大数据分析在网络安全态势感知中的应用

周　鉴
南昌数字经济产业有限公司　江西南昌　330038

摘　要：网络安全态势感知是保障网络空间安全的核心手段，随着网络规模扩大与攻击手段复杂化，传统感知方法难以应对海量异构数据处理与精准威胁识别需求。大数据分析技术凭借海量数据处理、深度关联挖掘、智能趋势预测等优势，为网络安全态势感知提供了新的解决方案。本文构建基于大数据的网络安全态势感知体系架构，从数据采集预处理、存储计算、分析呈现三个层面阐述体系运行逻辑；系统分析大数据分析在态势理解、评估、预测各阶段的具体应用；深入探讨应用过程中面临的数据质量、分析智能水平、隐私保护等关键挑战。研究旨在为提升网络安全态势感知的精准性与时效性提供技术参考，推动大数据技术在网络安全领域的深度应用。

关键词：大数据分析；网络安全；态势感知；体系架构

引言

在数字经济快速发展背景下，网络成社会生产生活核心基础设施，其安全稳定运行关乎国家安全、经济发展与公众利益。当前，网络攻击规模化、隐蔽化、智能化，传统基于特征匹配的防护手段难应对复杂网络威胁，网络安全态势感知作为主动防御核心受行业关注。它通过采集、分析网络多源数据，掌握安全状态，识别潜在威胁，预测风险趋势，为安全决策提供支撑。随着物联网等技术普及，网络数据量爆炸式增长且类型复杂，传统数据分析技术处理海量异构数据效率低、精度不足，制约态势感知能力提升。大数据分析技术可突破局限，快速处理与深度挖掘海量多源数据，提取有价值安全信息。将其应用于网络安全态势感知，能提升威胁识别精准度、态势评估全面性与风险预测前瞻性。因此，深入研究大数据分析在网络安全态势感知中的应用，构建科学架构，解决关键挑战，对提升网络安全防御能力、保障数字经济健康发展有重要现实意义。

一、基于大数据的网络安全态势感知体系架构

（一）多源异构安全数据的采集与预处理层

多源异构安全数据的采集与预处理是态势感知体系的基础，该层的核心目标是全面收集网络中的安全相关数据，并通过清洗、转换、整合等操作，为后续分析提供高质量数据支撑。数据采集范围覆盖网络基础设施、终端设备、应用系统等多个层面，采集的数据类型包括网络流量数据、日志数据、漏洞数据、威胁情报数据等。网络流量数据通过流量采集设备获取，包含数据包的源地址、目的地址、端口号、协议类型等信息，能够反映网络通信状态；日志数据来自路由器、防火墙、服务器、终端等设备，记录设备的运行状态、操作行为、异常事件等内容；漏洞数据包括网络设备与应用系统的已知漏洞信息，为威胁识别提供依据；威胁情报数据来自安全厂商、科研机构等第三方平台，包含最新的攻击手段、恶意代码特征等信息。

数据预处理环节针对采集数据的异构性、冗余性、噪声等问题，开展一系列数据治理操作。数据清洗去除重复数据、缺失数据与异常数据，避免无效数据对分析结果的干扰；数据转换将不同格式的数据统一转换为标准化格式，实现数据的互通互用；数据整合按照一定规则将来自不同数据源的数据进行关联组合，形成结构化的数据集。通过采集与预处理，将分散、杂乱的多源数据转化为统一、规范、高质量的数据资源，为后续的存储计算与分析呈现奠定基础[1]。

（二）面向态势理解的数据存储与计算层

数据存储与计算层是态势感知体系的核心支撑，负责对预处理后的海量数据进行高效存储与并行计算，为态势理解提供强大的算力支持。该层采用分布式存储架构，结合关系型数据库、非关系型数据库、数据仓库等多种存储技术，满足不同类型数据的存储需求。关系型数据库用于存储结构化数据，如设备基本信息、漏洞详情等，具有数据一致性高、查询便捷的优势；非关系型数据库适用于存储半结构化与非结构化数据，如日志数据、流量原始数据等，能够应对海量数据的快速写入与灵活查询；数据仓库用于整合多源数据，构建面向态势

分析的主题数据模型，支持复杂的数据分析与统计查询。

计算层采用分布式计算框架，具备强大的并行计算能力，能够快速处理海量数据。通过MapReduce、Spark等计算模型，将大规模数据处理任务分解为多个子任务，分配到不同的计算节点上并行执行，大幅提升数据处理效率。同时，结合流计算技术，实现对实时数据的动态处理，能够及时捕捉网络中的安全事件与异常变化，为态势的实时感知提供保障。

