软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software [doi: 10.13328/j.cnki.jos.004794]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

大数据隐私保护密码技术研究综述*

黄刘生[1,2], 田苗苗[1,2], 黄 河[3]

[1](中国科学技术大学 计算机科学与技术学院,安徽 合肥 230026)

[2](中国科学技术大学 苏州研究院,江苏 苏州 215123)

[3](苏州大学 计算机科学与技术学院,江苏 苏州 215006)

通讯作者: 田苗苗, E-mail: miaotian@mail.ustc.edu.cn

摘 要: 大数据是一种蕴含大量信息、具有极高价值的数据集合.为了避免大数据挖掘泄露用户的隐私,必须要对大数据进行必要的保护.由于大数据具有总量庞大、结构复杂、处理迅速等新特点,传统的保护数据隐私的技术很多都不再适用.从密码学的角度,综述了近年提出的、适用于大数据的隐私保护技术的研究进展.针对大数据的存储、搜索和计算 3 个重要方面,分别阐述了大数据隐私保护的研究背景和主要研究方向,并具体介绍了相关技术的最新研究进展.最后指出未来大数据隐私保护研究的一些重要方向.

关键词: 大数据;隐私;存储;搜索;计算

中图法分类号: TP309

中文引用格式: 黄刘生,田苗苗,黄河.大数据隐私保护密码技术研究综述.软件学报 http://www. jos.org.cn/1000-9825/4794.htm

英文引用格式: Huang LS, Tian MM, Huang H. Preserving privacy in big data: A survey from the cryptographic perspective. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/4794.htm

Preserving Privacy in Big Data: A Survey from the Cryptographic Perspective

HUANG Liu-Sheng[1,2], TIAN Miao-Miao[1,2], HUANG He[3]

[1](School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China)

[2](Suzhou Institute for Advanced Study, University of Science and Technology of China, Suzhou 215123, China)

[3](School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: Big data is a type of data set, which has rich information and significant value. To avoid the leakage of the privacy of users during big data mining, we should take some necessary measures on big data. Since big data has a huger volume, a more complex structure and has less tolerance of delays, traditional privacy preserving technologies are mostly unsuitable for preserving privacy of the big data. From a cryptographic perspective, this paper surveys the recent progress of privacy preserving technologies for big data. By investigating into storage, search and computation such three basic problems in big data, we respectively elaborate their research backgrounds and the main research directions. We also present the latest research progress of privacy preserving technologies for these problems. Finally, we point out some important directions of privacy preserving technologies for big data to fill the current research gaps.

Key words: big data; privacy; storage; search; computation

大数据(big data)是一个新兴的概念,主要是指大量的、非结构化的数据.这些数据的产生主要是由于近几年传感技术、社会网络和移动设备的快速发展和大规模普及,导致数据量以指数形式快速增加并且数据的类型和相互关系也变得更加复杂多样.根据IBM的统计,现在世界上每天大约产生 250 亿亿字节的数据[1];2012 年EMC/IDC的调查显示,世界上的数据总量在过去两年翻了一番达到 2.8 ZB(1ZB=1024^4GB)[2].大数据体量庞大、

* 基金项目: 国家自然科学基金(61170058, 61202407, 61202028, 613032067)

收稿时间: 2014-01-20; 修改时间: 2014-04-15; 定稿时间: 2014-11-25; jos 在线出版时间: 2015-02-02

CNKI 网络优先出版: 2015-02-02 15:32, http://www.cnki.net/kcms/detail/11.2560.TP.20150202.1532.007.html

增长迅速,而且来源广泛、类型繁多.根据这些特点可以知道,与以往的大型数据集相比,一方面通过挖掘大量的、相互关联的大数据能够得到更多有价值的信息,另一方面由于数据量及数据类型的急剧增加,现有的数据处理技术很难在合理的时间内对大数据进行有效的处理.

目前对大数据的认识,比较有代表性的是IDC的 4V定义[3],即大数据具有数量庞大(volume)、处理迅速(velocity)、类型多样(variety)和价值量高(value)这 4 个特点.数量庞大是大数据的基本特征,是指大数据的数据含量从传统的大型数据集的TB级增加到至少PB级;处理迅速是大数据的典型需求,是指为了最大化大数据的价值,对它的处理通常必须很快;类型多样是大数据的内在特性,是指大数据包含的数据类型非常多样,包括文本、图片、音频、视频等;价值量高是大数据的终极意义,是指大数据中原本碎片化的信息能够通过整合而得到更多有价值的信息.大数据蕴含的巨大价值得到了产业界、学术界和政府部门的高度关注与重视,纷纷开展相关的研究来挖掘大数据带来的巨大价值(具体实例参见文献[4]的第 2 节).然而我们在使用大数据挖掘出各种各样的信息、享受大数据带来的便利时,我们的隐私也不可避免地受到大数据的严重威胁.因此,如何在充分利用大数据的同时不泄露用户的隐私,是一个非常重要的现实问题,关系到大数据研究的发展前途.

大数据的保护隐私问题本质上是一种数据隐私保护问题, 而数据隐私是指数据拥有者不愿意被披露的敏感数据或者数据所表征的特性[5]. 因此, 保护大数据隐私最根本的是保护敏感数据不被泄露, 也就是说大数据的隐私问题本质是大数据的泄露问题. 在大数据的整个生命周期内,可能发生数据泄露的领域目前来看主要包括大数据的存储、搜索和计算.与传统的数据隐私保护不同,大数据的存储、搜索和计算 3 个方面所面临的隐私保护问题都是新型的隐私保护问题,是由大数据规模大、增长速度不可预知等特点带来的.具体来说,由于大数据体量很大且增长速度不可预知,导致传统的存储模式不再适用于大数据.云计算[6,7]作为一种新型的商业模式,其提供的服务之一——存储服务,具有专业、经济和按需分配的特点,正好适合大数据的存储需求.因此,大数据一般存储在云上,由云存储服务提供者进行管理.虽然将大数据存放在云上极大地方便了数据的拥有者,但是云存储服务提供者并不完全可信(如实例[8]),这导致,(1) 数据拥有者必须验证存储在云上数据的完整性,防止数据被破坏;(2) 数据可能将以密文形式存储,所以数据拥有者需要高效的密文搜索算法来搜索存储在云上的加密数据;(3) 数据拥有者需要安全地利用云上的数据进行计算.图 1 展示了大数据隐私保护协议的整体框架.

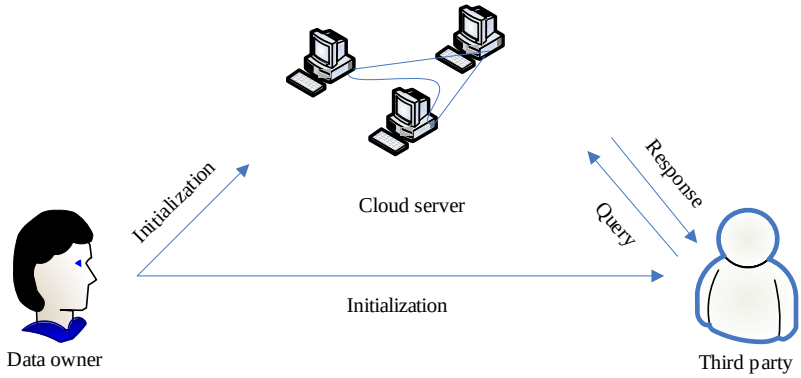

Fig.1 Framework of big data privacy preserving protocol

图 1 大数据隐私保护协议架构

在大数据的存储方面,虽然目前已经有很多协议可以较为高效地验证存储在云上的大数据的完整性[9−18],但是这些协议都需要数据拥有者亲自验证(文献[18]中的协议REDACT允许第三方验证数据的完整性,但是该协议每次仅能验证一个数据块,效率较低).限于数据拥有者的专业水平、计算能力等原因,令其频繁地验证大数据的完整性是不切实际的.因此在大数据完整性验证协议中引入第三方审计机构是一个自然的选择.然而,这种依赖于第三方的大数据完整性验证方法可能会将数据拥有者的数据泄露给审计机构.这就是说,大数据在存储阶段面临隐私泄露风险的主要原因是大数据的完整性验证协议采用了第三方审计机构.因此,大数据存储方面的主要隐私保护问题是如何设计一种安全高效的、能够阻止数据拥有者的数据泄露给第三方审计机构的大数

据完整性验证协议.

