_{\min(u, v)}) = E_{pk}(s_u) * \theta$ , 其中  $\theta = \delta' * E_{pk}(\alpha)^{(N - \bar{r})}$ 。否则, 计算  $E_{pk}(s_{\min(u, v)}) = E_{pk}(s_v) * \theta$ 。

在 SMIN 协议中, 主要观察点是, 如果函数  $F : u > v$ , 则总有  $\min(u, v)_i = (1 - \alpha) * u_i + \alpha * v_i, 1 \leq i \leq l$ 。另一方面, 如果函数  $F : v > u$ , 则总有  $\min(u, v)_i = \alpha * u_i + (1 - \alpha) * v_i$ 。相同的结论可以用于描述  $s_{\min(u, v)}$ 。在这里强调, 使用类似的公式也可以设计一个 SMAX 协议, 用来计算  $\max(u, v)$  和  $E_{pk}(s_{\max(u, v)})$ 。同样, 可以将  $u$  和  $v$  的多个秘密作为 SMIN 和 SMAX 的输入 (以加密的形式)。例如, 设  $s_u^1$  和  $s_u^2$  分别为  $u$  和  $v$  相关联的秘密信息。则 SMIN 协议将  $([u], E_{pk}(s_u^1), E_{pk}(s_u^2))$  以及  $([v], E_{pk}(s_v^1), E_{pk}(s_v^2))$  作为  $P_1$  的输入, 并得到输出  $[\min(u, v)], E_{pk}(s_{\min(u, v)}^1)$  以及  $E_{pk}(s_{\min(u, v)}^2)$ 。

$P_1$ 选择函数 $F : v > u$ , 其中 $u = 55, v = 58$												
$[u]$	$[v]$	$W_i$	$\Gamma_i$	$G_i$	$H_i$	$\Phi_i$	$L_i$	$\Gamma_{i'}$	$L'_i$	$M_i$	$\lambda_i$	$\min_i$
1	1	0	$r$	0	0	-1	$r$	$1 + r$	$r$	$r$	0	1
1	1	0	$r$	0	0	-1	$r$	$r$	$r$	$r$	0	1
0	1	1	$-1 + r$	1	1	0	1	$1 + r$	$r$	$r$	-1	0
1	0	0	$1 + r$	1	$r$	$r$	$r$	$-1 + r$	$r$	$r$	1	1
1	1	0	$r$	0	$r$	$r$	$r$	$r$	1	1	0	1
1	0	0	$1 + r$	1	$r$	$r$	$r$	$r$	$r$	$r$	1	1

表 4.1: SMIN 协议中间结果

除了  $M_i$  列, 其他所有的列都是加密形式。另外  $r \in_R \mathbb{Z}_N$ , 每行每列不同

例 2. 简单起见, 设  $u = 55, v = 58, l = 6$ ,  $s_u, s_v$  分别为  $u$  和  $v$  相关联的秘密。假设  $P_1$  持有  $([55], E_{pk}(s_u))$  和  $([58], E_{pk}(s_v))$ 。另外, 我们假设  $P_1$  考虑如下的随机置换函数。不失一般性, 设  $P_1$  选择函数  $F$  为  $v > u$ 。基于 SMIN 协议的各种中间结果如表 1 所示, 根据表 4.1 的内容, 我们可以观察到:

i	=	1	2	3	4	5	6
		↓	↓	↓	↓	↓	↓
$\pi_1(i)$	=	6	5	4	3	2	1
$\pi_2(i)$	=	2	1	5	6	3	4

- $H$  中最多的意向为  $E_{pk}(1)$ , 即  $H_3$ 。剩余的项为 0 的加密值或者为  $\mathbb{Z}_N$  中一个随机数的加密值。



- $u$  和  $v$  比特位中第一次出现不同的位置为索引  $j = 3$  处。
- 由于  $H_3$  等于  $E_{pk}(1)$  则  $_3 = E_{pk}(0)$ ，同样由于  $M_5 = 1$ ，则  $P_2$  设置  $\alpha$  为 1。
- $E_{pk}(s_{\min(u,v)}) = E_{pk}(\alpha * s_u + (1 - \alpha) * s_v) = E_{pk}(s_u)$ 。

