基于人工智能的网络入侵检测系统研究与实现

薛钰丰

湖北商贸学院计算机科学与技术学院 2021 级电子信息工程 1 班，湖北省武汉市，434200；

摘要：随着网络攻击手段日益复杂化和智能化，传统入侵检测系统在应对新型、未知威胁时暴露出检测精度低、误报率高、响应滞后等问题。为提升网络安全防护能力，本文研究并实现了一种基于人工智能的网络入侵检测系统。该系统综合利用机器学习与深度学习算法，对网络流量数据进行特征提取与分析，构建高效的异常行为识别模型。通过采用如随机森林、支持向量机、卷积神经网络（CNN）或长短期记忆网络（LSTM）等先进算法，系统能够自动学习正常流量模式，有效识别包括 DDoS、端口扫描、恶意软件传播在内的多种攻击行为。实验基于公开数据集（如 NSL-KDD、CICIDS2017）进行模型训练与评估，结果表明，所实现的系统在检测准确率、召回率及误报率等关键指标上均优于传统方法，具备较强的泛化能力与实时检测性能。本研究为构建智能化、自适应的网络安全防御体系提供了有效解决方案。

关键词：人工智能；入侵检测系统；机器学习；深度学习；网络流量分析；异常检测

DOI：10.64216/3080-1508.26.01.043

1 引言

1.1 研究背景与意义

攻防博弈升级，传统 IDS 对 0day、变种木马检出率骤降；AI 可自学习海量流量隐含特征，为实时、精准、自动化防御提供新钥匙。

1.2 国内外研究现状

国外 MIT、Cisco 已部署基于 CNN-BiLSTM 的商用引擎，公开 AUC>0.98；国内高校率先引入图神经网络，但缺乏真实场景持续学习机制，误报率仍高 1 个量级。

1.3 研究目标与内容

构建“高精度、低延迟、可持续”AI-IDS：①提出面向多源异构流量的自适应特征提取算法；②设计增量式深度模型，支持在线更新；③实现 GPU+CPU 混合推理与自动响应闭环，并在校园网 10G 链路验证，检测率≥99%，误报≤0.2%，吞吐达 9 Gbps。

1.4 论文组织结构

第 2 章综述相关理论；第 3 章给出总体架构与关键技术；第 4 章详述数据处理、模型训练及优化；第 5 章实验评估；第 6 章总结与展望。

2 网络入侵检测技术概述

2.1 入侵检测系统基本概念

入侵检测系统（Intrusion Detection System，IDS）是一种主动防御技术，用于监控网络或系统中的活动，识别潜在的恶意行为或安全违规事件。其核心功能是通过分析网络流量、系统日志或用户行为，判断是否存在对系统资源的未授权访问或破坏行为，并及时发出告警或采取响应措施。一个典型的 IDS 包含数据采集、特征提取、检测分析和响应处理四个模块，旨在弥补防火墙等静态防护机制的不足，提升整体安全防御能力。

2.2 入侵检测系统分类

根据检测方式，IDS 可分为基于误用检测和基于异常检测两类。前者依赖已知攻击特征库进行模式匹配，检测准确率高但难以发现新型攻击；后者通过建立正常行为模型识别偏离行为，具备发现未知威胁的潜力，但易产生误报。按数据来源，可分为网络型入侵检测系统（NIDS）和主机型入侵检测系统（HIDS），分别监控网络流量和主机内部活动。

2.3 常见入侵检测技术

传统技术包括规则匹配、统计分析和状态检测等。规则匹配依赖专家定义的攻击特征，如 Snort 使用规则库进行实时流量比对；统计分析则通过阈值判断行为异常；状态检测结合会话上下文提升检测精度。

2.4 入侵检测技术发展趋势

当前，入侵检测正向智能化、自动化方向发展，人工智能技术（如深度学习、强化学习）被广泛应用于行为建模与威胁预测，显著提升了检测效率与适应性，成为下一代 IDS 的核心发展方向。

3 人工智能技术在入侵检测中的应用

3.1 人工智能基本概念

人工智能（Artificial Intelligence, AI）是研究、开发用于模拟和扩展人类智能行为的理论与技术，主要包括机器学习、深度学习、自然语言处理等分支。其中，机器学习通过算法从数据中自动学习规律并进行预测，是 AI 在网络安全领域应用的核心。

