•科学前沿•

DOI:10.15961/j.jsuese.201700726

大数据安全保护技术

陈兴蜀[1], 杨　露[2], 罗永刚[1*]

(1.四川大学 网络空间安全研究院, 四川 成都 610065; 2.四川大学 计算机学院, 四川 成都 610065)

摘　要:大数据技术的发展和应用对国家的治理模式、企业的决策架构、商业的业务策略以及个人的生活方式都产生了深远影响。但是,大量数据的汇集不仅加大了用户隐私泄露的风险,而且大数据中包含的巨大信息和潜在价值吸引了更多的潜在攻击者。此外,大数据的应用是跨学科领域集成的应用,引入了很多新的技术,可能面临更多更高的风险。作者回顾了大数据的定义和特征,提出大数据架构和大数据安全体系,在此基础上分析大数据安全在法律法规、标准、数据生命周期保护和大数据平台4个方面的研究进展。梳理美国、欧盟、中国等在大数据安全方面的法律法规现状和国际标准化组织、美国、中国等大数据安全标准化研究现状。大数据在生命周期过程中需要大数据平台为其提供支撑,以实现大数据的收集、传输、存储和分析等功能。从大数据生命周期和大数据平台两个维度分析大数据面临的安全问题和关键技术研究现状。生命周期包括收集、存储、使用、分发和删除5个阶段。收集阶段的数据质量决定了数据价值,提升数据质量的技术手段主要有数据与模型不一致性的检测、数据清洗两类。大数据分发将处理后的大数据传递给外部实体,隐私保护或敏感信息保护至关重要,相关的关键技术有数据匿名化、支持隐私保护的数据检索和分析等。大数据的管理主要包含元数据管理、数据血缘管理等方面,可以为有效使用大数据和确保大数据安全提供支持。大数据平台安全主要解决大数据组件之间的身份认证、数据隔离、数据加密存储、大数据平台边界保护和审计,主要的关键技术有身份认证、访问控制、数据加密和审计等。目前,在国际上仍缺乏完善的大数据安全标准体系,在隐私保护、数据共享和数据跨境传输等方面缺乏标准的规范和指导。大数据分析技术仍处于快速发展阶段,很难预测今后的大数据关联分析对隐私保护和敏感信息保护带来的问题,因此,现有的数据脱敏技术和隐私保护技术有待进一步研究。数据同态加密实现了分析数据时不暴露数据隐私和敏感信息,现有的同态加密算法还远未成熟。现有的大数据平台的身份认证、数据加密、访问控制仍采用的传统技术,不能适应大数据面临的数据规模大、处理逻辑复杂、用户量大等新环境。一些大数据安全关键技术在性能和可用性方面还值得深入研究,以期可早日投入实际应用。另外,使用大数据处理技术研发安全态势感知、网络安全入侵检测、威胁情报分析等安全应用,利用大数据技术抵御针对大数据的攻击威胁也已成为大数据安全领域新的研究热趋势。大数据安全的发展需要法律法规、标准和关键技术的共同支撑和推动。

关键词:大数据; 安全; 身份认证; 访问控制; 隐私保护

中图分类号:TP391.4　　**文献标志码**:A　　**文章编号**:2096-3246(2017)05-0001-12

Big Data Security Technology

CHEN Xingshu[1], *YANG Lu*[2], *LUO Yonggang*[1*]

(1.Cyber Security Research Inst., Sichuan Univ., Chengdu 610065, China; 2.College of Computer Sci., Sichuan Univ., Chengdu 610065, China)

Abstract: The development and application of big data technology has a deep influence on the national governance model,corporate decision-making architecture,business strategy and personal lifestyle.The data aggretation not only increases the risk of user privacy leaks,but the huge information and potential value contained in big data also attract more potential attackers.Moreover,the big data application is a cross-disciplinary application,which introduces not only a lot of new technologies but more and higher risks.The defini-

收稿日期:2017 – 06 – 26

基金项目:国家自然科学基金资助项目(61272447)

作者简介:陈兴蜀(1968—), 女, 博士, 教授, 博士生导师. 研究方向: 云计算及大数据安全; 可信计算.E-mail: chenxsh@scu.edu.cn

* 通信联系人 E-mail: iamlyg98@163.com

网络出版时间:2017 – 09 – 14 11 : 27 : 14　　**网络出版地址**: http://kns.cnki.net/kcms/detail/51.1773.TB.20170914.1127.001.html

tion and characteristics of big data is reviewed,and the big data architecture and big data security system are put forward in this paper.Based on this system,the security challenges facing the current big data and research progress of big data security technologies are analyzed from four perspectives:laws and regulations,standards,data life cycle protection and big data platform key technology.Laws and regulations in America,European Union,China and the research status of big data security standarlization of International Organization for Standardization,America,China and so on was introduced.Big data platform is needed to realize the collection,transmission,storage and analysis and so on in big data lifecycle.In this paper,the security problems and key technologies of big data are analyzed from two dimensions of big data lifecycle and big data platform.The lifecycle includes collection,storage,usage,distribution and deletion five phases.Data value is determined by the data quality of the collection phase.Data and model inconsistency detection and data cleaning are the main technical means to improve data quality.The processed big data is transmit to external entities in big data distribution phase,so the protection of privacy and sensitive information is essential.The retalted key technologies are data anonymity,privacy-protecting data retrieval and analysis.The big data management support the effective use of big data and ensure big data security,which mainly contains metadata management and data lineage.The problems of authentication,data isolation,data encryption storage,big data platform border protection and audit between big data components can be solved by the big data platform security with the key technologies such as authentication,access control,data encryption and audit.At present,a perfect big data security standard system is still lacking in the world.The norms and guidance for privacy protection,data sharing,cross-border data transmission from standards are urgent needed.With the rapid development of big data analysis technology,it's difficult to predict the challenge of privacy protection and sensitive information protection from big data association analysis in the future.The existing data masking and privacy protection technology will face a great challenge.The data analysis without exposure to data privacy and sensitive information can be achieved by data homomorphic encryption,but the existing homomorphic encryption algorithm is far from mature.The current authentication,data encryption and access control in the big data platform use the traditional technology,which can't adapt to the new environment with large scale of data,complex processing logic and huge amount of users.Some of the big data security key technologies are also worthy of indepth study in the performance and availability for early practical application.In addition,using big data pro-cessing technology to develop security applications such as network security situation perception,intrusion detection and network threat intelligence analysis,and using big data technology to resist attacks against big data have become a new research trend in the field of big data security.The development of big data security requires the united support and promotion of laws and regulations,standards and key technologies.