当然，只有  $P_1$  知道  $[\min(u, v)] = [u] = [55]$  以及  $E_{pk}(s_{\min(u,v)}) = E_{pk}(s_u)$ 。

**n 数中的最小安全值**  $P_1$  连同加密值考虑私有输入  $([d_1], \dots, [d_n])$ ，其中  $0 \leq d_i \leq 2^l$ ， $[d_i] = \langle E_{pk}(d_{i,1}), \dots, E_{pk}(d_{i,l}) \rangle, 1 \leq i \leq n$ 。  $P_2$  考虑  $sk$ 。此协议的主要目标是在不向  $P_1$  和  $P_2$  揭露任何有关  $d'_i$  的信息同时，计算  $[\min(d_1, \dots, d_n)] = [d_{\min}]$ 。另外，计算和目标最小值对应的信息加密值，用  $E_{pk}(s_{d_{\min}})$  表示。这里利用 SMIN 协议作为构建块，构建一个新的  $SMIN_n$  协议，此协议为一个迭代方式，以分层的方式计算所需的输出。在每一次迭代过程中，计算两两之间的最小值以及秘密对应的最小值，并将结果作为下一次迭代的输入。因此，将会以自底向上的方式生成一个二叉查询树。在结束阶段，仅有  $P_1$  知道最终的结果  $[d_{\min}]$  以及  $E_{pk}(s_{d_{\min}})$ 。

算法2描述了  $SMIN_n$  中涉及一些步骤。首先， $P_1$  将  $[d_i]$  以及  $E_{pk}(s_{d_i})$  分别分配给临时向量  $[d'_i]$  和变量  $s'_i$ ， $1 \leq i \leq n$ 。另外，他/她创建一个全局变量  $num$ ，并初始化至  $n$ ，其中  $num$  代表每轮迭代中向量的个数。由于  $SMIN_n$  在二叉查找树中以自底向上的方式执行的，因此我们有  $\lceil \log_2 n \rceil$  次迭代，每次迭代中，向量的个数都不相同。在第一轮迭代中 ( $i = 1$ )， $P_1$  同输入  $(([d'_{2j-1}], s'_{2j-1}), ([d'_{2j}], s'_{2j}))$ ， $P_2$  同 SMIN 协议中涉及的  $sk$ ， $1 \leq j \leq \lfloor \frac{num}{2} \rfloor$ 。在第一次迭代的末尾，仅有  $P_1$  知道  $[\min(d'_{2j-1}, d'_{2j})]$  以及  $s'_{\min(d'_{2j-1}, d'_{2j})}$ ， $1 \leq j \leq \lfloor \frac{num}{2} \rfloor$ 。其中， $P_2$  不会得到任何信息。另外， $P_1$  分别存储  $[d'_{2j-1}]$  中的结果值  $[\min(d'_{2j-1}, d'_{2j})]$ ， $s'_{\min(d'_{2j-1}, d'_{2j})}$  以及  $s'_{2j-1}$ 。另外， $P_1$  分别更新  $[d'_{2j}]$  的结果值， $s'_{2j}$  为 0,  $num$  为  $\lceil \frac{num}{2} \rceil$ 。

在第  $i$  次迭代中 ( $2 \leq i \leq \lceil \log_2 n \rceil$ )，协议 SMIN 仅考虑非零向量 (连同相同的加密信息)。例如，在第二次迭代中 (即  $i = 2$ )，仅考虑  $([d'_1] \square s'_1), ([d'_3], s'_3)$  等。在每轮迭代中，输出值仅  $P_1$  可知，并将  $num$  更新为  $\lceil \frac{num}{2} \rceil$ 。在  $SMIN_n$  协议末尾阶段， $P_1$  将分配全局最小值的最终加密二进制向量分配给  $[d_{\min}]$ ，即  $[\min(d_1, \dots, d_n)]$ 。同