在大数据的搜索方面,我们知道大数据可能以明文或者密文两种方式存储,按照这两种不同的存储方式分类,大数据的搜索可分为明文搜索和密文搜索两种模式.如果大数据以明文方式存储,则相应的大数据搜索问题即为传统的数据查询和数据发布问题,否则大数据的搜索即为对密文的搜索.大数据以密文方式存储主要是因为这类大数据较为机密,如果大数据的机密性较低或者机密数据因为某些原因被解密,比如某用户得到数据拥有者的授权而获取某些机密数据等,此时大数据将(或者可以看作)以明文方式存储.以明文形式存储的大数据,如果对搜索不加以限制和防范必然会泄露数据的隐私,因此研究人员设计了很多保护隐私的明文搜索算法[5,19−38].然而这类搜索算法在大数据环境下很多都不适用,因为保护隐私的明文搜索算法往往需要对敌手的背景知识给出较为理想的假定,而这种假设恰恰与大数据具有的高度相关性矛盾,所以大多数保护隐私的明文搜索算法在大数据环境下是不安全的.本文考虑的大数据隐私保护搜索算法主要是密文搜索算法,这类算法比明文搜索算法更加安全.虽然密文搜索算法已不是新鲜事物,但是这类算法的安全性仅仅依赖于敌手的计算能力,与敌手的背景知识无关,所以更适用于大数据的隐私保护搜索.此外,对密文搜索算法的综述文章也鲜见于文献.在保护隐私的大数据密文搜索问题中,由于机密数据以密文形式存储,因此可以忽略数据泄露的问题而将隐私保护的重点放在如何设计满足大数据实际需求的密文搜索算法上.

另外,在大数据环境下,数据拥有者或者其他用户通常希望利用存储在云上的大数据,因此他们可能会需要云服务提供商计算特定的大数据并将计算结果返回.然而作为计算输入的大数据或者计算结果可能是非常机密的,所以保护隐私的大数据计算问题就是指如何能够安全的计算大数据而不泄露机密数据或计算结果给第三方.显然对于机密数据而言,直接将其提交给云服务提供商进行计算是不明智的,因此在大数据的计算中,数据和计算结果都需要以密文形式保存.在这种情况下,处理大数据的计算问题的一个有效的方法就是使用同态加密方案[39−41].实际上,能够完全保护大数据计算隐私的技术也几乎仅限于同态加密方案(虽然某些同态加密方案的派生方案也能保护大数据的计算隐私,但是由于它们本质上完全依赖于同态加密方案,所以本文不予讨论).由于大数据中用户较多,各用户所要求的计算问题可能也比较多,简单的同态加密方案是不适用的,所以本文关注的保护隐私的大数据计算方法是指完全同态加密方案,它可以对密文做任意复杂的计算,能够在理论上完美的解决大数据的隐私保护计算问题.

本文余下部分根据上述分类,首先从大数据的存储、搜索和计算 3 个基本方面分别介绍相关隐私保护算法的最新研究进展,然后指出下一步需要解决的若干重要问题.

1 大数据完整性审计协议

本节首先给出大数据完整性审计协议的基础知识,然后介绍相关协议的研究进展.

1.1 基础知识

大数据完整性审计协议有 3 个参与者:数据拥有者(data owner)、云存储服务提供者(cloud server)和第三方审计机构(auditor).协议的核心算法包括:系统建立算法、挑战和应答算法以及验证算法.下面对具体算法进行详细的介绍.

(1) 系统建立算法:该算法对应系统的初始化阶段.输入系统的安全参数,该算法将输出一些公共参数以及数据拥有者的公私钥对.然后数据拥有者将原始数据进行分块,用其私钥计算每个数据块的同态验证标签(homomorphic verifiable tag,简称 HVT).最后数据拥有者将 HVT 和相应的数据块一同存储在云服务器上,并将其公钥发送给第三方审计机构.

(2) 挑战应答算法:该算法是协议的主要部分.审计机构选择合适时机验证存储在云服务器上的大数据的完整性.当进行验证时,审计机构发送一个随机挑战(challenge)至云服务器,该挑战包括随机选择的部分数据块的标示符以及数据块所对应的随机数.云服务器接到挑战后,利用其存储的 HVT 和相应的数据块计算证明(proof)并将其返回.

(3) 验证算法:该算法用来验证响应的正确性.审计机构利用公共参数和数据拥有者的公钥来验证云服务

器发送的响应是否正确.如果应答正确,审计机构能以一定的概率确认数据是完整的,否则数据一定出现了损坏.如果审计机构希望以更高的概率确认数据是否完整,它可以重复多次运行挑战应答算法和验证算法.

为了降低协议的通信开销,在实际的协议设计中,云服务器发送的不是单个数据块的线性变换,而是一组数据块的线性组合,其中组合系数是由审计机构确定的.如果审计机构获得足够多云服务器对挑战的应答,那么审计机构可以轻松的通过解线性方程组而得到数据拥有者的数据.因此保护隐私的第三方审计协议的目的就是在完成验证数据完整性的前提下确保数据拥有者的数据不泄露给审计机构.

1.2 协议研究进展

目前,验证大数据完整性的第三方审计协议按照是否允许恢复原始数据分类,总体上可以分为只能验证数据完整性的PDP(proof of data possession)协议[42]和允许恢复数据的POR(proof of retrievability)协议[12]两类.这两类协议验证数据完整性的算法基本相同,主要区别是POR协议在验证数据完整性的基础上加入了纠错编码技术以便恢复原始数据.按照其他标准,例如协议是否允许数据动态变化,是否允许第三方验证,是否允许无限次的挑战或者是否保护数据拥有者的隐私等,这两类协议均可以再细分.由于大数据的完整性审计协议要求第三方在没有数据拥有者私钥的情况下能够独立地验证大数据的完整性,所以本文只关注允许第三方验证的PDP协议和POR协议.下面介绍相关协议的研究进展,其中重点关注协议是否适用于大数据以及是否具有隐私保护功能.

在PDP协议方面,首个支持第三方验证的PDP协议ABCH$^+$07 是由Ateniese等人[42]基于RSA困难问题设计的.之后Ateniese等人[43]提出了从任意同态认证协议构造允许第三方验证的PDP协议的一般框架,并首次具体实现了一种基于因子分解问题的PDP协议.以上协议都允许无限次挑战,能够避免频繁的系统建立过程,但是均不具有隐私保护功能也不支持数据的动态变化.Hao等人[44]通过将数据块的标签公开,设计了一种新的基于RSA问题的支持第三方验证的PDP协议HZY11,该协议允许无限次挑战、能够保护数据拥有者的隐私并且还支持数据的动态变化.基于因数分解问题和RSA问题设计的PDP协议所需的通信开销和存储开销都较大,Wang等人[45]采用双线性配对技术基于离散对数问题(discrete logarithm problem,简称DLP)提出了一种支持第三方验证的PDP协议.Wang等人声称该协议允许无限次挑战询问并且具有隐私保护功能,但是Xu等人[46]指出恶意的云存储服务提供者可以在数据受到破坏的情况下依然能够通过该PDP协议的检查.基于BLS短签名[47]技术,Hao等人[48]设计了一种安全的支持第三方公开验证的新PDP协议HY10.该协议可以进行无限次挑战询问并且能够验证多重副本数据完整性,但是不能保护数据拥有者的隐私也不支持对数据的动态操作.Zhu等人[49]基于双线性配对技术和Index-hash表设计了一种支持数据动态变化的第三方审计PDP协议,该协议允许无限次挑战询问并能够保护数据拥有者的隐私.考虑到数据拥有者可能将文件存放在多个云服务器上,Zhu等人[50]基于同态可验证响应(homomorphic verifiable response,简称HVR)和分层的哈希索引(hash index hierarchy,简称HIH)提出了一种适用的PDP协议ZHAY12,该协议允许无限次挑战询问并且支持第三方审计机构验证数据的完整性.通过引入一个可信组织,ZHAY12 协议也能够保护数据拥有者的隐私,然而却不支持对数据的动态操作.此外,以上两个协议都需要较高的计算和通信开销.为了解决这些问题,Yang等人[51]利用双线性配对的特点设计了一个高效的保护隐私的第三方审计PDP协议YJ13.该协议能够对多个数据拥有者存储在多个云服务器上的数据进行批量审计,并且支持数据的动态操作也允许审计机构进行无限次挑战询问.表 1 列出了几种PDP协议的比较结果.