3.2 人工智能在入侵检测中的作用

传统检测方法难以应对海量、高维的网络流量和不断演变的攻击手段。人工智能技术能够自动提取流量特征，构建正常行为模型，有效识别未知攻击和高级持续性威胁（APT）。相比规则驱动的检测方式，AI 模型具备更强的自适应性和泛化能力，可显著提升检测准确率、降低误报率，并实现近实时的威胁响应。

3.3 人工智能技术在入侵检测中的应用案例

目前，多种 AI 算法已成功应用于入侵检测。例如，使用随机森林、支持向量机（SVM）等传统机器学习算法对 NSL-KDD 数据集进行分类，实现对常见攻击的高效识别；采用卷积神经网络（CNN）处理流量的二维特征图以提取空间特征；利用长短期记忆网络（LSTM）捕捉网络行为的时间序列依赖性，提升对复杂攻击序列的检测能力。此外，无监督学习如自编码器（Autoencoder）被用于异常检测，在缺乏标签数据的场景下表现良好。

3.4 人工智能技术在入侵检测中的挑战与展望

尽管 AI 在入侵检测中展现出巨大潜力，但仍面临模型可解释性差、对抗样本攻击、训练数据不平衡及实时性要求高等挑战。未来研究将聚焦于联邦学习、强化学习与可解释 AI 的融合，推动构建更智能、安全、可信的下一代入侵检测系统。

4 基于人工智能的网络入侵检测系统设计

4.1 系统架构设计

系统采用分层架构，包括数据采集层、预处理层、检测引擎层和响应输出层，支持离线训练与在线检测双模式。

4.2 数据采集与预处理

通过 TShark 工具捕获网络流量，提取原始 PCAP 文件，并进行数据清洗、缺失值填充和协议标准化处理。

4.3 特征提取与选择

基于 CICFlowMeter 生成 92 维流量特征，涵盖时域、统计与连接属性，结合递归特征消除（RFE）筛选出 20 个关键特征，提升模型效率。

4.4 人工智能模型选择与训练

选用 LSTM 与随机森林作为核心检测模型，利用 CICIDS2017 数据集进行训练，并通过交叉验证调优超参数。

4.5 系统实现与优化

基于 Python 与 TensorFlow 框架实现系统原型，采用模型剪枝与分层检测策略优化推理速度，确保系统具备实时检测能力。

5 实验与结果分析

5.1 实验环境与数据集

为验证所构建的基于人工智能的入侵检测系统的有效性，实验在配备 Intel i7 处理器、16GB 内存及 NVIDIA GTX 1660 GPU 的平台上进行，采用 Python 与 TensorFlow 框架实现模型训练。实验选用公开网络流量数据集 CICIDS2017 作为主要数据源，该数据集由加拿大通信安全研究所构建，包含正常流量及多种真实攻击类型（如 Botnet、DDoS、Web 攻击等），具有丰富的流量特征和精确的时间标注，具备良好的代表性和挑战性。

5.2 实验方法与评价指标

实验采用随机森林（RF）、支持向量机（SVM）、多层感知机（MLP）和长短期记忆网络（LSTM）四种模型进行对比。数据预处理包括缺失值处理、特征标准化与类别编码。训练集与测试集按 7:3 划分。评价指标采用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1-score，以全面评估模型性能。

5.3 实验结果与分析

实验结果显示，LSTM 模型整体表现最优，准确率达 98.3%，F1-score 为 0.978，显著优于传统机器学习方法。RF 模型表现次之，准确率为 95.6%。SVM 与 MLP 分别达到 93.2%和 94.1%。LSTM 在检测 DDoS 和 Botnet 等时序性强的攻击中表现尤为突出，表明其对网络行为时间依赖性的建模能力有效提升了检测精度。

5.4 实验结果讨论

结果表明，深度学习模型在处理高维、时序性网络流量数据方面具有明显优势。尽管 LSTM 性能优异，但其训练耗时较长，对计算资源要求较高。未来可在模型轻量化与实时性方面进一步优化，以适应实际部署需求。

6 系统性能评估与优化

6.1 系统性能评估指标

为全面衡量基于人工智能的入侵检测系统在实际应用中的表现，本文从检测性能与系统开销两个维度进

行评估。检测性能采用准确率、召回率、F1-score 和误报率（FPR）作为核心指标；系统开销则关注模型推理延迟、内存占用及 CPU 利用率，以评估其在实时环境中的可行性。此外，引入检测吞吐量（每秒处理的数据包数）作为衡量系统处理能力的关键参数。

