网络入侵检测技术综述

蹇诗婕[1,2], 卢志刚[1,2], 杜 丹[1], 姜 波[1,2*], 刘宝旭[1,2]

[1] 中国科学院信息工程研究所, 北京 中国 100093
[2] 中国科学院大学网络空间安全学院, 北京 中国 100049

摘要 随着互联网时代的发展, 内部威胁、零日漏洞和 DoS 攻击等攻击行为日益增加, 网络安全变得越来越重要, 入侵检测已成为网络攻击检测的一种重要手段。随着机器学习算法的发展, 研究人员提出了大量的入侵检测技术。本文对这些研究进行了综述。首先, 简要介绍了当前的网络安全形势, 并给出了入侵检测技术及系统在各个领域的应用。然后, 从数据来源、检测技术和检测性能三个方面对入侵检测相关技术和系统进行已有研究工作的总结与评价, 其中, 检测技术重点论述了传统机器学习、深度学习、强化学习、可视化分析技术等方法。最后, 讨论了当前研究中出现的问题并展望该技术的未来发展方向和前景。本文希望能为该领域的研究人员提供一些有益的思考。

关键词 网络空间安全; 入侵检测; 机器学习; 深度学习; 强化学习; 可视化分析
中图法分类号 TP393.08 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2020.07.07

Overview of Network Intrusion Detection Technology

JIAN Shijie[1,2], LU Zhigang[1,2], DU Dan[1], JIANG Bo[1,2*], LIU Baoxu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract With the development of the Internet era, attacks such as internal threats, zero-day vulnerabilities and DoS attacks are increasing, and network security is becoming more and more important. Intrusion detection has become an important means of network attack detection. With the development of machine learning algorithms, researchers have proposed a large number of intrusion detection techniques. This article reviews these studies. Firstly, it briefly introduces the current network security situation, and gives the application of intrusion detection technology and system in various fields. Then, the existing research work of intrusion detection related technologies and systems is summarized and evaluated from three aspects: data source, detection technology and detection performance. Among them, the detection technology focuses on traditional machine learning, deep learning, reinforcement learning, visual analysis techniques. Finally, the problems in the current research are discussed and the future direction and prospects of the technology are expected. This article hopes to provide some useful thinking for researchers in this field.

Key words cyber security; intrusion detection; machine learning; deep learning; reinforcement learning; visual analysis

1 引言

随着信息技术的不断发展, 网络空间[1]成为继陆地、海洋、天空和太空之后的“第五前沿”。作为国家关键信息基础设施之一的互联网已在政治、军事、经济、交通等领域发挥着巨大的作用。与此同时, 网络中攻击事件出现的频率和规模也呈现出逐年增加的趋势。比如, 美国赛门铁克在 2019 年发布的《互联网安全威胁报告》[2]中指出: 1)在恶意软件方面, 恶意软件攻击频率略微下降, 勒索软件攻击数量自 2013 年来, 首次出现 20%的总体下滑, 但是针对企业的攻击数量上升了 12%。2018 年, 赛门铁克拦截的挖矿劫持攻击数量是 2017 年的四倍。2)在移动设备方面, 移动设备上的勒索软件的感染数量与 2017 年相比, 增加了 33%, 美国是移动恶意软件的重灾区, 占总量的 63%, 紧随其后的是中国(13%)和德国(10%)。3)在 Web 攻击方面, 2018 年端点上的总体 Web 攻击数量增加了 56%, 12 月, 赛门铁克每日

通讯作者: 姜波, 博士, 副研究员, Email: jiangbo@iie.ac.cn。
本论文得到国家重点研发计划(No.2018YFB0803602，No.2019QY1303), 北京市科技计划(No.Z181100002718005), 中国科学院战略性先导 C 类(No.XDC02040100), 国家自然科学青年基金(No.61702508, No.61802404)的资助。这项工作也得到了中国科学院网络评估技术重点实验室和北京市网络安全与保护技术重点实验室的部分支持。
收稿日期: 2019-10-17; 修改日期: 2019-12-09; 定稿日期: 2020-06-19

可在端点机器上拦截超过 130 万次的 Web 攻击。4)在目标性攻击方面, 供应链攻击和离地攻击成为网络犯罪的主流, 2018 年供应链攻击增加了 78%。5)在物联网方面, 物联网攻击数量在 2017 年大幅增加, 平均每月攻击次数达到 5200 次, 在 2018 年趋于稳定。除此之外, DDoS 攻击、挖矿活动、Web 攻击、利用系统漏洞攻击等攻击行为始终存在于互联网中, 这些攻击行为已经造成巨大的经济损失, 甚至严重威胁到国家安全和社会的稳定发展。因此, 如何有效地防护来自网络中的攻击行为已成为当前亟需解决的问题。

目前已有的安全防护方案有防火墙、数据加密、入侵检测系统等技术。其中, 入侵检测系统(Intrusion Detection System, IDS)[3]是一种积极主动的安全防护技术, 通过对网络进行实时监视, 能够有效感知网络攻击, 为安全管理人员提供响应决策。当前, 入侵检测在军事、医疗、交通、物联网安全、工业控制系统等领域均有广泛应用。

