 DOI: https://doi.org/10.12349/iser.v7i5.10534

Research on Data Security and Privacy Protection Technology in Big Data Environment

Jian Xu Zhe Wu Mingzi Tao

Xingcheng City Tax Bureau, State Taxation Administration China, Xingcheng, Liaoning, 125100, China

Abstract

Against the backdrop of the rapid evolution of big data environments, data scale has exhibited explosive growth, data types have become increasingly diverse, and data mobility has been significantly enhanced, making issues of data security and privacy protection more prominent.The complexity of data processing chains and multi-stakeholder participation models exposes data to multi-level security threats during collection, transmission, storage, and sharing processes.Focusing on the technical requirements of data security and privacy protection, this study systematically reviews the mechanisms and synergistic relationships of key technologies such as data encryption, access control, data desensitization, and secure computation from the perspective of the data lifecycle.In conjunction with the risk characteristics in the process of data circulation, it analyzes the effectiveness of different technical approaches in preventing data leakage, resisting unauthorized access, and safeguarding privacy rights.Furthermore, from the perspectives of system construction and technological integration, it explores the implementation paths for enhancing data security protection capabilities and privacy protection levels, providing technical support for building a secure and trustworthy data application environment.

Keywords

big data; data security; privacy protection; data encryption; access control

大数据环境下数据安全与隐私保护技术探究

徐健 吴哲 陶洺梓

国家税务总局兴城市税务局，中国·辽宁 兴城 125100

摘要

在大数据环境快速演进的背景下，数据规模呈现爆发式增长，数据类型不断丰富，数据流动性显著增强，数据安全与隐私保护问题日益凸显。复杂的数据处理链条与多主体参与模式，使得数据在采集、传输、存储与共享过程中面临多层次安全威胁。围绕数据安全与隐私保护的技术需求，从数据生命周期视角出发，系统梳理数据加密、访问控制、脱敏处理及安全计算等关键技术的作用机制与协同关系。结合数据流转过程中的风险特征，分析不同技术路径在防范数据泄露、抵御非法访问及保障隐私权益方面的应用效果。进一步从体系构建与技术融合角度，探讨提升数据安全防护能力与隐私保护水平的实现路径，为构建安全可信的数据应用环境提供技术支撑。

关键词

大数据；数据安全；隐私保护；数据加密；访问控制

1 引言

数字技术持续深化应用，数据已成为支撑经济社会运行的重要基础资源，数据价值不断释放的同时也带来了更为复杂的安全挑战。多源异构数据的融合处理推动了信息流动效率的提升，但也使数据边界逐渐模糊，传统以边界防护为主的安全模式难以适应动态化的数据流通环境。数据在跨平台、跨系统乃至跨区域流转过程中，面临身份认证不足、权限控制失效以及隐私信息暴露等多重风险。与此同时，数据应用场景不断拓展，对安全性与可用性的平衡提出了更高要求。如何在保障数据价值挖掘的前提下实现精细化安全管控，成为信息技术领域亟需解决的关键问题。围绕数据安全与隐私保护的技术体系进行系统梳理，

【作者简介】徐健（1982–），男，满族，中国辽宁人，本科，工程师，从事大数据环境下数据安全与隐私保护技术研究。

对于提升整体防护能力具有重要意义。

2 大数据环境下数据安全与隐私保护的内涵界定与发展特征

2.1 大数据环境的技术特征与数据形态演进

大数据环境以高体量、高速度与高多样性为核心特征，数据来源由传统结构化数据逐步扩展至半结构化与非结构化数据，文本、图像、视频及传感器数据共同构成复杂的数据体系。数据规模由GB级向PB乃至EB级跨越，单一业务系统日均数据增量可达数十TB，数据处理架构由集中式向分布式演进，形成以分布式存储和并行计算为基础的处理模式。数据流动路径呈现跨平台与跨区域特征，数据在不同节点间频繁迁移，带来更高的流动效率，同时也加剧了安全边界模糊问题。数据形态的动态变化使得数据生命周期管理更加复杂，数据实时性与价值密度差异显著，促使安全防护从静态防御向动态感知转变，安全体系需要适应高并发访问与海量数据处理的运行环境。

2.2 数据安全与隐私保护的基本内涵界定

数据安全主要指在数据全生命周期中保障数据的机密性、完整性与可用性，确保数据在采集、传输、存储及使用过程中不被非法获取、篡改或破坏。隐私保护则侧重对涉及个人或敏感信息的数据进行控制，避免个体身份及行为特征被识别或滥用。随着数据价值的不断提升，数据安全逐渐从传统的信息保护扩展至数据资产保护范畴，强调数据在流转过程中的安全状态控制。隐私保护的内涵也由单一信息遮蔽转向多维度保护，涵盖身份匿名、行为轨迹隐藏及数据关联分析风险防控。两者在实际应用中呈现高度耦合关系，既要保证数据可用性，又需防范隐私泄露风险，在安全与利用之间形成动态平衡[1]。

