文章类型：论文|刊号（ISSN）：2972-4236(P) / 2972-4244(O)

# 计算机技术在大数据安全中的应用与挑战

李润启
内蒙古医科大学计算机信息学院
DOI:10.32629/acair.v3i4.17900

[摘　要] 大数据背景下的信息安全受到越来越严重的威胁,这主要是由于在大数据时代的数据规模和难度更加巨大。计算机技术起到了提升大数据安全防护强度、减少风险检测时长的重要作用。本文基于大数据安全的相关原理、技术,强调了数据在高维环境下加密效率低和入侵检测误差大等显著问题;介绍了基于同态加密实现安全计算、基于深度学习实现高效智能的入侵检测和基于差分隐私实现匿名化数据采集,以期为建立高效和可扩展的大数据安全保障体系奠定技术基础,促进计算机技术在此方面的应用实践。
[关键词] 大数据安全；计算机技术；同态加密
中图分类号：G633.67　文献标识码：A

## The Application and Challenges of Computer Technology in Big Data Security

Runqi Li
School of Computer Information, Inner Mongolia Medical University

[Abstract] Information security in the context of big data is facing increasingly serious threats, mainly due to the greater scale and difficulty of data in the era of big data. Computer technology has played a significant role in enhancing the security protection strength of big data and reducing the duration of risk detection. Based on the relevant principles and technologies of big data security, this article emphasizes significant issues such as low encryption efficiency and large intrusion detection errors of data in a high-dimensional environment. This paper introduces the realization of secure computing based on homomorphic encryption, efficient and intelligent intrusion detection based on deep learning, and anonymous data collection based on differential privacy, with the aim of laying a technical foundation for establishing an efficient and scalable big data security guarantee system and promoting the application practice of computer technology in this aspect.
[Key words] Big Data security; Computer technology; Homomorphic encryption

## 引言

大数据环境下数据规模不断扩大及数据流通方式的变化,导致安全的范围会不断变化,而安全风险也会随着变化。大数据环境下,数据在采集、存储及分析环节存在多种攻击风险,攻击手段也在不断更新,并且变得具有复杂性。由此可知,计算机技术的应用不再以单纯的安全计算、智能监测及隐私保护等技术为主,而是需要更多的技术来应对全新挑战。大数据安全的研究开始由局部防护向全面管控转变,技术路径也逐渐变得多样化。本文主要是对关键安全挑战方面的论述,并对复杂数据环境提出部分有效的解决方案。

## 1 大数据安全的理论基础与计算机技术应用概述

### 1.1大数据安全的基本概念与内涵

大数据安全是保证在采集、传输、存储和分析处理大量、多样、异构数据的过程中,数据不泄露、不会被篡改、可管理的一系列理论和技术体系。具体内容包括数据本体安全、数据处理安全、隐私安全以及系统运行可靠性保障。大数据具有维度高、传递速度快、价值密度高等特点,导致安全问题更为多样化、复杂化,其潜藏的风险可能会从恶意攻击、数据滥用、越权操作或模型推断攻击等方面出现。大数据安全应着重对数据整个生命周期进行有效的管理和控制,并利用安全技术和策略保证风险识别、访问限制、安全计算和溯源审核等各个环节的有序运行。大数据安全除了注重技术的防御能力外,还注重安全策略、管理制度与法律合规性在数据环境中的协同作用,以实现对整个数据环境的全方位安全防护。

### 1.2计算机技术在大数据安全中的应用分类

计算机技术在大数据安全中的应用可分为底层的基础支撑技术和高层的智能技术两个方面。对于基础层而言,加密算法、访问控制、漏洞分析、安全审计等技术可保证数据在存储或传

输过程中具备原始的数据安全性；其中加密技术保证数据机密性,访问控制限定权力边界,安全审计形成行为轨迹。从智能化层面来看,机器学习、深度学习、知识图谱的智能化技术已在异常监测、行为分析、攻击预警方面应用并解决复杂性威胁防护问题；同态加密、安全多方计算、联邦学习等隐私保护计算技术解决多个机构合作的数据处理与分析的隐私问题;差分隐私、匿名化等更注重数据发布和共享中各个阶段的隐私管理问题。计算机技术在大数据安全应用中具备综合性、多样性的趋势。

1.3当前主流架构与系统模型

现如今,大数据安全架构主要包括分层式、联邦式以及无信任架构三种流行类型。分层式安全架构将数据源层、通信层、存储层、解析层、使用层分别建立安全控制点,通过分层保护及多层防御实现安全管理模式,更适用于大规模的数据处理框架。联邦式安全架构专注于多主体数据合作条件下的安全互信机制,利用联邦计算框架、共享验证方法及跨领域访问控制,确保数据在区域间的安全协作,不离开本地,其常用于跨机构的数据融合项目中。无信任架构基于持续证实方法及最小权责原则,摒弃传统的边界防御模式,在身份验证、行为分析和访问授权方面实施动态化策略,以应对云计算及分布式场景下不断演进的风险环境。隐私保护计算架构、区块链可信赖数据架构等新兴模型也逐渐应用于大数据安全环境中,形成了多模型并行发展的新局面。

