

**本科毕业设计（论文）**

|  |  |
| --- | --- |
| **题目：** | 关于隐私保护分布式统计的算法研究 |

|  |  |  |
| --- | --- | --- |
| **学号** | ： | 1800710238 |
| **姓名** | ： | 王智坚 |
| **学院** | ： | 数学与计算科学 |
| **专业** | ： | 信息与计算科学 |
| **指导教师** | ： | 张必山 |
| **指导教师职称** | ： | 副教授 |

2022年5月16日

摘 要

在分布式统计计算中，若自身的设备无法处理庞大的计算量，可交由第三方服务计算商代为处理。若数据为隐私、机密数据，直接发送可能会造成数据的泄露，引发不可估量的后果。所以将数据加密后，第三方服务计算商对加密后的密文进行运算，接着将结果返回这样保证了数据的隐私性、正确性和机密性。本文通过全同态加密法加密数据，分别为一数据提供商与多第三方计算服务商、多数据提供商与一第三方服务计算商的情况下，构造了两个不同的方案。文章给出了方案的复杂度，正确性，安全性。

**关键词**：同态加密；安全多方计算；安全计算协议；全同态加密

Abstract

In distributed statistical computing, if its own equipment cannot handle the huge amount of computation, it can be handled by a third-party service computing provider. If the data is private and confidential, sending it directly may result in data leakage and incalculable consequences. Therefore, after encrypting the data, the third-party service computing provider operates on the encrypted ciphertext, and then returns the result, which ensures the privacy, correctness and confidentiality of the data. In this paper, data is encrypted by fully homomorphic encryption, and two different schemes are constructed for the case of one data provider and multiple third-party computing service providers, and multiple data providers and one third-party computing service provider. The article gives the complexity, correctness and security of the scheme.

**Key words**：Homomorphic encryption; secure multi-party computation; secure computing protocol; fully homomorphic encryption

目录

[引言 1](#_Toc99940051)

[1 相关工作和理论 2](#_Toc99940052)

[1.1 数据通过何种方式加密。 2](#_Toc99940053)

[1.2 数据的传输安全。 2](#_Toc99940054)

[1.3 本文结构 2](#_Toc99940055)

[2 理论基础 2](#_Toc99940056)

[2.1 背景知识 2](#_Toc99940057)

[2.2 典型的统计计算公式变换与同态计算 3](#_Toc99940058)

[2.2.1 统计计算公式变换 3](#_Toc99940059)

[2.2.2 加法同态性与统计计算 4](#_Toc99940060)

[2.3 通用安全多方计算 4](#_Toc99940061)

[3 安全的计算协议 5](#_Toc99940062)

[3.1 刘锵，唐春明[8]基于多租赁用户模型下的安全外包计算 5](#_Toc99940063)

[4 方案的分析 6](#_Toc99940064)

[4.1 正确性 6](#_Toc99940065)

[4.2 安全性 6](#_Toc99940066)

[4.3 局限性 6](#_Toc99940067)

[5 结束语 6](#_Toc99940068)

[参考文献 7](#_Toc99940069)

# 引言

信息大爆炸的今天，海量的数据扑面而来。面对这些激增的数据，如果自己的设备无法处理，那么就可以交由有能力的第三方代行计算，这时就涉及到了数据的隐私与机密的问题。例如，一款软件若要在本地处理数据，若用户设备性能低，在本地处理会大大消耗设备资源，造成使用卡顿，用户体验性极差。而将这些数据提交到云服务器上去计算，则能大大解决这一问题。这时迎来了有一个问题，若将用户的隐私数据直接提交给云服务器，就会造成数据的泄露，这就是大数据下的隐私问题。而将数据进行加密，再安全传输协议发送至云服务器，由云服务器代为处理，这也许是一个好的办法。

针对上诉问题，Rivest等人在20世纪70年代首先提出的“同态加密”能实现基本的加密操作之外，还能实现密文间的多种计算功能，即先计算后解密可等价于先解密后计算。这使得数据能由多方提供给云服务器进行密文计算，对数据的机密性，完整性和隐私性得到了有效的保护。

