基于深度学习的网络入侵检测研究综述

苏书宾[1], 肖利民[2,3], 李书攀[5], 黄兴旺[1], 谢书童[1], 吴 博[4]

[1]集美大学 计算机工程学院 厦门 中国 361021
[2]北京航空航天大学 软件开发环境国家重点实验室 北京 中国 100091
[3]北京航空航天大学 计算机学院 北京 中国 100091
[4]南昌航空大学 软件学院 南昌 中国 330063
[5]郑州大学 信息工程学院 郑州 中国 450001

摘要 入侵检测是网络系统安全继防火墙之后的第二道防线, 在网络入侵的防护中发挥着重要的作用。深度学习具有强大的自动学习能力、良好的可移植性、模型容量大等突出优点。使用深度学习方法构建入侵检测系统可以实时监测网络流量、识别更复杂的入侵行为, 以及自适应检测新型攻击模式, 是网络安全领域一个重要的研究方向。本文首先介绍了网络安全的当前形式, 网络入侵的危害和分类, 并总结了入侵检测系统的分类、评估方法、常用的机器学习方法。另外, 深度学习是一种数据驱动的方法, 数据集对于深度学习至关重要, 因此本文对入侵检测领域的重要数据集和预处理方法也进行了详细介绍。然后, 回顾了自 2010 年以来关于基于深度学习方法研究入侵检测系统的代表性文献, 并以数据类型作为主要的分类标志对代表性方法进行分类总结。同时对深度学习在入侵检测应用中面临的挑战进行总结, 分析如何更好地将基于深度学习的入侵检测系统应用到实际环境中, 对此除了考虑深度学习准确率的相关指标外, 本文还重点分析了深度学习模型的时间效率与可解释性的重要性。最后, 对基于深度学习的入侵检测系统未来的发展进行总结, 随着网络技术的发展和应用入侵检测依然面临着一系列挑战、深度学习是入侵检测的一个有效技术、优化现有深度学习技术和研究新的深度学习方法是未来提高入侵检测性能的重要研究方向。

关键词 机器学习; 深度学习; 入侵检测; 网络安全
中图法分类号 TP311 DOI 号 10.19363/J.cnki.cn10-1380/tn.2026.01.07

State-of-the-Art Survey of Network Intrusion Detection Technology based on Deep Learning

SU Shubin[1], XIAO Limin[2,3], LI Shupan[5], HUANG Xingwang[1], XIE Shutong[1], WU Bo[4]

[1] College of Computer Engineering, Jimei University, Xiamen 361021, China
[2] State Key Laboratory of Software Development Environment, Beihang University, Beijing 100091, China
[3] School of Computer Science and Engineering, Beihang University, Beijing 100091, China
[4] School of Software, Nanchang Hangkong University, Nanchang 330063, China
[5] School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China.

Abstract Intrusion detection system is called the second strobe after firewall of cyber security, and plays a crucial role in preventing network intrusions. Deep learning has the outstanding advantages such as powerful automatic learning ability, good portability, and large model capacity. Using deep learning to build intrusion detection system can monitor network traffic in real time, identify more complex intrusion behaviors, and automatically detect the new attack patterns, making it an important research direction in the field of cyber security. This work first introduces the current situation of cyber security, the harm and classification of network intrusion, and summarizes the classification, evaluation methods, as well as commonly used machine learning methods of intrusion detection systems. In addition, deep learning is a data-driven approach, and the data sets are crucial for deep learning methods. Therefore, we have also provided a detailed introduction to the important data sets and the data preprocessing methods in the field of intrusion detection. Then, the representative researches on intrusion detection system based on deep learning since 2010 are reviewed, and they are classified and compared with data types as the main classification marks. Furthermore, we summarize the challenges of deep learning in intrusion detection applications, and analyze how to better apply deep learning based intrusion detection systems to practice. Specifically, in addition to considering relevant indicators of deep learning accuracy, this work also focuses on analyzing the importance of time efficiency and interpretability of the deep learning models. Finally, we summarize the future de-

通讯作者: 吴博, 博士, 讲师, Email: wubo@nchu.edu.cn。
本课题得到福建省自然科学基金(No. 2023J01802), 福建省教育厅资助科技项目(No. JAT210216), 集美大学科学基金(No. ZP2022007), 国家自然科学基金项目(No. 62272026), 国家自然科学基金青年科学基金项目(A\B\C 类) (No. 62006096), 福建省自然科学基金青年项目(No. 2020J05146), 集美大学科研基金(No. ZQ2021024), 江西省自然科学基金青年基金项目(No. 20242BAB20049)资助。
收稿日期: 2024-03-02; 修改日期: 2024-07-18; 定稿日期: 2025-11-11

velopment of deep learning based intrusion detection systems. With the development and application of network technology, intrusion detection still faces a series of challenges. Deep learning is an effective technology for intrusion detection, optimizing the existing deep learning techniques and researching new deep learning methods are important research directions for improving intrusion detection performance in the future.

Key words machine learning; deep learning; intrusion detection; cyber security

1 引言

随着信息技术的发展, 网络在人们的生产和生活中的应用变得愈加广泛。根据中国互联网络信息中心(CNNIC)发布的第 52 次《中国互联网络发展状况统计报告》, 截至 2023 年 6 月, 我国网民规模达 10.79 亿, 较 2022 年 12 月增长 1109 万人, 互联网普及率达 76.4%。随着互联网经济的进一步发展, 预计 2024 年互联网普及率将会进一步增加。互联网在多个领域发挥着越来越重要的作用, 然而网络空间面临的安全威胁也随着急剧增加。根据 CNCERT 发布的网络安全信息与动态周报统计, 2023 年 CNCERT 协调云服务商、域名注册服务机构、应用商店、各省分中心以及国际合作组织共处理网络安全事件共 137533 起, 含跨境网络安全事件共 72804 起。其中, 协调境内外域名注册机构、境外 CERT 等机构重点处理 86616 起仿冒投诉事件。协调 1837 个提供恶意移动应用程序下载服务的平台开展移动互联网恶意代码处理工作, 共处理传播移动互联网恶意代码的恶意 URL 链接 38314 个[1]。另外, 从 2023 年 1 月 1 日至 10 月 20 日, 国家信息安全漏洞共享平台(CNVD)共发现安全漏洞 30177 个。各种网络安全事件给国家和人们带来了巨大的损失, 根据 Cybersecurity Ventures 最新发布的“2022 年网络犯罪报告”, 预计 2023 年全球因网络安全事件导致的损失将高达 8 万亿美元的损失。随着互联网的发展, 新型攻击层出不穷, 互联网面临的安全形势不乐观, 网络安全保护任重道远。

网络安全形势愈加严峻, 网络安全成为了一个重要的研究领域。网络安全技术主要包括主动式防御技术和被动式防御技术, 被动式防御技术主要包括反病毒软件和防火墙, 而主动式防御技术主要是指入侵检测系统(Intrusion Detection System, IDS), 他们通过协同工作来有效减少网络受外部和内部的攻击。防火墙技术, 通常称作网络安全的第一道防线, 通过监控和过滤不安全的网络流量来保护网络安全。而反病毒软件通过检测处理包含病毒的文件来防护网络系统的安全。但是, 随着黑客技术的发展进步, 反病毒软件和防火墙技术已经无法满足对网络安全的要求。因而, IDS 作为一项动态的、主动的安全防护手段, 广泛应用于各种重要的网络系统, 开启了网络系统安全的第二道防线。

IDS 起源于 Anderson[2]在 1980 年提出的用来处理用户审计数据的“计算机安全威胁监测和监视系统”。基于同样的原则, Denning[3]提出使用由审计数据生成的用户特征来识别入侵, 即从审计记录中获取主体相对于客体的行为的知识和检测异常行为的规则。这些开创性的工作定义了入侵检测的相关概念, IDS 是一种网络安全检测系统, 通过监控网络流量、系统日志等信息, 检测网络系统中的安全漏洞和各种异常行为。IDS 作为一种积极主动的防护技术, 能够充分利用软件和硬件, 通过对网络或系统进行监控以感知恶意活动并及时发出警报, 能够弥补被动式防御技术的不足、增强网络系统安全性的优势, 因而日益受到研究人员的重视。机器学习是一种人工智能的方法, 根据以往的经验和数据改善程序的性能。目前机器学习广泛应用于计算机视觉、机器翻译、信息检索、医疗诊断等众多领域[4-6], 许多的入侵检测系统采用机器学习方法来提高检测效果, 特别是对于零日攻击的检测[7-8]。自从 2010 年以来, 深度学习(Deep Learning, DL)方法开始兴起, 其极大地简化了传统机器学习的整体算法分析和学习流程, 在很多通用的领域刷新了传统的机器学习方法达不到的精度和准确率。深度学习的一个显著特点是模型结构深, 包含多个隐藏层。而传统的机器学习算法, 例如 SVM、KNN 等, 没有或仅含有一个隐藏层, 这些传统机器学习方法也称为浅层模型。根据训练时是否需要标签, 机器学习方法可以分为有监督学习和无监督学习。有监督学习在训练中需要标签, 可以从数据中获得更多信息, 常用于分类任务, 也是攻击检测中最常用的方法。但是, 有监督学习对于数据集的要求高, 数据获取成本高。无监督学习可以从无标签的数据中获得有用信息, 数据获取容易。但是, 在入侵检测中, 无监督学习方法必须借助外部知识或监督方法进行检测。

本文的目的是从一个新的视角, 对目前提出的基于深度学习的 IDS 进行分类和总结、抽象出将深度学习应用于网络入侵检测的核心思想, 并分析当前基于深度学习的 IDS 面临的挑战和网络入侵检测未来的发展, 为计划从事该领域学习研究的人员提

供一份比较完整的文献材料。对此, 本文主要选择自2010 年以来发表的具有代表性的文献, 这些文献可以充分地反映 IDS 当前研究的整体进展。文献[9-16]根据采用的机器学习算法的差异, 对现有 IDS 的研究工作进行了分类。但是, 这些文献主要分类介绍了应用于 IDS 的机器学习算法, 而忽略了网络安全的本质问题, 这可能更有利于机器学习研究人员的学习研究。因此, 这些工作没有直接回答使用深度学习解决网络入侵检测的原理性问题。另外, 深度学习是一种数据驱动的方法, 数据作为深度学习方法的输入, 数据的类型、规模和准确性等都将对深度学习的结果产生重大影响。为此, 本文选择根据数据类型的差异对网络入侵检测模型进行分类, 着力揭示网络入侵检测和基于深度学习算法的 IDS 技术的本质, 并对深度学习在入侵检测应用中面临的挑战与网络入侵检测未来的发展进行分析总结。

本文的其余部分按照以下方式进行组织: 第 2 节介绍了入侵检测的相关预备知识; 第 3 节介绍了 IDS 的数据集和数据预处理方法; 第 4 节介绍了 IDS 常用的机器学习方法; 第 5 节介绍了基于深度学习的入侵检测方法; 第 6 节讨论了基于深度学习的 IDS 未来的发展方向; 第 7 节总结全文。

2 预备知识

网络需要面临各种不同入侵的威胁, IDS 作为网络安全的第二道防线得到了广泛的研究和应用。本节我们首先对网络入侵, IDS 的类型和评估方法进行概述。

2.1 网络入侵

网络入侵也称网络攻击, 是指未经授权访问和操纵计算机系统、网络或数据的行为, 是导致网络安全威胁的主要原因。尽管网络入侵的方法很多, 但是其结果通常是造成机密性、完整性、可用性和真实性 4 种安全特性的破坏。

- 机密性: 机密性破坏是指允许攻击者对数据进行非授权访问。
- 完整性: 完整性破坏是指允许攻击者对系统状态以及流经或存储在系统中的任何数据进行非授权修改。
- 可用性: 可用性破坏是指使授权用户无法正常访问特定的系统资源, 包括在任何时间、地点, 以任何方式。
- 真实性: 真实性破坏是指使允许攻击者对信息的来源以及内容进行伪造。

根据对系统信息造成的破坏性差异, 网络攻击类型可以分为主动攻击和被动攻击。其中, 主动攻击会导致某些数据的篡改和虚假数据的产生, 包括欺骗攻击、拒绝服务攻击、木马植入等攻击手段。而在被动攻击中, 攻击者则不对数据信息做任何修改, 他们在未经用户认可的情况下截取或窃听网络中相关的数据信息。被动攻击包括窃听、网络嗅探、流量分析等攻击方式。如表 1 所示, 本文总结了目前网络各层中存在的安全缺陷以及对应的攻击技术。

表 1 网络协议中的安全缺陷及对应的攻击技术

Table 1 Security vulnerabilities and corresponding attack techniques in network protocols

网络层次	网络协议	存在的安全缺陷	对应的攻击技术	破坏安全属性
网络接口层	以太网协议	共享传输媒介并明文传输	网络嗅探攻击	机密性
	以太网协议	缺乏 MAC 身份认证机制	MAC 欺骗	真实性
	PPP 协议	明文传输	网络嗅探攻击	机密性
互联层	IPv4	缺乏 IP 地址身份认证机制	IP 地址欺骗	真实性
	ICMP	ICMP 路由重定向缺乏身份认证	ICMP 路由重定向	完整性、真实性
		广播地址对 Ping 的方法器效应	Smurf 攻击	可用性
传输层	TCP	3 次握手存在连接队列瓶颈	SYN 泛洪攻击	可用性
		会话对身份认证不够安全	会话劫持	真实性、可用性
应用层	DNS	DNS 验证机制不够安全	DNS 欺骗	完整性、真实性
	HTTP	URL 明文传输, 缺少完整性保护	U2R 攻, R2L 攻击	机密性、完整性和真实性
		内嵌链接滥用	特洛伊木马攻击	完整性

2.2 入侵检测系统的分类

IDS 日益受到广泛的关注并取得了丰富的成果, 到目前为止甚至有些学者宣称 IDS 可以完全取代防火墙[17-18]。现在, 很多的电子信息企业都推出了自己的入侵检测系统, 并且很多的开源社区也在维护着一些重要的入侵检测系统。如表 2 所示, 本文总结分

析了目前一些主流的入侵检测系统。

根据不同的划分标准, 入侵检测系统可以划分为不同的类别[19-20]。本文借鉴 IDS 常用的划分框架, 对 IDS 按照检测技术、数据来源和工作方式进行分类, 如图 1 所示。

