 DOI: https://doi.org/10.12345/bdai.v7i1.35275

Cybersecurity in the Big Data Era and Preventive Measures

Lulu Jia

Qi An Xin Technology Group Co., Ltd., Beijing, 100000, China

Abstract

Big data technology has facilitated the rapid development of government governance optimization, industrial digital intelligence, and precise social services in China, but it has also brought about security governance challenges with highly heterogeneous data in terms of scale, structure, and circulation speed. Based on the analysis of China's current laws and regulations, national regulatory practices, and industry emergency situations, this paper first identifies four typical challenges in network information security arising in the massive data environment of the big data era, and finally proposes several actionable and assessable preventive measures and implementation paths for the aforementioned risks for reference.

Keywords

big data era; network information security; challenges; preventive measures

大数据时代的网络信息安全及防范措施

贾璐璐

奇安信科技集团股份有限公司，中国·北京 100000

摘　要

大数据技术在中国促成了政务治理优化、产业数智化与社会服务精准化的快速发展，但同时带来了数据规模、结构与流通速率等高度异质性的安全治理难题。本文以中国现行法律法规、国家监管实践与行业应急态势为分析基础，首先识别大数据时代的海量数据环境下出现的四类典型网络信息安全所面临挑战，最后针对上述风险提出几项可操作化、可评估的防范对策及实施路径，以供参考。

关键词

大数据时代；网络信息安全；挑战；防范措施

1 引言

进入新时期得益于中国信息技术迅猛发展，大数据时代应运而生，使人民群众的日常生产生活获得了极大的便利。但与此同时，网络信息安全问题也需要受到越来越多的重视，若不能有效保障网络信息安全，将有可能导致网络用户的安全和权益受到严重侵害，所以在当前大数据时代针对网络信息安全有必要提出合理的防范措施 [1]。

2 大数据时代的网络信息安全挑战

2.1 数据边界模糊与跨域流动风险

在中国大数据环境下，数据边界呈现组织内外与跨地域多层次重叠，业务系统、云服务与第三方平台之间常通过 API 与数据中台实现高速联通，导致数据主权、归集责任与访问控制边界不清晰。跨域流动既包含省内政务数据在云资源间迁移，也涉及跨境同步与海外服务调用，这种流转加剧了重要数据与个人信息在传输环节的暴露面，同时对于“重要数据”“出境评估”与合规判定产生实操性难题，监管尺度与企业合规路径存在脱节，链式依赖使得一处脱敏或接口错误可导致连带风险扩散，进而影响数据资产可追溯性与责任认定。

2.2 海量异构数据下的隐私外泄与重识别风险

海量多源数据融合是行业应用核心，但不同来源数据在结构、粒度与语义层面存在显著差异，去标识化处理常依赖规则化脱敏或简单的哈希处理，这在高维数据融合与机器学习场景下容易被辅助信息逆向重识别。尤其是健康、出行与消费等领域的小样本稀疏属性，与公开可得的轨迹或社交数据一经交叉比对即可能恢复个体身份或构建敏感画像。此外，数据共享链路上存在的非对称访问控制与日志稀疏，使得事后溯源与责任归属变得困难，给隐私权保护与合规证明带来实务挑战。

2.3 云服务与物联网带来的攻击面扩大

企业在迁移至公有云、混合云与微服务架构时，原有的边界式入侵检测与终端防护模式需调整以适配动态弹性资源。与此同时，物联网设备在城市、工业与消费场景

【作者简介】贾璐璐（1995–），女，中国山西太原人，硕士，工程师，从事网络安全研究。

的广泛部署带来大量低成本、低安全配置的终端，这些终端常作为僵尸网络与侧向渗透的入口 [2]。攻击者利用云端 misconfiguration、API 权限误置或物联网设备弱口令实施横向攻击，形成对后端数据湖与分析平台的直接威胁。在若干公开态势报告中，可见利用云主机进行的挖矿、横向渗透与大流量 DDoS 攻击呈现上升态势，表明攻击面扩展已成为影响大数据平台可用性与完整性的关键因素。

2.4 算法与模型层面的安全与合规风险

结合实践来看，大数据时代下网络信息安全中算法在大规模数据驱动的应用中承担决策与推荐功能，但模型训练数据质量、标签偏差与样本不均衡会引入系统性偏见，模型可解释性不足则增加监管审查难度。与此同时，模型权重与推理接口成为新的攻击面，模型窃取、对抗样本与数据投毒可在不改变原始数据库的前提下破坏模型行为或泄露训练数据特征。再者，算法部署涉及伦理合规、可追溯性与备案要求，企业在模型生命周期管理、日志留存与审计能力方面常面临治理盲区，使得算法风险既是技术问题也是合规与治理问题。