- 军事领域: 无线传感器网络在军事系统中具有广泛的应用[4], 但是该网络易受到路由攻击, 针对可重配置路由协议的网络攻击有洪水攻击[5]、黑洞攻击[6]、选择性转发攻击[7]等, 这些攻击行为会导致设备性能下降, 对军事任务产生较大的威胁。基于分布式跨层的机器学习异常检测系统[8]能有效检测军事网络攻击行为, 有效保护军事网络的安全。
- 医疗领域: 在医疗领域中, 医疗设备和医疗网络对于辅助医生决策, 保存患者信息起着非常重要的作用[9]。然而目前无线人体传感器网络和医疗设备仍然会受到很多攻击的影响, 例如 Sybil 攻击、洪水攻击等。这些攻击行为会对医生的医疗决策产生严重干扰, 并且会泄漏患者的健康状况信息。基于心电图和肌电图分析的入侵检测技术[10]能够提升医疗异常检测效率, 维护医疗网络的安全。IDS 能够检测医疗系统的未授权行为, 并加以预防, 因此被广泛用于医疗领域。
- 交通领域: 智能车辆在智能互联交通系统的发展中发挥着重要作用, 现代智能车辆是由传感器, 电子控制单元和执行器组成的复杂系统[11]。电子控制单元通过控制器局域网进行通信, 控制器局域网由于缺乏消息认证和广播传输, 容易受到各种网络攻击[12], 例如通过 CD 播放器, 蓝牙等攻击媒介对车辆进行攻击, 导致远程控制车辆[13], 可能引发严重交通事故, 基于神经网络算法的车辆入侵检测系统[14]能够检测车辆异常情况, 并及时做出响应。入侵检测系统可以检测车载网络中潜在的网络攻击, 已引起了极大的关注。
- 物联网领域: 物联网(Internet of Things, IoT)[15]是指将互联网集成到人类社会不同领域的实体。攻击者通过入侵物理设备、监视区域等来获取信息, 其中攻击的主要类型有拒绝服务攻击[16]、路由攻击等, 基于支持向量机的轻量级攻击检测策略[17]能够有效检测这些入侵行为。由于入侵检测机制被认为是信息和通信技术保护的主要来源, 因此被广泛应用于物联网领域。
- 工业控制领域: 工业控制系统(Industrial Control Systems, ICS)[18]: 主要负责实时数据采集, 系统监视以及工业过程的自动控制和管理。近几年, 针对 ICS 的网络攻击数量迅速增加, 例如 Stuxnet 恶意软件事件[19]、BlackEnergy 入侵电网事件[20]等等。基于混合增强设备指纹识别方法的入侵检测系统[21]能够有效检测攻击, 并及时做出响应。入侵检测通过收集和分析网络流量、日志等数据, 被广泛认为是维护 ICS 安全的重要手段。

尽管入侵检测技术在某些安全场景中发挥了一定的作用。但是, 随着黑客攻击手段的不断升级及所面临的海量网络数据, 基于传统的机器学习方法已不再适用于新的网络入侵检测场景。近年来, 随着大数据技术的出现, 深度学习、强化学习、可视化等技术得到了广泛的应用, 并已在自然语言处理、图像识别、视频检测等领域取得了很大的成功。同时, 在网络安全领域, 已有大量的研究工作利用这些技术进行网络的入侵检测, 并取得了一定的效果。然而, 已有的入侵检测综述侧重于早期的技术, 例如入侵检测系统的特征选择算法综述[22-23]、常用的传统机器学习技术在入侵检测领域的应用[24-25]、基于网络的入侵检测数据集综述[26]、基于主机数据的入侵检测综述[27]、入侵检测网络威胁分类[28]等。虽然这些工作都提供了有价值的信息, 但是当前并未有相关的文献对最新的入侵检测技术进行梳理与总结。因此, 本文通过梳理已有入侵检测的相关工作, 将其划分为基于误用的入侵检测和基于异常的入侵检测这两个方面。本文首先对基于误用的入侵检测技术进行简单概括, 然后重点介绍基于异常的入侵检测技术, 涵盖传统机器学习、深度学习、强化学习、可视化分析等方面。通过比较与分析, 重点归纳并总结当前最新的入侵检测技术的优劣, 以便于该领域研究人员快速且直观地了解当前的研究动态。

本文的组织结构安排如下: 第 2 节描述了入侵检测系统的详细分类; 第 3 节总结了传统机器学习技术在入侵检测领域的应用现状; 第 4 节阐述了深

度学习方法在入侵检测领域的应用情况; 第 5 节讨论了强化学习方法; 第 6 节对可视化分析技术在入侵检测的应用进行阐述; 第 7 节是讨论与展望, 对未来研究方向和挑战进行展望; 第 8 节是结束语。

2 入侵检测系统分类

入侵检测系统是监视和分析网络通信的系统, 通过主动响应来识别异常行为。按照不同的划分标准, 可以将入侵检测系统分为不同的类别。本文借鉴通用入侵检测系统的划分框架, 对入侵检测系统按照数据来源和检测技术进行划分。具体划分框架如图 1 所示。