---

**Algorithm 2**  $SMIN_n((([d_1], E_{pk}(s_{d_1})), \dots, ([d_n], E_{pk}(s_{d_n}))), \rightarrow ([d_{min}], E_{pk}(s_{d_{min}})))$ 


---

```

1: Require:  $P_1$  有  $(([d_1], E_{pk}(s_{d_1})), \dots, ([d_n], E_{pk}(s_{d_n})))$ ,  $P_2$  有  $sk$ .
2:
3:  $P_1$  :
4:   (a).  $[d'_i] \leftarrow [d_i]$  且  $s'_i \leftarrow E_{pk}(s_{d_i}), 1 \leq i \leq n$ 
5:   (b).  $num \leftarrow n$ 
6: for  $i = 1$  to  $\lceil \log_2 n \rceil$ :
7:   (a). for  $1 \leq j \leq \lfloor \frac{num}{2} \rfloor$ :
8:     if  $i=1$  then:
9:        $\bullet ([d'_{2j-1}], s'_{2j-1}) \leftarrow SMIN(x, y)$ , 其中  $x = ([d'_{2j-1}], s'_{2j-1})$  且
10:         $y = ([d'_{2j}], s'_{2j})$ 
11:        $\bullet [d'_{2ij-1}] \leftarrow 0$  且  $s'_{2ij-1} \leftarrow 0$ 
12:     else
13:        $\bullet ([d'_{2i(j-1)+1}], s'_{2i(j-1)+1}) \leftarrow SMIN(x, y)$  其中
14:         $x = ([d'_{2i(j-1)+1}], s'_{2i(j-1)+1})$  且  $y = ([d'_{2ij-1}], s'_{2ij-1})$ 
15:        $\bullet [d'_{2ij-1}] \leftarrow 0$  且  $s'_{2ij-1} \leftarrow 0$ 
16:   (b).  $num \leftarrow \lceil \frac{num}{2} \rceil$ 
17:  $P_1 : [d'_1]$  且  $E_{pk}(s_{d_{min}}) \leftarrow s'_1$ 

```

---

样,  $P_1$  将  $s'_1$  分配给  $E_{pk}(s_{d_{min}})$ 。例 3. 假设  $P_1$  持有  $\langle [d_1], \dots, [d_6] \rangle (n = 6)$ 。简单起见, 这里假设没有和  $d_i$  相关联的秘密。然后基于  $SMIN_n$  协议, 通过二叉执行树计算  $[min(d_1, \dots, d_6)]$ , 结果如图一所示, 注意, 这里初始化  $[d'_i] = [d_i]$ 。

**频率安全** . 考虑这样一种情况,  $P_1$  持有输入  $(\langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle, \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle)$ ,  $p_2$  持有密钥  $sk$ 。协议 SF 的目标是安全的计算  $E_{pk}(f(c_j))$ ,  $1 \leq j \leq w$ 。这里  $f(c_j)$  表示元素  $c_j$  在列表  $\langle c'_1, \dots, c'_k \rangle$  出现的次数, 即频率。这里假设  $c'_i \in c_1, \dots, c_w$ ,  $1 \leq i \leq k$ 。

输出值  $\langle E_{pk}(f(c_1)), \dots, E_{pk}(f(c_w)) \rangle$  仅  $P_1$  可知。在 SF 协议期间,  $P_1$  和  $P_2$  既没有揭露  $c'_i$  也没有揭露  $c_j$ , 同样,  $P_1$  和  $P_2$  也不知道  $f(c_j)$ , 其中  $1 \leq ik, 1 \leq j \leq w$ 。

算法3描述了 SF 协议涉及的整个步骤。开始阶段,  $P_1$  计算一个加密向量  $S_i$ , 有  $S_{i,j} = E_{pk}(c_j - c'_j)$ ,  $1 \leq j \leq w$ 。其次,  $P_1$  随机化  $S_i$  的分量得到  $S'_{i,j} = E_{pk}(r_{i,j} * (c_j - c'_j))$ , 其中  $r_{i,j}$  为  $\mathbb{Z}_n$  中的一个随机数。之后,  $P_1$  使用一个随

