基于大数据云计算网络环境的数据安全问题研究

李锐

兴安盟大数据中心，内蒙古兴安盟乌兰浩特市，137400；

摘要：大数据和云计算技术的快速发展已经成为现代社会的关键驱动力，为企业和组织提供了前所未有的数据处理和存储能力。然而，这个数字化浪潮也伴随着潜在的网络数据安全威胁，威胁着敏感信息的保护和隐私。网络攻击者日益巧妙地寻找新的方法来入侵和破坏数据，使数据安全问题成为当今云计算环境的紧迫挑战。

关键词：大数据；云计算；网络安全环境

DOI：10.69979/3041-0673.26.05.019

在大数据与云计算深度融合的今天，数据已成为关键生产要素，但其安全问题也日益突出。结合当前技术环境与实践挑战，最核心的安全风险集中在数据泄露、非法访问、系统漏洞及生命周期管理薄弱等环节。

1 大数据云计算环境下的主要安全问题

1.1 数据泄露与隐私暴露风险加剧

数据泄露风险加剧的核心原因，多租户架构下的隔离失效，云计算采用多租户模式提升资源利用率，但若虚拟化隔离机制存在漏洞（如侧信道攻击、虚机跳跃），攻击者可突破边界访问其他租户数据。例如，黑客通过获取某一客户的虚拟机权限，利用共享物理服务器的内存或存储残留信息，间接窃取相邻客户敏感数据。数据全生命周期防护不足，从采集、存储、传输到销毁，每个环节都可能成为泄露入口：采集阶段：非法抓取公开信息结合 AI 分析，可推断出原本匿名化的敏感数据；存储阶段：部分云平台未对敏感数据加密或使用弱加密算法（如 MD5），导致数据一旦被访问即明文暴露；传输过程：明文传输、API 接口暴露、开发测试环境数据未脱敏等问题频发；销毁不彻底：数据删除后仍可通过恢复技术还原，形成“数据残留”隐患。AI 工具滥用加剧泄露风险，将内部文件“投喂”给开源 AI 模型处理时，数据会被永久存入训练集，开发者或黑客均可远程调取。此类行为无异于将机密信息置于公共空间。更严重的是，若涉及国家秘密，可能直接威胁国家安全。隐私暴露的深层挑战，大数据关联分析导致“去匿名化”，即使单条数据不包含身份信息，通过跨平台数据聚合与行为建模，仍可精准识别个体。例如，谷歌街景中模糊房屋反而引发“此地无银”效应，成为犯罪目标。这说明传统匿名化手段在智能分析面前已失效[1]。用户隐私控制权弱化，用户在使用服务时被迫让渡数据所有权，而数据经多次流转后权利边界模糊，个人难以追踪和控制其使用路径。企业过度收集、违规使用个人信息现象普遍。远程办公与无线通信扩大攻击面，非涉密设备处理敏感信息、即时通讯工具传输文件、无线信号被监听等，使泄密场景高度日常化。移动终端易遭木马植入，远程开启麦克风、摄像头，变成“贴身间谍”。

1.2 虚拟化与多租户环境带来的隔离失效

多租户环境下隔离失效的主要表现，数据泄露与非法访问，不同租户的数据若未实现物理或逻辑隔离，攻击者可能通过漏洞访问其他租户的敏感信息。例如，Gmail 曾因应用程序漏洞导致用户可越权查看他人邮件。类似地，存储资源回收时若未彻底清除数据，后续租户可能恢复前任用户残留的敏感文件。网络通信被监听或篡改，多租户间的网络隔离若配置不当（如 VLAN、VPC 设置错误），可能导致流量被非法监听或注入恶意代码，造成身份冒充或数据劫持。虚拟机逃逸与横向渗透，虚拟机管理程序（Hypervisor）若存在漏洞，攻击者可突破虚拟机边界，获取宿主机控制权，进而影响同一物理机上的所有虚拟机。此类“虚拟机跳跃”攻击已在实际案例中发生。容器间隔离不严，在云原生环境中，容器共享操作系统内核，若隔离策略不严密，攻击者可能从一个容器入侵宿主机或其他容器。关键防护策略与最佳实践，强化租户间物理与逻辑隔离，采用“租户 ID+业务字段”作为分片键，确保数据集中在独立分片。使用虚拟私有云（VPC）实现网络层隔离，限制跨租户通信。实施细粒度访问控制，推行多因素认证（MFA）和基于角色的访问控制（RBAC），最小化权限分配。对敏感操作（如跨分片查询）触发二次验证，如短信验证码。加强虚拟化与容器安全，定期扫描虚拟机镜像与容器镜