2.1 基于数据来源划分

基于数据来源, 可以将入侵检测系统划分为基于主机的入侵检测和基于网络的入侵检测。

图 1 入侵检测系统构建框架

Figure 1 The Framework of Intrusion Detection System

2.1.1 基于主机的 IDS(HIDS)

基于主机的 IDS(Host-based Intrusion Detection System, HIDS)[29]通常是位于被监视系统上的软件组件, 监视主机系统的操作或状态, 检测系统事件, 例如未经授权的访问或安装, 主要分析与操作系统信息相关的事件。通过浏览日志、系统进行的调用、文件系统的修改及其他状态和活动来检测入侵[30]。优点是能够在发送和接收数据前通过扫描流量活动来检测内部威胁[31], 缺点是只监视主机, 需要安装在每个主机上[32], 且无法观测到网络流量, 无法分析与网络相关的行为信息。

基于主机的入侵检测系统中, 大多数系统可以分为程序级 IDS 或操作系统级 IDS。程序级 IDS 专注于监视单个应用程序, 使用源代码, 字节代码, 静态或动态控制流以及有关应用程序状态的其他信息。操作系统级 IDS 监视整个系统状态, 并可监视所有进程的组合行为, 以区分操作系统级别的正常和异常行为, 可能涉及系统日志, Windows 注册表数据等内容[27]。常用的基于主机的入侵检测公开数据集有 ADFA [33]、Blue Gene / L[34]数据集等。

2.1.2 基于网络的 IDS(NIDS)

基于网络的 IDS(Network-based Intrusion Detection System, NIDS)[35]观察并分析实时网络流量和监视多个主机, 旨在收集数据包信息, 并查看其中内容,

以检测网络中的入侵行为[36]。NIDS 的优点是只有一个系统监视整个网络, 节省了在每个主机上安装软件的时间和成本。缺点是 NIDS 难以获取所监视系统的内部状态信息, 导致检测更加困难。其中, 基于网络的数据分为基于流和基于包的数据, 基于流的数据只包含网络连接相关的元信息, 而基于包的数据还包含有效负载。常用的基于网络的入侵检测公开数据集有DARPA1998[37]、DARPA1999[38]、KDD1999[39]、NSL-KDD[40]、Gure-KDD[41]、CIDDS-001[42]、CICIDS2017[43]、ISCX2012[44]、Kyoto2006+[45]、UNSW-NB15[46]数据集等。

2.2 基于检测技术划分

基于检测技术可以将入侵检测系统划分为基于误用的入侵检测和基于异常的入侵检测。

2.2.1 基于误用的 IDS(MIDS)

基于误用的 IDS(Misuse-based Intrusion Detection System, MIDS)[47]也称为基于签名的入侵检测, 是将传入的网络流量与已有签名进行匹配, 一旦命中即发出告警信息来预示异常行为。MIDS 为每种攻击设计一个签名, 检测准确率较高。现有的 MIDS 包含如下三种方法:

- 状态建模[48]: 是指将攻击编码为有限自动机中的许多不同状态。通过在流量配置文件中进行观察, 检测攻击行为。
- 字符串匹配[49]: 是指一种获取知识的过程, 通过字符串模式匹配的方式查找攻击签名, 从而判断是否存在入侵行为。
- 专家系统[50]: 是根据一套用于描述系统已知的攻击情形的规则对审计数据进行分类。具体包括首先从训练数据中识别出不同的属性和类别。然后, 推导出一组分类规则或者程序。最后, 对审计数据进行分类。

早期的 IDS 大多使用基于误用的技术来检测入侵行为。但是, 基于误用的入侵检测系统高度依赖已有的签名知识库, 难以检测未知攻击, 无法适应新的智能攻击行为[51]。因此, 基于异常的入侵检测是目前学者们研究和开发的重点。

2.2.2 基于异常的 IDS(AIDS)

基于异常的 IDS(Anomaly-based Intrusion Detection System, AIDS)[52]是对系统异常的行为进行检测, 当检测行为与正常行为偏离较大时, 发出告警信息。早期的一些学者利用统计学习的方法进行异常的入侵检测, 这些方法使用正常活动的统计特性, 捕获网络流量活动并创建表示其随机行为的记录。文献[53]通过将参数建模为独立的高斯随机变量的方式, 对单变量定义值范围, 与值范围偏离较大的数据则标识为攻击行为。文献[54]提出了一种多变量质量控制技术, 在信息系统中建立正常活动的长期规范记录并通过使用该记录对入侵行为进行检测, 实验结果表明, 对于相关措施的组合, 能够获得更好的检测效果, 误报率更低。文献[55]提出使用时间序列技术来构建模型, 该模型使用了间隔计时器、事件计数器, 考虑了流量的观察顺序和到达间隔时间对应的数值, 如果观察到的流量实例在给定时间内发生的概率很低, 则将其标记为异常。