---

**Algorithm 3**  $SF(\Lambda, \Lambda') \rightarrow \langle E_{pk}(c_1), \dots, E_{pk}(f(c_w)) \rangle$ 


---

```

1: Require:  $P_1$  有  $\Lambda = \langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle, \Lambda' = \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle$  且  $\langle \pi_1, \dots, \pi_k \rangle, P_2$ 
   有  $sk$ 
2:  $P_1$  :
3: (a). for  $i = 1$  to  $k$  do:
4:      $T_i \leftarrow E_{pk}(c'_i)^{N-1}$ 
5:     for  $j = 1$  to  $w$  do:
6:          $S_{i,j} \leftarrow E_{pk}(c_j) * T_i$ 
7:          $S'_{i,j} \leftarrow S_{i,j}^{r_{i,j}}$ , 其中  $r_{i,j} \in_R \mathbb{Z}_N$ 
8:          $Z_i \leftarrow \pi_i(S'_i)$ 
9: (b). 将  $Z$  发送给  $P_2$ 。
10:  $P_2$ :
11: (a). 从  $P_1$  端接收  $Z$ 
12: (b). for  $i = 1$  to  $k$  do:
13:     for  $j = 1$  to  $w$  do:
14:         if  $D_{sk}(Z_{i,j}) = 0$  then  $u_{i,j} \leftarrow 1$ 
15:         else  $u_{i,j} \leftarrow 0$ 
16:          $U_{i,j} \leftarrow E_{pk}(u_{i,j})$ 
17: (c). 将  $U$  发送给  $P_1$ 
18:  $P_1$ :
19: (a). 从  $P_2$  接收  $U$ 
20: (b).  $V_i \leftarrow \pi_i^{-1}(U_i), 1 \leq i \leq k$ 
21: (c).  $E_{pk}(f(c_j)) \leftarrow \prod_{i=1}^k V_{i,j}, 1 \leq j \leq w$ 

```

---

机排列函数  $\pi_i$  (随机函数  $\pi_i$  仅  $P_1$  可知) 随机排列  $S'_i$  的分量。输出值  $Z_i \leftarrow \pi_i(S'_i)$  值发送给  $P_2$ 。 $P_2$  接收到输出值后, 解密  $Z_i$  的分量, 计算向量  $u_i$  并执行以下步骤:

- 如果  $D_{sk}(Z_{i,j}) = 0$ , 则设置  $u_{i,j} = 1$ , 否则  $u_{i,j} = 0$ 。
- 通过观察, 由于  $c'_i \in c_1, \dots, c_w$ , 则在向量  $Z_i$  中肯定有一个值为 0 的加密值, 剩下的项为随机数加密值。进一步的表明  $Z_i$  的解密值确切有一个为 0, 其余为随机数。更精确有, 若  $u_{i,j} = 1$  则  $c'_i = c_{\pi^{-1}(j)}$ 。
- 计算  $U_{i,j} = E_{pk}(u_{i,j})$  并发送给  $P_1$ , 其中  $1 \leq i \leq k, 1 \leq j \leq w$ 。

再后,  $P_1$  在  $u_i$  上执行行的逆排列, 得到  $V_i = \pi_i^{-1}(U_i), 1 \leq i \leq k$ 。最终,  $P_1$  计算

$$E_{pk}(c_j) = \prod_{k=1}^{i-1} V_{i,j}, \quad 1 \leq j \leq w.$$

在第三章中所有协议的输出都是以密文的形式输出的，并且只有  $P_1$  可知，另外， $P_1$  或得的中间结果要么是随机数要么是伪随机数。第三章提及的各种协议中，SMIN 协议最为复杂，因此本节将在标准假设下对 SMIN 协议进行安全性分析。当然相同的方法可用于分析在半可信模型下其他协议的安全性。

