大数据环境下网络信息安全防护探析

红　丽

国家计算机网络应急技术处理协调中心内蒙古分中心 内蒙古 呼和浩特 010000

【摘　要】：大数据环境下网络信息资产属性和风险敞口同时增大，海量多源数据汇聚、流转、挖掘在推动经济社会发展的同时，也把传统意义上的网络边界消解成模糊地带。安全防护面对攻击面急剧扩大、威胁智能化、潜伏期拉长这些新的变量，而现有的防护体系大多局限在静态边界防御以及合规清单式的管理当中，不能很好地应对数据全生命周期里的动态风险。本文从数据属性变化出发，分析访问控制、加密隔离和行为分析在大数据架构下失效的原因，然后从零信任重构、数据安全治理、智能威胁狩猎和协同运营四个方面提出可以落地的防护路径。意图在数据可用性和保密性之间找到韧性平衡，把安全基因植入到大数据流转的每一个节点里，让防护体系从事后补救变成伴随式免疫。

【关键词】：大数据；网络信息安全；防护体系；零信任

DOI:10.12417/3041-0630.26.07.041

数据成了核心生产要素的时代，网络运行轨迹、用户指纹、交易流水、设备日志等被交织成庞大数据画像。信息在流动中产生价值，在流转中暴露出问题。传统的防护手段以筑墙封堵为要义，在面对跨云、跨终端、跨业务链的数据渗透的时候会显得无能为力。攻击者对大量的旁路数据进行聚合分析之后，就可以清楚地看到防御拓扑的漏洞所在，内部越权以及配置错误在庞大的系统中就会被放大成致命的缺口。探索适应大数据特性的防护机制，不是简单的修补旧有策略，而是在信任基础、可视能力、处置闭环三方面做系统的重构。只有在数据汇聚的源头就植入安全规则，在动态访问的过程中不断对信任等级进行校验，才能把混乱的信息流纳入有序的防护体系之中，让安全成为数据价值释放的内在条件。

1 数据属性变化引发的安全异动

1.1 流转边界模糊化与攻击面延展

大数据架构天生就强调数据的贯通和复用，业务接口、分析工具、数据中台以及外部合作链条互相纠缠在一起，导致原本可以明确划分出来的网络边界被分解成了许多微边界。每一个数据交换节点、每一个 API 调用都是一个入侵的跳板。微服务化、容器编排平台的出现使得服务之间的通信密度急剧上升，东西向流量远远大于南北向，传统的防火墙和入侵检测系统依靠南北向流量过滤来感知东西向数据漂移，已经不再适用。大量的公开 API 端点没有进行严格的字段级权限控制，攻击者可以通过枚举接口参数来遍历出超出设计范围的数据集，将正常的查询通道变成数据收割工具。数据湖把结构化信息和非结构化信息一起放在一起，一份原始日志经过多道加工分支之后，它的副本分散在测试环境、备份系统以及分析沙箱里，任何一个环节的权限失控都会造成全链崩溃[1]。

1.2 数据聚合引致的隐性风险挖掘

孤立片段信息没有任何意义，但是大数据所具有的关联和推导能力，可以把分散在各个地方的低敏感记录拼凑成一个高度敏感的完整轮廓。攻击者利用长时间潜伏来搜集电网负荷、物流单据、员工出行等旁系数据，从而推断出企业的核心产线排产计划或者重要人员的活动规律。此类攻击不是利用破解加密算法来实现的，而是通过滥用平台合法的查询和分析功能，用拼图的方式绕过传统的 DLP 策略的监控。不同的来源的数据一旦进入分析沙箱之后，匿名化处理就会被重识别攻击所破坏，原本去标识化的个人记录又会和具体的主体一一对应起来。更隐蔽的差分攻击就是通过聚集查询结果的小差别来反推个体记录，把统计发布当作信息泄露的通道。数据在地理分布上跨区域流动，跨境流动合规风险以及数据被第三方法庭强制披露的风险都会增大，使得数据主权和安全策略的执行时刻处于拉扯之中[2]。

2 防护体系在大数据环境中的结构性缺陷

2.1 静态信任模型与动态风险的脱节

主流的安全架构依然采用“一次认证，持续信任”的思想，用 IP 地址、账号口令等静态的凭证来决定是否放行。大数据分析平台会把各种各样的异构数据源都汇集起来，一个合法的查询账号在数据抽取的过程中可能会被恶意脚本劫持或者用它的权限去浏览超出业务需求的数据集。微服务之间内部调用依靠简单的服务账户令牌，横向移动几乎没有限制。攻击者只需要攻破一个低权限节点，就可以用信任做跳板，顺着数据管道直接来到核心库。内部人员用合法的身份进行数据窃取、篡改，在静态模型下几乎不能被实时拦截，因为每一次的操作系统都认定它“可信”。云原生环境中的短生命周期容器和无服务器函数使得基于 IP 的信任策略经常出现错误，凭证转储和滥用也更难被追踪。当一个账号被攻破之后，攻击者就会在数