3 大数据时代的网络信息安全防范措施

3.1 构建基于分层分域的数据治理与标识体系

针对数据边界模糊与跨域流动风险，其防范措施主要在于以下几个方面：第一，在组织与行业层面应建立系统化的数据分类与标识流程，采用规则化识别引擎结合人工复核，形成“重要数据”“敏感个人信息”“普通业务数据”等标签集，并将标签写入数据资产目录与元数据层，同时制定版本化的分类标准与变更流程，保证数据在全生命周期内的可追踪性与可辨识性，且与组织风险目录及监管要求定期对齐，定期以抽样与自动化扫描校验资产标识。第二，构建分层分域治理框架，将数据按业务边界、风险等级与合规属性划分为域，明确域内数据域主、域管理员与合规负责人职责，基于最小权限与细粒度访问控制实现域间隔离，使用策略引擎实现跨域策略下发与一致性校验，配套实时审计、行为分析与异常告警以量化边界风险，并与 SIEM 联动形成态势感知，通过沙箱与标签驱动的访问治理降低敏感数据暴露面，并将隔离效果纳入日常合规考核。第三，对跨域与跨境流动实施基于规则的准入与评估，先行进行自动化合规检查、风险评分与隐私影响评估，依据评分触发分级加密、动态脱敏或拒绝传输，并优先采用符合国密要求的加密与密钥管理实践，在出境场景中嵌入合规合同、评估报告与可溯源审批链条，同时构建交互式审计面板支持事后取证与责任定位，确保在异常情形下可迅速回溯，并将处置流程纳入应急演练与备案 [3]。第四，完善元数据管理与数据目录建设，采用统一元数据模型支撑数据发现、敏感识别与脱敏策略的动态生成，结合密钥管理系统、加密网关与数据标记技术实现静态与实时防护协同，推动数据目录与身份认证、权限管理系统联动，利用自动化规则生成脱敏模板并通过度量化指标如覆盖率、策略执行率与异常响应时延持续评估与闭环改进，并结合国家与行业标准持续修订控制措施。

3.2 强化隐私保护技术链与最小必要原则的技术实现

在数据采集与共享环节，应构建基于业务场景的字段白名单与权限矩阵，前端以最小必要字段集为准并通过静态权限预校验与运行时埋点拦截机制禁止超量采集，后端将采集请求纳入策略引擎以进行目的匹配与字段级脱敏判定，策略引擎支持基于属性的访问控制与时间窗约束，同时以不可篡改日志记录采集溯源与同意时效以便合规审计，并将所有字段与用途在数据目录中登记以支持隐私影响评估流程。在隐私增强技术工程化方面，应构建可组合的隐私计算流水线，将联邦学习模型的参数同步、聚合与差分隐私噪声注入作为流水线节点，同态加密或安全多方计算用于敏感中间态保护，采用可信执行环境与硬件隔离保障边缘汇聚安全，建立误差预算与性能回归测试以评估对业务精度的影响，并将隐私参数纳入 CI/CD 测试与模型上线审批以实现持续回归与合规审计。在数据脱敏与重识别风险管理方面，应部署智能化敏感字段识别器、字段哈希与伪标识映射机制及脱敏编排工具，按照风险评分对外部发布数据实行掩码、泛化或置换等分级脱敏，并以红队仿真、多模型重识别攻击与合成样本对脱敏后数据开展定期压力测试，自动生成脱敏效果量化报告、残余风险矩阵与可逆回滚路径以支持分析需求变更与监管检查。在数据留存与销毁治理方面，应依据数据分级制定分层保留周期与销毁策略，关键数据采用不可逆化或分片加密并在多方见证下执行销毁操作，销毁行为保留可核验审计链与证据并形成销毁凭证，同时对历史数据二次利用要求再授权且施加差分隐私、沙箱化访问或最小化查询接口以控制新算法环境下的再识别风险，并要求法务与审计参与变更审批以形成技术—流程—合规闭环。

3.3 推进云原生安全与物联网安全加固实践

在大数据时代推进云原生安全与物联网攻防加固，应围绕四个技术域展开落地措施以符合产业环境与合规要求。第一，推行“云安全即代码”治理，将基础镜像加固、依赖项供应链溯源、镜像最小化与 IaC 模板合规扫描全部纳入版本控制，CI/CD 流水线内置 SCA、SAST 与 DAST 检测器与策略即代码引擎，触发配置漂移检测后自动运行基线校验、策略回滚与镜像重建，并对制品实施签名、时间戳与可核验元数据管理以保障制品完整性。第二，构建基于最小权限的细粒度授权体系，融合 RBAC 与 ABAC 并通过集中化 IAM 实现统一身份目录，规范短期令牌与多要素认证、密钥轮换与密钥保管服务，采用结构化云审计与不可篡改日志存储支持端到端访问链路回溯与安全取证。第三，针对物联网终端制定供应链安全与部署标准，强制硬件安全模块或可信根植入、出厂唯一证书与密钥托管、固件签名与差分补丁分发、OTA 回滚与回归测试流程，并建立设备入网认证、持续脆