基于统计的异常检测技术能够明确地表示和处理信息系统活动中所涉及的变化和噪声, 并且能从观察中学习到系统的预期行为。但是, 容易被攻击者训练, 使得攻击期间的流量被误以为是正常流量, 并且对常态假设很敏感, 如果测量数据不服从正态分布, 会产生较高的误报率。除此之外, 单变量的技术只为信息系统中的一种活动量度建立统计规范记录, 对于多种活动的入侵检测, 存在偏差。

从模型学习的角度, 基于异常的 IDS 还包括基于传统机器学习、基于深度学习、基于强化学习、基于可视化分析等技术。

3 基于传统机器学习的入侵检测

在本节中, 将讨论用于入侵检测领域最流行的传统机器学习相关技术。下面, 基于传统机器学习方法的入侵检测系统将从入侵数据处理、监督和无监督机器学习技术三个方面进行详细阐述。

3.1 入侵数据处理

通常, 大规模入侵检测数据集包含了基本特征、内容特征、流量特征。基本特征是指 TCP 连接的特征, 例如持续时间、协议类型、源字节数目、目的字节数目等; 内容特征是指从数据包提取的特征, 例如登陆失败次数、创建的文件数目、获取的文件数目等; 流量特征是指与流量传输相关的特征, 例如源主机数目、目的主机数目等。不同的入侵检测数据集包含的特征也存在差异, 例如 UNSW-NB15 数据集中还包括了附加特征等。

入侵检测数据中既存在连续数据又存在离散数据, 并且数据中不同特征属性之间数量级相差较大。为了构建合理的数据集, 通常需要对数据集进行预处理。数据预处理阶段通常包含符号特征数值化和数值特征归一化两部分内容。符号特征数值化是指使用独热编码等编码方式对数据进行映射处理; 数值特征归一化是指使用 Min-Max、Z-score 等归一化

方法将数据数值缩小至[0, 1]的范围内, 消除尺寸大小对特征属性的影响。

同时, 入侵检测数据集中存在大量冗余、噪声数据以及不相关数据, 因此对数据集进行特征选择或特征提取能够去除冗余数据、降低特征维度、减小计算开销, 提升分类器的泛化能力和检测性能。在入侵检测领域, 特征选择通常包括三种方案, 分别是过滤式方法、封装式方法、嵌入式方法。常用的过滤式方法有信息增益、相关系数等; 常用的封装式方法有遗传算法(Genetic Algorithm, GA)等; 常用的嵌入式方法有正则化方法等, 例如 LASSO 回归。入侵检测领域常用的特征提取方法有线性变换方法, 例如主成分分析方法(Principal Component Analysis, PCA)和线性鉴别分析方法等, 以及非线性变换方法, 例如基于核方法的 PCA 等。特征选择和特征提取都是通过减少数据集中的特征数量, 来提升模型的检测性能, 希望达到的效果是一致的, 只是采取的方式存在差异。

对引用论文进行综合分析后发现, 决定入侵检测性能的重要特征通常是基本特征和流量特征, 例如协议类型、源到目标的字节数、到同一目标 IP 地址的连接总数、到同一目标端口的连接总数等。

3.2 监督机器学习技术

3.2.1 生成方法

生成模型的特点是从统计角度表示流量数据的分布情况, 能够反映同类流量数据的相似度。

1) 朴素贝叶斯

由于朴素贝叶斯算法(Naive Bayes, NB)逻辑简单且易于实现, 经常被应用于入侵检测系统中。文献[56]对 KDD1999 数据集进行 Z-score 归一化、PCA 特征提取、离散化等处理, 得到网络路由检测的输入特征, 再利用朴素贝叶斯对处理后的网络路径数据集进行分类。实验结果显示, 该方法能够有效检测特洛伊木马攻击、假消息攻击、拒绝服务攻击和远程用户未经授权的访问攻击, 检测率(*Detection Rate, DR*)达到 87%~97%。文献[57]提出一种基于粒子群的加权朴素贝叶斯入侵检测模型, 结合粗糙集理论和改进的粒子群算法来提升检测能力。朴素贝叶斯方法存在难以解决流量高维度的问题, 文献[58]使用松散假设的隐藏朴素贝叶斯方法(Hidden Naïve Bayes, HNB)对流量数据进行多分类。该方法采用最小化熵离散化方法和 k 间隔比例离散化方法对 10%KDD1999 数据集中的连续特征进行离散化处理, 攻击流量的识别准确率(*Accuracy, ACC*)为 93.72%, 高于朴素贝叶斯方法(78.32%)。

2) 贝叶斯网络

贝叶斯网络(Bayesian Network, BN)具有强大的推理能力, 因此, 被广泛应用于入侵检测领域。BN 分类器通常选择次优模型的启发式方法来训练数据, 为了缓解此类问题, 文献[59]通过贝叶斯模型在最佳 k 个 BN 分类器上构建贝叶斯平均分类器, 使用了类别属性权变系数离散化方法(Class-Attribute Contingency Coefficient, CACC)对 NSL-KDD 数据集中的连续特征数据进行离散化, 并选择了 12 个特征用于训练模型, 所提方法能够识别正常流量和攻击流量, 检测准确率为 96.92%, 优于朴素贝叶斯模型。文献[60]使用连续时间贝叶斯网络对基于网络的入侵检测和基于主机的入侵检测分别进行建模。对于 NIDS, 为网络数据包跟踪构建了一个分层模型, 并使用 Rao-Blackwellized 粒子滤波来学习参数。对于 HIDS, 开发了一种新颖的学习方法来处理系统日志文件时间戳的有限分辨率。