据平台上长时间地隐藏自己，并且逐渐增加自己的权限，系统缺少对认证之后的行为进行持续评价的能力。边界信任的崩溃就是要把校验渗透到每一次访问、每一笔数据操作中去，使信任判定跟随上下文不断更新，把假定的安全变成时刻求证的安全。

2.2 数据生命周期防护的断链

大数据系统中信息从采集、传输、存储、加工、共享到最后的销毁，都经过了很长的过程，并且各个环节都很复杂。现实生活中大部分防护措施都是针对存储加密和数据备份展开的，而采集与加工阶段的防护却成了空白。日志采集器会把敏感字段以明文的形式保存在中间表中，在 ETL 过程中间表磁盘写入没有做加密脱敏处理，分析结果输出到非受控终端的时候也没有水印溯源机制。数据销毁大多采用简单的删除标记，磁盘残留数据可以轻易恢复，云端弹性存储的快照和副本也使得彻底擦除变得困难。加密技术的应用大多只停留在存储层，不能保证内存计算和使用态数据的安全，CPU 缓存和内存转储成了攻击者获取明文的捷径。密钥分散在各个应用系统当中，没有形成统一的密钥生命周期策略，共享流转时很难做到细粒度的动态授权，数据一离开存储卷，安全属性就会大大降低。大数据平台长久保存历史数据来满足重训模型以及审计的需求，客观上拉长了敏感信息暴露的窗口期，遗忘权执行同存储成本压力交织在一起。碎片化的分段管控，就数据贯通的大环境而言，就是一道道可以被悄无声息地跳过去的矮墙[3]。

2.3 告警过载与响应滞后的恶性循环

大数据平台每天产生大量的操作日志、系统告警，安全运维人员被淹没在误报、低价值告警的海洋里。传统的 SIEM 系统依靠规则匹配来工作，不能很好地理解异常行为模式，很多合法但是偏离常规的数据调用被错误地当作风险，真正的高级持续性威胁由于特征隐蔽而沉在噪音之中。告警的分级处理没有实现自动化编排，从发现故障到切断电源的窗口时间被人为操作延长了。告警上下文一般碎片化呈现，分析人员不能很快知道攻击所影响的数据资产范围、被影响的业务路径以及横向移动的轨迹，决策效率不断下降。攻击者在大数据环境的掩护之下可以进行慢速渗透，把敏感信息分批外传，就像是针尖取水一样，等到发现的时候数据已经完成转移。运维人员告警疲劳不断累积，处理效率持续下降，重要的信号被大量的杂音所覆盖，安全运营中心变成了故障记录站，没有了主动狩猎和预测威胁的核心功能[4]。

3 面向大数据特征的动态韧性防护构建

3.1 零信任架构在数据平面的落地

打破内外网划分的旧思维，用“永不信任，始终验证”的原则来代替它，就要把身份认证、权限判定下放到每一个数据访问请求上。基于属性的动态访问控制加上实时风险评分，给用户、设备、网络环境以及数据资源创建起多维匹配方案，哪怕环境或者行为出现最细微的变动，都会引发重新认证或者权限缩减。大数据平台内可以采用微隔离技术对不同的业务数据进行划分，跨区访问必须经过策略执行点再做认证。服务网格的 sidecar 代理可以对东西向数据流量进行透明拦截和策略判决，在同一个集群内也不能出现不受控制的数据漂移。策略引擎会不断地接收到终端检测、身份分析以及数据分类标签所发出的遥测信息，并据此不断对访问风险展开测算工作，从而由原来的粗放式的准入判定方式转变为现在的细颗粒度的按需授权机制。在数据使用的过程中，根据访问者的角色以及安全等级，对数据进行动态脱敏以及差分隐私处理，从而达到“恰如其分的可用”的目的，将过多的特权压缩到最小。策略判定点同数据节点紧密结合，每一次查询都会被即时的身份、设备状况以及数据标签所校验，从而避免后台进行绕行操作。

