 DOI: https://doi.org/10.12345/bdai.v5i9.22281

Application of Big Data Technology in Network Security Analysis

Shengcheng Xu Jianwei Lv*

Shanghai Maoyuan Shengyou Information Technology Co., Ltd., Shanghai, 200001, China

Abstract

With the rapid development of information technology, network security issues have become increasingly complex, and traditional security measures are no longer able to cope with the ever-changing and complex security threats in modern network environments. The paper investigates the specific application of big data technology in network security analysis, and studies its role in threat detection, risk identification, event response, and other aspects. It analyzes the application of real-time streaming data processing and machine learning technology in threat detection, and mentions the importance of big data technology in improving network security event response speed and abnormal behavior analysis. The paper aims to demonstrate the crucial contribution of big data technology to modern network security analysis and provide reference for research and practice in related fields.

Keywords

big data technology; network security; threat detection; event response

网络安全分析中的大数据技术应用

许盛诚 吕建伟 *

上海茂源盛悠信息技术有限公司，中国·上海 200001

摘 要

随着信息技术的飞速发展，网络安全问题变得日益复杂，传统的安全防护措施已无法应对现代网络环境中多变且复杂的安全威胁。论文结合大数据技术在网络安全分析中的具体应用进行了研究，对其在威胁检测、风险识别、事件响应等多个方面的作用进行了研究，分析了实时流数据处理和机器学习技术在威胁检测中的应用，提到了大数据技术在提升网络安全事件响应速度及异常行为分析中的重要性。论文旨在展示大数据技术对现代网络安全分析的关键性贡献，并为相关领域的研究与实践提供参考。

关键词

大数据技术；网络安全；威胁检测；事件响应

1 引言

网络攻击的规模和频率在不断增加，黑客手段也变得更加高级和隐蔽，导致大量企业和机构频繁遭受网络攻击。随着互联网设备和数据量的爆炸性增长，网络攻击的目标不再局限于单一系统，而是覆盖到整个网络生态系统，这给网络安全带来了前所未有的挑战。面对海量的安全事件和数据，传统的网络安全分析方法已经难以满足高效、精准的需求，单纯依靠规则和签名库的防护策略在面对复杂的高级威胁时显得力不从心。大数据技术因其在处理、分析和挖掘海量数据中的独特优势，逐渐成为网络安全领域的新兴解决方案。

【作者简介】许盛诚（1981–），男，中国上海人，硕士，工程师，从事计算机系统集成、应用软件研发、网络应用研发、信息安全等研究。

【通讯作者】吕建伟（1986–），男，中国甘肃白银人，本科，工程师，从事信息系统集成研究。

2 大数据技术在网络安全分析中的总体应用

2.1 大数据技术的基础概念与网络安全的结合

大数据技术对大规模的多样化和快速生成的数据进行有效处理，并能从中提取出有价值的信息，从而在网络安全领域为安全分析不仅依赖预设的规则和签名库，而且能够通过对历史数据和实时数据的深入分析来发现潜在的威胁和异常活动，而传统的安全技术往往由于数据规模大难以有效工作，而大数据技术在分布式存储和并行计算的基础上，对海量数据能够在短时间内进行有效处理。所以具有很大的应用价值 [1]。对数据处理能力的增强为网络安全带来了新的解决思路，即借助大数据分析来追溯历史攻击模式并对潜在攻击进行预测和预警，从而帮助安全团队在尚未发生攻击的情况下，为防患于未然而采取相应措施。这样，安全团队就可以提前发现潜在的威胁而防患于未然。

2.2 网络安全中的数据来源及处理方式

网络安全中的数据来源涵盖了从网络流量、日志文件到设备通信等。随着物联网设备的普及和云计算的广泛应用，数据的产生速度和规模大幅增加，给传统的安全系统带来了前所未有的挑战。这些数据来源于不同的网络层次和设备类型，包括防火墙日志、入侵检测系统（IDS）数据、端点安全日志、操作系统和应用程序日志、网络流量数据以及来自社交媒体和暗网的信息等。为了从这些庞大的数据源中获取有用的安全信息，大数据技术的处理方式依赖于分布式数据存储、实时流数据处理和数据清洗技术。通过这些处理方式，能够对来自不同来源的异构数据进行快速的整合与分析，使得网络安全团队能够从多维度的安全信息中发现潜在的威胁和风险。

2.3 大数据分析对网络安全风险识别的作用

对历史资料用户行为模式网络流量进行深度挖掘的大数据分析技术可以帮助企业和组织提前发现和预测潜在安全威胁，尤其是高级持续性威胁和各种复杂网络攻击。传统的规则和签名匹配方式往往容易受到绕过，而大数据分析能够识别那些隐藏在正常活动下的潜在威胁。基于行为分析模型的大数据分析，使安全团队能够对网络中不符合常规模式的活动进行检测，并将其标记为异常行为，从而实现对未知威胁的早期预警，从而在安全事件发生之前采取相应的应对措施 [2]。大数据技术对多维度的安全信息进行综合分析后，能形成较为全面的威胁情报图谱，既对当前可能存在的风险进行识别，又能对今后可能发生的攻击行为进行预测，从而为企业的安全防护提供更为精细和精准的决策支撑，使企业有效降低了遭受网络攻击的可能性和损失，提高了安全防范能力，增强了企业的市场竞争力。