3) 隐马尔可夫模型

隐马尔可夫模型(Hidden Markov Models, HMM)能够较好地捕获连续序列的依赖性, 因此, 被用于解决入侵检测中的字节序列等问题。对于 Web 应用程序的安全, 在字节级别对 HTTP 有效载荷的分析是有效的, 文献[61]通过将有效载荷表示为字节序列, 再使用 HMM 进行分析, 该方法对 XSS 和 SQL 注入的检测非常有效, 但是, 该方法没有考虑有效载荷的长度。文献[62]提出了使用 HMM 开发自适应的网络入侵检测系统, 用于软件定义网络技术。针对车载网络中的 IDS 存在检测时间过长, 损耗较大的问题。文献[63]提出了一种基于 HMM 的新型滤波器模型, 将车载网络中每辆车的状态模式建模为 HMM, 以快速过滤来自车辆的信息, 所提方法在检测率和开销方面均有较好性能。

3.2.2 判别方法

判别模型的特点是寻找不同类别的最佳分类面, 能够反映正常流量数据和异常流量数据之间的差异。

1) K 近邻算法

由于 K 近邻算法(K-Nearest Neighbor, KNN)不需要进行参数估计, 能够解决多分类问题, 因此被广泛应用于入侵检测领域。文献[64]采用 PCA 进行特征提取, 结合模糊技术得到样本对象属于每个类别的程度, 再使用 KNN 对攻击类别进行划分。但是, 随着数据量的增加, 该算法的准确率逐步降低。为了检测大规模网络数据中的攻击行为, 文献[65]使用最小依赖最大重要性算法(Minimum Dependence Maximum Significance, MDMS)从 KDD1999 数据集中选择出 6 个特征, 使用 KNN 来预测网络数据, 所提方

法能较好地识别探针攻击和拒绝服务攻击。文献[66]使用 k-means 对数据进行预处理, 再用 KNN 算法进行分类, 其具体建模过程如图 2 所示。

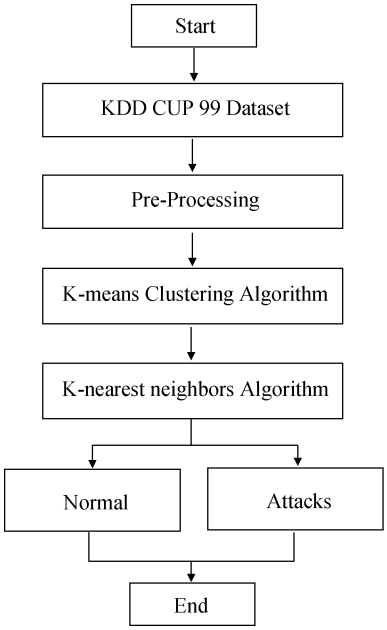


图 2　文献[66]的建模过程

Figure 2　The modeling process of [66]

为了解决入侵检测系统中的用于训练的攻击数据不足的问题, 文献[67]采用模糊 C 均值算法对训练数据集的聚类中心进行软计算和优化, 引入距离加权的 KNN 分类器, 与德姆斯特·谢弗理论相结合, 以评估与每个已知类相关联的输入数据的信念函数和概率值, 基于所获得的概率值及其熵函数来实现两阶段的入侵检测方案。

2) 决策树

决策树(Decision Tree, DT)计算复杂度低, 构造的规则易于理解, 因此, 在入侵检测领域同样具有广泛应用。文献[68]提出了一种基于 CART 决策树的智能电网高级计量基础设施 IDS, 在 CICIDS2017 数据集上的实验准确率为 99.66%。KDD1999 数据集不包含现阶段的新型网络攻击, 因此, 文献[69]使用 J48 决策树方法对 Kyoto2006+数据包进行分类。为了弥补传统决策树算法的不足, 文献[70]提出一种基于相对决策熵的决策树算法, 结合粗糙集属性约简技术来删除冗余特征。

将决策树与其他分类算法结合来构造入侵检测模型同样能够带来检测精度的提升。文献[71]组合了 REP 树, JRip 算法和 Forest PA 三种不同的分类器, 在使用 Min-Max 归一化处理后的 CICIDS2017 数据集上的最高准确率为 96.665%, 最低误报率(*False Alarm Rate, FAR*)为 1.145%, 准确率高于朴素贝叶斯(74.528%)等算法, 能够有效识别正常流量和异常流量。由于网络复杂流量的不断变化, 入侵检测需要自适应机制。因此, 文献[72]提出一种协同自适应入侵检测的方法, 使用二分类支持向量机和决策树方法, 来设计入侵检测模型。

