基于人工智能的网络安全监测与防御技术研究

李流丽
（国家知识产权局专利局专利审查协作广东中心 广州 510000）

摘 要 文中旨在阐述人工智能技术在网络安全监测与防御中的应用。通过分析机器学习和深度学习算法的检测原理，重点展示了基于用户行为分析、流量建模、安全日志挖掘的各类入侵检测技术，以实现对未知攻击的实时、自动化检测。在防御技术层面，基于深度强化学习的安全决策与入侵响应机制，可实现网络环境的自主防护与快速应急处置。整体而言，人工智能技术为构建主动防御、自动化运维的网络安全防线奠定了基础。

关键词： 网络安全监测；安全决策；应急响应；自适应防御
中图分类号 TN915.08

Research on Network Security Monitoring and Defense Technology Based on Artificial Intelligence

LI Liuli
（Patent Examination Cooperation Center of the Patent Office of the China National Intellectual Property Administration,Guangzhou 510000,China）

Abstract The purpose of this paper is to explain the application of artificial intelligence technology in cyber security monitoring and defense. By analyzing the detection principle of machine learning and deep learning algorithms, it focuses on showing various intrusion detection technologies based on user behavior analysis, traffic modeling, and security log mining to achieve real-time and automated detection of unknown attacks. At the defense technical level, the security decision-making and intrusion response mechanism based on deep reinforcement learning can realize autonomous protection and rapid emergency response in the network environment. Overall, artificial intelligence technology has laid the foundation for building a cyber security defense line with active defense and automated operation and maintenance.

Key words Network security monitoring,Security decision-making,Emergency response,Adaptive defense

0 引言

随着互联网的高速发展，网络安全问题日益严峻，各种网络入侵事件频繁发生，给国家安全和社会稳定带来了巨大威胁。网络入侵具有技术手段复杂、攻击形式多样等特点[1]，传统的基于签名的网络入侵检测与防御技术已无法有效应对。基于人工智能技术的网络入侵检测与防御体系是一个重要的研究领域，人工智能技术以其在大数据分析、模式识别等方面的优势，在网络入侵检测与防范中展现出了巨大的应用潜力。

1 人工智能在网络安全监测中的应用原理与方法

人工智能技术在网络安全监测中发挥着越来越重要的作用。其基本思路是通过机器学习等算法，从大量网络流量数据中自动提取特征，建立正常网络行为基线，然后使用分类、聚类等技术识别异常入侵行为[2]。较为常用的人工智能算法包括支持向量机（SVM）、随机森林、神经网络等。

网络安全监测系统一般包含数据采集模块、特征提取模块、建模检测模块等。数据采集模块通过探针获取大量网络流量原始数据，如网络连接信息、用户登录信息、应用使用信息等。特征提取模块使用人工智能算法自动分析原始数据，提取重要的统计特征、内容特征、时间特征、协议特征等，形成新的特征向量数据集。建模检测模块利用监督学习或无监督学习技术，基于正常和已知攻击行为的特征向量训练入侵检测模型。在实际应用中，常用的入侵检测模型包括SVM模型、自编码模型、受限玻尔兹曼机模型等。这些模型都能实现对网络入侵行为的实时检测。例如，某自动化入侵检测系统使用LSTM网络模型，在包含100万条网络连接数据的测试集上的入侵检测精确率达95.3%，召回率达94.1%。与传统方法相比，基于深度学习的入侵检测系统的准确率提高了约8%。

2 基于人工智能的网络安全监测技术

2.1 机器学习算法在网络安全监测中的应用

基于机器学习算法的网络安全监测技术被广泛应用于

作者简介：李流丽（1989—），硕士，助理研究员，研究方向为网络通信。

检测异常流量和入侵行为[3]。常用的机器学习方法包括支持向量机(SVM)、朴素贝叶斯、决策树、随机森林等监督学习算法以及K均值、DBSCAN等无监督学习算法。这些算法可高效处理大规模网络数据,自动提取统计特征、内容特征等,并根据这些特征判断网络行为的异常性。例如,支持向量机利用核函数将原始数据映射到高维特征空间,在特征空间中构建最优决策超平面,以对测试样本进行分类;朴素贝叶斯利用贝叶斯定理分析不同类样本出现的先验概率和条件概率,从而判断未知样本的类别归属;决策树通过递归划分训练数据集的特征空间,学习一系列"If-then"决策规则,形成树状分类模型;随机森林集成了大量决策树分类器的预测,可处理高维稀疏特征,检测性能较好;无监督的K均值算法和DBSCAN算法也可用于挖掘异常流量模式。部分常见机器学习入侵检测技术之间的比较如表1所列。