Key words: big data;security;authentication;access control;privacy protection

大数据时代是安全与发展并重、机遇与挑战并存的网络时代[1]。2016年11月，第十二届全国人民代表大会常务委员会通过了《中华人民共和国网络安全法》[2]，于2017年6月1日正式实施，明确个人信息保护义务，支持网络安全技术的研究、开发、应用和推广。2016年12月，国家互联网信息办公室发布《国家网络空间安全战略》[3]，提出“实施国家大数据战略，建立大数据安全管理制度，支持大数据、云计算等新一代信息技术创新和应用”。大数据目前已经成为国家重要战略，安全是大数据发展的重要基石，在充分发挥大数据价值的同时，解决大数据安全面临的问题和挑战也同样重要。

在产业界和学术界，对大数据安全的研究已经成为热点。国际标准化组织、产业联盟、企业和研究机构等都已开展相关研究以解决大数据安全问题。2012年，云安全联盟（CSA）成立了大数据工作组，旨在寻找大数据安全和隐私问题的解决方案。2016年，全国信息安全标准化技术委员会正式成立大数据安全标准特别工作组，负责大数据和云计算相关的安全标准化研制工作。在Engineering Village以“big data”和“security”为关键字检索期刊论文：2013年有54篇，2014年124篇，2015年214篇，2016年达到265篇。大数据安全相关的期刊论文篇数呈逐年增长的趋势，这反映出学术界对大数据安全的研究越来越多。

作者从法律法规、标准、大数据生命周期和大数据平台4个方面阐述了大数据安全发展现状，提出一种分层的大数据安全体系，阐述相关研究现状及一些开放问题。

1 大数据基本概念

1.1 大数据的定义与特征

2011年5月，美国麦肯锡全球研究院发布了《大数据：创新、竞争和生产力的下一个前沿》报告[4]，“大数据”一词被正式提出，自此其成为科研、金融和商业等众多领域的热门话题。麦肯锡提出“大数据是指其大小超出了典型数据库软件的采集、储存、管理和分析等能力的数据集”。大数据的定义可用4Vs特征表示，典型的有两类：1）国际数据公司（International Data Corporation, IDC）的大数据定义：使用种类、速度、体量和价值（variety、velocity、volume、value）定

义大数据。其中：种类（variety）包括结构化、半结构化和非结构化等各种类型的数据；速度（velocity）意味着大数据的采集、处理等环节必须快速及时，以便最大化大数据的价值；体量（volume）表示数据量大；价值（value）指大数据具有很大的社会价值。2）美国国家标准与技术研究院（NIST）的大数据定义：将IDC的4Vs特征中的“value”替换为“variability”，即“变化”这一特征[5]，突出数据随时间发生变化的特点。充分理解大数据的定义和特征，可以更好地理解大数据面临的各种问题。

1.2　大数据架构

从数据和技术两个角度可以将大数据架构划分为两层：大数据生命周期和大数据平台，如图1所示。

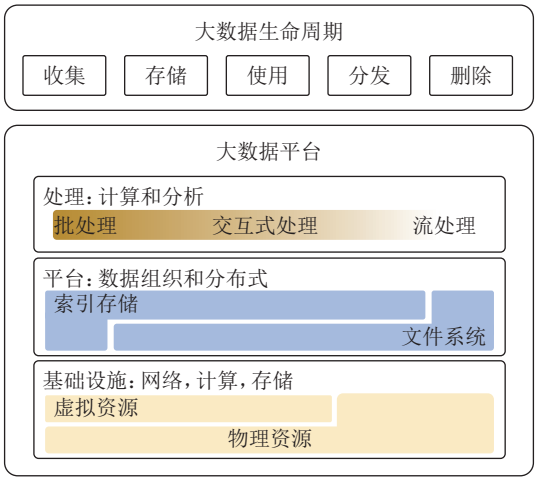

图 1　大数据架构

Fig.1　Big data architecture

1）大数据生命周期

大数据生命周期包含收集、存储、使用、分发和删除各环节。数据通过收集进入大数据平台并进行存储，通过使用发掘其潜在价值，通过分发传递和共享数据或分析结果，最后删除不再需要的数据。因此可以说大数据生命周期是数据转换为价值的过程。

2）大数据平台

大数据平台提供大数据生命周期各环节所需的基础设施、存储和处理平台以及数据分析的算法等，是整个大数据架构中的技术支撑。

2　大数据安全体系

在大数据架构的基础上，作者提出一种分层的大数据安全体系，如图2所示。1）法律、法规及标准层：法律、法规是约束或规制大数据各环节中行为的基础。大数据安全标准是引领和指导大数据安全工作落实的规范。大数据安全相关法律、法规和标准的制定不仅给予数据充分有效的保护，同时也能促进数据的开放、共享，推动大数据应用的发展。2）大数据生命周期层。主要涉及数据保护的相关技术：数据质量、数据生命周期管理、数据权属和隐私保护。3）大数据平台层。主要涉及大数据平台安全保护的相关技术：身份认证、访问控制、数据加密和审计。

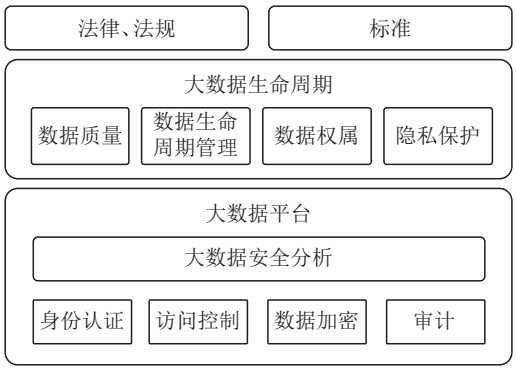

图 2　大数据安全体系

Fig.2　Big data security system

2.1　大数据安全相关法律、法规

随着大数据的安全问题越来越引起人们的重视，包括美国、英国、欧盟和中国在内的很多国家和组织都制定了大数据安全相关的法律法规和政策以推动大数据应用和数据保护。

2009年，美国发布《开放政府的指令》要求政府通过网站发布数据等方式公开政府信息；2012年5月，出台《数字政府：构建一个21世纪平台以更好地服务美国人民》支撑美国电子政府发展[1]。欧盟早在1995年就发布了《保护个人享有的与个人数据处理有关的权利以及个人数据自由流动的指令》（简称《数据保护指令》），为欧盟成员国保护个人数据设立了最低标准。2015年，欧盟通过《通用数据保护条例》（GDPR），该条例在《数据保护指令》的基础上进行了大刀阔斧的改革，对欧盟居民的个人信息提出更严的保护标准和更高的保护水平。巴西、韩国和日本等国家也都发布了《个人信息保护法》，对个人信息保护提出明确要求。

中国高度重视大数据安全问题，近几年发布了一系列大数据安全相关的法律法规和政策。2013年7月，工业和信息化部公布了《电信和互联网用户个人信息保护规定》[6]，明确电信业务经营者、互联网信息服务提供者收集、使用用户个人信息的规则和信息安全保障措施要求。2015年8月，国务院印发了《促进大数据发展行动纲要》[7]，提出要健全大数据安全保障体系，完善法律法规制度和标准体系。2016年3月，第十二届全国全国人民代表大会第四次会议表决通过了《中华人民共和国国民经济和社会发展第十三个五年规划纲要》提出把大数据作为基础性战略资源，明确指出要建立大数据安全管理制度，实行数据资源分类分级管理，保障安全高效可信