2.2.1 基于检测技术的分类

根据采用的检测技术的差异, 可以把 IDS 分为基于异常的入侵检测系统(Anomaly based IDS, AIDS)、基于规则的入侵检测系统(Rule based IDS, RIDS)和基于协议的入侵检测系统(Protocol based IDS, PIDS)。

AIDS 记录系统中用户的正常活动, 并学习正常活动的特征, 当用户的行为偏离正常活动特征时就被检测为攻击行为。AIDS 可以学习用户的行为习惯, 具有较高的检出率, 可以检测出各种未知的攻击。但是, AIDS 对复杂的网络环境适应性不强, 没有准确的判断准则。因此, AIDS 会经常出现虚报情况, 且入侵者可以通过逐步训练使入侵行为接近正常活动的特征而穿透检测系统。

RIDS 是基于已知的攻击模式, 将审计记录与已有的模式特征进行匹配来判断其是否就是入侵行为, 在一些场景下也称为特征检测。RIDS 的检测结果通常取决于事先定义好的攻击模式, 因而对于新型未知的攻击行为往往会束手无策。

表 2 主流入侵检测系统的对比分析

Table 2 Comparative analysis of mainstream intrusion detection systems

系统	厂商	主要功能	特色
SolarWinds Security Event Manager	SolarWinds	入侵检测、日志分析和合规性审计; 自动监控安全事件	分析日志文件; 可以处理实时数据; 与 Snort 兼容; 自动修复
CrowdStrike Falcon	CrowdStrike	通过查看每台计算机上运行的进程监视各个端点上的活动	端点保护平台; 检查日志文件; 在云端处理数据; 基于云的控制台
ManageEngine EventLog Analyzer	ManageEngine	管理和分析由标准应用程序和操作系统生成的日志文件	管理和分析日志文件; 审核数据保护标准合规性
Snort	Cisco	开源网络入侵检测; 实时分析网络流量	行业领先的 NIDS; 思科系统支持
OSSEC	趋势科技	通过监视所有日志文件的校验和签名以检测可能的干扰	日志文件分析器; 免费政策; 警报系统
Suricata	开源社区	检查不同网络应用程序(包括 FTP、HTTP 和 SMB)的实时流量	应用层操作; 根据实时数据进行操作
Zeek	开源社区	入侵检测; 网络监视	签名检测; 异常分析
Sagan	开源社区	分析来自各种 IDS 工具(如 Snort、Suricata 等)的日志数据	日志分析; 网络流量分析
Security Onion	开源社区	检测入侵行为; 监视网络的安全性	综合基于网络和主机的检测技术; 日志文件篡改警报
AIDE	开源社区	完整数据包捕获; 基于网络和基于主机的入侵检测系统(分别为 NIDS 和 HIDS)	创建配置基准; 回滚未经授权的更改
RG-IDP 1000E V2.0	锐捷	僵尸计算机侦测; 丰富的上网行为管理; 全面的日志报表功能和方便的集中管理	深度内容检测; 安全防护; 上网行为管理

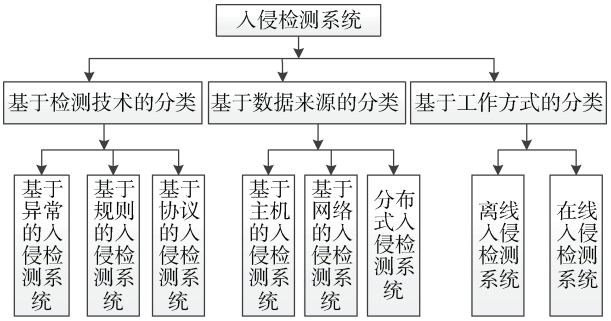

图 1 入侵检测系统分类

Figure 1 Classification of intrusion detection systems

PIDS 根据网络协议的高度规则性快速来快速检测网络攻击行为。PIDS 对于基于协议漏洞的入侵行为往往具有较高的准确度, 但是不适用于其他类型的入侵检测。除了上述三大类外, 近年来一些学者也开始研究混合检测。混合检测既分析系统的正常行为, 又可以比对已知的攻击模式, 还可以判断是否符合网络协议的规则性, 所以系统的检测更全面、准确, 但是对不同类型的技术进行整合应用是一个比较复杂的过程。

2.2.2 基于数据来源的分类

根据数据的不同来源可以把 IDS 分为基于主机的入侵检测系统(Host based IDS, HDIS)、基于网络的入侵检测系统(Network based IDS, NIDS)和分布式入侵检测系统(Distributed IDS, DIDS)。

HDIS 从被监测的主机系统收集数据, 通过收

集、分析用户的活动信息来判断各种入侵行为。HDIS的关键是能否收集到表示各种入侵行为的审计记录。HDIS 的优点是方法直接便捷、准确度较高, 缺点是需要每一个系统安装且只能监测主机上的一些特定应用。

NIDS 从网络中侦听采集数据, 通过解析网络数据包来分析是否存在网络攻击行为。NIDS 的优点是不需要在每一个主机上安装系统程序, 通过系统就可以对整个网络进行监视, 缺点是对于加密的传输信息、较长分析时间和较大计算量的检测难以实现。

DIDS 将 HDIS 和 NIDS 进行结合。DIDS 通常包括各个被监测主机上的传感器、局域网上的网络管理器和中央处理器三部分。主机上的传感器收集被监测主机上的审计信息, 网络管理器收集网络审计信息, 中央处理器用收集到的数据进行入侵检测。DIDS 综合考虑了入侵行为在主机和网络上的表现, 但系统会占用一定的网络资源。

2.2.3 基于工作方式的分类

根据系统的工作方式, 可以把 IDS 分为离线入侵检测系统和在线入侵检测系统。它们的本质区别是系统工作的实时性。

离线入侵检测系统通常是在事后对事件进行分析审计, 进而判断是否为入侵行为。离线检测一般由网络管理员来完成, 他们运用自身比较专业的网络安全知识, 根据系统对用户操作的历史审计记录进行分析, 判断用户是否存在入侵行为。管理员在实现此类网络安全分析通常是定期或者不定期的, 不具有实时性。

在线入侵检测系统是对网络数据包和主机状态参数进行实时的审计分析。具体的, 在线检测系统根据用户的历史行为模型、专家知识库和神经网络学习模型对用户的当前操作进行监测, 判断是否有入侵的行为, 当系统发现有入侵事件时开始收集证据、切断连接和恢复数据。在线检测系统的这个入侵防护过程是实时并不断循环的。

2.3 入侵检测系统的评估方法

目前, 主要采用二分类算法的评估方法来对入侵检测系统的性能进行评估。因此, 入侵检测系统有很多衡量指标, 在大部分研究中, 通常使用多个指标对系统的性能进行全面分析说明。

- 准确率(*Accuracy*): 定义为正确分类样本数量占样本总数的比例, 反映分类效果, 其计算方式如式(1)所示。当不同类别样本平衡时, 准确率是一个很好的度量标准。但在实际网络环境中, 正常样本数量要远高于攻击样本, 仅用准确率度量并不合适。

$$Accuracy=\frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

- 精度(*P*): 定义为真正例数量占分类器预测为正样本的比例, 反映检索的正确率, 其计算方式如式(2)所示。

$$P=\frac{TP}{TP+FP} \tag{2}$$

- 召回率(*TPR*): 在入侵检测中也称为检测率。定义为真正例数量占总正例的比例, 其计算方式如式(3)所示。检测率反应系统对攻击样本的检测能力, 是度量入侵检测系统的一个重要指标。

$$TPR=\frac{TP}{TP+FN} \tag{3}$$

- F_1-Measure(F_1): 定义为精度和召回率的调和平均数, 其计算方式如式(4)所示。该指标能够全面反映检测效果。

$$F_1=\frac{2*P*R}{P+R} \tag{4}$$

- 漏报率(*NPR*): 定义为未检测出正例占总正例比例, 其计算方式如式(5)所示。入侵检测中表示为未检测出攻击占攻击总数的比例。

$$NPR=\frac{FN}{TP+FN} \tag{5}$$

- 误报率(*FPR*): 定义假正例占分类器预测为正样本的比例, 其计算方式如式(6)所示。入侵检测中表示被错误检测出的攻击占检测出攻击的比例。

$$FPR=\frac{FP}{TP+FP} \tag{6}$$

其中, 真正例(*TP*)对应于分类器正确预测正例的数量, 假负例(*FN*)对应于分类器错误预测正例的数量, 假正例(*FP*)对应于分类器错误预测负例的数量, 真负例(*TN*)对应于分类器正确预测负例的数量。入侵检测的目标是发现攻击, 一般将攻击样本视为正例, 将正常样本视为负例。在入侵检测中, 准确率、误报率和漏报率是最广泛使用的。

3 入侵检测系统的数据集和数据预处理方法

深度学习的任务是从可用的数据中获取有价值的信息, 而数据的质量决定了深度学习方法获得的结果。因此, 对于深度学习方法而言, 数据是十分重要的。另外, 为了提高 IDS 的有效性和高效性, 通常

还需要对数据集进行预处理。

3.1 入侵检测系统的数据集

对于入侵检测系统, 使用的数据要求可容易获取并且能够反映主机或网络的状态。随着入侵检测技术的发展, 出现了一些标准的数据集。数据集的建立是十分复杂且耗时, 但是当一个标准数据集建立, 可以方便地供多个研究者反复使用。除了方便之外, 使用标准数据集进行入侵检测实验还有两点好处。首先, 标准数据集有权威性, 实验结果能够让人信服。其次, 大量的研究工作在标准数据集进行, 可以与之前的研究成果进行比较。一般而言, IDS 使用的数据可分为数据包、网络流、日志和其他数据。目前, IDS 常用的数据集一般由一些大学或企业的科研团队提供。

3.1.1 DARPA1998

DARPA(Defense Advanced Research Projects Agency)1998[21]是由 MIT Lincoln 实验室构建的数据集, 是一个广泛使用的 IDS 标准数据集。从网络中收集了 9 周的流量, 将 7 周作为训练集, 2 周作为测试集。该数据中包含了原始流量和标注信息。数据的类别分为 5 类:

- 正常: 非攻击的数据, 数据集中大部分数据都是正常的。
- DOS: 拒绝服务攻击, 特点是向单一目的 IP 发生大量报文, 数据集中此类攻击比较常见。
- Probe: 探测攻击, 特点是向大量目的 IP 发生报文, 数据集中此类攻击比较常见。
- U2R: 来自远程机器的非法访问, 特点是一对一通讯, 数据集中此类攻击很少见。
- R2L: 普通用户对本地超级用户特权的非法访问, 特点是一对一通讯, 数据集中此类攻击很少见。

DARPA1998 包含数据包, 其数据格式不能直接用于传统机器学习, 还需要特征工程提取特征。为了克服这个缺点, 研究人员提出了 KDD99 数据集。

3.1.2 KDD99

KDD99[22]是基于 DARPA1998 的数据, 通过特征工程, 提取 41 维特征, 是目前使用最广泛的 IDS 标准数据集之一。其数据的类别与 DARPA1998 相同。特征分为 4 种: 基本特征、内容特征、基于时间特征和基于主机特征。但 KDD99 仍然存在问题:

- 数据重复和冗余很多。
- 数据严重偏斜, 容易导致分类算法的结果偏向多数类。
- 数据过于庞大, 大部分研究者仅使用一个子集。数据杂乱, 研究者在使用时都要进行筛选处理。大家都是使用的一个子集, 导致各个文献之间的准确率没有可比性。
- 数据集使用 Tcpdump 工具进行数据包捕获, 存在丢包问题, 可能会导致重要信息丢失。
- 数据集过于陈旧, 不能反映当前的网络环境。

3.1.3 NSL-KDD

为了克服 KDD99 的问题, NSL-KDD[23]被提出。在 KDD 的基础上, NSL-KDD 进行了以下改进:

- 删除训练集中的冗余记录和重复记录, 所以分类器不会偏向多数类。
- 删除测试集中的重复记录, 使得检测率更为准确。
- 根据不同类别, 对数据集进行采样, 使得各个类别的样本数量尽量均衡。这使得对的准确评估更有效。
- 训练和测试中的记录数量设置是合理的, 使得可以在完整数据集上进行实验。因此, 不同研究工作的评估结果是一致的和可比较的。

NSL-KDD 通过筛选 KDD99 数据集, 一定程度上缓解了数据冗余和数据偏斜问题。但是, 并没有引入新数据, 对于稀缺类别样本数据仍然不足, 数据集陈旧问题也没有解决。

3.1.4 UNSW-NB15

UNSW-NB15[24]由南威尔士大学构建, 配置了 3 台虚拟服务器, 收集网络流量, 通过 bro 工具统计信息, 提取 49 维特征。相比于 KDD99, 该数据集包含更多的新型攻击, 特征种类也更丰富。数据类别包括正常、Fuzzers、Analysis、Backdoors、DOS、Exploits、Generic、Reconnaissance、Shellcode、Worms 等 10 种类型。特征包括流量特征、基本特征、内容特征、时间特征、附加特征和标签特征。

UNSW-NB15 是新数据集的代表, 在一些最近研究中被使用。尽管其影响力不如 KDD99 数据集, 但构造新数据集对于基于深度学习方法的入侵检测是有必要的。

3.1.5 其他数据集

除了上述的数据集外, 在一些文献中还使用了 UNB ISCX2012[25]、CICIDS[26]、LitNet2020[27]、IoT-23[28]、CTU-13[29]。但是, 这些数据集也没有包含基于流的特征, 因而它们都不适用于实时检测的功能。对此, 在文献[30]中, 作者介绍了一个新的数据集, 数据集使用 Netflow 版本 9[31]和 nProbe 工具[32]收集, 包括侦察攻击、拒绝服务攻击和僵尸网络攻击 3 种攻击类型。数据集引入了新型异常-拒绝服务攻

击, 同时在一些字段中引入了时间变量, 因此其适用于实时检测。

3.2 入侵检测数据预处理方法

将原始数据转换为可理解格式的过程称为数据预处理[33]。数据预处理旨在减少数据的大小, 找到数据之间的关系, 对数据进行归一化, 去除异常值, 并提取数据的特征。由于真实世界的数据是非结构化的、不一致的、有噪声的和冗余的, 因此数据预处理有助于将原始数据转换为合适的、经过处理的格式。大量的研究表明, 在IDS中, 数据预处理的工作量通常将达到整个入侵检测过程工作量的60%~80%[34]。入侵检测的数据预处理包含很多的内容, 主要包括数据清洗、数据集成、数据转换和数据规约技术, 而其中研究的比较多的数据预处理技术主要有数据清洗和数据规约技术:

(1) 数据清洗: 分析"脏数据"的产生原理和存在形式, 利用有效的技术方法处理"脏数据", 提高数据集的质量。

- 重复记录清洗: 由于可能存在的输入错误、数据格式和拼写差异等问题, 数据集中可能存在多条记录指向同一个实体。
- 消除噪声数据: 噪声数据是指通过监测获取变量的值相对于真实值出现了偏差或者错误, 主要包括错误数据、假数据和异常数据。为了避免这些噪声数据对数据集分析造成干扰, 需要消除数据集中的噪声数据。
- 缺失值清洗: 由于人为忽视、传输数据包的丢失, 或者目标个人不愿公布某敏感信息等, 存在不完整、包含缺失值的数据是真实大数据集的一个共同点。这些包含不完整的数据会直接影响从数据集中抽取模型和导出规则的准确性, 直接导致 IDS 的性能出现不同程度的降低。当前用于缺失值清洗的方法主要包括忽略不完整的数据和基于填充技术两种。

(2) 数据集成: 消除数据集的冗余数据, 并把数据集在数据库、文件或者数据仓库中统一存储, 形成一个完整的数据集。

- 模式集成方法: 将多个数据源包含不同模式的数据集集成为统一的模式, 用户能够按照统一的模式访问各个数据集。
- 数据复制方法: 将数据源的数据复制到与其相关的不同数据源上, 并保证不同数据源数据的一致性, 可以有效提高数据信息共享利用的效率。
- 综合性集成方法: 将模式集成方法和数据复制方法综合使用, 通常是尽力通过数据复制方式在本地数据源或者单一数据源上满足用户的访问要求, 而对于无法通过数据复制方式来满足的复杂的用户请求则使用模式集成方法。

(3) 数据转换: 对数据进行规格化操作, 将数据源的数据转化为目标格式, 以满足不同的技术或者应用要求。

- 简单函数转换: 应用 x^k、$\log x$、$1/x$、$\sin x$ 等简单的数学函数对数据集的各个属性值进行计算转换。
- 规范化: 对数据集的各个属性数据按比例进行缩放, 是属性值的大小在一个较小的区间, 通常是[0, 1]。

(4) 数据归约: 在尽量保持数据原貌的前提下, 通过减少维度和数据量最大程度地精简数据集, 降低数据集的规模。

- 特征选择: 从一组高维度的特征空间中选择一部分有效的特征来表示原始的特征空间。在特征选择中, 需要判定对标签贡献最大的最佳特征, 并消除没有贡献或者贡献微小的特征[35]。对于成功的高维数据挖掘, 特征选择是重要的一步。目前的特征选择算法可以划分为过滤式、嵌入式和包裹式三类[36]。
- 特征编码: 大多数的机器学习方法和深度学习方法能够可以更好地处理数值数据。对于由分类特征和数字特征组成的数据集, 需要用特征编码对分类特征进行编码[37]。
- 特征缩放: 数据集的不同属性往往会有不同的量纲, 为了避免大量纲的属性对模型学习会有决定性的作用, 需要对数据集的特征进行缩放, 使数据集的所有特征处于统一的范围内。

4 入侵检测系统常用的机器学习方法

入侵检测作为信息安全的重要防护手段, 提供了有效的网络入侵检测措施, 保护网络安全。然而传统的入侵检测系统存在许多不足, 基于机器学习的入侵检测方法能够实现对网络攻击的智能检测, 提高入侵检测的效率[38-39]。

4.1 传统机器学习方法

传统机器学习方法核心思想是利用已知数据来训练算法模型, 并利用该模型预测未知数据。传统机器学习方法研究的时间比较长, 方法相对成熟。传统机器学习算法在入侵检测中的应用, 包括 SVM、KNN、朴素贝叶斯、逻辑回归、决策树、聚类和混合算法等。在实际环境中, 入侵检测系统需要解决的

不仅只是检测模块, 还会注重一些实际运行的问题, 例如效率问题、数据获取问题。但是, 传统的机器学习方法研究的重点在于算法固有的问题进行改进, 例如贝叶斯算法的条件独立性假设, SVM 算法的 RBF 核的效率问题等。

传统机器学习算法相对简单, 计算速度快, 易于理解和实现, 在一些应用场景能够取得较好的效果。典型的传统机器学习算法的优缺点如表 3 所示:

表 3 传统模型的对比分析

Table 3 Comparative analysis of traditional models

算法	优点	缺点	改进思路
SVM[40]	小数据集学习到很好的效果, 模型泛化能力强	大规模数据和多分类任务表现不够理想; 对于带有核函数的 SVM, 参数设置是一个问题	通过粒子群算法对参数进行优化
KNN[41]	适应于大规模数据的分类; 适用于非线性数据的分类; 训练时间短; 对噪声点不敏感	稀有类分类准确率低; 模型计算量大测试时间长; k 值选取问题	使用三角不等式减少比较次数; 利用粒子群算法优化参数
朴素贝叶斯[42]	分类效果稳定, 对噪声不敏感; 可以进行增量学习	对于不满足条件独立性假设的数据, 分类效果不好	引入隐变量, 放宽条件独立性假设的限制
逻辑回归[43]	容易实现, 训练高效; 自动对特征进行缩放	非线性数据分类不好; 容易过拟合	引入正则化项防止过拟合
决策树[44]	自动选择特征; 训练速度快; 可解释性好	分类结果容易偏向多数类; 忽略特征之间的相关性	通过 SMOTE 增加少数类比例; 引入隐变量

4.2 深度学习方法

深度学习是一种基于神经网络的算法, 它模拟人脑的神经元网络进行学习和预测。使用深度学习解决实际问题一般分为以下 6 个步骤: 数学问题抽象、数据获取、数据预处理、模型训练、模型测试、模型部署。

- 数学问题抽象: 明确需要解决的问题, 将实际问题转化为数学模型。对于 IDS, 其目标一般是分类、回归、离群点检测等问题。
- 数据获取: 从实际环境中获取数据, 在大部分文献中, 一般使用标准数据集, 例如 DARPA1998、KDD99 等。数据集对于深度学习研究十分重要, 详细内容会在第 3 节介绍。
- 数据预处理: 数据预处理包括数据清洗、数据集成、数据归一化等。
- 模型训练: 使用特定的深度学习算法对于数据进行拟合, 并通过参数调整的方式来优化模型。
- 模型测试: 使用测试数据对模型效果进行评估, 需要注意的是测试数据不能与训练数据重叠。
- 模型部署: 将训练好的模型部署到实际运行环境, 不仅要求模型的准确率高, 还要具有执行效率高、可扩展性、可移植性等特点。IDS 有实时性需求, 需要模型具有很高的运行效率。

从 2010 年至今, 基于深度学习的 IDS 研究数量迅速增加。IDS 中常用的深度学习算法可以分为两种主要类型: 有监督学习和无监督学习。监督学习依赖于标记数据中的有用信息。分类是监督学习中最常见的任务(也是 IDS 中最常用的); 然而, 手动标记数据既耗时又耗费人力。因此, 缺乏足够的标记数据是监督学习的主要瓶颈。相比之下, 无监督学习从未标记的数据中提取有价值的特征信息, 从而更容易获得训练数据。然而, 非监督学习方法的检测性能通常不如监督学习方法。IDS 中使用的常见深度学习算法如图 2 所示。

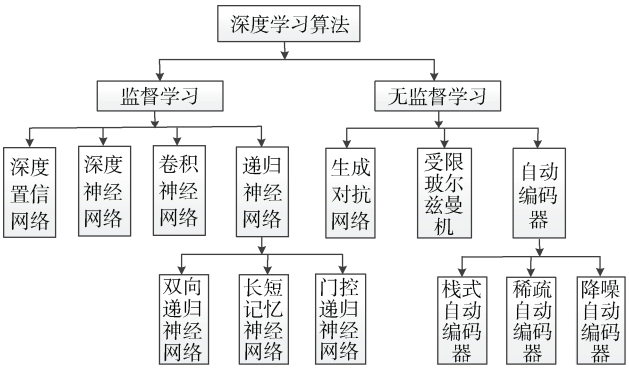

图 2 IDS 中常见深度学习模型

Figure 2 Common deep learning models in IDS

深度学习模型由各种深度网络组成。其中, 深度置信网络(DBN)、深度神经网络(DNN)、卷积神经网络(CNN)和循环神经网络(RNN)是有监督学习模型, 而自动编码器(Auto-encoder)、受限玻尔兹曼机(RBM)和生成对抗网络(GAN)是无监督学习模型。深度学习模型直接从原始数据(如图像和文本)学习特征表示, 无需手动特征工程。因此, 深度学习方法可以以端到端的方式执行。对于大型数据集, 深度学习方法比传

统机器学习方法具有显著优势。在深度学习的研究中，网络结构、超参数选择和优化策略是研究的重点。各种深度学习模型的比较如表 4 所示。

表 4 深度学习模型的对比分析

Table 4 Comparative analysis of deep learning model

模型	数据形式	是否监督模型	功能
Auto-encoder	特征向量	无监督	特征提取; 特征降维; 去噪
RBN	特征向量	无监督	特征提取; 特征降维; 去噪
DBN	特征向量	监督	特征提取; 分类
DNN	特征向量	监督	特征提取; 分类
CNN	特征向量; 矩阵	监督	特征提取; 分类
RNN	特征向量; 时序数据	监督	特征提取; 分类
GAN	特征向量	无监督	数据增强; 对抗训练

自动编码器是一种无监督的神经网络，一般用于特征提取。自动编码器包含两个对称的部分，编码器和解码器，如图 3 所示。编码器从原始数据中提取特征，解码器从特征重构数据，经过训练使得编码器的输入与解码器的输出差异尽量小。解码器能够从编码器提取的特征成功重构出原始数据，说明编码器提取的特征可以反映数据的本质。需要注意的是，整个过程不需要标注信息。自动编码器的有很多变种，代表性有降噪自动编码器[45-46]和稀疏自动编码器[47]等。自动编码器能够用于减少特征空间，而特征表示则位于工作流的下游，以训练不同类型的模型。自编码器在捕获输入特征空间的复杂多元分布方面也做得非常好。因此，自动编码器被广泛用于异常检测任务中。

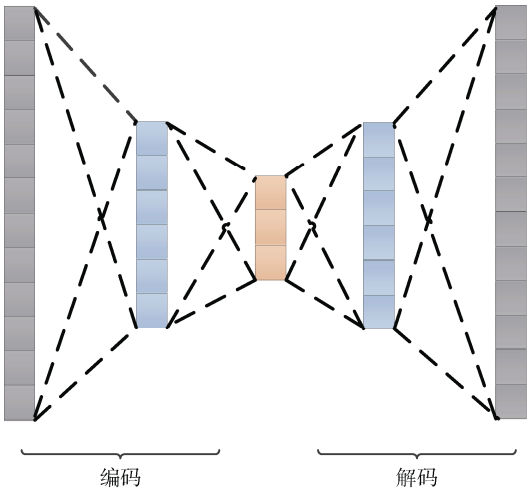

图 3 自动编码器架构

Figure 3 Auto-encoder architecture

受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)是一种可通过输入数据集学习概率分布的随机生成神经网络，包含一层可见层和一层隐藏层，网络各个节点的取值符合玻尔兹曼分布。如图 4 所示在同一层的神经元之间是相互独立的，而在不同的网络层之间的神经元是相互连接的。神经元之间的连接是双向的以及对称的。在网络进行训练以及使用时信息会在两个方向上流动，而且两个方向上的权值是相同的。RBM 不区分前向和反向。常用的 RBM 一般是二值的，无论隐藏层还是可见层，它们的神经元的取值只为 0 或 1。受限玻尔兹曼机是无监督模型，通过贪心的方式训练，一般用于特征提取与数据降噪。受限玻尔兹曼机可以在数据降维的过程中降低特征属性之间的相关性，以获得入侵检测数据的最优低维表示，从理论上提升入侵检测准确率。

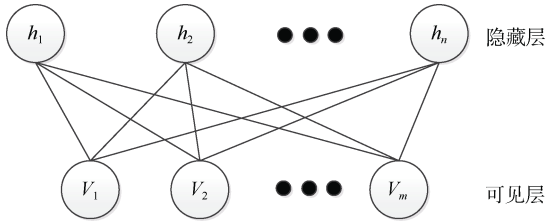

图 4 受限玻尔兹曼机架构

Figure 4 Restricted boltzmann machine architecture

深度置信网络(Deep Belief Network, DBN)使用无监督预训练和监督微调技术来构建模型[48-49]。DBN 含有多个受限玻尔兹曼机层，最后接入一个分类层，如图 5 所示。RBM 是概率生成模型，通过无监督的方式学习数据的联合概率分布。首先通过贪婪的逐层学习算法，一次学习 RBM 的一层。然后通过

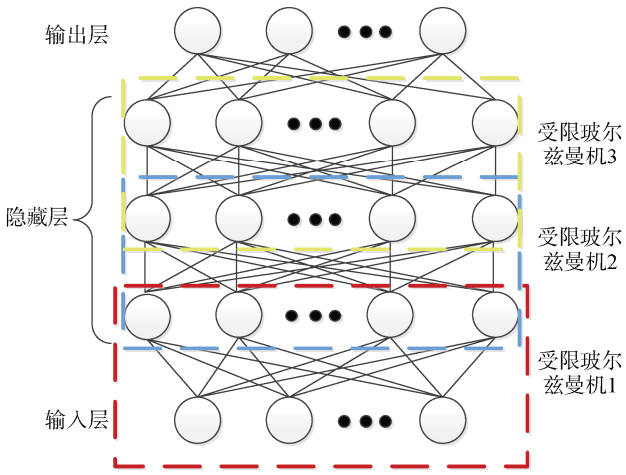

图 5 深度置信网络架构

Figure 5 Deep belief network architecture

带标签的数据进行学习分类层的权重。DBN 既可以用于特征提取又可以用于分类。

卷积神经网络(Convolutional Neural Network, CNN)是一种根据人类视觉原理设计的深度神经网络, 特别适合于处理视觉问题[50-52]。CNN 的结构由卷积层和池化层交替堆积构成, 卷积层用于提取特征, 池化层用于提高特征泛化能力, 如图 6 所示。CNN 通过卷积的权值共享及池化, 来减少网络参数。由于 CNN 适合于处理二维数据, 在进行入侵检测时, 需要将输入数据转为矩阵的形式。CNN 通过提取特征的局部相关性从而提高特征提取的准确度, 通过多层“卷积层-下采样层”的处理对网络中正常行为和异常行为的特征进行深度刻画, 最后通过多层感知机进行正确分类。