1985年，ElGamal [1]基于有限域上的离散对数困难假设设计了ElGamal加密算法，该加密方法具有乘法同态性，并且满足选择明文不可区分(IND-CPA)安全，但是ElGamal的一个不足之处是它的密文成倍扩张。应用最为广泛的当属Paillier[2]加密系统，基于高阶合数度剩余类困难问题，且具有IND-CPA安全。Goldwasser-Micali [3]加密系统属于异或同态加密系统，该加密系统基于二次剩余困难问题，虽具有IND-CPA安全，但每次只能加密单比特，因此加密效率会比较低。 1999—2005年间出现了不少浅同态加密方案，其中最著名浅同态加密方案当属Boneh[4]等基于理想成员判定困难假设设计的加密方案。该方案能执行一次乘法和若干次加法运算，Boneh [4]等虽然用它成功解决了2NF问题，但是该方案在解密时需要搜索解密，因此基于此方案的2NF保密计算协议效率很低。2009年9月，身处IBM的研究员Craig Gentry[5]发表一篇论文于STOC，他解决了一项棘手的数学问题，该问题自几十年前公钥加密发明以来，一致困扰着科学家们。他基于“理想格idea lattice”的数学对象，使人们可以充分操作加密状态的数据，即可以在不解密的情况下对加密数据进行任何可以在明文上的计算，对加密后的信息仍能深入和无限的分析，保证数据的隐私性、完整性和保密性。这个加密技术被称为全同态加密(full homomorphic encryption)。

通过上述可以了解同态加密的大致发展过程，目前支持密文计算的同态加密方案或许不支持多种密文混合计算，或者多次密文计算，亦或是复杂度高，那以实现，总之还较难应用于实际。针对上诉问题，设计一种方案，能使其适应分布式统计，保证其计算精度并保护其数据隐私。

# 相关工作和理论

对数据的加密与传输协议进行研究。

## 数据通过何种方式加密。

杨攀、桂小林、姚婧[6]等数据加密方案算法研究的CESIL方案，对数据进行加密。密钥的维度n越大，攻击者通过穷举攻击获得密钥的可能性越低。CESIL方案解决了已有加密方案不支持多次乘法及加法、乘混合运算的问题。但明文空间的限制使其只能计算正整数类型数据。针对上述问题， 引用文献[7]将分布式统计的常规计算公式进行转换，变形，使其能适应Craig Gentry的全同态加密方案，解决密文的计算问题。接着通过安全的传输通道，将加密后的数据传输至第三方服务计算商让其处理。

## 数据的传输安全。

文献[8]基于Shamir的密钥共享方案提出了安全的外包计算协议，将数据分成个秘密分块，通过秘密通道传输。在协议中，攻击者想通过截取秘密分块来重构数据是不可行的。Asharov等人基于文献[9-10]中的FTH方案给出了相对比较高效的TFHE方案的改造。该协议可以被证明在半恶意模型下安全，即模型中敌手基本遵照协议进行，但是可能会根据自己的视图恶意的产生运算中使用的随机数。

## 本文结构

第一部分本文的相关理论基础。第二部分分析计算协议对数据的安全。第三部分对方案正确性，安全性，复杂性以及局限性进行分析。

# 理论基础

## 背景知识

同态加密是一种加密形式，具有额外的评估能力，可以在不访问密钥的情况下计算加密数据。这种计算的结果仍然是加密的。同态加密可以看作是公钥密码学的扩展。同态是指代数中的同态，加密和解密函数可以被认为是明文空间和密文空间之间的同态。同态加密的方案中，有密钥生成算法，加密算法，解密算法