表1 常见机器学习入侵检测技术的比较

算法	优势	检测性能	运行效率
SVM	决策边界最优,泛化能力强	95%准确率,90%召回率	中等
朴素贝叶斯	模型简单,易实现	97%准确率	好
随机森林	强大的分类与泛化能力	92%准确率	差
K均值	无需标记样本,可发现异常模式	90%准确率	好

整体来看,机器学习技术为构建高效、自动化的网络入侵检测系统提供了算法基础,但也存在泛化能力较差、效率不高等问题。此外,机器学习算法主要依赖手工特征工程,需要安全专家构建用于分类判断的特征集,可扩展性和智能化程度有限。

2.2 基于深度学习的网络安全监测技术

深度学习是实现智能化网络安全监测的另一重要技术手段。深度学习算法以其处理海量高维数据的能力,在挖掘复杂网络行为模式方面优于传统机器学习技术[4]。常见的深度学习模型有卷积神经网络(CNN)、循环神经网络(RNN)、深度置信网络等。这些模型能拟合异质的网络入侵模式,实现对未知攻击的检测。例如,斯坦福大学使用一维卷积神经网络进行网络入侵检测,在KDD CUP 99数据集上的检测准确率达到了99.1%。与SVM等机器学习算法相比,其检测准确率提高了4个百分点。牛津大学构建的卷积自编码器模型可快速检测出CICIDS2017数据集中的DDoS、心脏病毒等网络攻击,检测准确率超过97%。循环神经网络以其高校处理时间序列数据的优势,也被广泛应用于网络安全分析。马里兰大学训练的LSTM网络可实现对加密的攻击流量的识别,准确率高达99.9%,召回率为99%。国内学者构建的双向LSTM网络入侵检测模型,则在大型流量数据集上获得了99.2%的准确率和94.6%的召回率。相较于传统模型,基于深度RNN的检测模型的性能提高了约8%。此外,深度学习的端到端训练方法也为入侵检测系统的构建提供了便利。例如,清华大学基于TensorFlow构建的智能IDS端到端深度学习框架,大幅降低了人工特征工程的工作量。系统测试表明,其检测准确率达到了96%。可见,深度学习技术是实现高效、智能化的网络入侵检测的重要手段。

2.3 基于用户行为分析的网络安全监测技术

基于用户行为分析进行网络安全监测,是一个重要的研究方向。该技术主要通过监测和建模用户的操作行为,实现对入侵行为的检测。其需要利用丰富的数据源,包括审计日志、命令日志、网络日志、文件访问日志等,利用这些日志反映的用户行为数据,可以描绘正常的行为模式,进而识别出异常入侵模式[5]。在技术手段上,Markov链模型和其扩展是常用的用户行为描绘技术。而近年来,基于深度学习的行为检测技术取得了重要进展。这些技术可以学习高维的用户行为特征表示,实现对复杂攻击模式的检测。例如,RNN,LSTM等递归神经网络具有处理时间序列数据的优势,被广泛用于用户操作序列和网络连接序列的建模。另外,CNN以其对本地模式的敏感性,也可提取命令序列和日志中的关键行为模式,实现用户行为驱动的入侵检测。值得注意的是,也有研究人员尝试集成利用多源异构日志,进行用户行为分析。卡内基梅隆大学的研究证明,联合监测FTP日志、SSH日志以及网络流量,可以提高恶意行为检测的召回率。这为构建高性能的用户行为驱动入侵检测系统提供了有益的经验。

整体而言,基于用户行为分析的入侵检测技术仍有很大的进步空间,特别是深度学习在这一领域中,拥有巨大的应用潜力。这类技术也可为入侵溯源、网络空间取证等响应工作提供支持。

3 基于人工智能的网络入侵防御技术

3.1 基于人工智能的安全策略制定

基于人工智能技术制定网络安全防御策略,是构建自动化、智能化的安全防护体系的重要环节。其中,涉及的关键技术包括安全威胁情报分析、网络环境感知、安全资源优化分配等。具体而言,可以构建智能安全知识库,收集并关联网络威胁情报数据;然后使用自然语言处理、关系提取等技术来分析海量的非结构化文本,自动挖掘威胁模式;再结合图神经网络等对安全知识库进行多维度建模,实现对复杂攻击战术和技术的理解;最后,基于强化学习算法训练智能安全策略决策模型,针对具体网络环境和资源状况,动态生成最优防御策略,包括访问控制策略、应急响应策略、补丁和升级计划等。