在POR协议方面,Juels和Kaliski[12]提出的首个POR协议JK07 仅支持数据拥有者自己验证数据的完整性.首个允许第三方审计机构验证数据完整性的POR协议SW08 是由Shacham和Waters[52]基于BLS短签名设计的.该协议在随机预言模型下是安全的,允许无限次挑战询问但不能保护数据拥有者的隐私也不支持数据的动态变化.Bowers等人[53]提出了一个POR协议的理论框架,并对已有的JK07 协议和SW08 协议进行了优化.Dodis等人[54]采用困难放大HA(hardness amplification)技术设计了一种比SW08 协议更加高效的允许第三方审计的POR协议DVW09,该协议在标准模型下是安全的并且允许无限次挑战询问.但是与SW08 协议相同,DVW09 协议不具有隐私保护功能也不支持对数据的动态操作.为了支持对数据的动态操作,Wang等人[55]基于BLS短签名和Merkle哈希树构造了一个允许数据动态变化的第三方审计POR协议WWLR$^+$09.在此基础之上,Wang等人[56]利

用聚合签名技术[57]进一步将WWLR+09 协议扩展为允许对多个数据拥有者的数据进行批量审计的POR协议WWRL+11,但是这些协议都不能保护数据拥有者的隐私.表 2 列出了几种POR协议的比较结果.

Table 1 Comparison of several PDP protocols

表1 几种PDP协议比较

Protocol	Preserve privacy	Dynamic operation	Multiple replica	Security assumption
ABCH+07[42]	No	No	No	RSA
HZY11[44]	Yes	Yes	No	RSA
HY10[48]	No	No	Yes	DLP
ZHAY12[50]	Yes	Yes	Yes	DLP
YJ13[51]	Yes	Yes	Yes	DLP

Table 2 Comparison of several POR protocols

表2 几种POR协议比较

Protocol	Preserve privacy	Dynamic operation	Multiple replica	Security assumption
SW08[52]	No	No	No	DLP
DVW09[54]	No	No	No	DLP
WWLR+09[55]	No	Yes	No	DLP
WWRL+11[56]	No	Yes	No	DLP

2 大数据密文搜索算法

本节首先给出密文搜索算法的基础知识,然后分别对可搜索的对称加密算法和公钥加密算法的研究进展进行介绍.

2.1 基础知识

大数据的密文搜索算法有 3 个参与方:数据拥有者(data owner)、云存储服务提供者(cloud server)和检索人(searcher),其中检索人可能是数据拥有者.他们涉及的具体算法概括下来包括下面 4 种.

(1) 系统建立算法:该算法由数据拥有者运行,主要用来生成系统参数和数据拥有者的密钥.输入安全参数、数据拥有者生成系统的公开参数和自己的私钥.当考虑的是可搜索的公钥加密算法时,公开参数也包括数据拥有者的公钥.

(2) 数据加密算法:该算法用来加密可搜索的数据.输入公开参数和数据明文,该算法输出相应的密文(有时该算法还会输出加密的数据关键词的索引表).如果是可搜索的对称加密算法,则该算法还需输入数据拥有者的私钥.

(3) 令牌生成算法:当检索人需要搜索数据时,检索人需要向数据拥有者提交搜索请求,然后数据拥有者运行该算法对请求进行相应.该算法输入检索条件和数据拥有者的私钥,输出一个令牌(token)或称为陷门.

(4) 数据检索算法:检索人利用该令牌逐一测试密文或索引是否满足指定的检索条件,仅当满足条件时该算法才输出相应的密文或者索引.

如果大数据加密之后存储,由于大数据是以密文的形式存在,所以在密文搜索阶段不太可能会泄漏大数据的隐私.然而在这种情况下,为了保证大数据的可用性,必须要求对密文能够进行有效的检索和查询,所以本阶段的主要问题是如何设计能够满足大数据特点的可搜索的加密算法,即如何设计安全、运行效率高且允许对一般数据进行复杂搜索请求的密文搜索算法.下面分别介绍可搜索的对称加密算法和公钥加密算法的研究进展.

2.2 可搜索的对称加密算法

Song等人[58]首先考虑了在加密数据上搜索目标数据的问题,并设计一种基于对称加密的密文搜索算法SWP00,但是该算法的搜索效率和安全性都不高.Goh[59]和Chang等人[60]分别给出了可搜索的对称加密算法更强的安全定义并且分别基于布隆过滤器[61]和伪随机函数提出了改进的可搜索对称加密算法.Curtmola等人[62]针对以往可搜索的对称加密算法的安全定义相对较弱的问题,提出了一种更强的安全模型并基于树结构设计了高效的算法CJKO06.此外,Curtmola等人在文献[62]中也首次给出了允许多个用户搜索的可搜索对称加密算

法的定义以及一个具体的算法,但是当用该算法处理频繁升级大数据时将非常耗时.为了解决这个问题,Van Liesdonk等人[63]设计了一种新的可搜索的对称加密算法VSDH+10,该算法具有较高的搜索效率并且允许数据系统快速升级.以上可搜索的对称加密算法仅考虑了被动敌手的攻击,然而在大数据环境下数据拥有者存放在云服务器上的加密数据可能会被云存储服务提供商主动的删除,因此能够抵抗敌手主动攻击的可搜索的对称加密算法更加适用.针对这个问题,Kurosawa等人[64]给出了抗主动敌手攻击的可搜索的对称加密算法的定义以及一个通用可组合(universally composable,UC)[65]安全的高效算法KO12.Kamara等人[66]针对之前的可搜索的对称加密算法的搜索效率、动态性和安全性不能完全兼顾的情况,提出了一种动态的可搜索的对称加密算法KPR12,该算法达到当时已知的最高安全性并且具有很高的搜索效率.之后Kamara等人[67]改进了文献[66]中的方案,改进方案具有更高的安全性和搜索效率,特别地,改进方案还允许并行操作.Chase和Kamara[68]考虑了更加一般的可搜索的对称加密问题,即对结构化的数据加密后的搜索问题,并且在文献[68]中给出了可搜索的结构化对称加密算法的定义、模型以及一些具体方案.此外,为了满足多关键词的搜索需求,Golle等人[69]给出了允许多关键词搜索的对称加密算法的安全模型和具体方案GSW04.随后Ballard等人[70]利用秘密分享和双线性对技术提出了性能更高的改进方案BKM05.

由于上述可搜索的对称加密算法都只允许简单的精确匹配,也就是说用户输入的搜索仅仅是针对某一具体密文的,而对于更加复杂的搜索请求,比如区间搜索,这类算法是无效的.为了处理这一问题,Agrawal等人[71]引入了保序加密(order-preserving encryption,简称OPE)的概念并给出了一个具体的算法AKSX04.OPE能够保证密文顺序和对应明文顺序的一致性,便于云服务器对大数据的管理也能处理区间搜索请求.然而早期的OPE算法在安全性和效率方面都不尽人意.例如,Agrawal等人提出的OPE算法AKSX04 需要输入所有数据才能对数据进行加密并且他们也没有给出算法正式的安全性证明.Boldyreva等人[72]重新回顾了OPE这一概念,正式定义了OPE的安全模型并利用伪随机函数和超几何分布设计了一个可证安全的OPE算法BCLO09.该算法一经提出便在云计算的数据隐私保护方面发挥了巨大的作用[73,74].随后Boldyreva等人[75]对OPE算法BCLO09 的安全性进行了更加深入的分析,并且提出了高效的有序加密(efficiently orderable encryption,简称EOE)的概念以及一个具体算法BCO11.EOE扩展了OPE,泛指任意可以实现区间搜索的对称加密算法.Popa等人[76]针对以往保序加密方案的安全性较差的问题,提出了一种安全性更高的保序加密算法PLZ13.

表 3 对可搜索的对称加密算法进行了总结.