应用。2016年11月，全国人民代表大会常务委员会发布了《中华人民共和国网络安全法》（以下简称《网络安全法》），于2017年6月1日正式实施。《网络安全法》中明确规定了网络运营者对个人信息的保护要求，要求其采取数据分类、重要数据备份和加密等措施。网络安全法是中国第一部全面规范网络空间安全管理方面问题的基础性法律，是中国网络空间法治建设的重要里程碑[8]。

2.2 大数据安全标准

大数据安全与隐私保护已成为国际标准化的热点和焦点，目前有多个标准化组织都正在开展大数据和大数据安全相关的标准化工作。

ISO/IEC JTC1 WG9是ISO/IEC JTC1于2014年11月成立的大数据工作组，目前正在开展《信息技术 大数据参考架构》（ISO/IEC 20547）国际标准的编制。ISO/IEC 20547为多部分标准，包括《信息技术 大数据参考架构 第1部分：框架和应用过程》（ISO/IEC TR 20547–1）、《信息技术 大数据参考架构 第2部分：用例和衍生需求》（ISO/IECTR 20547–2）、《信息技术 大数据参考架构 第3部分：参考架构》（ISO/IEC20547–3）、《信息技术 大数据参考架构 第4部分：安全与隐私保护》（ISO/IEC 20547–4）、《信息技术 大数据参考架构 第5部分：标准路线图》（ISO/IEC TR 20547–5）。其中，《信息技术 大数据参考架构 第4部分：安全与隐私保护》（ISO/IEC 20547–4）标准编制项目转交给了ISO/IEC JTC1 SC27下属的WG4和WG5共同负责，并任命中国专家担任项目的编辑。

NIST于2012年6月启动了大数据相关基本概念、技术和标准需求的研究，2013年5月成立了NIST大数据公开工作组（NBG-PWG），2015年9月发布了《NIST大数据互操作性框架》（NIST SP 1500）系列标准（第1版），包括7个分卷，即：《第1卷：定义》（NIST SP 1500–1）、《第2卷：大数据分类法》（NIST SP 1500–2）、《第3卷：用例和一般要求》（NIST SP 1500–3）、《第4卷：安全和隐 私保护》（NIST SP 1500–4）、《第5卷：架构调研白皮书》（NIST SP 1500–5）、《第6卷：参考架构》（NIST SP 1500–6）和《第7卷：标准路线图》（NIST SP 1500–7）。目前，在研的该标准第2版中加入了《第8卷：参考框架接口》（NIST SP 1500–8）和《第9卷：采纳和现代化》（NIST SP 1500–9）。

全国信息安全标准化技术委员会在2016年 4月成立了大数据安全标准特别工作组（以下简称“特别工作组”，），主要负责制定和完善中国大数据安全领域标准体系。目前，正在制定《信息安全技术 个人信息安全规范》、《信息安全技术 大数据服务安全能力要求》、《信息安全技术 大数据安全管理指南》等国家标准。其中，一些标准已进入报批或公开征求意见阶段，将为中国大数据安全的管理、技术和应用提供重要支撑。

2.3 大数据生命周期安全

大数据生命周期的安全以数据为中心，重点考虑大数据生命周期各环节中的数据安全问题。

数据质量对分析结果及基于数据的决策的准确性具有重要影响。大数据来源丰富，数据质量良莠不齐，因此在数据收集环节应关注数据质量问题。大数据生命周期包含众多环节，涉及数据规模大，导致数据非授权访问的风险增加，因此大数据生命周期管理是保障数据安全的重要措施。数据的交换与共享有助于发掘其潜藏的价值，数据权属不清是妨碍数据交换与共享的重要因素，数据权属是保障数据交易合法性、规范大数据应用秩序等的先决条件。大数据最具挑战性和最重要的问题是隐私保护。通过大数据关联分析可能发现隐私信息，且大数据关联分析技术还处于不断发展过程当中，因此针对大数据的隐私保护是一个研究难题。

2.4 大数据平台安全

传统的数据处理手段无法满足大数据应用对海量数据进行高速处理的需求，因此涌现出了很多新的技术，如分布式存储和处理架构、非关系型数据库等。处理模式和应用场景的改变给传统安全保护技术带来巨大挑战。

在大数据场景中，访问控制的安全问题主要体现在：1）用户数量庞大，增加了对主体描述的难度；2）数据的结构种类繁多，如半结构化数据、非结构化数据，增加了客体描述的困难；3）大数据分析应用种类很多，访问需求复杂且动态变化，如何准确按需制定访问控制策略也是一大难题。因此大数据环境下的访问控制需要更灵活的主客体描述方式及更细粒度的访问控制策略。加密是数据保护的重要措施，在大数据环境下，加密技术存在以下问题：1）对海量数据进行加密对加密算法的性能提出极大挑战；2）大规模的秘钥管理极其困难；3）直接对加密数据进行处理的技术还处于发展阶段。

3 大数据的数据保护关键技术

3.1 数据质量

数据质量直接影响大数据分析的结果，影响基于数据的决策结果，因此针对大数据的数据质量问题一直备受学术界和产业界热切关注。数据质量的一个核心技术问题是识别数据中的错误，即数据的不一致性检测。Fan等[9]在2010年提出一种分布式数

据中违反条件函数依赖(CDF)的检测方法; 2014年[10], Fan等做出了改进, 提出增量的分布式数据的CDF违反检测, 该文献实验证明能有效捕获分布式数据中的错误。分析错误的原因可能是数据不一致, 或模型错误。因此数据质量研究的另一个方向是判断导致数据质量问题的原因是数据不一致还是约束不一致[11–12]。对错误数据的清洗研究目前主要都采用基于主数据、编辑规则[13–15]的方式, 通过与主数据每个属性上的定义域进行比较, 挖掘编辑规则, 然后根据编辑规则进行修复。

另外, 产业界也开发出众多ETL(extraction-transformation-loading)工具解决实际生产中的数据质量问题。Hadoop、Hive等都可用作数据清洗工作, 除此之外, 还有专门的数据清洗工具, 如: Sqoop是一种数据导入工具; Informatica是具有图形界面的数据集成工具; Kattle是一种开源的ETL工具, 支持多种数据源及不同数据源之间的连接; DataX是阿里巴巴公司开发的用于异构数据源离线同步的开源项目, 除了提供数据快照搬迁功能, 还提供了丰富的数据转换功能。

3.2 数据管理

目前, 在Hadoop开源社区中出现了两个数据生命周期管理的开源项目: Apache Atlas和Apache Falcon。Apache Atlas通过定义元数据对象的模型表示Hadoop和外部组件的元数据对象, 并进行分类, 帮助Hadoop栈内外的工具间进行元数据交换; Apache Atlas与Apache Ranger项目结合提供安全策略, 并可记录授权、数据访问、拒绝访问等事件, 支持对这些事件的索引功能; Apache Atlas提供了可视化数据血缘关系的能力, 提供了在许多分析引擎(如Storm、Kafka和Hive)上移动数据的完整视图。Apache Falcon定义了数据采集、处理和导出的数据管道, 使用管道自动生成Oozie工作流; 其本质是将数据和处理过程的配置信息转化为业务处理流程; Apache Falcon提供可视化的数据管道系统, 可跟踪数据管道的审计日志, 查看数据管道的血缘关系。Apache Atlas侧重元数据管理, 而Apache Falcon更侧重于数据生命周期管理。