1）密钥生成算法。根据明文空间的随机数生成密钥。

2）加密算法。表示明文空间，表示密文空间，利用密钥加密明文，返回密文。

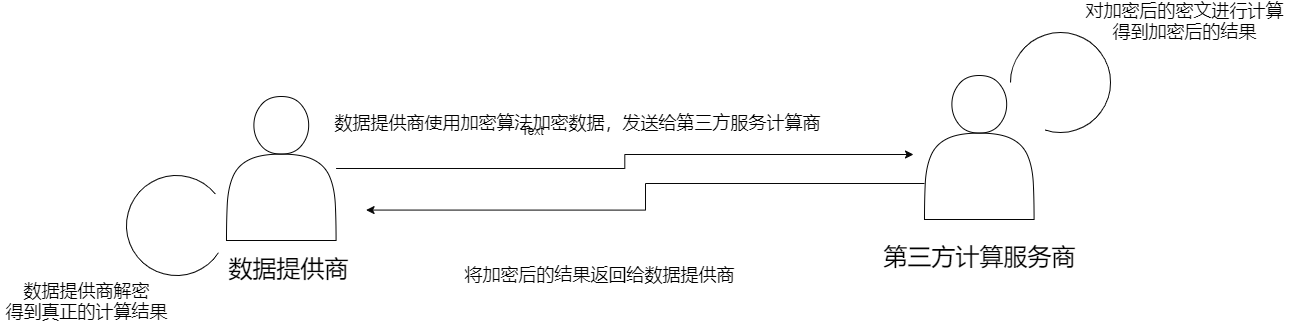
3）解密算法。利用密钥解密密文，返回明文

对于明文空间上的多项式



及数据，使用密钥生成算法，数据经过加密算法加密后，得到密文。这时候第三方对密文进行计算处理，得到加密后的结果，第三方将结果返回给数据拥有者，数据拥有者通过私钥，结合解密算法就可以将结果还原为明文形式，即





图表 1 同态加密在安全计算下的流程

加密后进行加法运算得到的结果解密后与明文运算的一致，称为加法同态加密。同理满足减法、乘法的称为减法同态加密，乘法同态加密，而能够对密文进行任意计算的称为全同态加密。

## 典型的统计计算公式变换与同态计算

### 统计计算公式变换

(1)算术平均：假设样本空间为,算数平均可以表示为



(2)方差：方差表征变量取值得散度，其变换后得计算公式如下



(3)线性回归：设有数据集：,一元线性回归的目的是找到线性方程去拟合这个数据集，最常采用最小二乘法来确定参数与，等价变换后公式如下





(4)相关系数：变量与的相关系数是用来衡量两者之间线性关系的强度与方向，其变换后的计算公式如下





### 加法同态性与统计计算

令中间结果： ，，，，。用密钥分别加密，，，，。根据全同态加密的加法同态和乘法同态，有一下等式成立







## 通用安全多方计算

安全多方计算（Secure Muti-party Computation ，简称MPC，也称为SMC或SMPC）问题是由中国计算机科学家、图领奖获得者姚启智教授于1982年首先提出的，即著名的百万富翁问题：两个胜负欲极强的有钱人Alice和Bob街上相遇。 如何在不暴露财富的情况下比较谁更富有？

后经Goldreich、Micali以及Wigderson等人的发展，成为现代密码学中非常活跃的研究领域，即安全多方计算，其数学描述为，“ 有个参与者，要以一种安全的方式共同计算一个函数，这里的安全是指输出结果的正确性和输入信息、输出信息的保密性。具体地讲，每个参与者，有一个自己的保密输入信息，个参与者要共同计算一个函数



计算结束时，每个参与者Pi只能了解,不能了解其他方的任何信息。” 简单来说，安全多方计算协议作为密码学的一个子领域，其允许多个数据所有者在互不信任的情况下进行协同计算，输出计算结果，并保证任何一方均无法得到除应得的计算结果之外的其他任何信息。换句话说，MPC技术可以获取数据使用价值，却不泄露原始数据内容。

# 安全的计算协议

本节分析参考文献中外包计算协议的好坏，然后选择适合于分布式统计的协议，设计出方案，使分布式统计的数据得到隐私保护。

## 刘锵，唐春明[8]基于多租赁用户模型下的安全外包计算

当数据拥有者选择一个不受信任的第三方服务计算商时，称为单租赁用户模型，也可以选择多个不受信任的第三方服务计算商时，称为多租赁用户模型。文献[8]在Kamara和Raykova结合门限密钥共享方案的安全协议上，提出了基于非门限存取结构的安全外包计算协议，原理如下。