3) 支持向量机

相比于其他算法, 支持向量机(Support Vector Machine, SVM)可以更好地解决小样本问题, 具有强大的泛化能力, 被认为是更有效的入侵检测算法。文献[73]使用 Z-score 归一化 KDD1999 数据, 并使用压缩采样方法进行特征压缩, 再结合 SVM 对压缩结果进行分类, 所提方法假阳性率(*False Positive Rate*, *FPR*)较低, 能够有效检测拒绝服务攻击、探针攻击等攻击行为。文献[74]采用了基于 SVM 的数据挖掘方法, 使用粗糙集理论进行特征选择, 减少了手动分析任务的需要。文献[75]提出了一种改进的遗传算法优化支持向量机的入侵检测方法, 并设计了一种基于分类准确率、误报率、数据特征维度的适应度函数。文献[76]使用对数边际密度比(logarithms of the marginal density ratios, LMDRT)作为特征转换技术来构建基于 SVM 的 IDS。文献[77]使用 k-means 对数据进行划分, 形成两个集群, 再用 SVM 对集群进行分类。在实践中还没有可以较好地解决核函数构造问题的理论, 为了解决这一问题, 文献[78]将径向基函数核和多项式核的优点有效地整合到博弈论的概念中, 基本思想是利用 NIDS 中的博弈论来获得具有更好学习能力和泛化性能的 SVM 分类器。

4) 逻辑回归

逻辑回归算法(Logistic Regression, LR)具有简单高效、计算速度快、易于并行的特点, 因此, 在入侵检测领域有较多应用。由于还没有学者使用多项式逻辑回归算法来预测攻击行为, 并且仍不清楚与个别重大攻击相关的风险因素, 为了填补这些空白, 文献[79]使用仿真方法拟合了 3000 个多项式逻辑回归模型, 并确定了与攻击在统计上显着相关的 13 个风险因素, 然后将这些风险因素用于构建最终的多项式模型。文献[80]使用统计方法结合二分类逻辑回归方法开发了基于异常的检测模型, 通过统计检查 OSI 的三层内正常字段值的程度来检测远程用户攻击和用户根攻击。文献[81]使用二元逻辑回归的统计工具来检测攻击, 从路由层获取关键数据, 并将其用作分析传感器行为的基础。在存在网络层攻击的环境中的测试准确率达到 96%~100%。

此外, 一些工作使用逻辑回归算法对特征进行

处理来构建 IDS, 文献[82]提出了基于稀疏逻辑回归进行特征选择的 IDS, 将特征选择和分类组合到一个统一的框架中, 在 KDD1999 数据集上的实验准确率为 97.86%, 能够有效识别攻击数据。文献[83]使用逻辑回归选出每个类别的重要特征, 结合正则化技术改进这些值, 再对神经网络、决策树、线性判别分析等方法进行组合并对流量数据进行分类。

3.2.3 小结

入侵检测可以看作是一个分类问题, 即对主机数据和网络中流量数据进行二分类或多分类的判断, 监督机器学习技术能够有效地对数据进行类别划分, 因此被广泛用于入侵检测领域。基于监督机器学习技术的入侵检测优点是能够充分利用先验知识, 明确地对未知样本数据进行分类。缺点是训练数据的选取评估和类别标注需要花费大量的人力和时间。其中, 监督机器学习技术包含了生成方法和判别方法, 生成方法在入侵检测中的优点总结如下: 1)在数据不完整的情况下, 仍能检测异常。例如朴素贝叶斯方法, 在数据较少的情况下, 仍然能够较好地对未知数据进行分类。2)可以学习存在隐变量的模型。例如贝叶斯网络, 对于不确定性问题具有强大的推理能力且鲁棒性较好。3)收敛速度快, 即当样本数据较多时, 可以更快地收敛于真实模型。例如隐马尔可夫模型, 能够有效处理入侵检测领域中的序列相关问题, 且模型收敛速度快。但是生成模型在入侵检测中也存在一些缺陷, 虽然生成方法能够为入侵检测提供很多信息, 但是也需要更多的计算资源, 仅用于分类任务时, 存在较多冗余信息, 且其学习和计算过程较复杂。判别方法在入侵检测中的优点总结如下: 1)直接面对预测问题, 准确率更高。例如 KNN 算法, 不需要对参数进行估计, 能够直接对多种攻击类型进行较好的预测。2)可以对输入数据进行各种程度的抽象, 从而简化学习问题。例如决策树方法和 SVM 方法, 决策树模型能够构造易于理解的规则, 在短时间内对大规模高维网络数据进行较好的处理, 基于结构风险最小化的 SVM 方法也能够较好地处理高维非线性数据。3)对于分类任务, 冗余信息少, 节省计算资源。例如逻辑回归模型, 具有简单高效的计算能力, 且计算速度快, 适合并行处理数据。但是判别方法在入侵检测中也存在一些缺点, 判别方法难以反映数据本身的特性, 且数据缺失或异常值对预测结果影响较大。