Table 3 Comparison of several searchable symmetric encryption algorithms

表3 几种可搜索的对称加密算法比较

Algorithm	Search pattern	Efficiency	Security	Dynamic
SWP00[58]	Single keyword	Low	Low	No
CJKO06[62]	Single keyword	Medium	Medium	No
VSDH+10[63]	Single keyword	High	Medium	Yes
KO12[64]	Single keyword	Medium	High	No
KPR12[66]	Single keyword	Medium	High	Yes
GSW04[69]	Multiple keywords	Low	Medium	No
BKM05[70]	Multiple keywords	Medium	Medium	No
AKSX04[71]	Range query	Low	Low	No
BCLO09[72]	Range query	Low	Medium	No
BCO11[75]	Range query	Medium	Medium	No
PLZ13[76]	Range query	Medium	High	No

2.3 可搜索的公钥加密算法

Boneh等人[77]将可搜索加密从对称密码体制转移到公钥密码体制中来,首次提出了可搜索关键词的公钥加密的概念并基于双线性对技术给出了几个具体的算法.然而Abdalla等人[78]指出Boneh等人的方案不满足一致性,在此基础上,Abdalla等人提出可搜索关键词的公钥加密算法的新定义,以及一个新方案和相关的扩展.此外,Abdalla等人在文献[78]中也给出了从基于身份的匿名加密方案设计可搜索关键词的公钥加密算法的一般方法.Baek等人[79]针对Boneh等人的方案[77]需要安全通道的问题,基于文献[80]的聚合签名技术提出了一个不

需要安全通道的可搜索关键词的公钥加密方案BSS08,该方案在随机预言模型下是可证安全的.Rhee等人[81]提出了一个安全性更高的不需要安全通道的可搜索的公钥加密方案,但该方案仍然是在随机预言模型下安全的.为了消除随机预言机,Fang等人[82]基于Gentry[83]标准模型下安全的基于身份的加密方案设计了首个标准模型下安全的不需要安全通道的可搜索的公钥加密方案FSGW09.针对大多数可搜索的公钥加密算法都依赖于双线性对的问题,Crescenzo等人[84]利用Cocks[85]的基于身份的加密方案设计了一种基于二次剩余问题的可搜索的公钥加密算法DS07.

为了设计搜索效率更高的可搜索的公钥加密算法,Bellare等人[86]提出了确定性加密(deterministic encryption,简称DE)的概念并且指出确定性加密是一类高效的可搜索的公钥加密方案(确定性加密是指对于同一个公钥和明文,确定性加密算法输出的密文相同).此外,Bellare等人也给出了在随机预言模型下可证安全的确定性加密方案BBO07.随后Bellare等人[87]利用单向陷门置换函数设计了标准模型下可证安全的确定性加密方案BFOR08,但是方案要求所加密的消息必须是随机独立的.Boldyreva等人[88]基于有损陷门函数(lossy trapdoor function,简称LTF)[89]也给出了标准模型下可证安全的确定性加密方案,方案所加密的消息虽然没有额外的限制但是方案基于的安全模型却较弱.Fuller等人[90]统一了确定性加密方案,指出确定性加密方案可以统一的利用陷门函数构造.由于在大规模系统中,敌手可能会得到用户的额外信息,而之前的确定性加密的安全模型没有体现这种情况,因此可能在实际使用中并不安全.为此,Brakerski和Segev[91]提出了具有额外输入的确定性加密的概念并设计了两个方案.第1个方案在d-linear Diffie-Hellman假设下可以保证即使在多用户的环境下也是安全的,而第2个方案在一般的子群不可区分假设下是可证安全的.Wee[92]提出了双投影哈希(dual projective has,简称DPH)的概念,指出Brakerski和Segev提出的两个具有额外输入的确定性加密方案可以统一的由双投影哈希构造.此外,Wee也基于格(lattice)上LWE假设[93]设计了一种具有额外输入的确定性加密方案W12.Xie等人[94]基于格上LWE假设也设计了一种标准模型下安全的具有额外输入的确定性加密算法XXZ12.Mironov等人[95]注意到确定性加密算法所加密的数据往往比较大,而数据之间的差别有时却很小,为了提高以往确定性加密算法的计算效率,Mironov等人提出了增量确定性加密(incremental deterministic encryption,简称IDE)的概念并给出了由普通确定性加密算法设计增量确定性加密算法的一般方法.此外,值得一提的是Cui等人[96]基于编码理论也设计了一种确定性加密方案.

以上可搜索的公钥加密算法仅允许关键词的精确匹配,为了支持更一般的搜索请求,Boneh等人[97]设计了允许关键词比较、子集查询以及任意合取连接词查询的可搜索的公钥加密方案BW07.针对BW07 算法的效率和安全性都较低的问题,Hwang等人[98]给出了一个改进方案HL07,并将其扩展到多用户搜索的场景中.Katz等人[99]利用更加复杂的双线性对技术提出了查询谓词可以为任意析取连接词、多项式和内积的可搜索关键词的公钥加密方案KSW08.

表4对可搜索的公钥加密算法进行了总结.

Table 4 Comparison of several searchable asymmetric encryption algorithms

表4 几种可搜索的公钥加密算法比较

Algorithm	Search pattern	Efficiency	Security model	Security ssumption
BSS08[79]	Single keyword	Low	ROM	DLP
FSGW09[82]	Single keyword	Low	SM	DLP
DS07[84]	Single keyword	Low	ROM	QR
BBO07[86]	Single keyword	Medium	ROM	RSA
BFOR08[87]	Single keyword	Medium	SM	Permutation
W12[92]	Single keyword	Medium	ROM	LWE
XXZ12[94]	Single keyword	Medium	SM	LWE
BW07[97]	Subset query	Low	SM	DH
HL07[98]	Subset query	Medium	ROM	DLDH
KSW08[99]	General query	Low	SM	DLP,RSA

3 大数据安全计算问题

在大数据的应用中,安全的计算用户的数据、保护用户的隐私是大数据面临的一个基本问题.如前文所述,

由于大数据中的计算问题非常复杂、多样,所以适用于特定计算情况的隐私保护算法通常不能满足大数据的需求.为此,必须选择一个功能较全面的方案来保护用户的计算隐私.完全同态加密(fully homomorphic encryption,简称FHE)算法是一个合适的选择(实际上其他类型的安全计算问题如安全多方计算[100]等都可以由完全同态加密实现[101,102]).本节首先给出完全同态加密的基础知识,然后介绍相关的研究进展.

3.1 基础知识

典型的完全同态加密方案包括两个参与者:用户(user)和云服务提供者(cloud server),其中用户是数据的拥有者.用户通常首先将其数据加密之后存放于云服务器上,当用户需要对云上的数据进行计算时,他发送通知给云服务器,然后云对数据进行相应的计算并最后将计算结果返回给用户.下面对完全同态加密的基础知识做一些介绍.

完全同态加密方案简单来说就是一种不需要密钥就能够实现对密文进行任意操作的加密方案,一般包括密钥生成(keygen)、加密(encrypt)、求值(evaluate)和解密(decrypt)4 种算法.

(1) 密钥生成算法:该算法输入安全参数,输出用户的公钥和私钥.

(2) 加密算法:该算法输入用户的公钥和明文数据,输出相应的密文.

(3) 求值算法:该算法输入用户的公钥、一个函数和一组密文,输出一个新密文.

(4) 解密算法:该算法输入用户的私钥和密文,输出对应的明文数据.

根据上面对算法的描述可知,在大数据的安全计算中,用户可以首先请求云调用求值算法对密文进行操作,然后云将计算结果返回给用户,最后用户利用私钥进行解密得到期望的结果.为了满足大数据计算应用对响应时间和安全性的要求,完全同态加密方案的研究主要集中于提高其运行效率和安全性,本节下面将着重对这两个方面进行介绍.

3.2 完全同态加密方案

同态加密方案不是新事物,实际上在Gentry[41]于 2009 年提出完全同态加密方案之前已经有了一些具有同态性质的加密方案,它们主要基于因数分解和离散对数问题.虽然那些方案的效率尚可,但是有两个致命的弱点使得它们不能用来保护大数据的计算隐私.其一是那些方案仅允许执行较为简单的加密操作,例如文献[39,103,104]中的同态加密方案只允许对密文进行加法操作、文献[40]中的同态加密方案只能对密文进行乘法操作,Boneh等人的方案[105]虽然能够对密文进行更复杂的操作,但也仅仅限于二次函数.大数据的实际应用需要对密文进行较为复杂的操作,以上同态加密方案显然不能满足这种需求.此外,考虑到大数据存在的长期性,从安全的角度来看上述同态加密方案也是不可取的,因为它们在未来的量子时代是不安全的[106].可喜的是,Gentry在文献[41]中基于格上困难问题设计了首个完全同态加密方案G09,它允许任意复杂的操作并且在量子时代也是安全的,从而解决了上述问题.Gentry和Halevi[107]完整地实现了G09 方案,运行结果显示该方案需要较大的时间和空间开销.Scholl和Smart[108]以及Stehle和Steinfeld[109]分别改进了G09 方案,得到了两个运行效率更高的完全同态加密方案.