3 实时威胁检测中的大数据应用

3.1 实时流数据处理技术的应用

每秒钟所产生的大量网络流量数据，随着网络环境的复杂化，要求系统实时处理的能力非常高效。传统的数据处理方式往往依赖于批量处理模式，无法满足实时威胁检测的需求，而批量处理模式又不能满足批量处理模式的需求。大数据技术中的流数据处理技术提供了新的思路，可以通过流处理框架对网络流量进行实时监控和分析，如 Apache Kafka、Apache Storm 和 Apache Flink 等。而不是等批量数据积累后再去分析，这一技术的核心在于数据到了就马上处理。Flow Data Processing 技术将网络流量分为连续的数据流，在不间断的分析过程中可以及时发现异常活动并采取应对措施。

流数据处理技术为了提高处理的精确度，经常结合大数据分析，利用各种数据源的数据进行交叉验证，进行关联分析。例如，在侦测潜在的网路攻击时，将网络日志、使用者行为日志、装置通信记录等各种资料来源结合起来，就可以综合判断流处理系统。通过伪造数据或隐藏攻击路径来避免攻击者逃避检测，这种多层次的数据交互和处理模式可以有效提高威胁检测的准确性和时效性。大数据的实时流数据处理技术不仅在威胁检测中发挥了重要作用，而且在事件响应和恢复过程中起到了决定性作用，能够在攻击发生的第一时间采取应对措施，缓解潜在的系统受损情况，因此，大数据的实时流数据处理技术

3.2 大数据与机器学习在威胁检测中的融合

机器学习依赖于大规模的数据样本式训练模型，而大数据技术能够提供广泛的数据源和充足的训练数据，在这些数据的基础上，机器学习能够提供攻击模式和威胁特征在网络安全领域具有高度的多样性和复杂性，对于未知的攻击，传统的基于规则的威胁侦测手段往往是无法应对的。并且通过大数据技术，在海量的网络数据中，机器学习模型能够自动学习并识别出新的攻击方式。基于这种融合，通过异常检测和分类模型，机器学习算法能够对潜在威胁进行实时识别。例如，网络中的正常行为可以通过无监督学习模型的使用来进行建模，任何活动与此行为模型相背离都有可能被看作是一种威胁 [3]。

传统的防护措施依赖于对规则库进行不断的更新，而现代以大数据和机器学习为基础的威胁检测系统则可以自主更新模型，基于实时数据对新的攻击模式进行训练和调整，并对实际检测到的攻击结果进行反馈和调整，以逐步优化检测规则和策略，从而在不依赖人为干预的情况下增强对新型攻击的防御能力。以增强学习模型为基础，系统在根据实际检测到的攻击结果进行反馈和调整的同时，对检测规则和策略进行不断优化，使网络安全系统对新型攻击的防御能力得到不断提高。基于大数据和机器学习的自适应威胁检测系统在为现代网络安全带来高度智能化和动态化的防护能力的同时，也极大地增强了对未知威胁的响应速度和精确性，在保障网络安全方面发挥着重要的作用。

3.3 案例分析：利用大数据技术预防 DDoS 攻击

美国东海岸地区 2016 年爆发大规模 DDoS 攻击，主要针对 DNS 服务商 Dyn，导致大量互联网服务中断，波及多家知名网站，包括 Twitter、Spotify 和 Netflix。攻击峰值流量达到 1.2Tbps，超越常规防护系统的承载能力。此次攻击事件暴露出面对大型 DDoS 攻击时传统网络防护手段的脆弱，而应用大数据技术展现出的强大侦测和防御能力则在此类攻击中一览无余。

企业为了应对类似大规模 DDoS 攻击开始引入大数据技术，通过分析和处理海量的网络流量数据，提早发现异常流量和攻击迹象。在本案例中，Dyn 公司引入了大数据平台，用于分析每秒产生的数 TB 级别的网络流量数据。通过实时监控流量，系统能够自动识别流量峰值的异常，并通过比对历史流量模式，判断是否存在 DDoS 攻击的可能。表 1 展示了攻击前后的网络流量变化情况。

表 1 攻击前后的网络流量变化情况

时间段	正常流量（Gbps）	攻击流量（Gbps）
10:00 AM—11:00 AM	400	1200
11:00 AM—12:00 PM	450	1350
12:00 PM—01:00 PM	430	1500

此次攻击爆发前，网络流量攀升迅速，远超正常范围，而通过实时和历史数据对比，大数据分析系统成功识别异常流量。系统判定这是一次 DDoS 攻击，依据流量峰值增长模式和 IP 地址分布，立即触发包括流量调度、流量限制措施在内的相应防御策略，有效缓解了部分业务的中断。