3 大数据环境下数据安全与隐私保护面临的关键问题

3.1 数据采集与传输环节的安全风险

数据采集阶段涉及多源设备与多种接口，终端设备安全能力参差不齐，容易成为数据泄露的源头。在高频数据采集场景中，数据传输速率可达Gbps级别，传输链路复杂，存在被监听与篡改的风险。无线传输环境下信号易受干扰，数据包丢失率可达到1%～3%，增加数据完整性风险。跨网络传输过程中，身份认证机制不完善会导致非法接入，数据在传输路径中可能被截获。加密机制缺失或配置不当也会削弱数据保护能力，部分系统仍采用低强度加密算法，存在被破解的隐患。数据在不同网络边界之间流转时，缺乏统一的安全策略，导致防护措施不连续，形成安全薄弱环节[2]。

3.2 数据存储与处理过程中的隐私泄露隐患

数据存储规模持续扩大，单一数据中心存储容量可达数PB，集中存储模式在提高管理效率的同时，也放大了数据泄露的影响范围。数据在处理过程中需要频繁访问与计算，权限控制机制若设计不合理，容易出现越权访问问题。日志记录不完整或审计机制缺失，会导致异常行为难以及时发现。数据处理过程中涉及多种算法模型，部分算法在训练阶段可能暴露原始数据特征，增加隐私泄露风险。缓存数据与临时文件未及时清理，也可能成为信息泄露的载体。多租户环境下资源共享程度较高，隔离机制不严密会导致数据交叉访问，增加隐私暴露概率。

4 大数据环境下数据安全与隐私保护的核心技术体系构建

4.1 数据加密与访问控制技术的协同应用

数据加密技术通过对数据内容进行算法处理，实现数据在存储与传输过程中的机密性保护。常用对称加密算法在大规模数据处理场景中具备较高效率，单次加密处理速度可达GB/s级别，非对称加密则用于密钥交换与身份认证，密钥长度通常为2048位或以上以保证安全强度。访问控制技术通过身份认证与权限管理实现数据访问限制，基于角色的访问控制模型可支持上万级用户权限配置。结合多因素认证机制，可将非法访问概率降低至0.1%以下。加密与访问控制协同应用，可在数据流转全过程中形成双重防护，既保护数据内容安全，又控制访问行为，提升整体安全防护能力。

4.2 数据脱敏与匿名化处理技术路径

数据脱敏技术通过替换、扰动或屏蔽敏感信息，实现数据在共享与分析过程中的隐私保护。常见脱敏方式包括掩码处理与随机替换，敏感字段可通过规则将真实数据替换为虚拟数据，使信息识别率下降至5%以下。匿名化技术则通过对数据进行泛化与聚合处理，降低个体可识别性，例如将精确位置数据转化为区域级信息，使定位精度由米级降低至千米级。k-匿名模型在应用中通常设置$k \geq 10$，可有效防止单条记录被识别。结合差分隐私机制，通过在数据中引入噪声，控制隐私泄露概率在可接受范围内，实现数据可用性与隐私保护的平衡。

4.3 安全计算与隐私保护算法的集成应用

安全计算技术通过在不暴露原始数据的前提下实现数据分析与处理，成为隐私保护的重要手段。多方安全计算可支持多个参与方在不共享原始数据的情况下完成联合计算，通信开销控制在原数据规模的2～5倍以内。同态加密技术允许在加密数据上直接进行运算，尽管计算效率较明文运算降低约10倍，但在高安全需求场景中具有重要价值。联邦学习通过分布式模型训练，实现数据不出本地的协同建模，单轮训练可支持上百节点参与。将多种安全计算与隐私算法进行集成应用，可在保障数据隐私的前提下实现数据价值挖掘，构建兼顾安全性与可用性的技术体系[3]。

5 大数据环境下数据安全与隐私保护的技术实施路径

5.1 基于全生命周期的数据安全防护机制构建

数据安全防护需贯穿数据生成、传输、存储、使用与销毁各阶段，形成闭环管理体系。数据在采集阶段通过身份认证与源头校验保证来源可信，在传输过程中依托加密隧道与链路完整性校验实现安全传递，在存储阶段通过分级分类策略对不同敏感等级数据进行差异化保护，敏感数据采用强加密与隔离存储方式，访问行为需进行细粒度权限控制。数据使用阶段强调操作留痕与审计追踪，日志记录粒度细化至操作级别，支持异常行为溯源。数据销毁阶段通过物理擦除与逻辑覆盖相结合方式，确保数据不可恢复。全生命周期防护机制以数据状态变化为主线，将安全策略嵌入各环节，实现持续防护与动态管理。