像中的漏洞与恶意代码。使用安全通道（如 TLS）保护虚拟机迁移过程，防止中间人攻击。数据全生命周期加密与审计，数据传输与存储均采用加密技术，日志通过区块链存证防篡改。启用完整审计日志，记录所有操作行为，支持安全事件回溯。自动化监控与响应，部署统一防火墙管理系统，自动封禁攻击 IP、优化冗余策略。利用 AI 驱动的安全策略推荐，实现动态适应与自动响应。

1.3 数据全生命周期防护不足

数据采集阶段：源头可信性难保障，大数据来源广泛，包括传感器、社交平台、交易系统等，数据在生成之初就可能被伪造或污染。例如，虚假评论、刷单数据等“脏数据”会干扰分析结果，导致决策失误。此外，缺乏对数据来源的验证机制，使得非法数据源难以识别，进一步加剧了数据污染风险。数据存储阶段：集中化带来高风险暴露，大数据通常采用分布式架构（如 Hadoop）和云存储技术进行多副本保存，虽然提升了可用性，但也导致数据高度集中，成为攻击者的主要目标。一旦攻击者突破防线，便可获取海量敏感信息。同时，部分平台缺乏有效的加密机制，数据以明文或弱加密形式存储，极易被窃取。数据传输阶段：边界模糊导致控制力下降，在将数据迁移至云端进行处理时，组织往往失去对数据的直接控制。尤其是在跨网络、跨平台传输过程中，若未采用强加密协议（如 TLS 1.3），数据可能在传输途中被截获或篡改。此外，内外网之间数据流动频繁，安全边界日益模糊，传统防火墙难以有效防护。数据使用阶段：访问控制滞后，隐私泄露频发，传统的基于角色的访问控制（RBAC）在复杂的数据共享场景下已显不足，角色爆炸导致权限管理混乱。相比之下，基于属性的访问控制（ABAC）虽更灵活，但落地仍不普及。此外，数据分析过程中，算法可能无意中暴露个人隐私，而现有技术在隐私保护与处理效率之间难以平衡。数据销毁阶段：残留风险长期存在，数据销毁不彻底是普遍隐患。即使执行删除操作，仍可通过数据恢复技术还原部分内容，造成“逻辑删除、物理留存”的尴尬局面。同时，缺乏统一的销毁标准和审计机制，导致部分敏感数据长期滞留系统中，形成潜在泄露源。其他关键风险补充，密钥管理薄弱：加密依赖密钥，但密钥常因硬编码、代码泄露或员工流动而暴露，尤其云访问密钥的泄露已成为重大安全隐患。供应链风险：云平台依赖多方软硬件组件，任一环节存在漏洞或后门，都可能波及整个系统安全。责任边界不清：用户与云服务商常对数据安全责任划分模糊，出现“谁都不管”的真空地带。

1.4 高级持续性威胁（APT）与新型攻击手段频发

高级持续性威胁（APT）：隐蔽而持久的“数字间谍”，APT 攻击是一种高度组织化、长期潜伏、目标明确的网络攻击形式，通常由具备国家级背景或强大资源支持的黑客组织发起。这类攻击不追求短期破坏，而是以长期窃取敏感数据、掌握关键系统控制权为目的，具有极强的隐蔽性和适应性。攻击特征：长期潜伏：APT 攻击者可在目标系统中潜伏数月甚至数年，逐步渗透、横向移动，避开常规安全检测。多阶段渗透：通常从钓鱼邮件、漏洞利用等“小切口”进入，再通过权限提升、内网渗透等方式扩大控制范围。伪装正常流量：攻击行为常与合法网络流量混合，难以被传统防火墙或 IDS/IPS 识别。典型攻击场景：针对政府机构、国防军工、高科技企业等关键基础设施，窃取国家机密或核心技术。利用开源代码仓库进行供应链投毒，通过恶意软件包污染开发环境，实现“一箭多雕”。结合勒索软件实施“双重勒索”，既加密数据又窃取信息，增加谈判筹码。新型攻击手段频发：技术演进催生新风险，随着云计算、大数据、人工智能的深度融合，攻击者也在不断升级战术，催生了一系列新型攻击模式：供应链攻击，攻击者不再直接攻击目标系统，而是通过入侵其依赖的第三方软件、服务或开发工具，实现间接渗透。例如：在开源代码库中植入后门，诱导开发者引入恶意依赖包。攻击云服务商的合作伙伴或运维接口，扩大攻击面。云配置错误导致的数据泄露，尽管云服务商提供强大的安全功能，但客户自身的配置失误仍是主要风险源。例如：云存储桶（如 S3）权限设置不当，导致数百万用户数据公开暴露。身份认证机制薄弱，未启用多因素认证（MFA），易被钓鱼攻击突破。DDoS 攻击规模与复杂度持续升级，分布式拒绝服务（DDoS）攻击正向应用层、混合型、AI 驱动方向演进：利用云计算资源或僵尸网络发起超大规模流量攻击，导致服务瘫痪。结合 AI 分析流量模式，动态调整攻击策略，绕过传统防护机制。内部威胁与恶意员工风险，无论是企业内部员工还是云服务商的运维人员，一旦拥有高权限且缺乏有效监控，都可能成为安全链条中最脆弱的一环。滥用权限访问敏感数据、导出机密信息。因疏忽或被社会工程学

