 DOI: https://doi.org/10.12349/iser.v6i11.8118

Research on Computer Network Security Situation Awareness Model Based on Big Data

Taiwen Xiao

Digital Government Operation Center, Haizhu District, Guangzhou, Guangdong, 510000, China

Abstract

In the context of the information age, computer networks have widely penetrated into all levels of society and become a key channel for information transmission. However, with the continuous development of network technology, network security issues have become increasingly prominent, and network attacks and security incidents occur frequently. In order to effectively solve these problems, timely and accurate monitoring and management of network security status is required. This study develops a network security situational awareness model utilizing big data technology. This model analyzes and processes massive network data to achieve real-time monitoring of network status and prediction of potential threats, providing solid technical support for network security defense. The model integrates cutting-edge data mining and machine learning technologies to conduct a comprehensive analysis of various data sources such as network traffic, log files, vulnerability information, etc., and establishes a multi-level

Keywords

Big data; Network security; Situational awareness; Data mining; Machine learning

基于大数据的计算机网络安全态势感知模型研究

萧太文

广州市海珠区数字政府运营中心，中国 · 广东 广州 510000

摘　要

在信息化时代背景下，计算机网络已广泛渗透至社会各个层面，成为信息传递的关键途径。但是，随着网络科技的不断发展，网络安全问题日益突出，网络攻击和安全事件频繁发生。为了有效解决这些难题，需要对网络安全状态进行及时、准确的监控和管理。本研究开发了一种利用大数据技术的网络安全态势感知模型。该模型通过分析处理海量网络数据，实现对网络状态的实时监控以及对潜在威胁的预测，为网络安全防御提供了坚实的技术支撑。模型整合了前沿的数据挖掘与机器学习技术，对网络流量、日志文件、漏洞信息等多样数据源进行全面分析，建立了一个多层次、多维度的网络安全态势感知框架。为检验模型的效能，此项研究以某一中职学校校园网为案例，执行了实验分析。结果显示，该模型大幅提升了网络安全态势感知的精确性和实时性，为网络安全管理供给高效的技术工具。

关键词

大数据；网络安全；态势感知；数据挖掘；机器学习

1 引言

在信息技术的促进下，计算机网络已成为现代社会生活的关键部分。然而，网络环境的开放性和复杂性也带来了许多安全问题，如网络攻击、数据泄露、恶意软件等。这些安全问题不仅危及个人隐私和财产安全，也对国家安全和社会稳定构成了严峻考验。为有效应对这些难题，我们需要对网络环境的安全状况进行即时、精确的感知和管理。网络安全态势感知正是此类技术，它通过即时监控、解析和预测网络环境中的安全事件和威胁，帮助我们及早发现和应对潜在的安全威胁。

随着网络攻击方式的不断演变和网络规模的持续扩展，传统的网络安全防护技术已难以适应当前的安全需求。传统的安全防护技术常依赖于静态的防护策略和规则，无法应对动态变化的网络环境和多元化的攻击方式。因此，探索基于大数据的网络安全态势感知模型，对于提高网络安全防护水平具有关键意义。大数据技术的进步为网络安全态势感知带来了新的机会。通过解析处理巨量的网络数据，我们可以更全面、深入地掌握网络环境的安全状况，从而制定更高效的安全防护策略。

【作者简介】萧太文（1988–），男，中国广东茂名人，硕士，工程师，从事计算机技术、网络工程、电子与通信工程研究。

2 相关工作

态势感知的理念最早由 Endsley 在 1988 年提出，其核心理念是通过感测环境、认知环境和预判环境来实现对态势

的全面理解。在网络安全领域，态势感知主要关注网络环境中的安全事件和威胁。近年来，随着大数据技术的发展，研究人员开始研究如何利用大数据分析技术增强网络安全态势感知的能力。一些研究已经获得了一定成果，譬如基于机器学习的异常检测算法、基于数据挖掘的攻击模式识别方法等。这些研究为本文提出的模型提供了关键参考和学习。

3 网络安全态势感知模型

3.1 模型架构

本文提出的网络安全态势感知模型主要包含数据采集层、数据处理层、态势分析层和决策支持层四个部分。数据采集层用于收集网络中的各类信息，包括网络流量、系统日志、安全日志等。这些数据构成模型进行分析的根基，也是模型能够精确感知网络安全态势的关键要素。数据处理层对采集的信息进行清理、转换和储存。由于网络数据通常具备庞大、复杂和异构等特性，因此数据处理层需采用高效的数据处理手段来确保数据的品质和可用性。态势分析层借助数据挖掘和机器学习技术对数据展开分析，抽取安全态势特征。这些特征映射出网络环境的安全状态，是模型执行态势感知的核心环节。决策支持层依据分析结果给出安全决策建议。这些建议可协助网络安全管理人员制定高效的安全防御策略，提升网络安全防护水准。