监督机器学习中的生成方法和判别方法在入侵检测领域均取得了很好的效果, 但两种方法的结合显示出更大的优越性[84]。已有一些研究将监督机器学习中生成方法和判别方法结合, 充分利用两种方法的优点, 并最大程度地减少它们的缺点, 但是部分研究对于攻击的检测率较低, 仍具有较大提升空间, 同时海量高维数据的增加仍是监督机器学习技术在入侵检测领域中面临的巨大困扰。

3.3 无监督机器学习技术

3.3.1 常用的无监督机器学习技术

1) k-means

k-means 算法可解释性强, 收敛速度快, 因此, 被广泛用于入侵检测领域。作为最常用的聚类算法之一, k-means 经常和其他分类算法结合使用, 来提升检测率。文献[85]使用改进的 k-means 算法来构建高质量的训练数据集, 组合使用 SVM 和极限学习机(extreme learning machine, ELM)算法来构建 IDS, 能够有效识别拒绝服务攻击。文献[86]和文献[87]分别将 k-means 与投影自适应共振理论和分类回归树算法结合来构建 IDS, 其中文献[86]的检测率高于文献[87], 但是文献[86]只使用了 10%KDD1999 数据。

通过对 k-means 算法模型进行改善, 也能提升检测精度。为了解决聚类结果随聚类中心的不同而波动的问题, 文献[88]提出一种改进 k-means 算法, 使用粒子群优化算法优化生成初始聚类中心, 得到全局最优的聚类结果, 但是该方法误报率较高。针对传统的 k-means 算法存在对聚类中心初始值敏感、容易受到噪声和孤立点的影响等问题, 文献[89]首先对数据集进行优化, 去除孤立点, 然后使用类内的最大相似距离和类间的最小相似距离来动态生成新类, 并结合最小支持树聚类算法对集群进行分割。传统的 k-means 方法的缺点是必须预先设置集群数 k 的值, 文献[90]使用了适应度函数来改进遗传 k-means 算法, 不需要预先设定便能找到最佳 k 值。

2) 层次聚类

层次聚类不需要预先制定聚类数目, 且能够发现类的层次关系, 因此被广泛用于数据处理。文献[91]使用层次聚类算法处理 KDD1999 数据集, 并结合了简单的特征选择方法和 SVM 方法。文献[92]使用聚集层次聚类方法来构造层次树, 并结合互信息理论, 来选择用于入侵检测的特征。文献[93]提出了一种用于多层无线传感器网络的基于多层聚类的入侵检测系统, 向下的入侵检测是对子节点的异常行为进行监视, 向上的入侵检测是对集群的头部信息进行检测。文献[94]研究了与恶意活动相关的流量数据的特征, 使用层次聚类方法对流量数据进行采样, 再进一步进行深度检测, 实验结果表明, 该方法检测准确率损失较小, 能够有效识别异常数据。

3) 高斯混合模型

高斯混合模型(Gaussian Mixture Model, GMM)可以近似任何分布的概率密度函数, 适用于解决同一个集合下的数据包含了多种不同分布的情况, 优势在于可以识别类似于正常行为分布的攻击行为。文献[95]提出一种使用决策树进行误用检测, 使用高斯混合模型进行异常检测的方法。使用决策树将已知的攻击样本和正常样本区分开, 然后在树的每个叶子节点上都训练高斯混合模型。文献[96]对NSL-KDD 数据集进行 Min-Max 归一化, 使用 GMM 对正常行为进行建模, 并提出一组用于评估样本每个特征概率的分类器, 再将此信息用作基于投票的聚合方法的输入, 从而识别异常数据。针对网络数据量增加, 入侵检测效率降低的问题, 文献[97]提出了一种基于 Hadoop 框架的分布式 GMM 模型, 使用两步 MapReduce 流程来实现此算法。结果表明, 所提算法能够有效减少消耗时间。

4) 主成分分析法

主成分分析法(Principal Component Analysis, PCA)作为最常用的降维方法, 能够降低所研究的数据空间的维数和计算开销, 因此被用于大量入侵检测工作中。文献[98]使用 PCA 和 Fisher 判别方法进行特征选择, 并使用概率自组织图对特征空间进行建模, 具体建模过程如图 3 所示。文献[99]提出了一种结合了信息增益(Information Gain, IG)和 PCA 的混合降维技术, 并使用基于支持向量机、基于实例的学习算法(Instance-Based Learning Algorithms, IBK)和多层感知机(Multilayer Perceptron, MLP)的集成分类器来构建入侵检测模型。实验果表明, 所提出的混合降维方法明显优于单个算法的评估结果, 能够有效识别正常流量和异常流量。文献[100]使用 PCA 去除噪声属性, 保留最佳特征子集, 来训练支持向量机分类器。实验结果证明所提方法能够大大缩短检测时间。

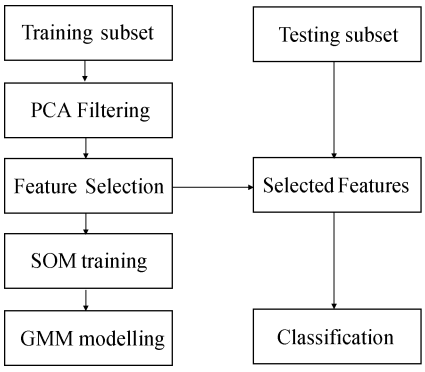


图 3 文献[98]的建模过程

Figure 3 The modeling process of [98]