3.2 全链路加密与数据安全治理的融合

加密保护要对数据传输过程中的每一个跳转、存储态和运行态都进行加密。可信执行环境技术的加入，使数据在内存计算时也处在加密隔离的状态之中，防止云平台管理员或者主机层面的窃视，机密计算成了保护使用态数据的重要支撑。创建统一的密钥管理服务，给每一个数据集、每一个分析任务赋予独一无二的密钥，借助属性加密达成细粒度的共享掌控，密钥的更换和销毁同数据生命周期保持同步。安全治理不能只依靠技术，必须把分类分级的标准嵌入到大数据开发的流水线之中。数据入库立即打上分级标签和溯源水印，以后所有的流转环节都要强制校验标签属性和操作权限是否匹配。数据脱敏策略是用标签为锚点自动执行的，从源头上防止敏感信息流向低安全环境。同态加密技术的发展使可以在密文中直接进行部分分析运算，虽然还受到性能限制，但是在隐私保护要求非常高的情况下，给问题的解决提供了一种新的思路。当安全策略同元数据管理深度嵌合的时候，防护的触发就不再依靠人工配置，而是在数据标签移动的过程中自动发生，形成起原生的安全数据流通骨架[5]。

3.3 人工智能驱动的威胁狩猎与异常感知

把机器学习模型部署到大数据安全分析中心，用用户和实体行为分析对海量日志做基线建模，可以发现隐藏的攻击。不断自适应调整的检测模型可以减少对固定规则、签名的依赖，能够发现未知的攻击工具以及内部滥用。杀伤链中侦查阶段一般表现为异常的数据查询量以及非业务时段的访问峰谷，智能体依靠对数据访问热力图的持续监测，再加上资源访问模式聚类，可以提前发现潜在的威胁行为，并给出风险评分。模型自身要对抗数据投毒、对抗样本的干扰，用输入验证、模型鲁棒性加固、训练数据来源追溯来保证检测分支本身不会被恶意数

据流所侵蚀。安全编排自动化及响应系统把告警分析、关联扩线、封禁处置编排成自动化剧本，缩减了由于人为主动干预所造成的滞后，把重复性的响应动作由人工的分钟级缩短到机器秒级。

3.4 协同化运营与实战检验的常态闭环

单个组织不能独立地对付各种各样的威胁，行业级大数据威胁情报共享机制借助联邦学习的技术手段，在没有使用原始数据的情况下，可以把攻击特征和防护策略共享给整个生态系统内部的成员，使得风险信号能够快速地在生态内部传播。定期举行以数据窃取为想定的红蓝对抗演练，模拟内部人员越权查询、API 滥用、供应链后门等情形，测量防护链的薄弱之处，迫使防御体系处在受控压力之下暴露短板。安全度量不能只看拦截率，还要关注平均检测时间、平均响应时间、数据影响半径等实战指标，并且把它们和团队绩效强硬挂钩。把软件物料清单加入到大数据组件的准入标准当中，持续监测开源依赖的已知漏洞和许可证风险，从供应链源头削减攻击者的注入途径。漏洞管理及配置合规检查要嵌入到大数据基础平台的每日构建流水线上，基础设施即代码的思想使安全基线可以版本化、自动化地进行验证，每一次的配置变更都会经过安全合规自动扫描。

4 结语

大数据环境把信息从静态的资产变成动态的血液，安全防护的思想也应该从圈地防守转变为伴随式守护。重新塑造信任链、加密保护数据全貌、把智能分析指挥起报警的洪流、创建协同作战常态化的机制，这并不是一个个孤立的技术选择，而是一个互相联系的生存方式。当每一个比特的数据都具有了身份、标签以及安全策略的时候，每一次的访问都会被即时进行可信度量，安全就变成了大数据价值释放的坚实基础。抵近实战的防护体系不能消灭所有的风险，但是可以让系统在受到攻击之后快速缩减损失、持续运行，为数字化进程构筑起可以依靠的底线。数据、威胁不断变化的长时期内，只有把安全能力埋入架构的根基里，用自适应和弹性来取代固化的守城观念，才能使创新与保护同步发展。安全治理由一种外在的约束变成系统本身的免疫系统的时候，大数据的潜力才能得到最大的释放。

参考文献：

[1] 陈斌.基于大数据技术的计算机网络信息安全防护策略分析[J].集成电路应用,2025,42(10):138-139.

[2] 郭锐,张征.新形势下移动网络信息安全防护对策研究[J].中国宽带,2025,21(11):82-84.

[3] 张亚辉.虚拟网络技术在网络信息安全防护中的应用[J].信息记录材料,2025,26(09):191-193.

[4] 申亚楠.现代化计算机网络信息安全影响因素及防护策略研究[J].科技与创新,2025,(16):197-199.

[5] 张喜亮.大数据时代网络信息安全及防护策略研究[J].数字传媒研究,2024,41(03):77-80.