## 2 计算机技术在大数据安全中应用的挑战分析

2.1高维大数据环境下的数据加密效率低

高维大数据环境下,加密过程是整个数据处理流程中的主要瓶颈。在数据特征不断扩大的环境中,加密算法需要处理更大量的输入数据,因此导致了加密操作的计算复杂性不断提高,加密计算的工作负荷一直较高。大量的高维数据本身具有稀疏和分层等性质,在执行过程中,数据划分、密钥匹配和遍历操作等工作经常会增加额外的运算开销,而且维度越大,加密的计算时间就越长。来源不同、形式不一致的数据会造成格式问题,在执行加密前还需要花费大量时间与精力对格式进行解析、特征提取和结构调整等预处理,这些因素会大幅增加整个流程的时延。在分布式系统中,加密操作一般是多个节点协同完成,但在高并发下会产生密钥同步、中间结果交换与状态管控等众多问题,增加大量的通信时延,进而增加整体计算量。

2.2入侵检测系统误报率高、响应延迟长

大数据环境中,入侵检测系统的识别精度和反应速度受到了严格考验。传统工作流程融入了剧烈的变化和噪音元素,不同环境下会有不同的表现形式,系统形成稳定的特征边界存在难度,正常流量和异常流量在高维空间交织,导致误报率随之增加。日志流、网络会话、用户操作记录在短时间内集中产生,大量特征解析、行为分析、模式匹配等任务随之而来,加大计算负荷,堵塞检测路径,延长响应时间。链式、阶段式具有不同特征的攻击,将事件片段分散在各个环节和时期,如果没有上下文关联,很难实时发现攻击序列,出现判定周期推迟的现象。

2.3数据匿名化技术破译风险增加

在大数据环境中,多源数据集合、跨领域数据整合会导致匿名化数据面临越来越大的重识别风险。数据源的自然关联性以及统计相关性等,导致匿名化特征模式在跨表匹配中容易暴露残余结构,攻击者可以根据特征相似性、分布一致性来逐渐缩小身份范围。高维度数据具有大量潜在标志,即便进行模糊化或者分组处理,依然存在不可消除的统计相关性,维度越大,组合路线越多,匿名化的特征空间越容易被推断重构。公共数据库、历史记录以及侧面通道的存在大幅削弱了匿名保护,攻击者可以利用来自多个渠道对照建立辅助比对合集,以减少重识别的成本。随着行为数据的不断累积,时间序列、操作轨迹及互动特征等将逐渐形成稳定独特的个人痕迹,即使单次匿名处理充分,在后续的再发布过程或者跨领域整合中都会泄露新的相关线索。

## 3 计算机技术在大数据安全中应用的解决策略

3.1构建基于同态加密的数据安全计算模型

构建基于同态加密的数据安全计算模型,需要以执行密文计算的能力作为关键,以模块化的结构实现整套安全计算。模型主要包括密文计算终端、密文计算引擎、分布式任务调度器、密钥管理中心和结果还原器。数据在本地通过公钥加密后进行发送,通过平台上的密文计算执行各种可支持的代数运算,以便使数据在运算过程中不会泄漏。为适配大规模数据的运算需求,系统将任务拆解成若干可以并行运行的密文计算块,以节点计算的形式提高密文处理效率,通过计算图调度协调这些密文块的处理顺序和各个节点的任务量,从而高效利用整体计算资源。密钥管理中心的任务是密钥生成、分发、升级和权限隔离,以确保所有参与者在计算过程中使用的密钥是独立的,实现安全边界的精细化控制。结果还原器通过私钥对密文结果进行解密,不同参与者仅能观测到最终结果,而不是任何过程数据。在跨组织协同应用环境中,该模型具有显著优势。例如,多家金融机构共建信用评估模型,将各方客户特征加密后,输送给协同计算平台,在“黑盒”内完成如权重设置、矩阵计算、梯度优化等关键步骤,得出共享模型的系数参数,但无法了解到其他金融机构的数据内容。该机制既满足了监管部门数据隐私保护的标准,又实现了模型训练中的协作能力,并为建立高安全等级的大数据系统提供了可操作的技术路径。