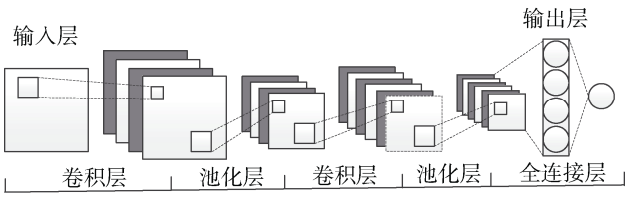

图 6　卷积神经网络架构

Figure 6　Convolutional neural network architecture

循环神经网络(Recurrent Neural Network, RNN)是一类用于处理序列数据的神经网络[53], 常用于自然语言处理[54-57]。序列数据的特点是数据前后有相关关系。网络在设计时, 每一单元不仅输入当前状态, 还有之前的状态。RNN 结构展开如图 7 所示。

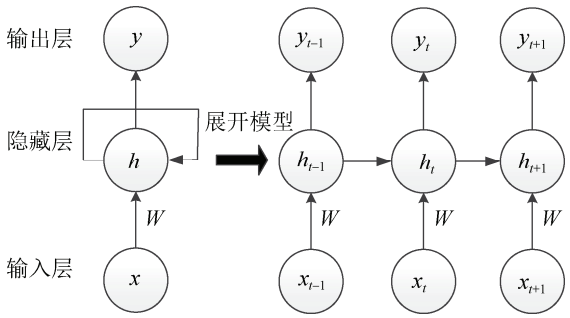

图 7　RNN 网络架构

Figure 7　RNN network architecture

如图 7 所示, 其中所有的权重 w 是相同的, 这导致 RNN 容易陷入梯度弥散和梯度爆炸问题。在实际应用中, 标准 RNN 能够处理的序列长度十分有限。RNN 有许多变种, 例如 LSTM[58]、GRU[59]、双向 RNN[60]等。网络中的流数据通常具有时序信息, 因此采用 RNN 模型能够实现流数据在时序上的相关性分析问题。

为了解决 RNN 的长期依赖问题, 由 Hochreiter 和 Schmidhuber 在 1997 年提出长短时记忆(Long Short Term Memory, LSTM)[58], 并在近期被 Alex Graves 进行了改良和推广。LSTM 在很多领域取得相当巨大的成功, 并得到了广泛的使用。LSTM 通过加入“门”的设计来避免长期依赖问题。每个 LSTM 单元都含有 3 个门, 遗忘门: 遗忘不需要记忆的信息; 输入门: 更新需要记忆的信息; 输出门: 组合长期记忆和短期记忆得到当前的记忆状态。深度学习的入侵检测研究中, LSTM 是使用相当广泛的算法。

生成对抗网络(Generative Adversarial Networks, GAN)包含两个网络部分: G(Generator)和 D (Discriminator)[61]。G 是一个数据生成的网络, 接收一个随机的噪声, 通过噪声生成数据。D 是一个判别网络, 判别数据是真实的还是合成的。在训练过程中, 生成网络 G 的目标是生成尽可能类似真实的数据去欺骗判别网络 D。而 D 的目标就是尽量区分 G 生成的图片和真实的图片。G 和 D 构成了一个动态的“博弈过程”。在最理想的状态下, G 可以生成十分接近真实的数据。于是得到了一个生成模型 G, 它可以用来生成数据。生产对抗网络是一个目前十分热门的研究领域, 在入侵检测中可以用来生成数据扩充数据集, 以缓解入侵检测领域中数据集短缺的问题。同时, 在训练数据中加入对抗样例, 即对抗学习, 可以提高入侵检测系统的鲁棒性。

4.3　入侵检测系统常用的机器学习方法总结

在深度学习广泛研究与应用之前, IDS 采用了传统的机器学习方法进行入侵检测。随着深度学习的盛行, 基于深度学习的 IDS 研究迅速得到了推广。但是, 传统机器学习由于其可解释性好、计算速度快、易于理解和实现等优点, 在很多的应用场景中可以取得很好的效果。而深度学习由于可解释性差、训练效率低、对小型数据集容易过拟合等缺点, 在一些应用场景中反而得不到理想的效果[9-13]。因此, 在实际应用中, 深度学习并不能完全取代传统机器学习方法。本文从理论和应用对传统的机器学习和深度学习的优缺点进行详细的对比, 具体的如表 5 所示。

5　基于深度学习的入侵检测方法

深度学习是一种数据驱动的方法, 数据的类型将对深度学习的结果产生重大影响。因此, 本文以数据类型作为基于深度学习的入侵检测技术的主要分类标志。对于网络入侵检测, 不同类型的数据反映了不同的攻击行为, 包括主机行为和网络行为。其中, 主机行为主要由系统日志反映, 网络行为主要由网

表 5 入侵检测系统常用的机器学习方法对比分析

Table 5 Comparative analysis of machine learning methods commonly used in IDS

特征	传统机器学习	深度学习
可解释性	模型通常具有较好的可解释性	较难解释, 深度神经网络的决策过程较为复杂
数据需求	需精心设计和提取特征	部分情况需要特征工程, 可从原始数据中学习高层次特征
算法复杂性	相对简单	参数量较大, 模型复杂
自动化程度	需手动选择和调整模型参数	自动学习特征表示和模型参数, 但需调整神经网络架构和超参数
适用场景	适用于中小规模数据, 特征明显且可解释性要求高的场景	适用于大规模数据, 复杂模式和高度抽象特征的场景
泛化能力	对于特定任务的泛化能力较好	在大规模数据上训练时具有较强泛化能力, 但对于小数据集容易过拟合
训练时长	训练速度较快	训练速度较慢, 特别是在深层网络和大规模数据集上
参数规模	模型参数较少	模型参数较多, 尤其是深层网络

络流量反映。但是, 实际网络环境中存在多种攻击类型, 每种类型都有一种独特的模式。因此, 需要根据攻击特征选择合适的数据源来检测不同的攻击。例如, DOS 攻击的一个显著特征是在很短的时间内发送许多数据包, 此时流数据适合用于检测 DOS 攻击; 隐蔽通道涉及两个特定 IP 地址之间的数据泄漏, 这样会话数据更适合进行这类攻击检测。在本节, 我们将详细分析基于深度学习面向不同数据类型的网络入侵检测方法。

5.1 基于数据包的入侵检测

数据包是网络通信的基本单元, 包含了每次通信的详细信息。数据包由二进制数据组成, 这意味着除非首先对它们进行解析, 否则它们是不可理解的。数据包由报头和应用程序数据两部分组成。报头是结构化字段, 用于指定 IP 地址、端口和特定于各种协议的其他字段。应用数据部分包含来自应用层协议的有效负载。使用数据包作为 IDS 数据源具有以下 3 个优点: 1) 数据包包含通信内容; 因此, 它们可以有效地用于检测提权(User to Root, U2L)攻击和远程用户(Remote to Login, R2L)攻击。2) 数据包包含 IP 和时间戳; 因此, 它们可以精确定位攻击源。3) 无需缓存即可即时处理数据包; 因此, 可以实时进行检测。然而, 单个数据包既不能反映完整的通信状态, 也不能反映每个数据包的上下文信息, 因此很难检测某些攻击, 例如 DDoS(Distributed Denial of Service)。基于数据包的检测方法主要包括包解析方法和负载分析方法。

网络通信中使用了各种类型的协议, 如 HTTP 和 DNS 等, 这些协议具有不同的格式。基于数据包解析的检测方法主要关注协议头字段。常用的方法是使用解析工具(如 Wireshark 或 Bro)提取头字段, 然后将最重要字段的值作为特征向量处理。

数据包作为一种非结构化数据, 有效载荷可以通过深度学习模型直接处理[62]。浅层模型依赖于手动功能和数据包中的私有信息, 容易产生较高人工成本和隐私泄漏问题。Min 等[63]利用基于文本的 CNN 来检测来自有效载荷的攻击。他们在 ISCX2012 数据集上进行了实验, 检测到具有统计和内容特征的攻击。统计特征主要来自数据包报头, 包括协议、IP 和端口, 内容特性则来自有效负载字段。他们首先将来自不同数据包的有效负载连接起来, , 并通过 skip-gram 字嵌入对连接的有效载荷进行编码。然后, 使用 CNN 提取内容特征。最后, 训练了一个随机森林模型来检测攻击, 实验结果表明模型的精度达到 99.13%。

结合各种有效载荷分析技术可以获得全面的内容信息, 从而提高入侵检测的效果。Zeng 等[64]提出了一种具有多个深度学习模型的有效载荷检测方法。他们采用 3 种深度学习模型(CNN、LSTM 和栈式自动编码器)从不同角度提取特征。其中, CNN 提取局部特征, RNN 提取时间序列特征, 栈式自动编码器提取文本特征。在 ISCX2012 数据集上, 这种组合方法的准确率可以达到 99.22%。

利用无监督学习提取有效载荷特征也是一种有效的检测方法。Yu 等[65]利用卷积自动编码器提取有效载荷特征, 并在 CTU-UNB 数据集上进行实验。其中, CTU-UNB 数据集包含 8 种攻击类型的原始数据包。为了充分利用卷积的特征提取能力, 他们首先将数据包转换成图像, 然后训练卷积自动编码器模型来提取特征, 最后使用学习到的特征对数据包进行分类。实验结果表明, 该方法在测试集上的准确率、召回率和 F_1-*Measure* 分别达到了 98.44%、98.40%和 98.41%。

为了增强 IDS 的鲁棒性, 对抗式学习已经成为一种新的研究方法。对抗式学习不仅可用于攻击 IDS, 同时也是一种提高入侵检测准确率的新方法。Rigaki 等[66]使用 GAN 改进恶意软件检测效果。为了逃避检测, 恶意软件应用程序试图生成与正常数据包类似

的数据包。以恶意软件流感为例, 命令与控制(C&C)数据包与 Facebook 生成的数据包非常相似。他们用主机、服务器和 IP 配置了一个虚拟网络系统, 然后启动了恶意软件流感并训练了一个 GAN 模型, 引导恶意软件生成类似 Facebook 的数据包。随着训练时间的增加, IP 阻止的数据包减少, 通过检测的数据包增加。由于通过 GAN 生成的恶意数据包与正常数据包更为相似, 因此通过分析生成的数据包提高了 IDS 的鲁棒性。

针对现有方法在有效识别和防御零日攻击的局限性, Liu 等[67]通过特征的提取和归一化, 引入模型进行训练、比较和改进, 提出了 4 种基于机器学习的入侵检测算法, 并比较了这 4 种算法的优缺点。然后, 设计了一种基于随机森林算法的入侵检测系统, 系统在 SQL(Structured Query Language)注入、命令执行、跨站点脚本攻击等测试中具有很好的识别效果, 能够有效区分异常流量和正常流量。实验结果表明系统的准确率可以达到 95%, 召回率可以达到 96%, F_1-*Measure* 可以达到 95%。在系统模型算法的选择上, 随机森林算法比决策树等其他算法能够更好地解决过拟合问题, 这也是减少新样本的假阴性和假阳性的关键。虽然系统对模糊和复杂的有效载荷具有很好的识别效果, 但在模型的特征提取和不同数据集适用性上还有很大的改进空间。

另外, Borgioli 等[68]提出了一种新的双自动编码器架构, 用于在网络边缘设备上实现基于异常数据包识别的实时入侵检测。其核心思想是利用现代自编码器技术对接收到的数据包进行重构, 并根据重构误差检测异常和攻击数据包。为此, 他们首先提出了两种基于最先进的序列自动编码器(autoencoder, AE)架构的解决方案。然后, 引入了多状态记忆 AE(MSM AE), 这是一种利用多个并行 LSTM 自动编码器的新型 AE 架构, 每个 LSTM 自编码器具有不同的嵌入尺寸, 用于提高检测精度。最后, 在真实网络的流量数据集上进行了广泛的实验评估。实验结果表明, 他们所提出的体系结构在检测各种常见攻击类型方面优于现有方法, 并能够以无监督的方式检测网络攻击且无需标记数据。

5.2 基于流数据的入侵检测

流数据由按时段分组的数据包组成, 是 IDS 最广泛的数据源。使用流数据检测攻击有两个好处: 1) 流数据可以代表整个网络环境, 能够检测大多数攻击, 尤其是 DOS 和 Probe。2) 与数据包的解析或会话重组相比, 流数据预处理更加简单。但是, 流数据忽略了数据包的内容, 因此其对 U2R 和 R2L 的检测效果难令人满意。另外, 提取流数据特征时数据包必须是缓存的数据包, 因此它涉及一些滞后。此外, 流数据具有较强的非均匀性可能导致检测效果差, 而流量分组是解决此问题的常用方法。基于流数据的入侵检测主要包括特征工程和深度学习方法。

深度学习方法可以直接处理原始数据, 它们在学习特征的同时可以进行分类。因此, 基于深度学习的检测方法能够自动学习特征, 并且以端到端的方式工作。Potluri 等[69]提出了一种基于 CNN 的检测方法, 并在 NSL-KDD 和 UNSW-NB 15 数据集上进行了实验。由于 NSL-KDD 和 UNSW-NB 15 数据集中的数据类型是特征向量, 且 CNN 擅长处理二维(2D)数据, 所以他们首先将特征向量转换为图像。该模型得到的特征是一个热编码, 特征尺寸从 41 增加到 464。然后, 将每个 8 字节块转换为一个像素, 并用 0 填充空白像素, 这样将特征向量转换为 8×8 像素的图像。最后, 构建了一个三层 CNN 来对攻击进行分类, 并将其模型与其他深度网络(ResNet 50 和 GoogLeNet)进行了比较。实验结果表明, 该模型在 NSL-KDD 和 UNSW-NB 15 上的准确率分别达到 91.14%和 94.9%。