在G09 方案的基础上,Smart和Vercauteren[110]利用中国剩余定理设计了一个密钥和消息长度都较小的完全同态加密方案SV10.Gentry等人在文献[111]中采用将明文打包的方法,基于R-LWE问题[112]设计了一个时间开销仅为多项式对数(polylog)的完全同态加密方案,随后在文献[113]中又通过将模设置为 2 的幂的近似值,得到了一个效率更高的完全同态加密方案GHS12.以上这些完全同态加密方案使用的明文打包技术仅限于R-LWE问题,Brakerski等人[114]利用Peikert等人[115]的打包技术设计了一种基于标准LWE问题的完全同态加密方案BGH13,该方案概念更加简单并且具有更高的安全性.

Gentry最初的完全同态加密方案G09 的安全性基于理想格中一种判定问题和稀疏子集求和问题(sparse subset sum problem,简称SSSP),而这两个问题都仅能规约到平均情况下的困难问题.为了将完全同态加密方案建立在更安全的基础之上,Gentry在文献[116]中设计了一种新的密钥生成算法,将完全同态加密方案的安全性建立在稀疏子集求和问题和理想格中一种最坏情况下的困难问题之上.然而方案仍然需要平均情况下困难的

稀疏子集求和问题.Brakerski和Vaikuntanathan[117,118]基于标准LWE问题和R-LWE问题分别设计了两个新的完全同态加密方案,由于LWE问题和R-LWE问题都可规约到最坏困难问题,所以新方案具有更高的安全性.随后Brakerski等人[119]对文献[117,118]中的两个完全同态加密方案的具体参数进行了改进.鉴于以往基于LWE和R-LWE问题的完全同态加密方案的安全性需要量子规约,Brakerski[120]利用尺寸不变性,设计了一种新的完全同态加密方案B12,该方案具有很高的效率,安全性可以使用经典技术规约到格上的困难问题.为了有效地对不同用户的加密数据进行计算,Lopez-Alt等人[101]在理想格中设计了一种允许多个密钥参与的完全同态加密方案,该方案比传统的完全同态加密方案更加灵活实用,是一次大的创新.然而其安全性却依赖于一个非标准的假设.Bos等人[121]采用Brakerski[120]提出的张量技术消除了这个非标准假设.

上述完全同态加密方案都基于格问题,描述较为复杂,不易理解.van Dijk等人[122]在整数环上设计了一个容易理解的完全同态加密方案VGHV10,其安全性依赖于近似最大公约数(approximate-greatest common divisor,简称A-GCD)问题.Coron等人[123]针对van Dijk等人方案的公钥过大的问题,给出了一个改进的方案.改进方案具有较短的公钥,但是安全性基于较强的近似最大公约数假设.Chen和Nguyen[124]针对这个较强的近似最大公约数假设给出了一个有效的攻击算法,指出Coron等人的方案[123]实际上是不安全.随后Coron等人[125]又提出了一个新的整数环上的完全同态加密方案CNT12,该方案的效率比以往的更高,安全性依赖于标准的近似最大公约数问题.为了更有效地处理整数环上的完全同态加密方案,类比于格上的方案,Cheon等人[126]设计了两个整数环上的批处理完全同态加密方案.第 1 个方案的安全性依赖于判定近似最大公约数问题,而第 2 个方案的安全性则依赖于无误近似最大公约数问题.

表 5 对完全同态加密方案进行了总结.

Table 5 Comparison of several fully homomorphic encryption schemes

表5 几种完全同态加密方案比较

Scheme	Efficiency	Security assumption
G09[41]	Low	SSSP
SV10[110]	Medium	SSSP
GHS12[113]	Medium	R-LWE
BGH13[114]	Low	LWE
B12[120]	Medium	GapSVP
VGHV10[122]	Low	A-GCD
CNT12[125]	Medium	A-GCD

4 总结和展望

本文介绍了大数据在存储、搜索和计算 3 个方面的隐私保护问题并分别介绍了相关的研究进展.这 3 个方面的隐私问题是所有类型的大数据所面临的主要隐私问题,然而对某一具体的大数据来说,这 3 个问题不一定全部都需要考虑.例如,有些公司的大数据是一些非机密数据,在这种情况下,该公司就不需要保护其大数据的搜索隐私.如果存储某公司大数据的云服务器是该公司自有,此时大数据的计算隐私问题也就自动消除.根据本文的介绍可以看出,虽然已经有一些工具能够处理大数据三个方面的隐私问题,但是这些工具仍有很多方面亟待改进.具体来说,有以下方向值得进一步研究.

(1) 在大数据的完整性审计协议方面,目前还没有能够保护数据拥有者隐私的支持数据动态变化的第三方审计 POR 协议.鉴于 POR 协议具有数据恢复功能,比 PDP 协议具有更高的实用价值,所以设计这类 POR 协议是大数据隐私保护研究的一个重要方向.

(2) 在大数据的安全搜索方面,虽然目前可搜索的对称加密算法和公钥加密算法在某些方面表现不错,但是仍然没有一个算法能够同时支持一般结构数据的动态变化和多关键词搜索.由于大数据的结构更复杂、动态性更强、搜索请求更多样,所以设计这类算法是非常重要的.

(3) 目前的完全同态加密方案可以很好地解决大数据的安全计算问题,并且已有的基于 LWE 问题的部分完全同态加密方案也能达到较为理想的安全性,但是这些完全同态加密方案的运行效率仍不高,不能满足大数据的实时性要求.因此设计运行效率更高的完全同态加密算法仍是一个重要的研究方向.

References:

[1] What is Big Data. http://www-01.ibm.com/software/data/bigdata

[2] Study: Only 1% of World's Data is Analyzed, Less Than 20% Protected. http://www.webhostmagazine.com/2012/12/study-only-1-of-worlds-data-is-analyzed-less-than-20-protected

[3] IIIS: The 'four Vs' of Big Data. http://www.computerworld.com.au/article/396198/iiis_four_vs_big_data/

[4] Li G, Cheng X. Research status and scientific thinking of big data. Bulletin of the Chinese Academy of Sciences, 2012,27(6): 647−657 (in Chinese with English abstract).

[5] Zhou SG, Li F, Tao YF, Xiao XK. Privacy preservation in database applications: A survey. Chinese Journal of Computers, 2009, 32(5):847−861 (in Chinese with English abstract).

[6] Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M. A view of cloud computing. Communications of the ACM, 2010,53(4):50−58.

[7] Feng DG, Zhang M, Zhang Y, Xu Z. Study on cloud computing security. Ruan Jian Xue Bao/Journal of Software, 2011,22(1): 71−83 (in Chinese with English abstract).

[8] Amazon's S3 down for several hours. http://www.pcworld.com/businesscenter/article/142549/amazons s3 down for several hours.html

[9] Ateniese G, Di Pietro R, Mancini LV, Tsudik G. Scalable and efficient provable data possession. In: Proc. of the 4th Int'l Conf. on Security and Privacy in Communication Netowrks. New York: ACM Press, 2008. 1−10.

[10] Erway C, Küpçü A, Papamanthou C, Tamassia R. Dynamic provable data possession. In: Proc. of the 16th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2009. 213−222.

[11] Curtmola R, Khan O, Burns R, Ateniese G. MR-PDP: Multiple-Replica provable data possession. In: Proc. of the 28th IEEE Int'l Conf. on Distributed Computing Systems (ICDCS). Beijing: IEEE Computer Society, 2008. 411−420.

[12] Juels A, Kaliski BS. PORs: Proofs of retrievability for large files. In: Proc. of the 14th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2007. 584−597.

[13] Shacham H, Waters B. Compact proofs of retrievability. In: Advances in Cryptology-ASIACRYPT 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 90−107.

[14] Dodis Y, Vadhan S, Wichs D. Proofs of retrievability via hardness amplification. In: Proc. of the 6th Theory of Cryptography Conference (TCC). Berlin, Heidelberg: Springer-Verlag, 2009. 109−127.

[15] Zheng Q, Xu S. Fair and dynamic proofs of retrievability. In: Proc.of the 1st ACM Conf. on Data and Application Security and Privacy. New York: ACM Press, 2011. 237−248.