弱性扫描与异常节点自动隔离策略以防止终端被并入僵尸网络或作为侧向跳板。第四，引入态势感知与行为分析平台，汇聚网络流、主机与应用日志及云审计数据，采用规则、统计与机器学习相结合的混合检测引擎并接入国内外威胁情报实现实时预警与自动化响应，配套可执行应急剧本对接 SOAR 实现自动隔离与补救，构建可量化的告警阈值与误报降噪流程，并通过定期红蓝对抗、演练与白盒审计评估检测覆盖度与响应时效并明确定责、量化指标与持续审计流程，并纳入年度合规检查 [4]。

3.4 建立算法治理与可审计的模型生命周期管理体系

在大数据时代构建可审计的模型生命周期管理体系，应首先就模型准入与上线审查形成书面化流程，明确算法分级与备案范围，要求提交训练数据来路证明、合规性材料、脱敏与最小化清单、数据血缘与元数据文档、训练参数与训练日志，并以不可篡改哈希或时间戳归档，审查由安全、合规与业务小组交叉复核，异议进入版本化复议流程以保证审计链路完整。其次，在评估阶段引入可解释性与公平性量化检测，采用局部与全局特征分解、对照敏感性分析、差异影响矩阵及群体公平性指标并保存可复现测试用例与对抗试验记录，超阈值偏差或鲁棒性失效须在变更管理中记录整改并重走审查。第三，运行期部署在线与离线联合监测与回溯机制，实时采集输入分布、预测置信度、响应延时与错误率，通过滑动窗口与概念漂移检测器识别退化，触发条件达到时自动降级或熔断并将事件、回滚与数据快照以加密审计日志保存供取证。第四，完善知识产权与访问控制策略，实施细粒度角色权限、最小权限与多因子认证，核心模型参数与推理接口置于密钥管理与可信执行环境隔离，对外发布推理结果时附带摘要式决策依据与可解释性片段，治理指标由安全、合规、业务及审计部门联合制定并纳入按版本可追溯的复审档案与处置闭环 [5]。形成 SLA 约束与定期复核，建立量化考核指标包括模型偏差率、漂移频次、平均恢复时间与审计完结率，并与安全事件响应流程对接，确保在发生疑义或合规调查时能迅速定位责任主体、回溯决策链并实施补救与问责。同时确立日志保全期与加密保存策略，定期引入第三方测评并定期调整以符合法规。

3.5 构建应急响应流程

针对大数据时代下网络信息安全防范还应构建应急响应流程，具体为：首先，建立基于流数据速率感知的快速隔离策略。部署流式探针并结合分布式追踪 ID 实现异常查询溯源，配置分层旁路与白名单机制，在检测到突增访问或非典型复杂联表时自动将相关请求导入隔离集群并对涉事会话做时间窗回放与短期回滚，随后通过版本化脚本逐表逐分区回退并保留审计镜像以便后续复核，同时设置差分权限锁定以禁止扩散性写操作。其次，实施面向数据湖与计算平台的取证快照与链证流程。采用列式快照格式并结合增量 WAL 记录保存不可变分区，所有快照写入前生成哈希摘要与时间戳并上链或存证机构备案，同时启用只读冻结索引以防止索引漂移，取证链按法务要求导出元数据、访问日志与计算快照以满足司法鉴定需要，并在隔离期间保留原始查询语句与临时表镜像。

4 结语

综上所述，大数据时代下中国网络信息安全治理既面临规模化数据流动与技术态势的挑战，也具备制度性与技术性协同创新的条件。实现风险可控依赖于制度原则的落地化转化、跨域数据治理的工程化实施与模型治理的持续审计，三者需在国家监管框架下与行业实践中并行推进。

参考文献

[1] 苏云川.大数据时代网络的信息安全及防范[J].信息产业报道, 2023(9):0012-0014.

[2] 李华龙,陈力超.大数据时代的计算机网络信息安全问题及防范策略[J].中文科技期刊数据库（全文版）工程技术, 2023.

[3] 刘继红,王佳慧,贺一峰.大数据时代信息安全风险防范探析[J]. 2023.

[4] 秦良斌,王力,张维利,等.论大数据时代下的信息安全及发展[C]// 内蒙古自治区通信学会2022年度学术年会论文集.2023.

[5] 李斌,李展.大数据时代网络信息安全保护及防护措施研究[J].移动信息, 2025(7).