无监督的深度学习模型也可以用于提取特征, 然后使用浅层模型进行分类。Zhang 等[70]使用稀疏自动编码器提取特征, 使用 XGBoost 模型检测攻击, 并使用 NSL-KDD 数据集进行验证分析。由于 NSL-KDD 数据集的不平衡性质, 他们使用 SMOTE 算法对少数类进行过采样, 并将多数类划分为许多子类, 这样可使每个类都是平衡的。同时, 稀疏自动编码器在原始自动编码器中引入稀疏约束, 增强其检测未知样本的能力。最后, 他们使用 XGBoost 模型对数据进行分类。实验结果表明, 他们的模型在正常、DOS、Probe、R2L 和 U2R 类上的精度分别为 99.96%、99.17%、99.50%、97.13%和 89.00%。

同样地, 图卷积神经网络能够有效提取非结构化的图数据特征, 而网络通常能够利用图结构的形式进行表示。因此, Lo 等[71]提出了一种基于 E-GraphSAGE 的新型网络入侵检测系统, 该模型将图卷积神经网络(GNNs)应用于网络入侵检测, 且在 UNSW-NB 15 和 NF-UNSW-NB 15 的数据集准确率分别达到了 98.15%和 96.17%。该模型初步展示了 GNNs 在网络入侵检测方面的潜力, GNNs 与现有的入侵检测系统相结合来提高检测准确率将成为网络入侵检测领域重要的研究方向之一。由于网络环境的动态性和时变性, 网络入侵数据被大量正常样本淹没, 导致模型训练和结果检测的样本不足。虽然深

度学习模型在大数据分析方面取得了巨大进步, 但是在小型数据集上它们的性能并不理想。对抗式学习方法可以提高小型数据集的检测精度, 对此, Zhang 等[72]使用 GAN 进行了数据扩充。由于 KDD99 数据集既不平衡又缺乏新数据, 这导致深度学习模型的泛化性较差。为了解决这些问题, 他们使用了 GAN 来扩展数据集。GAN 模型生成的数据与 KDD99 的流量数据相似, 将生成的数据添加到训练集中可以检测攻击变体。他们选取了 8 种类型的攻击, 并将对抗性学习在原始数据集与扩展数据集上的准确度进行了比较。实验结果表明, 对抗性学习在 8 种攻击类型中提高了其中 7 种的准确度。但是, 作者只是通过扩展数据集来缓解流数据的不平衡问题, 并没有从模型上解决流数据的不平衡问题。

为了缓解流数据的强非均匀性问题, Rahi 等[73]提出并分析了一种新的基于混合采样和 DHN 分组的入侵检测系统。他们使用 OSS(One-Side Selection)和 SMOTE(Synthetic Minority Oversampling Technique)为模型训练提供平衡的数据集, 这在一定程度上克服了不平衡数据集在模型训练存在的不足, 同时还可能将模型的训练时间缩短一半。另外, 针对复杂、多维的网络威胁, 他们还设计了一种适用于所提出的 DHN 模型的网络数据准备方法。然后, 使用堆叠 CNN 创建的分层网络模型, 对输入数据进行分类。该模型利用深度学习的特性, 通过递归多级学习自动提取特征。最后, 利用 UNSW-NB 15 和 NSL-KDD 两个数据集对所提出的方法的总体性能进行了测试, 实验结果表明该方法优于其他基于统计显著性检验的分类器。同样, Li 等[74]探讨了使用机器学习相关方法解决计算机网络的入侵检测和带宽预测问题。同时, 提出了一种多级分类模型, 巧妙地解决了样本类别不平衡的数据集分类问题, 并提出了一种用于网络入侵带宽消耗预测的 LSTM 模型。

流量包括一段时间内的所有流量, 其中许多类型的流量可能在攻击检测中充当白噪声。使用这些数据训练机器学习模型可能会导致过度拟合。一种自然的方法是对流量进行分组以减少异构性。分组方法包括基于协议的方法和基于数据的方法。不同协议的流量特性存在显著差异, 因而按协议分组流量是提高准确性的有效步骤。Teng 等[75]利用 KDD99 数据集的数据, 提出了一种基于协议分组的检测方法, 该方法涉及多种协议。他们首先根据协议类型划分数据集, 只考虑 TCP、UDP 和 ICMP 协议。然后, 根据这些不同协议的特点, 为每个子数据集选择特征。最后, 在 3 个子数据集上训练 CNN 模型, 平均准确率可达 89.02%。基于数据特征的分组是另一种流量分组方法, 聚类是一种典型的基于数据特征分组的方法。Ma 等[76]提出了一种基于 DNN 和光谱聚类的检测方法。由于流数据的不均匀性可能导致模型的分类精度降低, 他们首先将原始数据集划分为 6 个子集, 每个子集都是高度同质的。然后, 在每个子集上分别训练 DNN 模型。实验表明他们的方法在 KDD99 和 NSL-KDD 数据集上的准确率达到 92.1%。

在深度神经网络中引入注意力机制, 聚焦于流数据中的关键区域, 能够提高网络流量异常检测准确率。Sethi 等[77]提出了一种基于注意力机制的入侵检测系统。他们的系统在多个分布式代理中采用深度强化网络单元, 并使用注意力机制有效地检测和分类高级网络攻击。同时, 将去噪自动编码器(Denoising Autoencoder, DAE)与模型相结合, 进一步提高其鲁棒性。最后, 在 NSL-KDD 数据集上进行了验证, 实验结果表明模型在正常、DOS、Probe、R2L 和 U2R 类上的精度分别为 97.7%、92.25%、97.50%、96.40%和 87.30%。

对于流数据, 相邻的数据往往具有强关联性。另外, 针对相邻时刻数据具有相关性的问题, 强化学习可以有效地根流据数据当前时刻数据与前一时刻数据表示的状态和动作推断下一时刻可能出现的状态和动作。因此, 结合深度学习与强化学习, 利用上一时刻的深度学习预测模型和当前时刻的模型推断下一时刻可能出现的状态和动作, 是提高 IDS 效率的一种有效方法[78]。深度强化学习(Deep Reinforcement Learning, DRL)因其轻量级和自适应性, 是入侵检测研究中的一个极具潜力的方向。然而, Deep-RL 所依赖的神经网络很容易受到黑客的攻击, 通过对恶意流量应用经过精心地计算修改, 恶意示例可以逃避检测[79]。

同样的, Tao 等[35]也考虑了不同数据特征之间的关系, 并提出一种结合 Attention 和 BiLSTM-DNN 的入侵检测模型(ABD)。ABD 利用 Attention 对输入数据进行初步特征提取, 读取不同特征之间的关系, 通过 BiLSTM 提取长距离依赖特征, 并利用 DNN 进一步提取深层特征, 最后通过 SoftMax 分类器进行分类。与其他方法相比, ABD 的准确度得到了提高, 通过手动设置参数初始化的分布, 使模型的训练过程更加稳定。然而, 对于样本数据分布不均匀的数据集, ABD 在某些分类的检测率较低, 并且 ABD 存在训练时间较长的问题。另外, Alalmaie 等[80]提出了一种具有双向长短期记忆的注意力卷积神经网络(CNNBiLSTM), 该网络利用自动编码器瓶颈特征用

于网络入侵检测系统。他们首先利用自动编码器的压缩瓶颈功能, 同时使用 CNN 分析提取的特征之间的空间关系。在 CNN 上应用多头自注意模块来聚合特征, 并在下一层关注 BiLSTM 的 CNN 特征图中最重要的部分。然后, 使用两个 BiLSTM 层进行分类。为了减少数据不平衡的问题, 他们还建议使用平衡采样器对 CNN 进行预训练。实验结果表明, 对于 UNSW-NB 15 数据集, 他们提出的方法在 6 个和 10 个类别上的分类准确率分别为 89.79%和 91.72%, 优于现有方法。

为了扩展入侵检测的研究宽度, 提升 IDS 的性能, 在文献[81]中 Merzouk 等使用 NSL-KDD 数据集训练了最先进的深度强化学习检测代理, 并使用几种对抗性攻击方法评估了其性能。实验证明, 当使用对抗性攻击来干扰恶意数据包使其被归类为良性数据包时, 检测性能会显著下降。因此, 将深度学习与强化学习相结合应用于入侵检测是一个重要的研究方向, 但是如何防止恶意示例逃避检测是其面临的主要挑战。另外, Sun 等[82]从特征工程角度来提高 IDS 的性能。他们提出了一种混合多策略 Aquila 优化器(HMAO), HMAO 同时具有 Aquila 优化器和 Harris Hawks 优化器的优点。HMAO 可以获得卷积神经网络提取的特征的最优子集, 用于训练 IDS 分类器。同时, 他们介绍了一种将 CNN 和 HMAO 相结合的 IDS 模型, 并在 UNSW-NB 15 数据集上对所提出的 IDS 模型进行了评估。实验结果表明, HMAO 算法在寻找优秀解的能力上大大优于原算法。此外, 在 IDS 的应用中, HMAO 在准确率、召回率、误报率和 F_1-*Measure* 等方面优于现有相关工作中的几种特征工程方法。

5.3 基于会话数据的入侵检测

会话是两个终端应用程序之间的交互过程, 可以表示高级语义。会话通常根据 5 元组(客户端 IP、客户端端口、服务器 IP、服务器端口和协议)进行划分。使用会话进行检测主要有两个优点。1) 会话适用于检测特定 IP 地址之间的攻击, 例如隧道攻击和特洛伊木马攻击。2) 会话包含攻击者和受害者之间的详细通信信息, 这有助于定位攻击源。但是, 会话持续时间可能会有很大的变化。因此, 会话分析有时需要缓存许多数据包, 这可能会增加延迟。基于会话的检测方法主要包括基于统计的特征和基于序列的特征。

与流数据不同, 会话中的数据包具有严格的顺序关系。其中序列特征主要包括包长序列和时间间隔序列, 通过分析序列可以获得详细的会话交互信息。目前, 基于序列特征的检测大多采用 RNN 算法。

对原始数据进行编码是 RNN 方法的常见预处理步骤。单词袋(BoW)模型是一种常用的文本处理技术。Yuan 等[83]利用 UNB ISCX 2012 数据集提出了一种基于 LSTM 的 DDoS 检测方法。他们首先从数据包中提取 20 维特征, 并采用 BoW 编码。然后, 按顺序连接数据包, 得到一个大小为 $m \times n$ 的矩阵, 其中 m 是会话中数据包的数量, n 是数据包的维度, m 和 n 都是可变的。最后, 他们训练了一个 CNN 提取局部特征, 并训练了一个 LSTM 对会话进行分类。实验结果表明, 该方法的准确度、精密度、召回率和 F_1-*Measure* 分别达到 97.606%、97.832%、97.378%和 97.601%。

Bow 的缺点之一是它不能表示单词之间的相似性, 而单词嵌入方法克服了这个问题。对此, Radford 等[84]提出了一种基于 BiLSTM 的会话检测方法。由于 LSTM 在 NLP 方面取得了巨大的进步, 他们将会话数据表示为一种特定的语言, 并在 ISCX IDS 数据集上进行了实验。首先根据 IP 地址对数据包进行分组以获得会话, 然后用单词嵌入的方式对会话进行编码, 最后训练了一个 LSTM 模型来预测异常会话。为了利用上下文信息, 作者还采用 BiLSTM 模型从两个方向学习序列特征。

除了文本处理技术外, 字符级 CNN 是一种新的编码方法。Wang 等[85]提出了一种分层深度学习检测方法, 其中会话不仅包含数据包内容, 还包含数据包时间序列。具体的, 他们设计了一种分层深度学习方法, 使用 CNN 学习低级空间特征, 使用 LSTM 学习高级时间特征, 其中时间特征基于空间特征。最后, 他们在 DARPA 1998 和 ISCX2012 数据集上进行了 598 项实验, 首先应用 CNN 从数据包中提取空间特征, 然后将空间特征按顺序连接起来, 并使用 LSTM 模型提取时间特征。实验结果表明, 该模型的准确率介于 99.92%~99.96%, 检测率介于 95.76%~98.99%。

现有的研究通常针对特定领域存在的问题而提出特定的解决方法。为了了解各种技术在不同领域问题中的表现, Arikkat 等[86]将不同的机器学习和深度学习技术的分析扩展到 DDoS 攻击检测、恶意 URL 检测和 Tor 流量分类 3 个不同的问题领域。在他们的比较研究中, 利用 ISCXTor2016、CIC-DDoS2019 和 ISCX-URL2016 三个公开可用的数据集训练 6 种深度学习模型, 进行多类和二元分类。同时, 他们采用 K-Best 特征选择方法选取用于训练模型的最佳特征集, 并使用 F_1-*Measure* 来评估不同学习模型对网络

流量分类的性能。实验结果表明, 对于深度学习模型, 具有随机森林的自动编码器在多类和二进制问题上的 F_1-*Measure* 分别可以达到 89%和 100%。

5.4 基于日志数据的入侵检测

日志是操作系统或应用程序的活动记录, 它们包括系统调用、警报日志和访问记录。日志数据是丰富的信息源, 它跟踪系统中发生的几乎所有事件, 通常作为事件发生后取证调查的基础[87]。日志具有明确的语义, 在 IDSs 中将日志作为数据源有 3 个好处: 1) 日志记录了适用于检测 SQL 注入、U2R 和 R2L 攻击的详细内容信息。2) 日志通常包含有关用户和时间戳的信息, 可用于跟踪攻击者并显示攻击时间。3) 日志记录了整个入侵过程, 因此检测结果具有可解释性。然而, 一个问题是日志分析依赖于网络安全知识。此外, 不同应用程序的日志格式不一致, 导致可扩展性低。基于日志的攻击检测主要包括基于规则和机器学习的混合方法、基于日志特征提取的方法、基于文本分析的方法和基于异常检测的方法。

许多入侵检测系统存在较高的虚警率, 这会导致在许多无意义的警报中嵌入真正的攻击。通过机器学习模型对警报进行排序是一种可能的解决方案。为了降低虚警率, Meng 等[88]提出了一种 KNN 方法来过滤警报。他们在真实的网络环境中进行实验, 使用 Snort 生成警报, 并训练 KNN 模型对警报进行排序。该实验总共有 5 个威胁级别, 实验结果表明 KNN 模型将警报数量减少了 89%。