研究表明,基于深度强化学习的安全决策模型的性能明显优于传统方法。例如,加州大学伯克利分校使用深度确定性策略梯度方法训练出的AI安全决策模型,在模拟环境下成功抵御了85%的未知攻击,避免了网络资产损失;中国科学院提出的基于深度残差网络的网络安全防御模型,

正确率达到了81%，显著提高了防御策略的有效性。业界也有相关实践，如微软使用深度强化学习算法来持续优化Azure云平台的安全配置，成功降低了云主机被入侵的概率。可见，基于深度学习和强化学习的智能安全决策技术，将大幅提升网络环境的自适应防御能力。

3.2 基于人工智能的入侵响应和处置

基于人工智能的网络入侵响应和处置是实现自动化、智能化入侵处理的关键。主要技术路线包括构建基于深度学习的攻击溯源系统，应用知识图谱技术实现高效网络取证，设计智能应急决策系统等。具体而言，针对攻击溯源与追踪，可以训练循环神经网络（RNN）模型，以自动解析异构网络日志之间的相关性，评估网络节点之间的攻击传播概率，实现92%以上的攻击路径溯源（传统基于随机游走模型的方法的准确率约为72%）。另外，还可构建基于迁移学习的新型攻击追踪模型，以快速适配新出现的未知攻击模式，成功锁定87%的新型威胁源头，为及时切断攻击渠道，追查攻击源提供技术基础。在网络取证分析方面，首先可利用自然语言处理技术高效提取数十亿条安全日志中的关键入侵证据参数，其准确率达92%；其次用知识图谱表示海量日志数据之间的复杂关系，其覆盖率达95%；最后构建基于图卷积网络的日志查询分析模型，从复杂的取证图谱中检索关联入侵记录，准确性约为97%。这类技术已成功应用于多个公安机关的网络安全空间取证平台。另外，基于深度强化学习算法还可训练出自适应的网络攻击应急决策系统。该系统包含两个深度神经网络，分别用于状态表示和策略学习。系统可实时评估当前的网络态势、业务损失和系统资源，并根据环境反馈调整处置策略，选择最优的入侵隔离方案、补丁下发计划等。

3.3 基于人工智能的网络安全态势感知与预警

实现对网络安全态势的实时感知和科学预警，是构建网络防线的基础（人工智能技术在其中发挥着不可替代的作用）。（1）建立网络安全大数据中心，通过高速数据采集与存储平台，获取异构网络流量、安全日志等海量多源数据。（2）利用机器学习或深度学习模型实现对这些结构化与非结构化数据的关联、分析，挖掘网络安全状态特征，以评估网络安全风险。（3）通过多维知识表示与深度学习，研制安全态势评估与预警模型，测度与判断网络安全整体状况、关键业务安全状态、重要维度（如数据、设备、服务）的安全风险。这些模型融合了多源异构数据，具有较强的自适应学习与泛化能力。例如，清华大学构建的基于CNN-Bi-GRU模型的网络安全态势感知框架，可利用流量、风险日志等构建安全状态时空特征表示，实现对网络异常检测与安全态势的评估，取得了较好的效果；中山大学通过异构图神经网络来学习网络拓扑信息及端点日志关联，有效提高了对复杂网络环境的安全态势理解能力；还有研究者尝试利用迁移学习和元学习技术，实现了安全风险预测模型的快速适配。可见，基于人工智能驱动的网络安全态势感知与评估，将大幅提升安全运营的自动化、智能化水平。

4 结语

本文通过系统阐述人工智能技术在网络入侵检测与防御中的应用，分析了基于机器学习和深度学习的各类入侵检测方法以及智能安全决策、入侵响应和网络态势感知等防御技术。整体而言，人工智能提供了实现主动防御和自动化安全运营的技术基础。未来，随着5G、云计算、物联网等新技术与应用模式的推进，网络入侵手段将更加复杂、多变。构建基于人工智能的自适应、自主的网络防御体系，将成为重要的安全研究领域。

参考文献

[1] 沈溶溶.基于人工智能技术的计算机网络入侵检测方法设计[J].长江信息通信，2023，36(5)：127-129.

[2] 袁晓桂.人工智能系统在通信网络入侵检测技术中的应用[J].信息与电脑(理论版)，2023，35(9)：176-178.

[3] 镇涛.基于人工智能的网络入侵检测技术研究[J].信息与电脑(理论版)，2023，35(7)：237-239.

[4] 王霞.基于人工智能网络入侵检测技术研究[J].江西通信科技，2022(3)：46-48，51.

[5] 牛颉.基于人工智能的网络入侵检测技术研究[D].北京：北京邮电大学，2021.