诱导，成为攻击跳板。数据泄露与非法交易产业链猖獗，大数据本身的价值吸引了大量黑产关注，形成了完整的地下数据交易链条：黑客通过 APT、SQL 注入等方式窃取用户个人信息、金融数据等。数据在暗网中被批量售卖，用于诈骗、身份冒用等非法活动。

2 关键技术防护策略

全生命周期数据保护，分类分级：按敏感程度对数据进行动态分级（如医疗数据分三级），实现差异化防护。加密全覆盖：静态数据采用 AES-256 或国密 SM4 加密，传输过程使用 TLS 1.3 协议，确保“存储-传输-使用”全链路安全。销毁可审计：执行多次覆写+物理粉碎，并通过第三方审计确保不可恢复。零信任架构落地，推行“持续验证、永不信任”原则：多维度身份认证（生物特征+数字证书+行为画像），动态授权与微隔离技术，将网络划分为微段，限制横向攻击面，某制造业企业部署后，内部数据泄露事件下降 76%，AI 驱动的主动防御体系，UEBA（用户实体行为分析）：利用机器学习识别异常操作，如非工作时间批量导出数据，威胁情报融合：整合 MITRE ATT&CK 框架与内部日志，提升 APT 检测率，蜜罐技术：部署诱饵系统捕获攻击手法，日均处理 10 亿条日志，威胁响应时间压缩至 8 分钟，强化 API 与配置安全管理，对云服务接口实施严格访问控制，防止不安全 API 被滥用，自动化扫描配置错误（如开放端口、错误权限），及时修复误配置问题备份与灾难恢复机制，遵循“3-2-1”备份规则（3 份副本、2 种介质、1 份异地），结合云对象存储版本控制实现快速回滚，保障业务连续性。

3 管理与制度层面的应对措施

大数据与云计算环境下的安全问题日益突出，尤其在数据集中存储、共享频繁、技术架构复杂的背景下，安全风险已从单一的技术漏洞扩展到涵盖管理、制度、合规等多维度的系统性挑战。

3.1 建立健全数据安全管理制度

制定覆盖全生命周期的数据安全管理规程，明确数据分类分级、访问权限、加密策略、审计机制等操作规范，避免“无规可依”。制度应与业务流程深度融合，确保可执行、可追溯。

3.2 落实安全责任制与共担机制

明确云服务商、用户、第三方供应商的安全职责边界，通过服务协议（SLA）固化责任分工[2]。推行“谁使用、谁负责”“谁运营、谁保护”的原则，强化用户自身安全主体责任。

3.3 推动标准化与合规体系建设

参照国家等级保护制度、NIST、ISO/IEC 等国际标准，构建统一的安全控制框架。积极参与云计算服务安全评估（如我国的云计算服务安全评估办法），提升可控性和合规性。

3.4 加强安全组织与人员管理

设立专职安全团队，定期开展安全培训与意识教育，提升全员安全素养。实施背景审查、权限审批、行为审计等机制，防范内部人员违规操作或恶意行为。

3.5 建立持续监控与应急响应机制

构建安全运营中心（SOC），集成日志分析、SIEM、SOAR 等工具，实现安全事件的实时监测、自动响应与闭环处置。制定并演练灾难恢复计划（DRP），确保 RTO≤4 小时、RPO 达标。

3.6 强化供应链安全管理

对第三方产品和服务进行安全准入审查，建立供应商风险评估机制。推动核心软硬件国产化替代，提升供应链自主可控能力，满足信创要求。

总之，大数据云计算环境的数据安全问题一直备受关注，随着云计算的普及和大数据的快速增长，确保数据的机密性、完整性和可用性变得至关重要。

参考文献

[1]徐小静. 大数据背景下人工智能在网络技术中的应用. 2024.

[2]王源玉. 基于数据式审计模式的大数据智慧审计平台构建研究. 2023.