某些 IDS 执行类似于人机交互的功能, 其中警报通过深度学习进行排序, 以减少分析人员的工作量。McElwee 等[89]提出了一种基于 DNN 的警报过滤方法。他们首先收集了 McAfee 生成的日志, 然后训练了一个 DNN 模型, 用于在日志中找到重要的安全事件。接着, 安全专家对提取的重要事件进行分析, 并将分析结果作为训练数据对 DNN 模型进行增强, 从而形成一个互动和提升循环。这种形式的模型可以减少分析人员的工作量并加速安全分析。

入侵行为可能会留下系统调用的痕迹, 使用分类算法分析这些系统调用可以检测入侵。Tran 等[90]提出了一种 CNN 方法来分析系统调用。每个涉及操作系统的底层操作都将使用系统调用。因此, 分析系统调用路径可以重现完整的入侵过程。他们在 NGIDS-DS 和 ADFA-LD 数据集上进行了实验, 其中包括一系列系统调用。他们首先使用滑动窗口提取特征, 然后应用 CNN 模型进行分类。实验结果表明, CNN 善于发现局部特征关系, 并从系统调用中检测异常行为。

模型解释是另一个重要的研究方向, 引起了广泛的关注。Tuor 等[91]提出了一种可解释的深度学习检测方法, 使用来自 CERT 内部威胁数据集的数据, 该数据集由系统日志组成。他们首先使用滑动窗口提取了 414 个维度特征, 然后采用 DNN 和 RNN 对日志进行分类。DNN 根据日志内容检测攻击, RNN 根据日志序列检测攻击。实验表明, 所提出的方法将分析工作量减少 93.5%, 检测率达到 90%。此外, 他们还将异常分数分解为每个行为的贡献, 这是一个有用的分析。总之, 可解释的模型相较于不可解释的模型更有说服力。

日志通常规模较大, 针对每条日志信息进行标识显然不合理, 因此监督学习是不适用的。无监督学习方法通常用于未标记日志的特征学习。Bohara 等[92]提出了一种基于企业环境的无监督学习检测方法。他们在 WAST 2011 Mini Challenge 2 数据集上进行实验, 首先对 log 信息进行模板抽取用于构建 log key, 并采用 LSTM 模型对原始 log 信息进行特征学习。由于每个特征的影响不同, 作者使用皮尔逊相关系数选择特征。然后, 采用 K-means 和 DBSCAN 算法对日志进行聚类。通过测量显著的聚类特征, 聚类与异常行为相关联。最后, 手动分析异常集群以确定特定的攻击类型。

在基于日志的检测中, 常用的方法是从日志中提取文本特征, 然后进行分类。在分析文本时, 少量的关键词对整体有很大的影响, 因此, 网络安全领域的关键词有助于提高检测效果。Uwagbole 等[93]提出了一种物联网(IoT)的 SQL 注入检测方法。他们从真实环境中收集并标记日志。日志提供 SQL 注入攻击的上下文信息。他们首先从日志中提取了 479000 个高频词, 并添加了 862 个出现在 SQL 查询中的关键字从而组成一个字典。然后, 从日志中删除重复和丢失的记录, 并使用 SMOTE 平衡数据。最后, 他们使用 CNN 提取特征并进行分类, 准确度、查全率、召回率和 F_1-*Measure* 分别达到 98.6%、97.4%、99.7%和 98.5%。

在实际的网络环境中, 正常样本占大多数, 异常样本很少。One-class 分类是一种无监督学习方法, 只利用正常样本进行训练, 解决了异常样本不足的问题。Vartouni 等[94]提出了一种基于孤立深林模型的 web 攻击检测方法, 使用了 CSIC 2010 数据集的数据。他们首先利用 n-gram 从 HTTP 日志中提取了 2572 个维度特征, 然后利用自动编码器来删除不相关的特征, 最后训练了一个孤立森林模型来发现异常网络。实验结果表明, 所提出的方法准确率可达到 88.32%。

尽管基于日志数据的入侵检测方法取得了不错的研究成果, 尤其是对于已知攻击类型方面具有很高的效率, 但是对于未知的攻击或不存在签名的攻击变体往往会显得束手无策, 并且需要频繁更新知识库[95]。另外, 日志数据通常以多种不同的格式出现, 不同记录数据存在着关联关系, 并且日志事件之间的依赖性以及事件参数与应用程序复杂的底层工作流程相关, 这些都会给数据分析技术带来不同的挑战。基于异常的 IDS 通过利用自学习技术来缓解这些问题, 该技术自动生成系统正常行为的模型, 并将所有偏离此基线的行为检测为潜在的恶意行为[96]。在文献[97]中作者推出了开源工具 AMiner, 一种用于日志数据的模块化分析工具, 可以快速解析事件, 使用机器学习技术进行入侵检测, 并提供检测报告接口。具体的, AMiner 首先接收来自不同来源的日志, 并使用自定义和自动生成的解析器模块来处理数据; 其次, 利用检测器模块为正常行为创建模型并检测异常; 最后, 将检测结果传送给系统操作员。该文最大的贡献是提出开源工具 AMiner, 将基于日志数据的入侵检测方法按功能进行模块化设计, 但是需要进一步研究模型学习方法以提高 AMiner 的性能。

5.5 基于深度学习的入侵检测方法总结

深度学习作为机器学习的一个重要分支, 包含多种结构的深度网络, 其中 CNN、RNN、DNN、DBN 可以用于分类, Auto-encoder、RBM 可以用于特征提取、特征压缩、数据降噪和数据重构, GAN 可以用于数据增强。深度学习方法可以从多个层面提升入侵检测的准确度, 特别是图卷积神经网络(GNN)初步显示了在入侵检测应用中的优势, 为入侵检测提供了新思路。表 6 列出了不同方法最新研究成果的对比分析, 从表 6 可以发现各种方法都有各自的优缺点, 研究的重点在于如何克服算法固有的缺点, 提升检测的准确度与效率。总之, 使用深度学习方法进行入侵检测是一个可行的方法, 在不同的应用场景, 例如云计算、智能电网、大规模通信网络等都有突出的表现。

表 6 基于深度学习的网络入侵检测方法

Table 6 Network intrusion detection method based on deep learning

方法	代表论文	深度学习方法	优点	缺点
基于数据包的深度学习方法	文献[63-68]	CNN, LSTM , Auto-encoder 和 GAN	1) 可以用于检测 U2L 和 R2L 攻击 2) 可以精确定位攻击源 3) 无需缓存即可即时处理数据包	不能反映完整的通信状态
基于数据流的深度学习方法	文献[49,69-82]	CNN, Auto-encoder, XGBoost, GAN, DNN, Attention, BiLSTM 和 GCN	1)流数据可以代表整个网络环境 2) 预处理简单	具有的强非均匀性, 会导致检测效果差
基于会话数据的深度学习方法	文献[83-89]	LSTM 和 CNN	1) 会话适用于检测特定 IP 地址之间的攻击 2) 能够定位攻击源	需要缓存大量数据, 会增加延迟
基于日志数据的深度学习方法	文献[90-97]	CNN, DNN, RNN, K-means 和 DBSCAN	1) 适用于检测 SQL 注入、U2R 和 R2L 攻击的详细内容信息 2) 可以对攻击者进行跟踪 3) 具有一定的可解释性	日志信息通常规模大, 不具备标签信息, 无法使用监督的深度学习方法

深度学习在网络安全防护中取得了重大成功, 但是这还远远没有达到可以高枕无忧的状态。首先, DL 技术本身还存在着一些局限性, 当把 DL 技术应用于网络的入侵检测时也难免会存在各种问题。另外, 目前大部分研究注重 DL 的准确率、检测率和误报率等指标, 忽略时间开销、存储开销、可扩展性等性能指标。为了追求高准确率, 采用十分复杂的算法和预处理方法, 导致时间效率低下, 不适于实际检测。除了 DL 技术自身的局限性, 实际应用场景的复杂多样性也会带来各种挑战。随着计算机技术和互联网技术的发展, 各种网络的规模越来越大、各种新型的运用软件层出不穷、软件系统的规模越来越大、网络和软件系统的逻辑变得越来越复杂, 这些给计算机网络带来了更多可能的安全隐患, 也会给 IDS 提出更多的要求。而对于攻击者而言, 为了逃避基于 DL 的 IDS 对他们攻击行为的检测, 攻击者会研究相应的对抗性攻击策略, 这也会入侵检测系统带来严峻的考验[98]。具体的, 基于 DL 的 IDS 仍然存在着以下的挑战:

(1) 数据集稀缺。当前入侵检测应用最广泛的数据集是 DARPA/KDD, 但这些数据集比较陈旧且包含的入侵类型有限, 训练的模型不能有效检测各种新型的入侵行为, 需要新的数据集来替代。但是, 目前新提出的数据集还达不到 DARPA/KDD 数据集的影响力。另外, 构建新的数据集不仅需要丰富的领域知识, 还需要专家进行标注, 人工成本非常高。

(2) 数据样本不平衡。网络安全数据中正常样本

通常远远大于异常样本, 这样的结果是容易使训练出的模型在多数情况下会明显地偏向于正常样本, 从而严重影响 IDS 检测结果的准确性。

(3) 网络安全数据复杂且变化大。首先, 网络攻击者为了避免攻击行为被防御系统发现, 通常会不断采用新的、更复杂的、更隐蔽的攻击方法, 此时触发底层的网络流量特征发生变化的数量往往就会增加。因此, 入侵检测需要处理的数据往往达到上百维。另外, 当目标网络遭受攻击时, 体现在底层的网络流量特征的变化程度会存在明显的差异, 即不同的分类特征的权重不尽相同。

(4) 不可解释性。大部分深度学习方法是一个黑盒, 这些方法仅仅给出了检测结果而没有可解释的判断依据。但是, 在网络安全中每一个决策都是很谨慎的, 没有任何依据的结论是很难令人信服并成为做出决策的根据。

(5) 泛化能力较低。深度学习方法具有一定的泛化能力, 但是对于完全陌生的数据, 效果并不理想。大部分的研究都是在事先标注好的数据集上进行, 如果数据集不能完全覆盖全部典型样例, 即使在测试集上取得很好效果, 也不能保证在实际环境中得到理想的检测效果。

(6) 时效性问题。时效性主要体现在两方面, 首先是入侵检测的时效性, 海量高维的检测数据会影响检测系统的时间效率, 但是快速地检测入侵行为有利于用户及时执行各种防护措施, 降低入侵行为带来的危害。另外, 网络安全数据的时效性很强, 例如某个 IP 曾经属于一个僵尸网络, 过一段时间又成为正常 IP, 这就导致深度学习方法进行入侵检测需要进行增量学习, 不断更新模型。

(7) 效率较低。大部分研究为了提升模型的检测率、准确率等指标, 常常采用复杂的模型和预处理方法。为了保证模型的效果通常需要逐层训练, 这直接导致计算量大、训练速度缓慢。

(8) 深度学习自身的安全性问题。深度学习模型由各种深度网络组成, 当运用于入侵检测系统时其原理是透明的。此时, 攻击者容易实现一些可以绕过 IDS 的攻击策略, 实现对目标网络的攻击。

6 总结与展望

IDS 是网络系统重要的安全保障之一。深度学习因为其具有自动学习能力、模型容量大等特点, 在网络安全防护中取得了重大成功。但是, 随着攻击行为的不断升级和网络数据量的快速增长, 再加上近年来内部威胁、零日漏洞、加密攻击等行为的出现, 入侵检测依然面临着一系列挑战。

为了更好地将基于深度学习的 IDS 应用到实际环境中, 除了考虑深度学习准确率的相关指标外, 还需要重点考虑深度学习模型的时间效率、可解释性、泛化能力、自身的安全性等的重要性。通过对相关研究的总结分析, 本文对基于 DL 的 IDS 未来的发展趋势进行如下讨论:

(1) 提升无监督学习作用。首先, 无监督学习不需要大规模标注数据, 在数据集短缺的情况下可以提升监督的效果。另外, 还可以选择一些无监督学习方法用于数据集的标记, 为 DL 模型的训练提供有效的数据集。

(2) 解决样本数据的不平衡。解决数据样本的不平衡可以有效克服模型的偏向性问题, 从而提高 IDS 检测结果的准确性。对于深度学习而言, 解决数据样本不平衡问题可以从数据、算法设计、模型评价三方面出发。从数据角度而言, 可以扩大数据集、数据集重采样、人工产生数据样本, 甚至可以把微量样本当成异常数据进行处理; 从算法角度出发, 可以增加小样本数据的权值、对小样本进行过采样, 还可以重构 ID 模型; 而从模型评价方面, 可以选择合适的性能指标用于评估训练好的模型性能。

(3) 特征选取。首先, 不同的入侵行为会触发不同的底层网络流量特征。因此, 进行入侵数据收集时候需要选择合适的特征参数。其次, 当目标网络遭受攻击时, 体现在底层的网络流量特征的变化程度会存在明显的差异。因此, 计算不同的分类特征的权重可以有效提高模型的准确度。

(4) 可解释性研究。在网络安全中, 决策需要十分谨慎。在深度学习输出结果的同时, 需要给出相关的可解释依据[99], 使结果令人信服, 指导人们进行决策。对 DL 进行可解释性研究不仅可以提高模型的透明度, 还可以提高模型的可信度。针对 IDS 的应用需求, 学者一方面可以设计本身具有良好可解释性的模型, 另一方面可以利用可解释的方法对已设计好的模型进行解释为具体的应用决策提供依据。

(5) 对特定攻击的专业化检测。很多的研究都是针对非特定的攻击, 但也有少部分研究对于特定攻击进行检测, 例如 DOS 攻击[100-104]、钓鱼网站[105]等。这些方法通常会根据领域知识, 对特定问题进行专门化处理, 在这些特定攻击的检测上面效果更好。

(6) 模型结构的选择与优化。深度神经网络的结构对 IDS 最终的检测结果有很大的影响。在实际的应用中, 应该根据各种模型的特性, 结合具体的检测任务, 确定最优的模型结构。另外, 预训练与微

调[100-104]可以提升训练效果，mini-batch[105]、BN[83,106]、改进优化算法[107-109]加快模型收敛，dropout[110-112]用于防止过拟合。

(7) 提升模型的性能。首先，需要提升模型的训练效率。其次，需要提升模型的检测效率。入侵检测系统需要进行实时检测，才能最大程度降低危害。在实际应用中，通常需要在检测效果和效率之间找一个合适的折中。而并行计算[113-114]和 GPU 加速[115-116]是提升效率的重要研究方向。再次，与基于规则系统互补，提升检测结果准确度。基于规则检测方法的优点是误报率低、检测效率高，缺点是不能检测新型攻击。基于深度学习的检测方法的优点是漏报率低，可以检测新型攻击，缺点是误报率高。两者的优势是互补的，通过深度学习方法提升 snort[117-120]等基于规则方法，可以获得一种具有低误报率、低漏报率的系统。最后，扩展应用场景。基于深度学习的检测方法可以方便应用于很多新场景，例如云计算[121-123]、物联网[124-128]、智能电网[129,133]等。

(8) 提高安全性和易用性。IDS 作为保护网络安全的产品，其自身的安全也是至关重要的。因此在使用 IDS 时，尽可能避免把自身的安全问题带到网络系统中。另外，在实际环境中对于 IDS 的易用性的要求也越来越高。友好的人机交互界面、自动化的系统维护、多样化的报表和预警等，这些都是一个优秀的IDS应该具备的特性，也是IDS后续发展应该努力的目标。

7 结束语

入侵检测是网络系统重要的安全保障之一，基于深度学习的入侵检测是当前的一个研究热点。本文介绍了入侵检测系统的基本概念、数据集和数据预处理方法、评估方法、常用的机器学习方法，然后以数据类型作为主要的分类标志对最近的代表性研究进展进行分类总结，最后对基于深度学习的入侵检测面临的挑战和未来的发展进行分析总结。总之，随着网络技术的发展，入侵检测依然面临着一系列挑战，而深度学习是入侵检测的一个重要技术。本文旨在对基于深度学习的入侵检测的相关技术进行分析总结，为相关人员的后续开展工作提供帮助。

参考文献

[1] CNCERT/CC[Z]. https://www.cert.org.cn/publish/main/44/index.html.Jun. 2024.