对于任意函数，假设能被个值线性计算，即，则函数。因此，按照文献[11]中所得到的密钥共享方案，可以把分成个秘密分块，对于任意的，个秘密分块都可以线性计算。因此，我们能把函数转换为一个方函数



且任意的都能计算该函数。下面给出计算的外包计算协议。

假设用户和每个至今啊都存在一条秘密通道，任意可计算函数在个租赁用户中存在一个安全多方计算协议以及基于线性的线性密钥共享方案，则我们为计算提出一个外包计算协议，该协议首先的时非门限存取结构。

图表 2 安全的非门限外包计算协议

|  |
| --- |
| 1）用户根据算法把输入分成个秘密分块，每个秘密分块通过秘密通道发送给服务器，需要强调的是也许是一个向量。因为对于任意存取结构，不一定存在理想的密钥共享方案（若的秘密分块具有相同的长度，则他们是理想的密钥共享方案）。  2）计算参与方执行安全多方计算协议计算。执行的结果是每个得到输出的秘密分块。同理，也许是一个向量。  3）把秘密分块通过秘密通道发送给用户后，用户再使用算法从恢复出。 |

如果计算协议在动态和主动攻击者下是安全的，是一个安全的密钥共享方案，则协议在动态和主动攻击者下是安全的。其中中攻击者与协议中攻击者是一样的。

# 方案的分析

## 正确性

符合全同态加密的加法同态与乘法同态。由于加法的交换性，当把输入分成个秘密分块，对于多项式的计算是不会产生影响，保证其结算结果的正确性。

## 安全性

为所有参与方都是半诚实的，且使用的加密体制是安全的，计算参与方获取到的为加密数据的一部分，此时计算参与方是无法推算出原始数据，保证其安全性。

## 局限性

对于多次的加法与乘法，全同态加密仍无法做到加密后计算再解密的结果与直接计算的结果保持一致。此方案适用于将数据交由多个半诚实的计算参与方进行计算。

# 结束语

本文在半诚实模型下，关于数据在分布式计算下的隐私性提出了基于全同态加密的方案，并对方案进行了正确性，安全性与局限性的分析。能支持对数据的加密与解密， 保证数据的隐私不被泄露。

# 参考文献

* + - 1. T. Elgamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in IEEE Transactions on Information Theory, vol. 31, no. 4, pp. 469-472, July 1985, doi: 10.1109/TIT.1985.1057074.
      2. Paillier P . Public-key cryptosystems based on composite degree residuosity classes[J]. Advances in Cryptology Leurocrypt, 2004.
      3. Goldwasser, S.; Micali, S.; Rivest, R. L. (1988). "A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks". SIAM Journal on Computing. 17 (2): 281.
      4. D Boneh, EJ Goh, K Nissim (April 2006). "Evaluating 2-DNF Formulas on Ciphertexts"
      5. STOC '09: Proceedings of the forty-first annual ACM symposium on Theory of computingMay 2009 Pages 169–178https://doi.org/10.1145/1536414.1536440
      6. 杨攀,桂小林,姚婧,林建财,田丰,张学军.支持同态算术运算的数据加密方案算法研究[J].通信学报,2015,36(01):171-182.
      7. 马飞,蒋建国.具有隐私保护的分布式协作统计计算方案[J].计算机工程与设计,2015,36(09):2383-2387.DOI:10.16208/j.issn1000-7024.2015.09.013.
      8. 刘镪,唐春明,胡杏,张永强.多租赁用户模型下有效安全外包计算[J].信息网络安全,2013(09):17-21.
      9. Brakerski Z, Vaikuntanathan V. Effieient fully homomorphic eneryption from (standard) LWE [J]. SIAM Journal on Computing. 2014.43(2):831-871
      10. Brakerski Z, Gentry C. Vaikuntanathan V. (Leveled) fully homomorphic encryption without bootstrapping [C]//Proc of the 3rd Innovations in Theoretical Computer Science Conf.
      11. Tang CM, Gao SH, Zhang CL. The Optimal Linear Secret Sharing Scheme for Any Given Access Structure[EB/OL]. http://eprint.iacr.org/2011/147.ps,2011.