（三）支撑态势评估与预测的分析与呈现层

分析与呈现层是态势感知体系的核心应用层，负责对存储计算层输出的数据进行深度分析，并以直观的方式呈现态势评估与预测结果，为安全决策提供支持。分析环节融合大数据挖掘、机器学习、统计分析等多种技术，实现对网络安全态势的全面理解、精准评估与科学预测。通过关联分析挖掘不同数据之间的潜在联系，识别隐藏在数据背后的安全威胁；通过聚类分析将具有相似特征的数据归类，发现网络中的异常行为模式；通过机器学习算法构建预测模型，基于历史数据预测未来的安全风险演化趋势。

呈现环节采用可视化技术，将分析结果以图表、仪表盘、拓扑图等形式直观展示，使安全管理人员能够快速掌握网络安全态势。可视化内容包括网络安全总体态势、异常事件分布、威胁类型统计、风险等级评估等。网络安全总体态势以仪表盘形式展示关键安全指标，如异常事件数量、高风险漏洞数量、攻击次数等；异常事件分布通过地图或拓扑图呈现异常事件的发生位置与影响范围；威胁类型统计以柱状图、饼图等形式展示不同类型威胁的占比情况；风险等级评估通过颜色标识等方式直观呈现网络各区域、各设备的风险等级[2]。

二、大数据分析在态势感知各阶段的应用

（一）在态势理解阶段的数据融合与关联分析

态势理解阶段的核心任务是整合多源数据，挖掘数据之间的关联关系，明确网络安全事件的本质与上下文信息，实现对网络安全状态的全面认知。大数据分析技术在该阶段的核心应用是数据融合与关联分析，通过融合多源异构数据，打破数据壁垒，形成对安全事件的完整刻画；通过关联分析挖掘不同数据之间的因果关系、时序关系、空间关系等，揭示安全事件的内在逻辑。

数据融合技术整合不同数据源信息，消除数据不确定性与歧义性，提升信息可靠性与完整性。如融合网络流量与日志数据，结合异常通信特征与设备操作记录，准确判断恶意攻击；融合威胁情报与漏洞数据，依据攻击特征与漏洞信息，识别潜在威胁。关联分析技术构建关联规则，挖掘数据隐藏关系。如分析日志用户登录与网络流量数据传输行为，发现异常登录与恶意传输关联，识别账号被盗攻击；分析不同设备异常事件时间与特征，发现分布式攻击协同规律。

（二）在态势评估阶段的异常检测与攻击溯源

态势评估阶段的核心任务是对网络安全状态进行量化评估，确定威胁的严重程度与影响范围，为安全响应提供决策支持。大数据分析技术在该阶段的主要应用包括异常检测与攻击溯源，通过异常检测及时发现网络中的偏离正常行为的异常事件，通过攻击溯源明确攻击的来源、路径与目的，为威胁评估提供关键信息。

异常检测技术基于大数据分析构建正常行为模型，对比实际行为与正常模型差异识别异常事件。用机器学习算法训练历史数据，学习网络、用户、设备的正常模式，实际行为偏离正常模型阈值超设定范围则判定为异常。如分析网络流量建正常波动模型，流量异常时系统报警；分析用户登录等数据建正常行为模型，出现异常登录则识别为异常事件。攻击溯源技术分析多源数据，追踪攻击源头与传播路径。结合网络流量、日志、威胁情报数据构建攻击溯源图谱，明确攻击发起者、工具方法、目标与路径[3]。

（三）在态势预测阶段的风险演化与威胁预警

态势预测阶段的核心任务是基于历史数据与当前态势，预测未来网络安全风险的演化趋势，提前发出威胁预警，为主动防御提供时间窗口。大数据分析技术在该阶段的应用主要体现为风险演化分析与威胁预警，通过挖掘历史数据中的规律，构建预测模型，实现对未来安全态势的科学预测。

风险演化分析通过分析历史安全事件的时间、范围、路径等数据，挖掘安全风险演化规律。结合网络拓扑、设备配置、漏洞生命周期等因素，构建风险演化模型，预测安全风险传播路径与影响范围。如基于历史漏洞利用数据，分析漏洞利用周期，结合未修复漏洞分布，预测未来漏洞攻击范围与强度；分析历史DDoS攻击规律，结合流量趋势，预测潜在DDoS攻击风险。威胁预警基于风险演化分析结果和预设阈值，及时发出预警信息，明确威胁类型、风险等级、受影响设备系统及预计发生时间，为安全管理人员提供防御指引。