3.2构建融合深度学习的智能入侵检测机制

构建深度学习辅助的高级别特征学习与多类型数据融合、全流程模型辅助复杂网络行为识别能力较强的智能化入侵检测系统。该系统模型包括数据源采集模块、特征提取/预处理模块、深度学习检测引擎、告警解释模块、模型定期更新模块五大组成部分。数据源采集模块为入侵检测系统中的数据源,采集的数据来源于网络流量、主机日志、应用系统操作、API访问路径;特征提取模块将原始数据转化为模型所需的向量形式,特征提取可能涉及序列化、嵌入式学习、时间区间化、协议特性提取。深度学习检测引擎可以依据特定条件选择最优卷积神经网络实现深度学习功能,使用循环神经网络或Transformer模式来探索

行为序列的特性，也可以利用图神经网络来模拟攻击链中节点之间的连接关系，从而使得其在高维度的信息输入下能实现良好的异常区分检测。告警解析模块将检测引擎产生的告警数据进行信任度的评分、多个源告警数据的合并和报警严重性的确定，以打造更为高效的预警路径，从而避免出现大量的无效警告干扰。模型更新功能主要是将平台数据样本作为增量训练、迁移学习或参数修改的目标，从而保证平台在适应变化的数据集时检测性能不受影响。实际上深度学习入侵检测系统已经在云计算资源的安全管理上被广泛使用。例如，某大型服务器集群，根据虚拟机网络流量、容器操作日志及管理人员的登录路径等信息为模型输入训练数据，从而将是否具有端口扫描、横向移动或越权访问等风险进行自动区分，并将结果集中输出一条告警信息。该方法使得深度学习在入侵检测中具有实时学习、高维特征演示、多信息关联等优势，成为大数据环境下安全监测体系中智能化技术方法。

3.3优化匿名化算法并引入差分隐私保护策略

在大数据环境下对差分隐私和匿名化进行全面保护，就需要为发布的数据设置一系列的差分隐私与匿名的控制与监管。总体设计包括数据预处理、匿名化处理、差分隐私噪声注入算法、可用性评估和访问审计五个部分。数据预处理工作阶段确定数据中的敏感信息并删除不相关的特征信息，以便为后续的匿名化准备一个固定的输入模式。匿名化处理工作阶段进行属性扩展、类目划分、区间化处理、关联消除等，这是通过精确可识别分类的方法进行，其目的在于降低关键线索；差分隐私工作阶段是在特定的需求和查询情境中选择和确定一个特定的隐私级别，在统计结果或模型参数中插入噪声，在保留整体统计属性的同时无法再将一条记录单独识别。可用性评估模块还将通过信息损失度量、偏离度评估和工作负载验证等来保证匿名化和差分隐私处理后的数据具有研究和分析利用的价值。访问审计系统对数据调用过程进行日志记录、身份校验和权限控制，形成能够被追踪的数据流，以防止隐私再次暴露。这种一体化的方法能够在现实中被用来跨组织共享人口统计、公共健康或行业指标等。例如，某市的卫生部门把慢性病调研数据共享给多家研究机构，将这些数据根据年龄进行分层、根据地区编码进行混淆以及对病历进行结构化去敏处理，创建匿名数据集，并且对查询数据接口增加差异隐私噪声，使研究所获取的数据集能够包含关于群体分布特征、疾病发展趋势以及风险关联结果等，但无法推断出具体个人信息。利用匿名化结构优化与差分隐私保护联合设计，为复杂、多源数据的共享应用提供了可控、安全且持续有效的隐私保障路径。

## 4 结语

大数据安全的构建基于计算机技术的安全计算、智能监测和隐私保护三重合力，其体系化发展对于实现数据驱动社会的可靠运行有着重要价值。围绕提升加密效率、实现智能入侵检测以及深化隐私保护形成的技术路线，为实现跨行业数据合作、可信应用提供了基础。未来，随着大数据规模的提升以及情境复杂化程度的加深，还应继续提升安全计算能力、跨域协同能力和隐私保护能力。计算机技术将继续在大数据安全系统可靠运行中发挥关键性作用。

## [项目基金]

内蒙古医科大学2023年度高等教育教学改革研究与实践项目，大数据时代增强大学生信息安全素养的探索与研究(项目编号：NYJXGG2023058)。

## [参考文献]

[1]许亮，王伟平，吴照光.大数据时代计算机网络安全技术的优化策略[J].微型计算机，2024(9):61-63.

[2]郑利平，徐学锋.云计算技术在计算机大数据分析中的应用[J].通讯世界，2023，30(1):91-93.

[3]关昕，高沛鑫.大数据时代计算机网络安全技术应用分析[J].计算机产品与流通，2024(8):155-157.

[4]沈杏杏.大数据时代计算机网络安全技术探讨[J].前卫，2024(30):0183-0185.

## 作者简介：

李润启(1977--)，男，汉族，河北南宫人，硕士，讲师，研究方向：计算机技术、信息安全。