3.2 关键技术

3.2.1 数据采集技术

采集技术。数据采集技术主要包括网络流量捕获、日志收集和漏洞扫描等。

网络流量捕获技术的核心功能通过抓取网络接口数据包实现全流量镜像，支持从物理层到应用层的协议解析（如 HTTP/HTTPS、TCP/IP、工业 Modbus 等），技术优势是实时捕获与离线分析结合，支持毫秒级延迟处理，深度包解析（DPI）能力可识别加密流量特征，典型应用：异常流量检测、网络性能优化、工业控制协议监控

日志收集技术支持操作系统日志（Syslog）、应用日志（如 Apache/Nginx）、数据库审计日志的统一采集，关键技术：分布式架构（如 Fluentd、Logstash）实现高并发日志处理；

日志归一化技术将异构数据转为结构化格式，场景案例：安全事件溯源（如通过登录日志分析暴力破解行为）

漏洞扫描技术具备动态检测机制，主动探测目标系统端口与服务（Nessus 支持 65,000+ 漏洞检测项），结合 CVSS 评分量化风险等级，输出修复建议，扩展能力：支持与 CMDB 集成，实现资产 - 漏洞关联分析

采集工具。本文采用开源工具如 Wireshark、Syslog 和 Nessus 进行数据采集。这些工具具有功能强大、使用方便等特点，可以满足模型对数据采集的需求。

表 1 开源工具的技术特点与选型依据

工具	核心能力	适用场景	性能指标
Wireshark	支持 1,000+ 协议解析，实时流量可视化；过滤表达式（如 tcp.port==80）精准定位数据包	网络故障诊断、工业协议分析	10Gbps 网络吞吐量
Syslog	轻量级日志转发协议，支持 RFC 5424 标准；与 Rsyslog/Syslog-ng 集成实现日志聚合	集中化日志管理、合规审计	单节点 10 万条 / 秒处理
Nessus	插件化漏洞库（含 0day 漏洞检测），支持 SCAP 合规检查	系统安全基线验证、渗透测试	分钟级全网扫描

3.2.2 数据处理技术

数据管理技术包含数据清理、数据整合和数据保存等。本研究使用 Hadoop 生态系统中的 HDFS 和 MapReduce 进行海量数据处理。HDFS 是一种分散式文件系统，能够存储大量数据；MapReduce 是一种并行计算架构，可以高效处理分散数据。

3.2.3 态势分析技术

态势分析技术涵盖数据挖掘和机器学习方法，例如聚类方法、分类技术和关联规则发现等。本论文应用支持向量机 (SVM)、随机森林和神经网络等算法执行安全事件的分类与预测。这些算法具备良好性能和稳健性，能高效地分析网络信息并抽取安全态势特征。

态势分析技术通过整合多源异构数据（网络流量、日志、漏洞扫描结果等），结合数据挖掘与机器学习算法，实现对安全威胁的动态感知与预测。其核心流程包括：

数据层：整合实时流量（Wireshark 捕获）、系统日志（Syslog）、漏洞扫描结果（Nessus）等，形成多维度数据池。

特征层：关联规则挖掘：发现事件间隐含模式（如“端口扫描常伴随 SQL 注入攻击”）。

特征降维：通过主成分分析（PCA）处理高维数据，提升模型效率。

算法层：采用分类、聚类、回归等模型，实现安全事件标签化与风险量化。

表 2 算法性能对比

算法	适用场景	优势	局限性
SVM	小样本、高维数据分类	边界清晰、抗噪声能力强	计算复杂度随数据量激增
随机森林	大规模实时流量分析	并行处理高效、支持特征自动选择	模型解释性较弱
神经网络	复杂攻击模式识别	自适应学习、非线性关系建模能力强	依赖大量标注数据训练

以 SVM、随机森林和神经网络为核心的分析技术，通过多算法协同与动态优化机制，显著提升安全态势感知的精度与时效性。

4 实验设计与结果分析

4.1 实验环境

实验环境包括一台服务器和若干客户端，服务器配置为 Intel Xeon CPU E5-2620 v4，32GB 内存，客户端配置为 Intel Core i5 处理器，8GB 内存。操作系统为 CentOS 7.2，大数据处理框架为 Hadoop 2.7.3。