学术界也已对大数据管理开展了相关研究。Kanchi等[16]分析了大数据管理目前面临的问题和挑战, 并给出了相应的解决方案和最佳实践。Siddiqa等[17]提出一种分层的大数据管理处理工作流, 对每一层涉及的大数据管理技术现状进行详细阐述。Zhang等[18]则对大数据管理中的内存数据管理和处理的关键技术进行了深入调研。

3.3 隐私保护

隐私保护的研究大致可以分为两个方向: 基于数据发布的隐私保护研究和针对数据挖掘的隐私保护研究。

3.3.1 数据发布

Sweeney等[19]最早提出了κ-匿名规则以处理数据发布中的隐私泄漏问题。但κ-匿名模型并未考虑敏感属性的约束, 因此仍可能遭受同质性和背景知识等攻击。为解决该问题, Machanavajjhala等[20]提出了ℓ-多样性模型, 其核心思想是要求匿名后的每个准标识符分组中不同敏感属性值的个数要不少于l个。Lin等[20]指出ℓ-多样性模型不能抵御近似攻击和偏度攻击, 故提出t-接近模型, 要求等价类中敏感属性值的分布与敏感属性值在匿名化表中总体分布的差异不超过t, 还引入一种陆地移动距离(earth mover distance, EMD)函数来度量两个分布的距离。为了使匿名模型能抵御多种攻击, 提高匿名效率, 各种个性化的匿名模型被提出。

Wong等[22]提出(α, κ) -anonymity模型, 该模型在κ-anonymity模型的基础上还要求统一等价类中任何一个敏感属性值出现的概率不大于α($0<\alpha<1$)。Zhang等[23]提出(κ, e)-anonymity模型, 要求等价类中敏感属性取值范围不能小于给定的阈值e, 即等价类中敏感属性值的最大值与最小值之差至少是e。除了(α, κ)-匿名模型和(κ, e)-匿名模型外, 还有(κ, ℓ)-匿名模型[24]、基于时间序列的多样模式的(κ, P)-匿名模型[25]、基于熵分类的κ-匿名隐私保护算法[26]等。

3.3.2 数据挖掘

数据通过关联分析可能会产生个人敏感信息, 为解决该问题, 数据挖掘的隐私保护研究成为隐私保护领域的热点之一。

数据扰动技术在数据挖掘隐私保护研究中得到广泛应用。数据扰动的思想是: 对数据进行变换, 使其中敏感信息被隐藏, 只呈现出数据的统计学特征。Oliveira等[27]提出使用平移、缩放和旋转的数据变换方式, 该算法能保证变换前后数据之间的相对距离保持不变, 对聚类有较好的可用性, 但隐私保护效果不够理想。张鹏等[28]将数据干扰和查询限制相结合, 提出一种新的基于部分隐藏随机化回答(randomized response with partial hiding, RRPH)的隐私保护关联规则挖掘方法。许焕霞等[29]提出一种用户分类的隐私保护数据挖掘算法, 采用随机正交变换方法对原始数据进行处理。刘峰等[30]提出一种基于安全多方计算与随机干扰相结合的混合算法, 解决半诚实模型下水平分布数据的隐私数据保护问题。

4 大数据平台安全关键技术

4.1 身份认证

单点登录是解决复杂的云计算环境中统一身份认

证和管理的一种方案，Apache Knox Gateway就是一种单点登录方案。Celesti等[31]提出了一种3阶段的跨云联合模型，使用安全断言标记语言技术解决跨云间的单点登录问题。Powell等[32]提出了一种云间单点登录认证的基础架构，使用了分布式账户链接技术、Shibboleth管理技术[33]和代理证书库。Méndez等[34]通过扩展GSS-EAP机制提出了一种高效的组织内或联盟内云间单点登录身份认证的方案，减少了访问云服务的时间并节省了认证、授权和审计的基础设施。

使用用户ID和密码的传统验证方式不足以抵御云计算环境中复杂的攻击方式，多因子认证在传统标准安全凭证的基础上附加使用多种安全凭证，进一步加强认证的安全性。Lee等[35]使用两种不同的通信渠道，提出一种云计算中的双因子认证框架，使用基于PKI的Web认证和基于手机的带外（out of band，OOB）认证。Choudhury等[36]提出一种基于密码、智能卡与带外认证的双因子的云计算认证方案，可抵抗重放攻击、中间人攻击和拒绝服务攻击等。Banyal等[37]提出了一种用于云计算环境的多因子认证框架，该框架结合了传统的用户ID和密码的认证与基于动态多因子秘密分割的认证方法。Liu等[38]提出一种名为MACA的隐私保护的多因子身份认证系统，第1个认证因子是用户密码，第2个认证因子是汇总了用户行为的混合用户配置文件；其用户行为包括基于主机的特性和基于网络流的特性；为了保护用户行为的隐私安全，使用了完全同态加密和模糊散列技术确保用户行为数据不泄漏。

4.2 访问控制

目前，适用于大数据细粒度访问控制需求的方案主要有两种：基于属性加密的访问控制和基于角色的访问控制。

4.2.1 基于属性加密的访问控制

基于属性加密的访问控制是一种利用密文机制实现客体访问控制的方法，主要可以分为两种：基于密钥策略的属性加密（KP-ABE）和基于密文策略的属性加密（CP-ABE）。在KP-ABE中[39]，引入了访问结构，密文与属性集合相关联，密钥与访问策略关联，只有当用户提供的属性集可以达到密钥的访问结构时才能解密文件，KP-ABE主要用于访问静态数据。在CP-ABE中，密文由访问结构生成，密钥是用户的属性集合，只有当用户的属性满足密文中的访问结构时才能解密该段密文。CP-ABE使得数据拥有者可以灵活地控制哪些用户访问数据，因此也被广泛地用作云计算的访问控制方案[40]。

2011年，Waters[41]完善了CP-ABE的策略表示方式，使其支持LSSS描述方式，并在标准模型下证明该方案是选择性安全的。在通常的CP-ABE方案中，密文和密钥长度都与属性的个数线性相关，这增加了计算开销。为解决此问题，Emura等[42]实现了第一个CP-ABE定长密文方案，但该方案的策略仅限于（n，n）门限的访问策略，即用户私钥中的属性数与密文的访问策略必须相同。Heranz等[43]提出了另一种定长密文的方案，可处理（t，n）门限的访问策略。Yang等[44]提出了一种有效、安全的云计算多权限访问控制方案，并解决了属性撤销的问题。Sreenivasa和Ratna[45]提出了一种多权限分散的CP-ABE机制，利用最小授权集加密数据，因此密文大小与访问结构中的最小属性集呈线性关系，且在解密期间双线性配对操作数是不变的。Chen等[46]提出了一种用于云计算的具有定长密文的多权限CP-ABE访问控制方案，密文的长度和解密过程中的配对操作数都是不变的，与访问结构中设计的属性数也无关，在相对较强的安全模型中保持了高效率。