4 网络安全事件响应与大数据

4.1 大数据对安全事件响应速度的提升

数据处理的并行化和自动化分析能力主要体现在大数据技术的应用上。在海量数据面前往往有很大的滞后性，很容易错过事件处理的最佳时间窗，传统的安全事件反应依赖于人工的分析判断。大数据技术则可以在短时间内通过分布式数据处理架构快速整合分析海量网络日志、流量数据和设备记录等数据。Hadoop 和 Spark 等分布式计算系统可以在数分钟内处理数百 TB 的数据量，从而使安全团队可以对潜在威胁的全貌进行实时获取，从而使事件响应时间大大缩短。

事件响应速度的提高还表现在事件处理的后续步骤中，大数据系统能够迅速定位安全事件的根本原因，并通过数据关联性分析帮助重建事件发生过程，如分析网络流量用户行为和日志文件之间的关联性，快速识别攻击者的入侵路径使用的漏洞和攻击目标等，与传统手动分析相比，大数据技术能够在几分钟内完成复杂的分析任务，使安全团队能够更快地制定应对方案并及时堵住安全漏洞，防止损害的进一步扩散。通过大数据技术的运用，使安全事件的处理速度得到提高，为保护信息安全保驾护航。

4.2 异常行为分析在事件响应中的重要性

许多高级的网络攻击往往表现为异常的用户行为或网络活动，而并非明显的攻击行为。APT（高级持续性威胁）攻击的手法通常是通过长期潜伏和隐蔽渗透，逐步取得系统控制权，而不会立即进行破坏性操作 [4]。在这种情况下，传统的基于签名和规则的检测方法难以识别这些隐蔽的威胁。大数据技术使得异常行为分析成为可能，通过对大量正常行为模式的数据进行长期观察和建模，能够自动发现那些偏离常规模式的异常行为。

异常行为分析对安全团队来说是发现潜在威胁的有力武器，是为避免攻击者长时间潜伏而对系统造成破坏而采取的一项重要措施。例如，对某个用户的登录行为数据访问模式与其历史记录的偏差进行分析，大数据系统就可以识别出可能是账户劫持或内网攻击等可能的情况。另外，异常行为分析在事件响应的后期还可以对攻击源头及入侵路径进行快速定位和回溯分析，从而发现攻击者如何突破安全防线并渗透到系统内部的痕迹。通过分析行为异常发生时间点与网络活动的关联性，可以对攻击者的攻击方式及路径有一个比较清晰的认识。

4.3 案例分析：大数据在 APT 攻击溯源中的应用

知名安全公司 FireEye2017 年通过大数据技术成功溯源了一起持续近一年的 APT 攻击，该攻击针对一家全球性金融机构，攻击者通过复杂的社会工程学手段获取了该机构部分员工的账户凭据，并利用这些凭据进行长时间的隐蔽入侵。攻击者的目标是窃取该机构的敏感金融数据并获取对该机构内部网络的长期访问权限。FireEye 的安全团队利用大数据技术对数 TB 级的网络流量、用户行为日志以及系统操作记录进行了详细分析，最终成功定位了攻击者的活动路径和攻击时间线。

在此次溯源过程中，FireEye 采用了大数据异常行为分析技术，通过对该机构网络中数千个用户的正常行为模式进行建模，分析出了多个异常行为指标。某员工账户在多个时区的同时登录行为被识别为异常，且该账户在被盗用后访问了与其日常工作无关的敏感系统。表 2 展示了部分异常行为分析的数据。

表 2 PT 攻击溯源中的异常行为分析数据

用户 ID	异常登录时间	异常访问系统	异常行为评分
user1234	3:45 AM	金融核心数据库	87
user5678	2:30 AM	内部邮件系统	91
user9012	4:15 AM	研发资料库	85

FireEye 的团队通过大数据技术不仅发现了这些异常行为，还通过关联分析确定了攻击者使用的多条渗透路径。攻击者在长达近一年的时间内，利用被窃取的多个账户凭据，不断扩大对该机构内部网络的控制范围。

5 结论

在现代复杂的网络环境下，应对日益复杂的威胁，传统的安全防护手段已显得力不从心，而大数据技术的引入则提供了安全防护的全新解决方案，尤其是在应对 DDoS 攻击、APT 攻击等高级威胁方面优势明显。企业可以通过对大量异构数据的整合分析，对潜在威胁进行预先识别，并做出快速反应，从而有效降低损失。大数据技术将在帮助企业应对更复杂的安全挑战的未来网络安全防护中扮演越来越重要的角色。

参考文献

[1] 赵鹏.人工智能和大数据技术在计算机网络安全防御中的运用[J].通讯世界,2024,31(9):46-48.

[2] 曹纪磊.基于大数据技术的计算机网络安全态势感知透析[J].网络空间安全,2024,15(4):244-247.

[3] 崔馨月.基于大数据区块链的网络安全防护系统研究[J].网络空间安全,2024,15(4):248-251.

[4] 谌颂,张袖斌,肖斌.基于大数据的网络安全态势感知技术探讨[J].网络空间安全,2024,15(4):240-243.