[2] Anderson J P. Computer security threat monitoring and surveillance[R]. Technical report, James P. Anderson Company, 1980.

[3] Denning D E. An Intrusion-detection model[J]. *IEEE Transactions on Software Engineering*, 1987, SE-13(2): 222-232.

[4] Wang F Y, Wang X, Li L X, et al. Steps toward parallel intelligence[J]. *IEEE/CAA Journal of Automatica Sinica*, 2016, 3(4): 345-348.

[5] Parmar J, Chouhan S, Raychoudhury V, et al. Open-world machine learning: Applications, challenges, and opportunities[J]. *ACM Computing Surveys*, 2023, 55(10): 1-37.

[6] Murdock V. Mixed Methods Machine Learning[C]. *Companion of the 2023 International Conference on Management of Data*, 2023: 3-4.

[7] Apruzzese G, Laskov P, Montes de Oca E, et al. The role of machine learning in cybersecurity[J]. *Digital Threats*: *Research and Practice*, 2023, 4(1): 1-38.

[8] Devendiran R, Turukmane A V. Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy[J]. *Expert Systems with Applications*, 2024, 245: 123027.

[9] Buczak A L, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection[J]. *IEEE Communications Surveys & Tutorials*, 2016, 18(2): 1153-1176.

[10] Xin Y, Kong L S, Liu Z, et al. Machine learning and deep learning methods for cybersecurity[J]. *IEEE Access*, 2018, 6: 35365-35381.

[11] Agrawal S, Agrawal J. Survey on anomaly detection using data mining techniques[J]. *Procedia Computer Science*, 2015, 60: 708-713.

[12] Thakkar A, Lohiya R. A review of the advancement in intrusion detection datasets[J]. *Procedia Computer Science*, 2020, 167: 636-645.

[13] Hindy H, Brosset D, Bayne E, et al. A taxonomy of network threats and the effect of current datasets on intrusion detection systems[J]. *IEEE Access*, 2020, 8: 104650-104675.

[14] Wu Z J, Liang C, Li Y Q. Intrusion Detection Method Based on Deep Learning[C]. *2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications*, *Big Data & Cloud Computing*, *Sustainable Computing & Communications*, *Social Computing & Networking*, 2021: 445-452.

[15] Ferrag M A, Maglaras L, Moschoyiannis S, et al. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study[J]. *Journal of Information Security and Applications*, 2020, 50: 102419.

[16] Zipperle M, Gottwalt F, Chang E, et al. Provenance-based intrusion detection systems: A survey[J]. *ACM Computing Surveys*, 2022, 55(7): 1-36.

[17] Mirlekar S, Kanojia K P. A Comprehensive Study on Machine Learning Algorithms for Intrusion Detection System[C]. *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing*, 2022: 1-6.

[18] Amanoul S V, Abdulazeez A M. Intrusion Detection System Based on Machine Learning Algorithms: A Review[C]. *2022 IEEE 18th International Colloquium on Signal Processing & Applications*, 2022: 79-84.

[19] Zhang C Y, Jia D H, Wang L Y, et al. Comparative research on network intrusion detection methods based on machine learning[J]. *Computers & Security*, 2022, 121: 102861.

[20] Alomari D, Anis F, Alabdullatif M, et al. A Survey on Botnets Attack Detection Utilizing Machine and Deep Learning Models[C]. *The 27th International Conference on Evaluation and Assessment in Software Engineering*, 2023: 493-498.

[21] 1998 DARPA Intrusion Detection Evaluation Dataset. Lincoln Laboratory, Massachusetts Institute of Technolo-gy (MIT)[Z]. http://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset. Jun. 2024.

[22] KDD Cup 1999 Data. UCI Machine Learning Repository[Z]. http://www.ics.uci.edu/~kdd/databases/kddcup99/kddcup99.html. Jun. 2024.

[23] ISCX NSL-KDD Dataset. Canadian Institute for Cybersecurity, University of New Brunswick (UNB)[Z]. https://www.unb.ca/cic/datasets/nsl.html. Jun. 2024.

[24] Moustafa N, Slay J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)[C]. *2015 Military Communications and Information Systems Conference*, 2015: 1-6.

[25] Shiravi A, Shiravi H, Tavallaee M, et al. Toward developing a systematic approach to generate benchmark datasets for intrusion detection[J]. *Computers & Security*, 2012, 31(3): 357-374.

[26] Sharafaldin I, Lashkari A H, Hakak S, et al. Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy[C]. *2019 International Carnahan Conference on Security Technology*, 2019: 1-8.

[27] Damasevicius R, Venckauskas A, Grigaliunas S, et al. LITNET-2020: An annotated real-world network flow dataset for network intrusion detection[J]. *Electronics*, 2020, 9(5): 800.

[28] Iot-23 dataset. A labeled dataset with malicious and benign IoT network traffic. Avast-AIC laboratory, Stratosphere IPS, Czech Technical University (CTU)[Z]. https://www.stratosphereips.org/datasets-iot23. Jun. 2024.

[29] García S, Grill M, Stiborek J, et al. An empirical comparison of botnet detection methods[J]. *Computers & Security*, 2014, 45: 100-123.

[30] Komisarek M, Pawlicki M, Mihailescu M E, et al. A Novel, Refined Dataset for Real-Time Network Intrusion Detection[C]. *The 17th International Conference on Availability*, *Reliability and Security*, 2022: 1-8.

[31] Rohmad M S, Azmat F, Manaf M, et al. Enhanced Netflow Version 9 (E-Netflow V9) for Network Mediation: Structure, Experiment and Analysis[C]. *2008 International Symposium on Information Technology*, 2008: 1-6.

[32] Deri L, SpA N. nProbe: An open source netflow probe for gigabit networks[C]. *TERENA Networking Conference*, 2003: 1-4.

[33] Meena G, Dhanwal B, Mahrishi M, et al. Performance Comparison of Network Intrusion Detection System Based on Different Pre-Processing Methods and Deep Neural Network[C]. *The International Conference on Data Science*, *Machine Learning and Artificial Intelligence*, 2022: 110-115.

[34] Sandhu R, McLean J, Lee W K, et al. A framework for constructing features and models for intrusion detection systems[J]. *ACM Transactions on Information and System Security*, 2000, 3(4): 227-261.

[35] Tao Y C, Zhang J T, Wei L, et al. An Intrusion Detection Model with Attention and BiLSTM-DNN[C]. *The 2023 2nd Asia Conference on Algorithms*, *Computing and Machine Learning*, 2023: 78-83.

[36] Aladeemy M, Tutun S, Khasawneh M T. A new hybrid approach for feature selection and support vector machine model selection based on self-adaptive cohort intelligence[J]. *Expert Systems with Applications*, 2017, 88: 118-131.

[37] Leon M, Markovic T, Punnekkat S. Feature Encoding with Autoencoder and Differential Evolution for Network Intrusion Detection Using Machine Learning[C]. *The Genetic and Evolutionary Computation Conference Companion*, 2022: 2152-2159.

[38] Thakkar A, Lohiya R. A review on challenges and future research directions for machine learning-based intrusion detection system[J]. *Archives of Computational Methods in Engineering*, 2023, 30(7): 4245-4269.

[39] Apruzzese G, Laskov P, Montes de Oca E, et al. The role of machine learning in cybersecurity[J]. *Digital Threats*: *Research and Practice*, 2023, 4(1): 1-38.

[40] Kuang F J, Xu W H, Zhang S Y. A novel hybrid KPCA and SVM with GA model for intrusion detection[J]. *Applied Soft Computing*, 2014, 18: 178-184.

[41] Li Y, Guo L. An active learning based TCM-KNN algorithm for supervised network intrusion detection[J]. *Computers & Security*, 2007, 26(7/8): 459-467.

[42] Singh D M, Harbi N, Zahidur Rahman M. Combining naive bayes and decision tree for adaptive intrusion detection[J]. *International Journal of Network Security & Its Applications*, 2010, 2(2): 12-25.

[43] Wang Y. A multinomial logistic regression modeling approach for anomaly intrusion detection[J]. *Computers & Security*, 2005, 24(8): 662-674.

[44] Lee J H, Lee J H, Sohn S G, et al. Effective Value of Decision Tree with KDD 99 Intrusion Detection Datasets for Intrusion Detection System[C]. *2008 10th International Conference on Advanced Communication Technology*, 2008: 1170-1175.

[45] Vincent P, Larochelle H, Bengio Y, et al. Extracting and Composing Robust Features with Denoising Autoencoders[C]. *The 25th International Conference on Machine Learning*, 2008: 1096-1103.

[46] Vincent P, Larochelle H, Lajoie I, et al. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion[J]. *Journal of Machine Learning Research*, 2010, 11: 3371-3408.

[47] Deng J, Zhang Z X, Marchi E, et al. Sparse Autoencoder-Based Feature Transfer Learning for Speech Emotion Recognition[C]. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013: 511-516.

[48] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.

[49] Ranzato M A, Boureau Y L, LeCun Y. Sparse Feature Learning for Deep Belief Networks[C]. *The 21st International Conference on Neural Information Processing Systems*, 2007: 1185-1192.

[50] Razavian A S, Azizpour H, Sullivan J, et al. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition[C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014: 512-519.

[51] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification

with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.

[52] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: A convolutional neural-network approach[J]. *IEEE Transactions on Neural Networks*, 1997, 8(1): 98-113.

[53] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(10): 2222-2232.

[54] Graves A, Mohamed A R, Hinton G. Speech Recognition with Deep Recurrent Neural Networks[C]. *2013 IEEE International Conference on Acoustics*, *Speech and Signal Processing*, 2013: 6645-6649.

[55] Mikolov T, Karafiát M, Burget L, et al. Recurrent Neural Network Based Language Model[C]. *Interspeech 2010*, 2010: 1045-1048.

[56] Graves A, Jaitly N. Towards End-to-End Speech Recognition with Recurrent Neural Networks[C]. *The 31st International Conference on International Conference on Machine Learning - Volume 32*, 2014: II-1764-II-1772.

[57] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks[EB/OL]. 2014: arXiv: 1409.3215. https://arxiv.org/abs/1409.3215.

[58] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.

[59] Gurunarayanan A, Agrawal A, Bhatia A, et al. Improving the Performance of Machine Learning Algorithms for TOR Detection[C]. *2021 International Conference on Information Networking*, 2021: 439-444.

[60] Schuster M, Paliwal K K. Bidirectional recurrent neural networks[J]. *IEEE Transactions on Signal Processing*, 1997, 45(11): 2673-2681.

[61] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. *International Conference on Neural Information Processing Systems*, 2014: 2672-2680.

[62] Liu H Y, Lang B, Liu M, et al. CNN and RNN based payload classification methods for attack detection[J]. *Knowledge-Based Systems*, 2019, 163: 332-341.

[63] Min E X, Long J, Liu Q, et al. TR-IDS: Anomaly-based intrusion detection through text-convolutional neural network and random forest[J]. *Security and Communication Networks*, 2018, 2018(1): 4943509.

[64] Zeng Y, Gu H X, Wei W T, et al. -: A deep learning based network encrypted traffic classification and intrusion detection Framework[J]. *IEEE Access*, 2019, 7: 45182-45190.

[65] Yu Y, Long J, Cai Z P. Network intrusion detection through stacking dilated convolutional autoencoders[J]. *Security and Communication Networks*, 2017, 2017(1): 4184196.

[66] Rigaki M, Garcia S. Bringing a GAN to a Knife-Fight: Adapting Malware Communication to Avoid Detection[C]. *2018 IEEE Security and Privacy Workshops*, 2018: 70-75.

[67] Liu B C, Huang Z Q, Zhu Z G. Intrusion Detection System Based on Machine Learning[C]. *The 7th International Conference on Cyber Security and Information Engineering*, 2022: 121-125.

[68] Borgioli N, Thi Xuan Phan L, Aromolo F, et al. Real-Time Packet-Based Intrusion Detection on Edge Devices[C]. *Proceedings of Cyber-Physical Systems and Internet of Things Week 2023*, 2023: 234-240.

[69] Potluri S, Ahmed S, Diedrich C. Convolutional Neural Networks for Multi-Class Intrusion Detection System[C]. *Mining Intelligence and Knowledge Exploration*, 2018: 225-238.

[70] Zhang B A, Yu Y H, Li J. Network Intrusion Detection Based on Stacked Sparse Autoencoder and Binary Tree Ensemble Method[C]. *2018 IEEE International Conference on Communications Workshops*, 2018: 1-6.