4.2.2 基于角色的访问控制（RBAC）

Zhou等[47]提出一种将基于角色的访问控制与加密相结合的数据安全存储方案，即只有满足基于角色的访问控制策略的角色才可以解密和查看数据；该方案中角色具有层次结构，且解密密钥大小恒定，与用户分配的角色数无关。Tang等[48]提出一种云环境下的基于角色的访问控制，将角色分为用户角色和所有者角色；用户从所有者获取凭证，再与服务提供者通信，获得资源的访问权限。Zhou等[49]提出一种基于角色的访问控制系统的信任模型以评估角色的可信度；数据所有者可以使用该信任模型评估是否以某个特定角色将加密数据存储在云中；该信任模型的一个特点是考虑了角色继承和层次结构。在此方案基础上Zhou等[50]扩展了信任模型，形成拥有者–角色和角色–用户两种基于角色的访问控制信任模型，该模型可以帮助角色识别对其可信度造成不良影响的恶意用户。Luo等[51]提出一种基于安全性和可用性的信任关系的RBAC；角色是否分配给用户由用户的信任度决定，信任度由以下因素计算获得：用户使用的主机的安全状态和网络可用性、与角色相关的服务提供商的保护状态，并提供了量化信任度计算过程的数学公式。

目前，开源社区已出现针对大数据平台的访问控制开源项目，并被一些大数据企业用于实际大数据应用开发。例如，Apache Sentry是Cloudera公司发布的一种针对Hadoop组件的可插拔授权引擎，对Hadoop集群中的数据和元数据提供细粒度、基于角色的授权。Apache Ranger提供一个集中式安全管理框架，解决大数据平台的访问控制授权和审计问题。

4.3 数据加密

数据加密的一个重要问题是如何对密文数据进行处理。同态加密和可搜索加密为此问题提供了解决方案。

4.3.1 同态加密

同态加密是一种保护数据私密性的解决方案，其思想可用式（1）表示：

$$E(K,F(x_1,x_2,\cdots,x_n)) = G(K,F,(E(x_1),\cdots,E(x_n))) \quad (1)$$

其中，$E(K,x)$表示用加密算法E和秘钥K对x进行加密，F表示一种运算。如果存在有效算法G，使得式（1）成立，就称加密算法E对于运算F是同态的[52]。

加密算法的同态性质可让用户将数据委托给第三方进行处理，也避免了数据泄漏，解决了大数据的安全计算问题。

1978年，Rivest等[53]首次提出秘密同态的概念，但构造全同态加密体制一直是密码学领域的难题。直到2009年，Gentry[54]使用理想格构造了第一个真正的全同态加密方案，在理论上取得了很大的突破。Gentry的全同态加密方案有3个步骤：1）构造一个部分同态加密算法，该算法对密文进行低阶多项式运算时保持同态性质；2）降低解密多项式运算的阶数；3）使用Bootstrapping方法将部分同态加密算法转化成同态加密算法。但是，因为计算和内存的开销巨大，该方案还无法投入实际应用。Smart和Vercauteren[55]提出了一个可对密文进行单指令多数据流运算的同态加密方案，该方案的重加密过程可实现并行计算，提高了运行效率。2010年，van Dijk等[56]在欧洲密码年会上提出了一种基于整数的全同态加密体制，该体制不需要在理想格上运算，但公钥的长度仍然很长。Coron等[57]针对van Dijk方案[56]密钥规模过大的缺点，提出了一种改进密钥生成的算法，因为其采用简单的整数环作为基础代数结构，所以比基于理想格的全同态加密方案更容易理解；其思路是：在密钥生成过程中仅生成一部分公钥，在加密过程中通过对存储的此部分公钥进行乘法运算获取全部公钥；该方案虽然减小了密钥规模，但增大了加密算法的计算复杂度。汤殿华等[58]基于部分近似最大公因子难题，在整数范围内提出了一个较快速的全同态加密方案，降低了van Dijk等方案[56]的计算复杂度，提高了执行效率。Brakerski等[59]利用Peikert等[60]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。

同态加密可为大数据的处理过程提供数据保护的功能，但是，如何提高同态加密的安全性和运算效率以满足实际应用需求，还需要进一步的研究。

4.3.2 可搜索加密

采用一般的加密方式对数据进行加密很难为数据建立索引，导致数据可用性很低[61]。可搜索加密技术的出现实现了对密文的检索和查询，有效解决了此问题。目前，主要的可搜索加密技术可以分为两大类：对称可搜索加密和非对称可搜索加密。

2000年，Song等[62]使用流密码和伪随机数构建了一种基于对称加密算法的可搜索加密方案，首次提出了可搜索加密的概念。但该方案的计算量与文档大小呈线性关系，计算效率不高。为解决该问题，Goh[63]采用了布隆过滤器和伪随机函数建立安全索引实现可搜索加密，该方案对每个数据项生成一个加密索引，而不是为所有数据项的内容生成索引，明显提高了搜索效率，但无法抵御统计攻击。

2004年，Boneh等[64]首次将公钥体制引入可搜索加密方案，提出了第一个非对称可搜索加密算法，被称为PEKS方案；在加密邮件系统的场景下实现邮件网关对关键字的密文搜索。Baek等[65]指出PEKS方案中的密钥分发存在需要安全通道和不可抵御关键词统计攻击的不足之处，并提出了一种不需要安全信道的可搜索公钥加密方案。Zhang和Lmai[66]则提供了一种基于PEKS和tag-KEM/DEM方案[67]的通用构造方法，并在标准模型下证明安全。

自从Gentry在2009年成功实现全同态加密后，随着同态加密技术发展，陆续有学者将同态加密应用到可搜索加密，出现了基于同态加密技术的可搜索加密。2011年，Gahi等[68]利用全同态加密技术对加密数据库实现查询、更新、删除等操作操作，但其效率较低。魏占祯等[69]基于RSA乘法同态实现了一种数据库的密文检索方案。李宏霞等[70]基于对称加密结构的思路，结合基于非对称加密的同态加密算法，实现多关键字的分级搜索。

4.4 审　计

目前，大数据平台主要通过审计日记记录平台中所有数据操作。HDFS、MapReduce、Hive等Hadoop生态常用组件均可通过配置开启审计日志功能；除此之外，Apache Knox Gateway、Apache Sentry、Apache Ranger等访问控制和身份认证项目也提供审计功能，记录用户的访问行为和管理组件间的安全交互行为。

5 大数据安全的开放问题

5.1 大数据安全标准缺口

大数据安全的标准化研究与制定工作还处于发展阶段，国际标准化组织积极推出了相关研究项目，探寻大数据安全标准的缺口。ISO/IEC JTC1 SC27（信

息安全分技术委员会)在工作组会议中提出以下研究项目:云服务可信接入架构、物联网隐私保护指南、智慧城市隐私保护、隐私保护工程框架等。由此可见隐私保护是大数据安全标准化研究中的热点问题。