#### 4.5 证明 SMIN 安全性

通过执行 SMIN 证明 SMIN 在计算上是不可区分的，从而证明 SMIN 协议在半可信模型下是安全的。

执行过程通常包括消息的转换，以及来自这些消息的计算信息。因此，根据算法 1，设  $P_2$  的执行过程用  $\prod_{P_2}(SMIN)$  表示，由  $\langle \delta, s + \bar{r} \bmod N \rangle, \langle \Gamma'_{i, \mu_i + \hat{r}_i \bmod N} \rangle, \langle L'_i, \alpha \rangle$  给出。 $s + \bar{r} \bmod N$  和  $\mu_i + \hat{r}_i$  分别为  $\delta$  和  $\Gamma'_i$  的解密结果，其中  $1 \leq i \leq l$ 。模运算隐含在解密函数中。另外， $P_2$  接收来自于  $P_1$  的  $L'$ ， $\alpha$  表示  $L'$  的计算比较结果。不是一般性，假设  $P_2$  的模拟值为  $\prod_{P_2}^S(SMIN)$ ，通过  $\langle \delta^*, r^* \rangle, \langle s'_{1,i}, s'_{2,i} \rangle, \langle s'_{3,i}, \alpha' \rangle | 1 \leq i \leq l$  计算。这里  $\delta^*, s'_{1,i}$  以及  $s'_{3,i}$  是通过  $\mathbb{Z}_{N^2}$  随机生成的， $r^*, s'_{2,i}$  来自于  $\mathbb{Z}_N$  的随机生成数。另外， $\alpha'$  表示随机比特位。由于  $E_{pk}$  为语义安全加密方案，结果密文长度少于  $N^2$ ，因此  $\delta$  在  $\delta^*$  是计算不可区分的。同样， $\Gamma'_i$  和  $L'_i$  分别在  $s'_{1,i}$  和  $s'_{3,i}$  上是计算不可区分的。另外作为  $\mathbb{Z}_N$  的随机生成数  $\bar{r}$  和  $\hat{r}_i$ ， $s + \bar{r} \bmod N$  和  $\mu_i + \hat{r}_i \bmod N$  分别在  $r^*$  和  $s'_2$  上计算不可区分。而且，由于函数是  $P_1$  随机选择的（算法 1 的第一步）， $\alpha$  为 0 或 1 是等概率的。因此， $\alpha$  在  $\alpha'$  上计算不可区分。结合这些所有的特性，可以总结出  $\prod_{P_2}(SMIN)$  在  $\prod_{P_2}^S(SMIN)$  上市计算不可区分的。这就意味着在执行 SMIN 协议的时候， $P_2$  不需要知道任何关于  $u, v, s_u, s_v$  的信息以及实际的比较结果。直观的说，在执行 SMIN 协议期间， $P_2$  的信息要么是随机的要么是伪随机产生的，一次这些信息没有披露任何有关  $u, v, s_u$  以及  $s_v$  的相关信息。另外， $F$  仅  $P_1$  可知，实际的比较结果对  $P_2$  是隐藏的。

另一方面，用  $\prod_{P_1}(SMIN)$  表示执行  $P_1$ ， $\prod_{P_1}(SMIN) = M'_i, E_{pk}(\alpha), \delta' | 1 \leq i \leq l$ 。

$M'_i$  和  $\delta'$  是接收于  $P_2$  的加密结果值。设  $P_1$  的模拟结果由  $\prod_{P_1}^S(SMIN)$  表示,  $\prod_{P_1}^S(SMIN) = s'_{4,i}, b', b'' | 1 \leq i \leq l$ 。值  $s'_{4,i}, b'$  以及  $b''$  为  $\mathbb{Z}_{N^2}$  的随机生成数。由于  $E_{pk}$  结果密文长度少于  $N^2$ , 属于语义安全加密方案, 这意味着,  $M'_i, E_{pk}(\alpha)$  以及  $\delta'$  分别在  $s'_{4,i}, b'$  以及  $b''$  上是计算不可区分的。因此,  $\prod_{P_1}(SMIN)$  在  $\prod_{P_1}^S(SMIN)$  上是计算不可区分的。作为结果,  $P_1$  不可能知道任何有关  $u, v, s_u, s_v$  以及在执行 SMIN 协议期间的比较结果值信息。综合所有, 可以发现 SMIN 协议在半可信模型下是安全的。