[71] Lo W W, Layeghy S, Sarhan M, et al. E-GraphSAGE: A Graph Neural Network Based Intrusion Detection System for IoT[C]. *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, 2022: 1-9.

[72] Zhang H, Yu X R, Ren P, et al. Deep Adversarial Learning in Intrusion Detection: A Data Augmentation Enhanced Framework[EB/OL]. 2019: arXiv: 1901.07949. https://arxiv.org/abs/1901.07949.

[73] Rahi P, Dandotiya M, Anushya A, et al. An Effect of Stacked CNN for Network Intrusion Detection System[C]. *The 4th International Conference on Information Management & Machine Intelligence*, 2023: 1-9.

[74] Li X F, Li L Q. Research on Multi-Level Classification Models for Imbalanced Network Intrusion Dataset[C]. *The 2022 6th International Conference on Electronic Information Technology and Computer Engineering*, 2023: 967-972.

[75] Teng S H, Wu N Q, Zhu H B, et al. SVM-DT-based adaptive and collaborative intrusion detection[J]. *IEEE/CAA Journal of Automatica Sinica*, 2018, 5(1): 108-118.

[76] Ma T, Wang F, Cheng J J, et al. A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks[J]. *Sensors*, 2016, 16(10): 1701.

[77] Sethi K, Madhav Y V, Kumar R, et al. Attention based multi-agent intrusion detection systems using reinforcement learning[J]. *Journal of Information Security and Applications*, 2021, 61: 102923.

[78] Silaa J, Muyingi H, Gamundani A. A Review of Deep Learning IDS for DDoS Attacks in WLANs[C]. *The International Conference on Data Science*, *Machine Learning and Artificial Intelligence*, 2022: 74-79.

[79] Merzouk M A, Cuppens F, Boulahia-Cuppens N, et al. Investigating the practicality of adversarial evasion attacks on network intrusion detection[J]. *Annals of Telecommunications*, 2022, 77(11): 763-775.

[80] Alalmaie A, Nanda P, He X J. Zero Trust Network Intrusion Detection System (NIDS) Using Auto Encoder for Attention-Based CNN-BiLSTM[C]. *The 2023 Australasian Computer Science Week*, 2023: 1-9.

[81] Merzouk M A, Delas J, Neal C, et al. Evading Deep Reinforcement Learning-Based Network Intrusion Detection with Adversarial Attacks[C]. *The 17th International Conference on Availability*, *Reliability and Security*, 2022: 1-6.

[82] Sun W, Li Q M, Wang P C, et al. Evolving Convolutional Neural Networks for Intrusion Detection System Using Hybrid Multi-Strategy Aquila Optimizer[C]. *The Genetic and Evolutionary Computation Conference Companion*, 2022: 304-307.

[83] Yuan X Y, Li C H, Li X L. DeepDefense: Identifying DDoS Attack *via* Deep Learning[C]. *2017 IEEE International Conference on*

Smart Computing, 2017: 1-8.

[84] Radford B J, Apolonio L M, Trias A J, et al. Network Traffic Anomaly Detection Using Recurrent Neural Networks[EB/OL]. 2018: arXiv: 1803.10769. https://arxiv.org/abs/1803.10769.

[85] Wang W, Sheng Y Q, Wang J L, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection[J]. *IEEE Access*, 2018, 6: 1792-1806.

[86] Arikkat D, K A R R, Yerima S Y. Multi-Domain Network Traffic Analysis Using Machine Learning and Deep Learning Techniques[C]. *The 2022 Fourteenth International Conference on Contemporary Computing*, 2022: 305-312.

[87] Chuvakin A A, Schmidt K J, Partricia Moulder C P, et al. Logging and Log Management: The Authoritative Guide to Understanding the Concepts Surrounding Logging and Log Management[M]. Amsterdam: Syngress, 2013.

[88] Meng W Z, Li W J, Kwok L F. Design of intelligent knn-based alarm filter using knowledge-based alert verification in intrusion Detection[J]. *Security and Communication Networks*, 2015, 8(18): 3883-3895.

[89] McElwee S, Heaton J, Fraley J, et al. Deep Learning for Prioritizing and Responding to Intrusion Detection Alerts[C]. *MILCOM 2017 - 2017 IEEE Military Communications Conference*, 2017: 1-5.

[90] Tran N N, Sarker R, Hu J K. An Approach for Host-Based Intrusion Detection System Design Using Convolutional Neural Network[C]. *Mobile Networks and Management*, 2018: 116-126.

[91] Tuor A, Kaplan S, Hutchinson B, et al. Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams[EB/OL]. 2017: arXiv: 1710.00811. https://arxiv.org/abs/1710.00811.

[92] Bohara A, Thakore U, Sanders W H. Intrusion Detection in Enterprise Systems by Combining and Clustering Diverse Monitor Data[C]. *The Symposium and Bootcamp on the Science of Security*, 2016: 7-16.

[93] Uwagbole S O, Buchanan W J, Fan L. Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention[C]. *2017 IFIP/IEEE Symposium on Integrated Network and Service Management*, 2017: 1087-1090.

[94] Vartouni A M, Kashi S S, Teshnehlab M. An Anomaly Detection Method to Detect Web Attacks Using Stacked Auto-Encoder[C]. *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems*, 2018: 131-134.

[95] Khraisat A, Gondal I, Vamplew P, et al. Survey of intrusion detection systems: Techniques, datasets and challenges[J]. *Cybersecurity*, 2019, 2(1): 20.

[96] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. *ACM Computing Surveys*, 2009, 41(3): 1-58.

[97] Landauer M, Wurzenberger M, Skopik F, et al. AMiner: A modular log data analysis pipeline for anomaly-based intrusion detection[J]. *Digital Threats*: *Research and Practice*, 2023, 4(1): 1-16.

[98] Thanka D R, Jaspher G, Kathrine W, et al. Using Machine Learning for Cyber Security[M]. AI, Machine Learning and Deep Learning. Boca Raton: CRC Press, 2023: 169-190.

[99] Guo W B, Mu D L, Xu J, et al. LEMNA: Explaining Deep Learning Based Security Applications[C]. *The 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018: 364-379.

[100] Farahnakian F, Heikkonen J. A Deep Auto-Encoder Based Approach for Intrusion Detection System[C]. *2018 20th International Conference on Advanced Communication Technology*, 2018: 178-183.

[101] Al-Qatf M, Yu L S, Al-Habib M, et al. Deep learning approach combining sparse autoencoder with SVM for network intrusion detection[J]. *IEEE Access*, 2018, 6: 52843-52856.

[102] Van N T, Thinh T N, Sach L T. An Anomaly-Based Network Intrusion Detection System Using Deep Learning[C]. *2017 International Conference on System Science and Engineering*, 2017: 210-214.

[103] Nguyen K K, Hoang D T, Niyato D, et al. Cyberattack Detection in Mobile Cloud Computing: A Deep Learning Approach[C]. *2018 IEEE Wireless Communications and Networking Conference*, 2018: 1-6.

[104] Ding S, Wang G Y. Research on Intrusion Detection Technology Based on Deep Learning[C]. *2017 3rd IEEE International Conference on Computer and Communications*, 2018: 1474-1478.

[105] Aldwairi T, Perera D, Novotny M A. An evaluation of the performance of restricted boltzmann machines as a model for anomaly network intrusion detection[J]. *Computer Networks*, 2018, 144: 111-119.

[106] Xiao Y H, Xing C, Zhang T N, et al. An intrusion detection model based on feature reduction and convolutional neural networks[J]. *IEEE Access*, 2019, 7: 42210-42219.

[107] Ludwig S A. Intrusion Detection of Multiple Attack Classes Using a Deep Neural Net Ensemble[C]. *2017 IEEE Symposium Series on Computational Intelligence*, 2018: 1-7.

[108] Paul S, Banerjee C, Ghoshal M. A CFS–DNN-Based Intrusion Detection System[C]. *Advances in Communication, Devices and Networking*, 2018: 159-168.

[109] Kim J, Shin N, Yeon J, et al. Method of Intrusion Detection Using Deep Neural Network[C]. *2017 IEEE International Conference on Big Data and Smart Computing*, 2017: 313-316.

[110] Radford B J, Apolonio L M, Trias A J, et al. Network Traffic Anomaly Detection Using Recurrent Neural Networks[EB/OL]. 2018: arXiv: 1803.10769. https://arxiv.org/abs/1803.10769.

[111] Wang Z. Deep learning-based intrusion detection with adversaries[J]. *IEEE Access*, 2018, 6: 38367-38384.

[112] Lutscher P M, Weidmann N B, Roberts M E, et al. At home and abroad: The use of denial-of-service attacks during elections in nondemocratic regimes[J]. *Journal of Conflict Resolution*, 2020, 64(2/3): 373-401.

[113] Potluri S, Diedrich C. Accelerated Deep Neural Networks for Enhanced Intrusion Detection System[C]. *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation*, 2016: 1-8.

[114] Pektaş A, Acarman T. Deep learning to detect botnet *via* network flow summaries[J]. *Neural Computing and Applications*, 2019, 31(11): 8021-8033.

[115] Kuttranont P, Boonprakob K, Phaudphut C, et al. Parallel KNN and neighborhood classification implementations on GPU for network intrusion detection[J]. *Journal of Telecommunication, Electronic and Computer Engineering*, 2017, 9(2-2): 29-33.

[116] Shone N, Ngoc T N, Phai V D, et al. A deep learning approach to network intrusion detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41-50.

[117] Ammar A. A decision tree classifier for intrusion detection priority tagging[J]. *Journal of Computer and Communications*, 2015, 3(4): 52-58.

[118] Patel J, Panchal K. Effective intrusion detection system using data mining technique[J]. *Journal of Emerging Technologies and Innovative Research*, 2015, 2(6): 1869-1878.

[119] Khamphakdee N, Benjamas N, Saiyod S. Improving intrusion detection system based on snort rules for network probe attacks detection with association rules technique of data mining[J]. *Journal of ICT Research and Applications*, 2015, 8(3): 234-250.

[120] Ali Raza Shah S, Issac B. Performance comparison of intrusion detection systems and application of machine learning to snort system[J]. *Future Generation Computer Systems*, 2018, 80: 157-170.

[121] Peng K, Leung V C M, Huang Q J. Clustering approach based on mini batch Kmeans for intrusion detection system over big data[J]. *IEEE Access*, 2018, 6: 11897-11906.

[122] Peng K, Leung V C M, Zheng L X, et al. Intrusion detection system based on decision tree over big data in fog environment[J]. *Wireless Communications and Mobile Computing*, 2018, 2018(1): 4680867.

[123] He Z C, Zhang T W, Lee R B. Machine Learning Based DDoS Attack Detection from Source Side in Cloud[C]. *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing*, 2017: 114-120.

[124] Bansal R, Gaur N, Singh S N. Outlier Detection: Applications and Techniques in Data Mining[C]. *2016 6th International Conference - Cloud System and Big Data Engineering*, 2016: 373-377.

[125] Doshi R, Apthorpe N, Feamster N. Machine Learning DDoS Detection for Consumer Internet of Things Devices[C]. *2018 IEEE Security and Privacy Workshops*, 2018: 29-35.

[126] Mirsky Y, Doitshman T, Elovici Y, et al. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection[EB/OL]. 2018: arXiv: 1802.09089. https://arxiv.org/abs/1802.09089.

[127] Meidan Y, Bohadana M, Mathov Y, et al. N-BaIoT—Network-based detection of IoT botnet attacks using deep autoencoders[J]. *IEEE Pervasive Computing*, 2018, 17(3): 12-22.

[128] Diro A, Chilamkurti N. Leveraging LSTM networks for attack detection in fog-to-things communications[J]. *IEEE Communications Magazine*, 2018, 56(9): 124-130.

[129] Alkasassbeh M, Al-Naymat G, Hassanat A, et al. Detecting distributed denial of service attacks using data mining techniques[J]. *International Journal of Advanced Computer Science and Applications*, 2016, 7(1): 436-445.

[130] Niyaz Q, Sun W Q, Javaid A Y. A Deep Learning Based DDoS Detection System in Software-Defined Networking (SDN)[EB/OL]. 2016: arXiv: 1611.07400. https://arxiv.org/abs/1611.07400.

[131] Yadav S, Subramanian S. Detection of Application Layer DDoS Attack by Feature Learning Using Stacked AutoEncoder[C]. *2016 International Conference on Computational Techniques in Information and Communication Technologies*, 2016: 361-366.

[132] Nguyen S N, Nguyen V Q, Choi J, et al. Design and Implementation of Intrusion Detection System Using Convolutional Neural Network for DoS Detection[C]. *The 2nd International Conference on Machine Learning and Soft Computing*, 2018: 34-38.

[133] Bontemps L, Cao V L, McDermott J, et al. Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks[C]. *Future Data and Security Engineering*, 2016: 141-152.

苏书宾 于 2020 年在北京航空航天大学计算机系统结构专业获得博士学位。现任集美大学计算机工程学院讲师。研究领域为大数据、云计算、网络安全。Email: dreamsu@126.com

肖利民 于 1998 年在中国科学院计算所计算机专业获得博士学位。现任北京航空航天大学教授, CCF 杰出会员。研究领域为计算机体系结构和系统软件、高性能计算机和服务器系统、系统虚拟化与云计算、大数据存储和分布式文件系统、计算机系统安全。Email: xiaolm@buaa.edu.cn

李书攀 于 2020 年在北京航空航天大学计算机系统结构专业获得博士学位。现任郑州大学信息工程学院助理研究员。研究领域为计算机系统安全、人工智能和智能工业质检。Email: iespli@zzu.edu.cn

黄兴旺 于 2018 年在中国科学院大学信号与信息处理专业获得博士学位。现任集美大学计算机工程学院副教授。研究领域为计算智能。Email: huangxw@jmu.edu.cn

谢书童 于 2010 年在厦门大学测试计量技术及仪器专业获得博士学位。现任集美大学计算机工程学院副教授。研究领域为数据挖掘、机器学习和大数据分析。Email: shutong@jmu.edu.cn

吴博 于 2020 年在北京航空航天大学计算机系统结构专业获得博士学位。现任南昌航空大学软件学院讲师。研究领域为网络安全、智能运维和图神经网络。Email: wubo_buaa@126.com