全国信息安全标准化技术委员会大数据安全标准特别工作组最新发布的《大数据安全标准化白皮书(2017年)》中对大数据安全标准化工作提出了建议[1],其中,特别提到要“加快制定个人信息安全相关标准”“加快制定数据共享相关安全标准”“加快制定数据出境安全相关标准”“加快大数据安全审查支撑性标准研制”。这4个“加快”明确指出了下一步大数据安全标准化工作的重点。

目前,大数据安全的标准体系还未完善,应按照急用先行,成熟先上的原则快速推进标准制定。在研制大数据安全顶层标准的同时,加快对大数据应用的安全标准研制,解决目前最急迫的问题,如加快数据出境、数据交易和数据共享等的相关规范、制度的标准研究和制定工作。

5.2 大数据安全关键技术难点

当前的信息安全技术并不能完全满足大数据的安全需求,针对大数据应用中特有的安全风险,还有很多关键技术难点需要突破。

隐私保护是其中最受关注的问题之一。目前,已有一些解决方案。差分隐私保护通过添加噪声使数据失真,从而达到保护隐私的目的。但是,差分隐私保护算法的时间复杂度较高,实现效率并不理想。全同态加密方案适用于大数据场景中的隐私保护,但其性能较低的问题一直阻碍了同态加密技术应用于大数据环境。因此,设计高效的全同态加密方案值得深入研究。

加密是数据保护最基本也是最重要的手段之一。可搜索加密算法应用于大数据环境需要满足对多用户场景的支持和对不同加密算法的加密数据进行访问的要求,这对可搜索加密算法提出了新的研究方向。基于属性的加密方案因将访问控制策略直接嵌入到用户的私钥或加密数据中,不仅解决了公钥基础设施效率低下的问题,而且具有可扩展的密钥管理和灵活的数据分发的优势。目前,基于属性的加密方案主要采用椭圆曲线上的双线性映射构建,其中涉及计算成本昂贵的双线性配对操作,加之大数据规模庞大,因此难以应用到大数据平台[71]。

细粒度的访问控制是大数据安全领域一个新的热点问题。虽然目前开源社区已有一些解决方案,但细粒度的访问控制仍面临一些亟待解决的问题,如:对给定领域选择合理的访问控制粒度;当数据集增长到PB级的规模,访问控制方案的可扩展性问题;查询访问控制策略的效率问题等。

5.3 大数据安全分析的技术难点

大数据是一把双刃剑。它可能成为黑客实施攻击的手段,但将其有效利用,也可以成为防御安全攻击的盾牌。在开展大数据安全关键技术研究的同时,大数据中的批量和流式数据处理技术、交互式数据查询技术等可为网络安全与情报分析中的数据处理问题提供重要支撑,形成交互式可视分析、多源事件关联分析、用户实体行为分析等大数据安全应用。虽然大数据技术为网络信息安全提供了支撑,但其中仍存在许多问题亟待解决。隐蔽性和持续性网络通信行为检测、基于大数据分析的网络特征提取、综合威胁情报的高级网络威胁预测等关键技术有待实现突破,以提升网络信息安全风险感知、预警和处置能力[72]。

6 结 论

大数据作为国家基础性战略资源,已广泛应用于各重大行业,其安全问题得到学术界和产业界的高度重视和积极研究。本文介绍了大数据安全相关的法律法规和标准现状,分别从大数据生命周期安全和大数据平台安全两个角度分析目前大数据面临的安全问题,阐述大数据安全关键技术研究现状及其开源项目,最后提出了大数据安全在标准缺口、关键技术难点和大数据安全分析3个方面的开放问题。

大数据安全技术的发展,不仅是大数据产业发展所驱动的结果,还是国家部署的重要战略。加强研究大数据安全保护技术,可推动大数据的开放共享,有力支撑大数据产业的持续发展,更加增强国家网络空间安全的防御能力。

参考文献:

[1] 大数据安全标准化白皮书(2017)[R].北京:全国信息技术标准化技术委员会大数据安全标准特别工作组,2017.

[2] 中华人民共和国工业和信息化部.中华人民共和国网络安全法[EB/OL].(2016-11-08)[2017-06-12].http://www.miit.gov.cn/n1146295/n1146557/n1146614/c5345009/content.html.

[3] 中华人民共和国国家互联网信息办公室.国家网络空间安全战略[EB/OL].(2016-12-27)[2017-06-12].http://www.cac.gov.cn/2016-12/27/c_1120195926.htm.

[4] McKinsey & Company.Big data:The next frontier for innovation,competition,and productivity[EB/OL].[2017-06-12].https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf.

[5] National Institute of Standards and Technology.NIST big data interoperability framework:Volume 1,Definitions [EB/OL].(2015-09-16)[2017-06-12].https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-1.pdf.

[6] 中华人民共和国工业和信息化部.电信和互联网用户个人信息保护规定(工业和信息化部令 第24号)[EB/OL].(2016-04-07)[2017-06-15].http://www.miit.gov.cn/n1146295/n1146557/n1146619/c4700556/content.html.

[7] 国务院关于印发促进大数据发展行动纲要的通知[EB/OL].(2015-08-31)[2017-06-14].http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm

[8] 中华人民共和国国家互联网信息办公室.《网络安全法》解读[EB/OL].(2016-11-07)[2017-06-14].http://www.cac.gov.cn/2016-11/07/c_1119866583.htm

[9] Fan W,Geerts F,Ma S,et al.Detecting inconsistencies in distributed data[C]//Proceedings of the 2010 IEEE 26th International Conference on Data Engineering(ICDE).Long Beach:IEEE,2010:64–75.

[10] Fan W,Li J,Tang N.Incremental detection of inconsistencies in distributed data[J].IEEE Transactions on Knowledge and Data Engineering,2014,26(6):1367–1383.

[11] Beskales G,Ilyas I F,Golab L,et al.On the relative trust between inconsistent data and inaccurate constraints[C]//Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE).Brisbane:IEEE,2013:541–552.

[12] Chiang F,Miller R J.A unified model for data and constraint repair[C]//Proceedings of the 2011 IEEE 27th International Conference on Data Engineering(ICDE).Hannover:IEEE,2011:446–457.

[13] Fan W,Li J,Ma S,et al.Towards certain fixes with editing rules and master data[J].Proceedings of the VLDB Endowment,2010,3(1/2):173–184.

[14] Fan W,Ma S,Tang N,et al.Interaction between record matching and data repairing[J].Journal of Data and Information Quality,2014,4(4):16.

[15] Fan W,Li J,Ma S,et al.CerFix:A system for cleaning data with certain fixes[J].Proceedings of the VLDB Endowment,2011,4(12):1375–1378.

[16] Kanchi S,Sandilya S,Ramkrishna S,et al.Challenges and Solutions in Big Data Management—An Overview[C]//Proceedings of the 2015 3rd International Conference on IEEE Future Internet of Things and Cloud(FiCloud).Rome:IEEE,2015:418–426.

[17] Siddiqa A,Hashem I A T,Yaqoob I,et al.A survey of big data management:Taxonomy and state-of-the-art[J].Journal of Network & Computer Applications,2016,71:151–166.