三、应用大数据分析面临的关键挑战

（一）数据质量与处理效率的挑战

数据质量与处理效率是大数据分析在网络安全态势

感知应用中面临的首要挑战。网络环境中采集的数据具有海量性、异构性、动态性等特点，数据质量难以保证。部分设备因硬件故障或网络中断，导致数据采集不完整，出现缺失值；不同设备的日志格式不统一，数据标准不一致，造成数据异构性问题；网络攻击手段的隐蔽化导致部分异常数据被噪声数据掩盖，难以识别。低质量的数据会直接影响分析结果的准确性，导致态势感知出现误判或漏判。

同时，海量数据的快速增长对数据处理效率提出了更高要求。随着网络规模的扩大与设备数量的增加，安全数据量呈指数级增长，传统的数据处理技术难以在有效时间内完成海量数据的处理与分析。特别是在实时态势感知场景中，需要对实时产生的大量数据进行快速处理，及时发现威胁并发出预警，而数据处理效率不足会导致态势感知滞后，无法及时响应安全事件，错失最佳防御时机[4]。

（二）分析深度与智能水平的挑战

分析深度与智能水平不足是制约态势感知能力提升的重要挑战。当前，大数据分析在态势感知中的应用多集中于数据的浅层关联与统计分析，缺乏对复杂安全事件的深度挖掘与智能推理。网络攻击手段的智能化与隐蔽化趋势日益明显，攻击者通过伪造正常通信特征、利用零日漏洞等方式规避检测，传统的基于规则与特征的分析方法难以识别此类复杂攻击。

现有分析模型的智能学习能力有限，难以适应网络安全态势的动态变化。网络安全环境处于不断变化之中，新的攻击手段、漏洞与恶意代码不断涌现，分析模型需要具备持续学习能力，及时更新分析规则与特征库。但当前的分析模型多为静态模型，缺乏自适应调整能力，无法快速响应新的安全威胁。此外，不同安全事件之间的关联关系日益复杂，单一维度的分析难以全面把握事件的本质，需要从多维度、多层次进行深度分析与推理，而现有分析技术在多维度关联推理方面还存在不足。

（三）隐私保护与数据安全的挑战

隐私保护与数据安全是大数据分析在态势感知应用中面临的伦理与安全挑战。态势感知体系采集的网络数据中包含大量的用户隐私信息与企业敏感数据，如用户的个人身份信息、通信内容、企业的商业数据等。在数据采集、存储、分析与共享过程中，若缺乏有效的安全防护措施，可能导致隐私信息泄露与敏感数据被盗，引发严重的安全风险与法律责任。

一方面，数据集中存储增加了数据泄露的风险。态势感知体系将海量的多源数据集中存储在数据中心，一旦数据中心遭受攻击或出现内部人员泄露，将导致大量敏感数据泄露；另一方面，数据共享过程中的安全防护不足也会引发隐私问题。为提升态势感知能力，不同组织之间需要共享威胁情报、安全数据等资源，但数据共享过程中缺乏统一的安全规范与隐私保护机制，可能导致数据被滥用或泄露[5]。

结语

大数据分析为网络安全态势感知提供技术支撑，构建涵盖数据采集预处理、存储计算、分析呈现的体系架构，实现对网络安全态势的全面感知、评估与预测。其在态势理解、评估、预测各阶段的应用提升了态势感知深度与广度，支持网络安全主动防御。但大数据分析在态势感知应用中面临数据质量与处理效率、分析深度与智能水平、隐私保护与数据安全等挑战，制约态势感知能力提升。未来，需推进技术创新与实践探索，优化数据采集与预处理技术，提升数据质量与效率；融合人工智能等先进技术，增强分析模型智能学习与推理能力；建立隐私保护与数据安全机制，保障数据应用安全合规。解决关键挑战，发挥大数据分析优势，提升网络安全态势感知能力，保障网络空间安全，推动数字经济健康发展。

参考文献

[1]徐航，张冬冬.大数据技术在网络安全分析中的应用[J].计算机与网络，2022，40（1）：240-242.

[2]周金全，朱世伟，张建平.基于大数据和人工智能的网络安全态势分析方法研究[J].2022.

[3]王荣汉，彭添焕.分析大数据驱动网络安全的机遇与挑战[J].电子测试，2020（20）：3.

[4]郑磊，韩鹏军.大数据技术在网络安全分析中的应用研究[J].信息技术，2021.DOI：10.13274/j.cnki.hdzj.2021.01.029.

[5]田艳超.应用大数据分析的无人船通信网络安全态势识别模型[J].舰船科学技术，2021（22）：61-63.