4.2 实验案例

实验案例：某中职学校网络安全态势感知

实验时间：2023 年 5 月

实验地点：某中职学校网络中心

实验对象：该中职学校的校园网

实验目的：验证基于大数据的网络安全态势感知模型的有效性

4.3 数据采集

在该中职学校的校园网中，部署了 Wireshark 进行网络流量捕获，Syslog 收集系统日志和安全日志，Nessus 进行漏洞扫描。数据采集持续时间为一周，共收集到网络流量数据 1TB，系统日志和安全日志 500GB，漏洞扫描报告 100 份。

4.4 数据处理

应用 Hadoop 生态系统中的 HDFS 保存收集的数据，并借助 MapReduce 执行数据清理和格式处理。数据清理包含删除冗余数据、应对空缺值和离群值等。格式处理使源数据转变成适宜机器学习算法处理的格式。

4.5 态势分析

在处理安全事件的分类问题时，我们采用了支持向量机（SVM）这一强大的机器学习算法。通过精心设计的训练数据集，我们对 SVM 模型进行了细致的训练，以期达到最佳的分类效果。此外，为了深入挖掘安全事件之间的关联规则，我们利用了随机森林算法，它以其出色的性能和稳定性在处理此类问题上表现出色。同时，为了对未来的安全态势进行预测，我们采用了神经网络算法，这种算法能够通过学习历史数据来预测未来趋势。在模型训练完成后，我们使用独立的测试数据集对模型的性能进行了全面的评估，以确保模型的准确性和可靠性。

表 3 模型性能评估指标

评估指标	准确率	召回率	F1 分数
SVM 分类器	96%	94%	95%
随机森林关联规则挖掘	92%	90%	91%
神经网络预测	95%	93%	94%

4.6 决策支持

基于对当前安全态势的深入分析和评估，我们提出了一系列针对性的安全决策建议，旨在帮助组织机构更好地防范和应对潜在的网络威胁。例如，在面对日益频繁的分布式拒绝服务（DDoS）攻击时，我们建议采取加强网络带宽监控的措施，以便及时发现异常流量并采取相应措施。另外，通过应用高效的流量过滤方案，可以有效降低有害流量对网络资源的消耗，确保网络服务的稳定运行。对于数据库安全方面，我们着重强调了对 SQL 注入漏洞的防御。推荐对数据库进行全面安全加固，例如升级安全更新、设置访问控制列表以及执行周期性的安全审核。通过这些方法，可以大幅减少数据库受到 SQL 注入攻击的威胁，保障数据的完整性和安全。

4.7 结果分析

基于实验数据的详尽分析，我们能够清楚地观察到，本研究所开发的模型在多个关键性能指标上均体现了优于传统方法的优点。详细而言，该模型在精度、召回率以及 F1 得分这三个重要的评估指标上均实现了显著改进。在精度方面，模型的性能超过了 95%，这显示模型在检测网络攻击行为时拥有极高准确性。另外，模型的情境感知能力也获得证实，其精度达到 95% 以上，这表明模型不仅能精确检测当前网络威胁，还能预测潜在攻击行为，因此为网络安全防护提供强大支持。经由一系列实际案例的检验，我们进一步证实了该模型在现实应用中的效力。它能实时监控网络状态，迅速识别并应对各类潜在安全风险，因此大幅提升网络安全防护的整体效能。

表 4 模型在实际案例中的表现

安全事件类型	检测数量	实际数量	误报率	漏报率
DDoS 攻击	50	50	0%	0%
SQL 注入	30	30	0%	0%
恶意软件	20	21	4.76%	4.76%

注：漏报率 = 漏报数量 / 实际数量；误报率 = 误报数量 / 检测数量

5 结语

本文所阐述的依托于大数据技术的计算机网络安全态势感知模型，通过整合应用数据收集、数据整理、数据开采以及机器学习等多种前沿技术，有效达成了对网络安全态势的高效感知和解析。在测试阶段，我们选择了一所中等职业学校作为研究案例，通过多个实例案例对该模型的实用性进行了证实。测试结果明确显示，该模型在提高网络安全态势感知的精确度和即时性方面体现了明显的优点和潜力。未来展望，我们的核心任务将放在深度改进模型的架构规划以及加强模型的自我调整能力上，旨在使模型能够更灵活应对日趋复杂和变化多端的网络环境，从而为网络安全提供更强有力的技术支撑。

参考文献

[1] 王伟, 李大辉. (2017). 基于大数据的网络安全态势感知研究. 计算机科学, 44(S2), 372-375.

[2] 张勇, 李舟军. (2016). 网络安全态势感知系统研究综述. 计算机科学, 43(2), 11-17.

[3] 李晓瑜, 李涛, 吴礼发. (2015). 基于大数据的网络安全态势感知技术研究. 信息网络安全, (9), 1-7.