[18] Zhang H,Chen G,Ooi B C,et al.In-memory big data management and processing:A survey[J].IEEE Transactions on Knowledge and Data Engineering,2015,27(7):1920–1948.

[19] Sweeney L.*k*-anonymity:A model for protecting privacy[J].International Journal of Uncertainty,Fuzziness and Knowledge:Based Systems,2002,10(5):557–570.

[20] 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),DOI:10.1145/1217299.1217302.

[21] Li Ninghui,Li Tiancheng,Venkatasubramanian S.*t*-closeness:Privacy beyond *k*-anonymity and *l*-diversity[C]//Proceedings of the IEEE 23rd International Conference on Data Engineering(ICDE),2007.Istanbul:IEEE,2007:106–115.

[22] Wong R C W,Li Jiuyong,Fu A W C,et al.(α,k)-anonymity:An enhanced *k*-anonymity model for privacy preserving data publishing[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.Philadelphia:ACM,2006:754–759.

[23] Zhang Qing,Koudas N,Srivastava D,et al.Aggregate query answering on anonymized tables[C]//Proceedings of the IEEE 23rd International Conference on Data Engineering,2007.Istanbul:IEEE,2007:116–125.

[24] Li Zude,Zhan Guoqiang,Ye Xiaojun.Towards an anti-inference (k,ℓ)-anonymitymodel with value association rules[C]//Proceedings of the 17th International Conference on Database and Expert Systems Applications.Kraków:Springer-Verlag,2006:883–893.

[25] Shang Xuan,Chen Ke,Shou Lidan,et al.(*k*,*P*)-anonymity:Towards pattern-preserving anonymity of time-series data[C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management.Toronto:ACM,2010:1333–1336.

[26] Liu Jian.Research on *K*-Anonymity for privacy preserving[D].Shanghai:East China University,2010.[刘坚.*K*-匿名隐私保护问题的研究[D].上海:东华大学,2010.]

[27] Oliveira S R M,Zaïane O R.Privacy preserving clustering by data transformation[J].Journal of Information and Data Management,2010,1(1):53–56.

[28] Zhang Peng,Tong Yunhai,Tang Shiwei,et al.An effective method for privacy preserving association rule mining[J].Journal of Software,2006,17(8):1764–1774.[张鹏,童云海,唐世渭,等.一种有效的隐私保护关联规则挖掘方法[J].软件学报,2006,17(8):1764–1774.]

[29] Xu Huanxia,Shao Liangshan,Chu Lili.The application of random orthogonal transformations in privacy preserving association rules mining[J].Science Technology and Industry,2010,10(1):75–79.[许焕霞,邵良杉,褚丽丽.随机正交变换法在隐私保持关联规则挖掘中的应用[J].科技和产业,2010,10(1):75–79.]

[30] Liu Feng,Xue An'rong,Wang Wei.Hybrid algorithm for privacy preserving association rules mining[J].Application Research of Computers,2012,29(3):1107–1110.[刘峰,薛安荣,王伟.一种隐私保护关联规则挖掘的混合算法[J].计算机应用研究,2012,29(3):1107–1110.]

[31] Celesti A,Tusa F,Villari M,et al.Three-phase cross-cloud federation model:The cloud sso authentication[C]//Proceedings of the 2010 Second International Conference on Advances in Future Internet (AFIN).Venice:IEEE,2010:94–101.

[32] Powell C,Aizawa T,Munetomo M.Design of an SSO authentication infrastructure for heterogeneous inter-cloud environments[C]//Proceedings of the 2014 IEEE 3rd International Conference on,Cloud Networking (CloudNet).Luxembourg:IEEE,2014:102–107.

[33] Morgan R L,Cantor S,Carmody S,et al.Federated Security:The Shibboleth approach[J].Educause Quarterly,2004,27(4):12–17.

[34] Méndez A P,López R M,Millán G L.Providing efficient SSO to cloud service access in AAA-based identity federations[J].Future Generation Computer Systems,2016,58(C):13–28.

[35] Lee S,Ong I,Lim H T,et al.Two factor authentication for cloud computing[J].Journal of Information and Communication Convergence Engineering,2010,8(4):427–432.

[36] Choudhury A J,Kumar P,Sain M,et al.A strong user authentication framework for cloud computing[C]//Proceedings of the 2011 IEEE Asia-Pacific Services Computing Conference.Jeju:IEEE,2011:110–115.

[37] Banyal R K,Jain P,Jain V K.Multi-factor authentication framework for cloud computing[C]//Proceedings of the 2013 Fifth International Conference on Computational Intelligence,Modelling and Simulation (CIMSim).Seoul:IEEE,2013:105–110.

[38] Liu Wenyi,Uluagac A S,Beyah R.MACA:A privacy-preserving multi-factor cloud authentication system utilizing big data[C]//Proceedings of the 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).Toronto:IEEE,2014:518–523.

[39] Goyal V,Pandey O,Sahai A,et al.Attribute-based encryption for fine-grained access control of encrypted data[C]//Proceedings of the 13th ACM Conference on Computer and Communications Security.Alexandria:ACM,2006:89–98.

[40] Chen Yanli,Song Lingling,Yang Geng.Attribute-based access control for multi-authority systems with constant size ciphertext in cloud computing[J].China Communications,2016,13(2):146–162.

[41] Waters B.Ciphertext-policy attribute-based encryption:An expressive,efficient,and provably secure realization[C]//Proceedings of the 14th International Conference on Practice and Theory in Public Key Cryptography.Taormina:Springer-Verlag,2011:53–70.

[42] Emura K,Miyaji A,Nomura A,et al.A ciphertext-policy attribute-based encryption scheme with constant ciphertext length[C]//Proceedings of the 5th International Conference on Information Security Practice and Experience.Xi'an:Springer-Verlag,2009:13–23.

[43] Herranz J,Laguillaumie F,Ràfols C.Constant size ciphertexts in threshold attribute-based encryption[C]//Proceedings of the 13th International Conference on Practice and Theory in Public Key Cryptography.Paris:Springer-Verlag,2010:19–34.

[44] Yang Kan,Jia Xiaohua.Attributed-based access control for multi-authority systems in cloud storage[C]//Proceedings of the 2012 IEEE 32nd International Conference on Distributed Computing Systems.Macau:IEEE,2012:536–545.

[45] Rao Y S,Dutta R.Decentralized ciphertext-policy attribute-based encryption scheme with fast decryption[C]//Proceedings of the 14th IFIP TC 6/TC 11 International Conference onInternational Conference on Communications and Multi-

media Security.Magdeburg:Springer-Verlag,2013:66–81.

[46] Chen Yanli,Song Lingling,Yang Geng.Attribute-based access control for multi-authority systems with constant size ciphertext in cloud computing[J].China Communications,2016,13(2):146–162.

[47] Zhou Lan,Varadharajan V,Hitchens M.Enforcing role-based access control for secure data storage in the cloud[J].The Computer Journal,2011,54(10):1675–1687.

[48] Tang Zhuo,Wei Juan,Sallam A,et al.A new RBAC based access control model for cloud computing[C]//Proceedings of the 7th International Conference on Advances in Grid and Pervasive Computing.Hong Kong:Springer-Verlag,2012:279–288.