[16] Bowers KD, Juels A, Oprea A. HAIL: A high-availability and integrity layer for cloud storage. In: Proc. of the 16th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2009. 187−198.

[17] Deswarte Y, Quisquater J, Saidane A. Remote integrity checking. In: Proc. of the 6th Working Conf. on Integrity and Internal Control in Information Systems (IICIS). Berlin, Heidelberg: Springer-Verlag, 2004. 1−11.

[18] Chang EC, Xu J. Remote integrity check with dishonest storage server. In: Proc.of the 13th European Symp. on Research in Computer Security (ESORICS). Berlin, Heidelberg: Springer-Verlag, 2008. 223−237.

[19] Agrawal R, Srikant R. Privacy-Preserving data mining. In: Proc. of the ACM SIGMOD Conf. on Management of Data (SIGMOD). New York: ACM Press, 2000. 439−450.

[20] Warner SL. Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 1965,60(309):63−69.

[21] Fienberg SE, McIntyre J. Data swapping: Variations on a theme by dalenius and reiss. In: Proc. of the Privacy in Statistical Databases. Berlin, Heidelberg: Springer-Verlag, 2004. 14−29.

[22] Evfimievski A, Srikant R, Agrawal R, Gehrke J. Privacy preserving mining of association rules. Information Systems, 2004,29(4): 343−364.

[23] Kantarcioglu M, Clifton C. Privacy-Preserving distributed mining of association rules on horizontally partitioned data. IEEE Trans. on Knowledge and Data Engineering, 2004,16(9):1026−1037.

[24] Vaidya J, Clifton C. Privacy preserving association rule mining in vertically partitioned data. In: Proc. of the 8th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2002. 639−644.

[25] Vaidya J, Clifton C. Privacy-Preserving *k*-means clustering over vertically partitioned data. In: Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2003. 206−215.

[26] Jagannathan G, Wright RN. Privacy-Preserving distributed *k*-means clustering over arbitrarily partitioned data. In: Proc. of the 11th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2005. 593−599.

[27] Sweeney L. *k*-Anonymity: A model for protecting privacy. Int'l Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002,10(5):557−570.

[28] Sweeney L. Achieving *k*-anonymity privacy protection using generalization and suppression. Int'l Journal on Uncertainly, Fuzziness and Knowledge-Based Systems, 2002,10(5):571−588.

[29] LeFevre K, DeWitt DJ, Ramakrishnan R. Incognito: Efficient full-domain *k*-anonymity. In: Proc. of the ACM SIGMOD Conf. on Management of Data (SIGMOD). New York: ACM Press, 2005. 49−60.

[30] Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. *l*-Diversity: Privacy beyond *k*-anonymity. ACM Trans. on Knowledge Discovery from Data, 2007,1(1):1−52.

[31] Li N, Li T, Venkatasubramanian S. *t*-Closeness: Privacy beyond *k*-anonymity and *l*-diversity. In: Proc. of the 23rd IEEE Int'l Conf. on Data Engineering (ICDE). Istanbul: IEEE Computer Society, 2007. 106−115.

[32] Zhu Q, Zhao T, Wang S. Privacy preservation algorithm for service-oriented information search. Chinese Journal of Computers, 2010,33(8):1315−1323 (in Chinese with English abstract).

[33] Fung B, Wang K, Chen R, Yu PS. Privacy-Preserving data publishing: A survey of recent developments. ACM Computing Surveys, 2010,42(4):1−53.

[34] Dwork C. Differential privacy. In: Proc. of the 33rd Int'l Colloquium on Automata, Languages and Programming (ICALP). Berlin, Heidelberg: Springer-Verlag, 2006. 1−12.

[35] Dwork C. Differential privacy: A survey of results. In: Proc. of the 5th Int'l Conf. on Theory and Applications of Models of Computation (TAMC). Berlin, Heidelberg: Springer-Verlag, 2008. 1−19.

[36] Dwork C. The differential privacy frontier. In: Proc. of the 6th Int'l Conf. on Theory of Cryptography Conf. (TCC). Berlin, Heidelberg: Springer-Verlag, 2009. 496−502.

[37] Mironov I, Pandey O, Reingold O, Vadhan S. Computational differential privacy. In: Advances in Cryptology-CRYPTO 2009. Berlin, Heidelberg: Springer-Verlag, 2009. 126−142.

[38] Friedman A, Schuster A. Data mining with differential privacy. In: Proc. of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD). New York: ACM Press, 2010. 493−502.

[39] Paillier P. Public-key cryptosystems based on composite degree residuosity classes. In: Advances in cryptology—EUROCRYPT'99. Berlin, Heidelberg: Springer-Verlag, 1999. 223−238.

[40] ElGamal T. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Trans. on Information Theory, 1985,31(4):469−472.

[41] Gentry C. Fully homomorphic encryption using ideal lattices. In: Proc. of the 41st Annual ACM Symp. on Theory of Computing (STOC). New York: ACM Press, 2009. 169−178.

[42] Ateniese G, Burns R, Curtmola R, Herring J, Kissner L, Peterson Z, Song D. Provable data possession at untrusted stores. In: Proc. of the 14th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2007. 598−609.

[43] Ateniese G, Kamara S, Katz J. Proofs of storage from homomorphic identification protocols. In: Advances in Cryptology—ASIACRYPT 2009. Berlin, Heidelberg: Springer-Verlag, 2009. 319−333.

[44] Hao Z, Zhong S, Yu N. A privacy-preserving remote data integrity checking protocol with data dynamics and public verifiability. IEEE Trans. on Knowledge and Data Engineering, 2011,23(9):1432−1437.

[45] Wang C, Wang Q, Ren, K, Lou W. Privacy-Preserving public auditing for data storage security in cloud computing. In: Proc. of the 29th IEEE INFOCOM. San Diego: IEEE Communications Society, 2010. 1−9.

[46] Xu C, He X, Abraha-Weldemariam D. Cryptanalysis of Wang's auditing protocol for data storage security in cloud computing. In: Proc. of the 2012 Int'l Conf. on Information Computing and Applications (ICICA), Part II. Berlin, Heidelberg: Springer-Verlag, 2012. 422−428.

[47] Boneh D, Lynn B, Shacham H. Short signatures from the Weil pairing. In: Advances in Cryptology—ASIACRYPT 2001. Berlin, Heidelberg: Springer-Verlag, 2001. 514−532.

[48] Hao Z, Yu N. A multiple-replica remote data possession checking protocol with public verifiability. In: Proc. of the 2nd Int'l Symp. on Data, Privacy and E-Commerce. Buffalo: IEEE Computer Society, 2010. 84−89.

[49] Zhu Y, Wang H, Hu Z, Ahn GJ, Hu H, Yau SS. Dynamic audit services for integrity verification of outsourced storages in clouds. In: Proc. of the 2011 ACM Symp. on Applied Computing (SAC). New York: ACM Press, 2011. 1550−1557.

[50] Zhu Y, Hu H, Ahn GJ, Yu M. Cooperative provable data possession for integrity verification in multi-cloud storage. IEEE Trans. on Parallel and Distributed Systems, 2012,23(12):2231−2244.

[51] Yang K, Jia X. An efficient and secure dynamic auditing protocol for data storage in cloud computing. IEEE Trans. on Parallel and Distributed Systems, 2013,24(9):1717−1726.

[52] Shacham H, Waters B. Compact proofs of retrievability. In: Advances in Cryptology—ASIACRYPT 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 90−107.

[53] Bowers KD, Juels A, Oprea A. Proofs of retrievability: Theory and implementation. In: Proc. of the 2009 ACM Workshop on Cloud Computing Security. New York: ACM Press, 2009. 43−54.

[54] Dodis Y, Vadhan S, Wichs D. Proofs of retrievability via hardness amplification. In: Proc. of the 6th Theory of Cryptography Conf. (TCC). Berlin, Heidelberg: Springer-Verlag, 2009. 109−127.

[55] Wang Q, Wang C, Li J, Ren K, Lou W. Enabling public verifiability and data dynamics for storage security in cloud computing. In: Proc. of the 14th European Sym. on Research in Computer Security (ESORICS). Berlin, Heidelberg: Springer-Verlag, 2009. 355−370.

[56] Wang Q, Wang C, Ren K, Lou W, Li J. Enabling public verifiability and data dynamics for storage security in cloud computing. IEEE Trans. on Parallel and Distributed Systems, 2011,22(5):847−859.