## 4.6 本章小结

本章着重介绍一种能够满足加密数据上语义安全的  $k$  近邻分类器——PPKNNONED 方案。并介绍了协议中涉及到的一些子协议, 接着对子协议的安全性进行了相关的介绍, 并给出了相关的算法步骤。最后对子协议——最小值安全协议进行了安全性证明。

## 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  $D = \langle t_1, \dots, t_n \rangle$  以及  $m+1$  个属性，其中  $t_{i,j}$  表示记录  $t_i$  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  $q = \langle q_1, \dots, q_m \rangle$  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 4**  $PPKNNONED(D', q) \rightarrow c_q$ 


---

```

1: Require:  $C_1$  有  $D'$  以及  $\pi$ ;  $C_2$  有  $sk$ ; Bob 有  $q$ 
2: Bob:
3: (a). 计算  $E_{pk}(q_j), 1 \leq j \leq m$ 
4: (b). 将  $E_{pk}(q) = \langle E_{pk}(q_1), \dots, E_{pk}(q_m) \rangle$  发送给  $C_1$ 
5:  $C_1$  和  $C_2$ :
6: (a).  $C_1$  从 Bob 端接收  $E_p(q)$ 
7: (b). for  $i = 1$  to  $n$  do:
8:     •  $E_{pk}(d_i) \leftarrow SSED(E_{pk}(q), E_{pk}(t_i))$ 
9:     •  $[d_i] \leftarrow SBD(E_{pk}(c'))$ 
10: for  $s = 1$  to  $k$  do:
11:     (a).  $C_1$  和  $C_2$ :
12:         •  $([d_{min}], E_{pk}(I), E_{pk}(c')) \leftarrow SMIN_n(\theta_1, \dots, \theta_n)$  其中
13:              $\theta_i = ([d_i], E_{pk}(I_{t_i}), E_{pk}(t_{i.m+1}))$ 
14:         •  $E_{pk}(c'_s) \leftarrow E_{pk}(c')$ 
15:     (b).  $C_1$  :
16:         •  $\Delta \leftarrow E_{pk}(I)^{N-1}$ 
17:         • for  $i = 1$  to  $n$  do:  $\tau_i \leftarrow E_{pk}(i) * \Delta$ 
18:              $\tau'_i \leftarrow \tau_i^{r_i}, r_i \in_R \mathbb{Z}_N$ 
19:         •  $\beta \leftarrow \pi(\tau')$ ; 将  $\beta$  发送给  $C_2$ 
20:     (c).  $C_2$  :
21:         •  $\beta'_i \leftarrow D_{sk}(\beta_i), 1 \leq i \leq n$ 
22:         • 计算  $U', 1 \leq i \leq n$ :
23:             if  $\beta'_i = 0$  then  $U'_i = E_{pk}(0)$ 
24:             else  $U'_i = E_{pk}(0)$ 
25:         将  $U'$  发送给  $C_1$ 
26:     (d).  $C_1$  :  $V \leftarrow \pi_{U'}^{-1}$ 
27:     (e).  $C_1$  和  $C_2$ :
28:         •  $E_{pk}(d_{i,j}) \leftarrow SBOR(V_i, E_{pk}(d_{i,\gamma}))$ 
29:  $SCMC_k(E_{pk}(c'_1), \dots, E_{pk}(c'_k))$ 