[49] Zhou Lan,Varadharajan V,Hitchens M.Integrating trust with cryptographic role-based access control for secure cloud data storage[C]//Proceedings of the 2013 12th IEEE International Conference on Trust,Security and Privacy in Computing and Communications (TrustCom).Melbourne:IEEE,2013:560–569.

[50] Zhou Lan,Varadharajan V,Hitchens M.Trust enhanced cryptographic role-based access control for secure cloud data storage[J].IEEE Transactions on Information Forensics and Security,2015,10(11):2381–2395.

[51] Luo Jun,Wang Hongjun,Gong Xun,et al.A novel role-based access control model in cloud environments[J].International Journal of Computational Intelligence Systems,2016,9(1):1–9.

[52] Li Shundong,Dou Jiawei,Wang Daoshun.Survey on homomorphic encryption and its applications to cloud security[J].Journal of Computer Research and Development,2015,52(6):1378–1388.[李顺东,窦家维,王道顺.同态加密算法及其在云安全中的应用[J].计算机研究与发展,2015,52(6):1378–1388.]

[53] Rivest R L,Adleman L,Dertouzos M L.On data banks and privacy homomorphisms[J].Foundations of Secure Computation,1978,4(11):169–179.

[54] Gentry C.A fully homomorphic encryption scheme[D].Stanford:Stanford University,2009.

[55] Smart N P,Vercauteren F.Fully homomorphic SIMD operations[J].Designs,Codes and Cryptography,2014,71(1):57.

[56] van Dijk M,Gentry C,Halevi S,et al.Fully homomorphic encryption over the integers[C]//Proceedings of the 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques.French Riviera:Springer-Verlag,2010:24–43.

[57] Coron J S,Mandal A,Naccache D,et al.Fully homomorphic encryption over the integers with shorter public keys[C]//Proceedings of the 31st Annual Conference on Advances in Cryptology.Santa Barbara:Springer-Verlag,2011:487–504.

[58] Tang Dianhua,Zhu Shixiong,Cao Yunfei.Faster fully homomorphic encryption scheme over integer[J].Computer Engineering and Applications,2012,48(28):117–122.[汤殿华,祝世雄,曹云飞.一个较快速的整数上的全同态加密方案[J].计算机工程与应用,2012,48(28):117–122.]

[59] Brakerski Z,Gentry C,Halevi S.Packed ciphertexts in LWE-based homomorphic encryption[C]//Proceedings of the 16th International Conference on Practice and Theory in Public-Key Cryptography.Nara:Springer-Verlag,2013:1–13.

[60] Peikert C,Vaikuntanathan V,Waters B.A framework for efficient and composable oblivious transfer[C]//Proceedings of the 28th Annual International Conference on Advances in Cryptology.Santa Barbara:Springer-Verlag,2008:554–571.

[61] Wei Kaimin,Weng Jian,Ren Kui.Data security and protection techniques in big data:a survey[J].Chinese Journal of Network and Information Security,2016,2(4):1–11.[魏凯敏,翁健,任奎.大数据安全保护技术综述[J].网络与信息安全学报,2016,2(4):1–11.]

[62] Song D X,Wagner D,Perrig A.Practical techniques for searches on encrypted data[C]//Proceedings of the 2000 IEEE Symposium on Security and Privacy.Berkeley:IEEE,2000:44.

[63] Goh E J.Secure indexes[EB/OL].(2003-10-07)[2017-06-15].https://eprint.iacr.org/2003/216.

[64] Boneh D,Di Crescenzo G,Ostrovsky R,et al.Public key encryption with keyword search[C]//Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques,2004.Interlaken:Springer-Verlag,2004:506–522.

[65] Baek J,Safavi-Naini R,Susilo W.Public key encryption with keyword search revisited[C]//Proceedings of the International Conference on Computational Science and Its Applications,2008.Perugia:Springer-Verlag,2008:1249–1259.

[66] Zhang Rui,Imai H.Generic combination of public key encryption with keyword search and Public key encryption[C]//Proceedings of the 6th International Conference on Crypto-

logy and Network Security.Berlin:Springer-Verlag, 2007:159–174.

[67] Abe M,Gennaro R,Kurosawa K.Tag-KEM/DEM:A new framework for hybrid encryption[M].New York:Springer-Verlag,2008.

[68] Gahi Y,Guennoun M,El-khatib K.A secure database system using homomorphic encryption schemes[EB/OL].(2015-12-11)[2017-06-15].https://arxiv.org/abs/1512.03498.

[69] Wei Zhanzhen,Yang Yatao,Chen Zhiwei.Ciphertext retrieval in database based on RSA's multiplicative homomorphism[J].Journal of Harbin Engineering University,2013,34(5):641–645.[魏占祯,杨亚涛,陈志伟.RSA 乘法同态的数据库密文检索实现[J].哈尔滨工程大学学报,2013,34(5):641–645.]

[70] Li Hongxia,Pang Xiaoqiong.Searchable homomorphic encryption scheme supporting multi-keyword ranking[J].Computer Engineering and Applications,2016,52(22):93–98.[李宏霞,庞晓琼.支持多关键字分级的可搜索同态加密方案[J].计算机工程与应用,2016,52(22):93–98.]

[71] Samsudin A.Big data:Related technologies,security challenges,and research opportunities[EB/OL].[2017-6-22].http://repository.nauss.edu.sa/bitstream/handle/123456789/64343/002.pdf?sequence=1.

[72] Chen Xingshu,Zeng Xuemei,Wang Wenxian,et al.Big data analytics for network security and intelligence[J].Advanced Engineering Sciences,2017,49(3):1–12.[陈兴蜀,曾雪梅,王文贤,等.基于大数据的网络安全与情报分析[J].工程科学与技术,2017,49(3):1–12.]

陈兴蜀，女，博士，教授，博士生导师。四川大学网络空间安全研究院常务副院长，网络与可信计算研究所所长。全国信息安全标准化技术委员会委员，教育部信息安全教学指导委员会委员，中央网信办云计算服务安全专家组副组长，全国信息安全标准化技术委员会大数据安全工作组副组长，国际标准化组织ISO/IEC专家，四川省学术与技术带头人，四川省有突出贡献的优秀专家。

陈兴蜀主持了国家科技支撑项目、国家自然科学基金项目等20余项国家级、省部级科研课题。发表高水平论文100余篇，其中SCI/EI检索论文60余篇。主持编写GB/T 31167—2014《信息安全技术 云计算服务安全指南》国家标准，主编GB/T 31168—2014《信息安全技术 云计算服务安全能力要求》。获得四川省科技进步奖二等奖两项，荣获首届中国互联网发展基金会网络安全专项基金2016年网络安全优秀教师奖。

（编辑 赵 婧）

引用格式:Chen Xingshu,Yang Lu,Luo Yonggang.Big data security technology[J].Advanced Engineering Sciences,2017,49(5):1–12.[陈兴蜀, 杨露, 罗永刚. 大数据安全保护技术[J].工程科学与技术,2017,49(5):1–12.]