[57] Boneh D, Gentry C, Lynn B, Shacham H. Aggregate and verifiably encrypted signatures from bilinear maps. In: Advances in Cryptology—EUROCRYPT 2003. Berlin, Heidelberg: Springer-Verlag, 2003. 416−432.

[58] Song DX, Wagner D, Perrig A. Practical techniques for searches on encrypted data. In: Proc. of the IEEE Symp. on Security and Privacy (S&P). Berkeley, California: IEEE Computer Society, 2000. 44−55.

[59] Goh EJ. Secure Indexes. IACR Cryptology ePrint Archive, 2003. http://eprint.iacr.org/2003/216

[60] Chang YC, Mitzenmacher M. Privacy preserving keyword searches on remote encrypted data. In: Proc. of the 3rd Int'l Conf. on Applied Cryptography and Network Security (ACNS). Berlin, Heidelberg: Springer-Verlag, 2005. 442−455.

[61] Bloom B. Space/time trade-offs in hash coding with allowable errors. Communications of the ACM, 1970,13(7):422−426.

[62] Curtmola R, Garay J, Kamara S, Ostrovsky R. Searchable symmetric encryption: Improved definitions and efficient constructions. In: Proc. of the 13th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2006. 79−88.

[63] Van Liesdonk P, Sedghi S, Doumen J, Hartel P, Jonker W. Computationally efficient searchable symmetric encryption. In: Proc. of the Int'l Workshop on Secure Data Management (SDM). Berlin, Heidelberg: Springer-Verlag, 2010. 87−100.

[64] Kurosawa K, Ohtaki Y. UC-Secure searchable symmetric encryption. In: Proc. of the 16th Int'l Conf. on Financial Cryptography and Data Security (FC). Berlin, Heidelberg: Springer-Verlag, 2012. 285−298.

[65] Canetti R. Universally composable security: A new paradigm for cryptographic protocols. In: Proc. of the 42nd IEEE Sym. on Foundations of Computer Science (FOCS). Las Vegas: IEEE Computer Society, 2001. 136−145.

[66] Kamara S, Papamanthou C, Roeder T. Dynamic searchable symmetric encryption. In: Proc. of the 19th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 2012. 965−976.

[67] Kamara S, Papamanthou C. Parallel and dynamic searchable symmetric encryption. Proc. of the 17th Int'l Conf. on Financial Cryptography and Data Security (FC). Berlin, Heidelberg: Springer-Verlag, 2013. 258−274.

[68] Chase M, Kamara S. Structured encryption and controlled disclosure. In: Advances in Cryptology—ASIACRYPT 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 577−594.

[69] Golle P, Staddon J, Waters B. Secure conjunctive keyword search over encrypted data. In: Proc. of the Int'l Conf. on Applied Cryptography and Network Security (ACNS). Berlin, Heidelberg: Springer-Verlag, 2004. 31−45.

[70] Ballard L, Kamara S, Monrose F. Achieving efficient conjunctive keyword searches over encrypted data. In: Proc. of the 7th Int'l Conf. on Information and Communications Security (ICICS). Berlin, Heidelberg: Springer-Verlag, 2005. 414−426.

[71] Agrawal R, Kiernan J, Srikant R, Xu Y. Order-Preserving encryption for numeric data. In: Proc. of the ACM SIGMOD Conf. on Management of Data (SIGMOD). New York: ACM Press, 2004. 563−574.

[72] Boldyreva A, Chenette N, Lee Y, O'neill A. Order-Preserving symmetric encryption. In: Advances in Cryptology—EUROCRYPT 2009. Berlin, Heidelberg: Springer-Verlag, 2009. 224–241.

[73] Wang C, Cao N, Li J, Ren K, Lou W. Secure ranked keyword search over encrypted cloud data. In: Proc. of the 30th IEEE Int'l Conf. on Distributed Computing Systems (ICDCS). Genova: IEEE Computer Society, 2010. 253–262.

[74] Tang Q. Privacy preserving mapping schemes supporting comparison. In: Proc. of the 2010 ACM Workshop on Cloud Computing Security. New York: ACM Press, 2010. 53–58.

[75] Boldyreva A, Chenette N, O'Neill A. Order-Preserving encryption revisited: Improved security analysis and alternative solutions. In: Advances in Cryptology—CRYPTO 2011. Berlin, Heidelberg: Springer-Verlag, 2011. 578–595.

[76] Popa RA, Li FH, Zeldovich N. An ideal-security protocol for order-preserving encoding. In: Proc. of the 2013 IEEE Symp. on Security and Privacy (S&P). Berkeley: IEEE Computer Society, 2013. 463–477.

[77] Boneh D, Di Crescenzo G, Ostrovsky R, Persiano G. Public key encryption with keyword search. In: Advances in Cryptology—Eurocrypt 2004. Berlin, Heidelberg: Springer-Verlag, 2004. 506–522.

[78] Abdalla M, Bellare M, Catalano D, Kiltz E, Kohno T, Lange T, Malone-Lee J, Neven G, Paillier P, Shi H. Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions. In: Advances in Cryptology—CRYPTO 2005. Berlin, Heidelberg: Springer-Verlag, 2005. 205–222.

[79] Baek J, Safavi-Naini R, Susilo W. Public key encryption with keyword search revisited. In: Proc. of the Int'l Conf. on Computational Science and Its Applications (ICCSA). Berlin, Heidelberg: Springer-Verlag, 2008. 1249–1259.

[80] Boneh D, Gentry C, Lynn B, Shacham H. Aggregate and verifiably encrypted signatures from bilinear maps. In: Advances in Cryptology—EUROCRYPT 2003. Berlin, Heidelberg: Springer-Verlag, 2003. 416–432.

[81] Rhee HS, Park JH, Susilo W, Lee DH. Improved searchable public key encryption with designated tester. In: Proc. of the 4th ACM Int'l Symp. on Information, Computer, and Communications Security (ASIACCS). New York: ACM Press, 2009. 376–379.

[82] Fang L, Susilo W, Ge C, Wang J. A secure channel free public key encryption with keyword search scheme without random oracle. In: Proc. of the Int'l Conf. Cryptology and Network Security (CANS). Berlin, Heidelberg: Springer-Verlag, 2009. 248–258.

[83] Gentry C. Practical identity-based encryption without random oracles. In: Advances in Cryptology—EUROCRYPT 2006. Berlin, Heidelberg: Springer-Verlag, 2006. 445–464.

[84] Di Crescenzo G, Saraswat V. Public key encryption with searchable keywords based on Jacobi symbols. In: Progress in Cryptology —INDOCRYPT 2007. Berlin, Heidelberg: Springer-Verlag, 2007. 282–296.

[85] Cocks C. An identity based encryption scheme based on quadratic residues. In: Proc. of the 8th IMA Int'l Conf. on Cryptography and Coding (IMACC). Berlin, Heidelberg: Springer-Verlag, 2001. 360–363.

[86] Bellare M, Boldyreva A, O'Neill A. Deterministic and efficiently searchable encryption. In: Advances in Cryptology—CRYPTO 2007. Berlin, Heidelberg: Springer-Verlag, 2007. 535–552.

[87] Bellare M, Fischlin M, O'Neill A, Ristenpart T. Deterministic encryption: Definitional equivalences and constructions without random oracles. In: Advances in Cryptology—RYPTO 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 360–378.

[88] Boldyreva A, Fehr S, O'Neill A. On notions of security for deterministic encryption, and efficient constructions without random oracles. In: Advances in Cryptology—CRYPTO 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 335–359.

[89] Peikert C, Waters B. Lossy trapdoor functions and their applications. In: Proc. of the 40th Annual ACM Symp. on Theory of Computing (STOC). New York: ACM Press, 2008. 187–196.

[90] Fuller B, O'Neill A, Reyzin L. A unified approach to deterministic encryption: New constructions and a connection to computational entropy. In: Proc. of the 9th Theory of Cryptography Conf (TCC). Berlin, Heidelberg: Springer-Verlag, 2012. 582–599.

[91] Brakerski Z, Segev G. Better security for deterministic public-key encryption: The auxiliary-input setting. In: Advances in Cryptology—CRYPTO 2011. Berlin, Heidelberg: Springer-Verlag, 2011. 543–560.

[92] Wee H. Dual projective hashing and its applications--lossy trapdoor functions and more. In: Advances in Cryptology—EUROCRYPT 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 246–262.

[93] Regev O. On lattices, learning with errors, random linear codes, and cryptography. Journal of the ACM, 2009,56(6):1–40.