```

---

PPKNNONED 方案涉及的主要步骤由算法4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  $E_{pk}(q) = \langle E_{pk}(q_1), \dots, E_{pk}(q_m) \rangle$  并将  $E_{pk}(q)$  发送给  $C_1$ 。阶段 1 所涉及的主要步骤由算法 4 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,  $E_{pk}(t_i) = \langle E_{pk}(t_{i,1}), \dots, E_{pk}(t_{i,m}) \rangle, 1 \leq i \leq n$ 。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  $[d_i] = \langle E_{pk}(d_{i,1}), \dots, E_{pk}(d_{i,l}) \rangle$  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  $\langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle$ 。首先考虑第一次迭代。 $C_1$  和  $C_2$  共同计算  $d_1, \dots, d_n$  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  $(\theta_1, \dots, \theta_2)$  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  $d_1, \dots, d_n$  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  $\tau_i$  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  $\tau'$  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  $\pi$  ( $\pi$  仅  $C_1$  可知) 排列  $\tau'$  得到  $\beta = \pi(\tau')$  并发送给  $C_2$ 。

在收到  $\beta$  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  $\tau'$  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  $[d_i] = \langle E_{pk}(1), \dots, E_{pk}(1) \rangle$ 。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 5**  $SCMC_k(E_{pk}(c'_1), \dots, E_{pk}(c'_k)) \rightarrow c_q$ 


---

```

1: Require:  $\langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle, \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle$  仅  $C_1$  可知
2:       $sk$  仅  $C_2$  可知
3:  $C_1$  和  $C_2$ :
4: (a).  $\langle E_{pk}(f(c_1)), \dots, E_{pk}(f(c_w)) \rangle \leftarrow SF(\Lambda, \Lambda')$ , 其中  $\Lambda = \langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle$ 
5:       $\Lambda' = \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle$ 
6: (b). for  $i = 1$  to  $w$  do:
7:       $[f(c_i)] \leftarrow SBD(E_{pk}(f(c_i)))$ 
8:      end for
9: (c).  $([f_{max}], E_{pk}(c_q)) \leftarrow SMAX_w(\psi_1, \dots, \psi_w)$ , 其中
10:     $\psi_i = ([f(c_i)], E_{pk}(c_i)), 1 \leq i \leq w$ 
11:  $C_1$ :
12: (a).  $\gamma_q \leftarrow E_{pk}(c_q) * E_{pk}(r_q)$ , 其中  $r_q \in_R \mathbb{Z}_N$ 
13: (b). 将  $\gamma_q$  发送给  $C_2$ ,  $r_q$  发送给 Bob
14:  $C_2$ :
15: (a). 从  $C_1$  接收  $\gamma_q$ 
16: (b).  $\gamma'_q \leftarrow D_{sk}(\gamma_q)$ ; 将  $\gamma'_q$  发送给 Bob
17: Bob:
18: (a). 从  $C_1$  接收  $r_q, C_2$  接收  $\gamma'_q$ 
19: (b).  $c_q \leftarrow \gamma'_q - r_q \bmod N$ 

```

---

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  $E_{pk}(c'_1), \dots, E_{pk}(c'_k)$ 。

## 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  $c = \langle c_1, \dots, c_w \rangle$ 。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  $\langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle$  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  $\Lambda = \langle E_{pk}(c_1), \dots, E_{pk}(c_w) \rangle$ 、 $\Lambda' = \langle E_{pk}(c'_1), \dots, E_{pk}(c'_k) \rangle$ ,

以及  $C_2$  的密钥  $sk$  安全计算  $E_{pk}(c_q)$ 。这里  $c_q$  表示  $c'_1, \dots, c'_k$  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5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  $(\Lambda, \Lambda')$  计算  $E_{pk}(f(c_i)), 1 \leq i \leq w$  作为 SF 协议的输入。输出值  $\langle E_{pk}(f(c_1)), \dots, E_{pk}(f(c_w)) \rangle$  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  $[f_{max}] = [max(f(c_1), \dots, f(c_w))]$ ,  $c_q$  表示  $\Lambda'$  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  $\gamma_q$  发送给  $C_2$ ,  $r_q$  发送给 Bob。  $C_2$  收到  $\gamma_q$  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  $\gamma'_q$  后, Bob 计算  $q$  对应的输出类标签  $c_q = \gamma'_q - r_q \bmod N$ 。

### 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  $w \ll n$ , 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'82. 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]. *Teletronikk*, 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网络诈骗案，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