5.2 面向多场景的数据安全防护体系优化

不同应用场景对数据安全需求存在明显差异，金融、医疗与政务等领域对数据机密性与完整性要求较高，而互联网平台更强调数据流动效率与可用性。在多场景环境中，需根据业务特征构建分层防护体系，对核心数据、重要数据与一般数据实施差异化策略。高敏感数据在存储与传输过程中采用高强度加密与隔离机制，访问需经过多重认证，普通数据则在保证基本安全前提下提升处理效率。跨场景数据流转过程中，通过统一身份认证与权限管理机制，实现安全策略的连续性。结合容器化与虚拟化技术，对不同业务环境进行隔离部署，降低数据交叉访问风险，使安全体系具备灵活适配能力[4]。

5.3 融合智能化手段的数据安全动态监测机制

数据规模与访问行为的持续增长，使传统静态防护方式难以满足实时安全需求，动态监测机制成为关键支撑。通过对用户行为、访问频率与数据调用模式进行持续采集与分析，可建立行为特征模型，实现异常行为识别。系统在运行过程中对访问请求进行实时比对，一旦发现异常模式即触发预警机制，缩短安全响应时间。结合自动化策略，可对异常行为进行即时阻断或权限调整，减少潜在损失。动态监测机制还可通过日志分析与数据关联分析发现潜在风险点，对历史数据进行回溯分析，提升风险识别能力，使安全防护由被动响应向主动防控转变。

6 大数据环境下数据安全与隐私保护的协同治理与优化方向

6.1 技术驱动与制度约束的协同融合路径

数据安全与隐私保护的实现依赖技术手段与制度体系的协同作用，单一技术或管理方式难以应对复杂的数据环境。技术层面通过加密、访问控制与安全计算等手段提供基础保障，制度层面通过规范数据使用行为与明确责任边界形成约束机制。数据流转过程中需建立清晰的权限划分与责任追溯体系，使各参与主体在既定规则下开展数据活动。制度约束对技术应用形成导向作用，推动安全技术标准化与规范化发展，技术手段则为制度执行提供支撑，使安全管理具备可操作性与可验证性，从而实现安全治理体系的有效运行。

6.2 数据安全风险评估与分级管控机制构建

风险评估是数据安全管理的重要基础，通过对数据资产价值、敏感程度及潜在威胁进行综合分析，可形成多维度风险识别体系。数据按照敏感等级划分为不同类别，对高风险数据实施重点防护策略，在访问控制、加密强度与审计频率等方面设置更高标准。风险评估过程需结合动态监测结果进行持续修正，使评估结果能够反映实际风险变化。分级管控机制通过对不同风险等级数据实施差异化管理，实现资源的合理配置，避免安全投入与风险水平不匹配问题。通过定期评估与持续优化，可提升整体安全管理的科学性与精准性[5]。

6.3 数据安全与隐私保护持续优化的技术方向

随着数据规模与应用复杂度的不断提升，数据安全技术需要持续演进以适应新环境。未来发展方向将更加注重隐私计算与安全计算技术的融合，通过提升算法效率与降低计算成本，实现高安全性与高性能的平衡。分布式安全架构将进一步强化数据在不同节点间的保护能力，减少集中存储带来的风险。数据可追溯技术的发展将增强数据流转过程的透明性，使数据使用行为可监控、可审计。安全防护体系将向自动化与自适应方向演进，通过实时感知环境变化并动态调整策略，提升系统应对复杂威胁的能力，推动数据安全与隐私保护水平不断提升。

7 结语

在大数据持续深化应用的背景下，数据安全与隐私保护已成为保障数据价值有效释放的重要前提。围绕数据全生命周期构建系统化防护体系，通过多层次技术手段与动态管理机制的协同运行，可有效提升数据安全防护能力。面对复杂多变的风险环境，持续推进技术创新与治理优化，强化风险识别与精细化管控，有助于实现数据利用与安全保障之间的平衡，推动数据应用向安全可信与高效有序方向发展。

参考文献

[1] 王博,阮俊杰.大数据环境下档案信息安全风险防控策略分析[J].办公自动化,2026,31(06):69-71.

[2] 刘宏.大数据环境下医疗行业档案安全风险的动态识别与精准防控[N].山西科技报,2025-12-30(B07).

[3] 李春,王一.大数据环境下的计算机通信技术中网络安全与数据保护意识研究[J].中国信息界,2025,(12):133-135.

[4] 胡丽娜,李陆燕,王锐,胡冰,孟晔南.大数据环境下信息安全管理体系研究[J].中国管理信息化,2025,28(24):227-229.

[5] 田小娜.大数据环境下的信息安全隐私保护技术[J].办公自动化,2025,30(23):8-10.