[94] Xie X, Xue R, Zhang R. Deterministic public key encryption and identity-based encryption from lattices in the auxiliary-input setting. In: Proc. of the 8th Int'l Conf. on Security and Cryptography for Networks (SCN). Berlin, Heidelberg: Springer-Verlag, 2012. 1–18.

[95] Mironov I, Pandey O, Reingold O, Segev G. Incremental deterministic public-key encryption. In: Advances in Cryptology—EUROCRYPT 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 628–644.

[96] Cui Y, Morozov K, Kobara K, Imai H. Efficient constructions of deterministic encryption from hybrid encryption and code-based PKE. In: Proc. of the 18th Int'l Symp. on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (AAECC). Berlin, Heidelberg: Springer-Verlag, 2009. 159–168.

[97] Boneh D, Waters B. Conjunctive, subset, and range queries on encrypted data. In: Proc. of the 4th Theory of Cryptography Conf. (TCC). Berlin, Heidelberg: Springer-Verlag, 2007. 535–554.

[98] Hwang YH, Lee PJ. Public key encryption with conjunctive keyword search and its extension to a multi-user system. In: Proc. of the Int'l Conf. on Pairing-Based Cryptography (Pairing). Berlin, Heidelberg: Springer-Verlag, 2007. 2–22.

[99] Katz J, Sahai A, Waters B. Predicate encryption supporting disjunctions, polynomial equations, and inner products. In: Advances in Cryptology—EUROCRYPT 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 146–162.

[100] Yao ACC. Protocols for secure computations. In: Proc. of the Annual IEEE Symp. on Foundations of Computer Science (FOCS). Chicago: IEEE Computer Society, 1982. 160–164.

[101] Lopez-Alt A, Tromer E, Vaikuntanathan V. On-the-Fly multiparty computation on the cloud via multikey fully homomorphic encryption. In: Proc. of the 44th Annual ACM Symp. on Theory of Computing (STOC). New York: ACM Press, 2012. 1219–1234.

[102] Damgard I, Pastro V, Smart N, Zakarias S. Multiparty computation from somewhat homomorphic encryption. In: Advances in Cryptology—CRYPTO 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 643–662.

[103] Naccache D, Stern J. A new public key cryptosystem based on higher residues. In: Proc. of the 5th ACM Conf. on Computer and Communications Security (CCS). New York: ACM Press, 1998. 59–66.

[104] Okamoto T, Uchiyama S. A new public-key cryptosystem as secure as factoring. In: Advances in Cryptology—EUROCRYPT'98. Berlin, Heidelberg: Springer-Verlag, 1998. 308–318.

[105] Boneh D, Goh EJ, Nissim K. Evaluating 2-DNF formulas on ciphertexts. In: Proc. of the 2nd Theory of Cryptography Conf. (TCC). Berlin, Heidelberg: Springer-Verlag, 2005. 325–341.

[106] Shor P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal on Computing, 1997,26(5):1484–1509.

[107] Gentry C, Halevi S. Implementing Gentry's fully-homomorphic encryption scheme. In: Advances in Cryptology—EUROCRYPT 2011. Berlin, Heidelberg: Springer-Verlag, 2011. 129–148.

[108] Scholl P, Smart NP. Improved key generation for Gentry's fully homomorphic encryption scheme. In: Proc. of the 13th IMA Int'l Conf. on Cryptography and Coding (IMACC). Berlin, Heidelberg: Springer-Verlag, 2011. 10–22.

[109] Stehle D, Steinfeld R. Faster fully homomorphic encryption. In: Advances in Cryptology—ASIACRYPT 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 377–394.

[110] Smart NP, Vercauteren F. Fully homomorphic encryption with relatively small key and ciphertext sizes. In: Proc. of the Public Key Cryptography (PKC). Berlin, Heidelberg: Springer-Verlag, 2010. 420-443.

[111] Gentry C, Halevi S, Smart NP. Fully homomorphic encryption with polylog overhead. In: Advances in Cryptology—EUROCRYPT 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 465–482.

[112] Lyubashevsky V, Peikert C, Regev O. On ideal lattices and learning with errors over rings. In: Advances in Cryptology—EUROCRYPT 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 1–23.

[113] Gentry C, Halevi S, Smart NP. Better bootstrapping in fully homomorphic encryption. In: Proc. of the Public Key Cryptography (PKC). Berlin, Heidelberg: Springer-Verlag, 2012. 1–16.

[114] Brakerski Z, Gentry C, Halevi S. Packed ciphertexts in LWE-based homomorphic encryption. In: Proc. of the Public-Key Cryptography (PKC). Berlin, Heidelberg: Springer-Verlag, 2013. 1–13.

[115] Peikert C, Vaikuntanathan V, Waters B. A framework for efficient and composable oblivious transfer. In: Advances in Cryptology—CRYPTO 2008. Berlin, Heidelberg: Springer-Verlag, 2008. 554–571.

[116] Gentry C. Toward basing fully homomorphic encryption on worst-case hardness. In: Advances in Cryptolog—-CRYPTO 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 116−137.

[117] Brakerski Z, Vaikuntanathan V. Efficient fully homomorphic encryption from (standard) LWE. In: Proc. of the 52nd IEEE Annual Symp. on Foundations of Computer Science (FOCS). Palm Springs, California: IEEE Computer Society, 2011. 97−106.

[118] Brakerski Z, Vaikuntanathan V. Fully homomorphic encryption from ring-LWE and security for key dependent messages. In: Advances in Cryptology—CRYPTO 2011. Berlin, Heidelberg: Springer-Verlag, 2011. 505−524.

[119] Brakerski Z, Gentry C, Vaikuntanathan V. (Leveled) fully homomorphic encryption without bootstrapping. In: Proc. of the 3rd Innovations in Theoretical Computer Science Conf. New York: ACM Press, 2012. 309−325.

[120] Brakerski Z. Fully homomorphic encryption without modulus switching from classical GapSVP. In: Advances in Cryptology—CRYPTO 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 868−886.

[121] Bos JW, Lauter K, Loftus J, Naehrig M. Improved security for a ring-based fully homomorphic encryption scheme. In: Proc. of the 14th IMA Int'l Conf. on Cryptography and Coding (IMACC). Berlin, Heidelberg: Springer-Verlag, 2013. 45−64.

[122] Van Dijk M, Gentry C, Halevi S, Vaikuntanathan V. Fully homomorphic encryption over the integers. In: Advances in Cryptology—EUROCRYPT 2010. Berlin, Heidelberg: Springer-Verlag, 2010. 24−43.

[123] Coron J, Mandal A, Naccache D, Tibouchi M. Fully homomorphic encryption over the integers with shorter public keys. In: Advances in Cryptology—CRYPTO 2011. Berlin, Heidelberg: Springer-Verlag, 2011. 487−504.

[124] Chen Y, Nguyen PQ. Faster algorithms for approximate common divisors: Breaking fully-homomorphic-encryption challenges over the integers. In: Advances in Cryptology—EUROCRYPT 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 502−519.

[125] Coron J, Naccache D, Tibouchi M. Public key compression and modulus switching for fully homomorphic encryption over the integers. In: Advances in Cryptology—EUROCRYPT 2012. Berlin, Heidelberg: Springer-Verlag, 2012. 446−464.

[126] Cheon JH, Coron J, Kim J, Lee MS, Lepoint T, Tibouchi M, Yun A. Batch fully homomorphic encryption over the integers. In: Advances in Cryptology—EUROCRYPT 2013. Berlin, Heidelberg: Springer-Verlag, 2013. 315−335.

附中文参考文献:

[4] 李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考.中国科学院院刊,2012,27(6):647−657.

[5] 周水庚,李丰,陶宇飞,肖小奎.面向数据库应用的隐私保护研究综述.计算机学报,2009,32(5):847−861.

[7] 冯登国,张敏,张妍,徐震.云计算安全研究.软件学报,2011,22(1):71−83.

[32] 朱青,赵桐,王珊.面向搜索服务的数据隐私保护算法.计算机学报,2010,33(8):1315−1323.

黄刘生(1957－),男,安徽太湖人,教授,博士生导师,主要研究领域为信息安全,无线传感网络,大数据.

田苗苗(1987－),男,博士,主要研究领域为密码学,大数据安全,信息安全.

黄河(1983－),男,博士,副教授,主要研究领域为无线频谱资源分配,隐私保护.