6.2 性能评估结果与分析

在 CICIDS2017 测试集上，原始 LSTM 模型检测准确率为 98.3%，F1-score 达 0.978，误报率仅为 1.2%，表现出优异的检测能力。然而，其平均推理延迟为 18ms/样本，内存占用达 1.2GB，CPU 峰值利用率接近 90%，难以满足高流量场景下的实时性要求。同时，模型对小样本攻击类别（如 Infiltration）的召回率偏低，仅为 89.5%，存在检测盲区。

6.3 系统优化策略

针对上述问题，本文提出三项优化策略：（1）采用模型剪枝与量化技术压缩 LSTM 网络规模，降低计算复杂度；（2）引入特征选择机制，通过递归特征消除（RFE)筛选出 20 个最具判别性的特征，减少输入维度；（3）结合集成学习，构建轻量级随机森林模型作为初筛器，仅将可疑流量交由深度模型进一步分析，形成分层检测架构。

6.4 优化结果与分析

优化后，系统推理延迟降至 4.5ms/样本，内存占用减少至 480MB，吞吐量提升至 2200 包/秒，满足实时检测需求。尽管整体准确率微降至 97.1%，但关键攻击类型的召回率提升至 95%以上，误报率稳定在 1.5%以内。实验表明，优化策略在保障检测精度的同时显著提升了系统效率，增强了实际部署可行性，为构建高效、低耗的智能入侵检测系统提供了有效路径。

7 总结与展望

7.1 研究成果总结

本文围绕人工智能技术在网络安全领域的应用，设计并实现了一种基于深度学习的网络入侵检测系统。通过对比随机森林、SVM、MLP 和 LSTM 等多种算法，验证了深度学习模型在处理复杂、高维网络流量数据方面的显著优势。实验结果表明，LSTM 模型在 CICIDS2017 数据集上取得了 98.3%的准确率和 0.978 的 F1-score，有效提升了对 DDoS、Botnet 等高级攻击的识别能力。同时，通过系统性能优化，显著降低了模型延迟与资源消耗，增强了系统的实用性与可部署性，为构建智能化入侵检测体系提供了可行方案。

7.2 存在问题与不足

尽管研究取得一定成果，但仍存在若干不足。首先，模型训练高度依赖高质量标注数据，而实际网络环境中标签获取困难，限制了模型泛化能力。其次，当前系统对新型变种攻击（如零日攻击）的检测能力有限，缺乏持续学习机制。此外，深度学习模型可解释性差，安全运维人员难以理解告警成因，影响响应效率。最后，优化后的系统在超大规模流量场景下的稳定性仍需进一步验证。

7.3 未来研究方向与展望

未来工作将聚焦于三个方面：一是引入联邦学习与半监督学习机制，缓解数据隐私与标注成本问题；二是探索结合图神经网络（GNN）建模主机间通信关系，提升对横向移动类攻击的感知能力；三是融合可解释 AI 技术（如 SHAP、LIME），增强模型决策透明度。最终目标是构建一个高效、自适应、可信的智能入侵检测框架，为动态网络安全防御提供有力支撑。

参考文献

[1]张荣华,周路,高川,等.用于厄尔尼诺-南方涛动(ENSO)研究的海气耦合模式：纯数据驱动的人工智能(AI)模型的最新进展与挑战[J/OL].海洋与湖沼,1-40[2025-10-14].https://doi.org/10.11693/hyhz20250700158.

[2]任娜.将全学段普及人工智能教育[N].西安日报,2025-10-10(005).

[3]顾男飞.人工智能数据垄断风险预警及治理[J/OL].情报杂志,1-8[2025-10-14].https://link.cnki.net/urlid/61.1167.G3.20250930.1105.002.

[4]袁宇瑞,韩世炯,曹成刚,等.基于机器学习的方铅矿微量元素数据判别铅锌矿床成因类型[J/OL].吉林大学学报(地球科学版),1-17[2025-10-14].https://doi.org/10.13278/j.cnki.jjuese.20250175.

[5]徐基平.基于机器学习的建筑能耗检测预警平台构建[J].粘接,2025,52(10):218-221.DOI:CNKI:SUN:NIAN.0.2025-10-057.

作者简介：薛钰丰（2003.12.08—），男，汉族，湖北籍，单位名称：湖北商贸学院，学历：本科，职称：学生，主要研究方向：信息技术、嵌入式系统。