由于 PCA 对噪声和离群值敏感, 并且仅限于线性主成分。为了解决这一问题, 文献[101]比较了 PCA 和模糊 PCA 两种降维方法, 并结合 KNN 来构建 IDS。在检测用户到根攻击和拒绝服务攻击方面, 模糊 PCA 方法优于 PCA, 因为模糊 PCA 技术能够减小离群值的影响。

3.3.2 小结

基于无监督机器学习方法的入侵检测模型在训练过程中, 不需要带标签的数据集, 能够直接对主机数据和网络中流量数据进行处理, 并能发现数据中的结构性知识。基于无监督机器学习技术的入侵检测优点是不需要人为对数据标注类别信息, 减少了人为误差。缺点是需要对无监督处理结果进行大量分析。在现实世界中, 有标签的网络流量数据远远少于无标签的网络流量数据, 由于缺乏足够的先验知识, 难以进行人工类别标注, 并且人工类别标注所耗费的成本较高, 因此无监督机器学习算法在入侵检测领域的应用非常广泛。

在入侵检测领域中, 常用的无监督机器学习方法能够直接有效地对大规模网络数据进行分类, 不需要对数据进行类别标注, 从而减少复杂的计算开销。常用的入侵检测无监督降维能够有效解决高维数据集中的冗余和不相关问题, 并对高维数据集进行降维, 简化计算问题, 提升分析效率。

综上所述, 基于无监督算法的入侵检测模型能够有效处理网络中逐年增加的大规模流量数据问题, 降低计算开销, 提升检测准确性。因此, 随着网络中海量数据的增加, 无监督机器学习算法将会得到更加广泛的应用, 但是对于噪声和异常值敏感, 也是无监督机器学习算法在入侵检测领域面临的问题。

3.4 总结与讨论

传统机器学习技术的发展促进了入侵检测研究的进步, 基于监督机器学习和基于无监督机器学习的方法均在入侵检测领域有广泛的研究, 表 1 总结了传统机器学习代表性工作及方法特点的对应关系。

在传统机器学习领域中, 最常用的构建入侵检测系统的思路是, 首先对数据进行预处理, 例如对数据进行归一化、将符号特征转换为数值特征等, 然后进行特征工程(特征提取、特征选择)。再选择用于分类的传统机器学习算法, 并使用训练数据来训练模型, 使用测试数据进行预测。对于入侵检测的二分类问题而言, 分类器对数据的预测结果是正常数据或攻击数据, 对于多分类问题而言, 分类器的输出是预测该数据是正常数据或某一种具体类型的攻击, 最后使用评估指标对预测结果进行评估。

表 1　传统机器学习代表性工作

Table 1　Traditional machine learning representative work

文献	传统机器学习技术	数据预处理特征转换方法	特征选择/提取	数据集	任务类别	性能评价
[58]	NB	最小化熵离散化方法、k 间隔比例离散化方法	相关性方法、一致性方法、交互方法	10%KDD1999	多分类	*ACC*: 93.72%
[59]	BN	CACC 离散化方法	马尔可夫	NSl-KDD	二分类	*ACC*: 96.92%、*AUC*: 99.33%
[64]	KNN、PCA、模糊聚类	独热编码	PCA	NSl-KDD	多分类	*ACC*: 94% *DR*: 78.86%(Probe)、94.23%(DoS)、80.09%(U2R)、69.87%(R2L)
[65]	KNN	\	MDMS	KDD1999	多分类	*DR*: 99.51%(Probe)、99.29%(DoS)、71.62%(others)
[68]	DT	\	\	CICIDS2017	二分类	*ACC*: 99.66%、 *DR*: 99.3%
[69]	DT	\	信息增益	Kyoto 2006+	多分类	*ACC*: 97.23%、*Precision*: 96.5%、*Recall* : 97.2%、*F-measure*: 96.3% *Precision*: 99.7%(Normal)、96.1%(Attack)、65.5%(Unknown_attack) *Recall*: 99.7%(Normal)、99.7%(Attack)、10.7%(Unknown_attack) *F-measure*: 99.8%(Normal)、97.9%(Attack)、18.5%(Unknown_attack)
[71]	DT	Min-Max 正则化	\	CICIDS2017	二分类	*ACC*: 96.665%、*DR*: 94.475、*FAR*: 1.145%
[73]	SVM	Z-score 正则化	压缩采样	10%KDD1999	多分类	*DR*: 96.42%(Probe)、97.08%(DoS)、90.37%(U2R)、97.64%(R2L) *FPR*: 1.19%(Probe)、0.96%(DoS)、1.26%(U2R)、0.98%(R2L)
[74]	SVM	\	粗糙集理论	KDD1999 ISCX2012 HTTP-CSIC	二分类	KDD1999: *ACC*: 99.95%、*Precision*: 99.9%、*Recall*: 99.8% ISCX2012: *ACC*: 100%、*Precision*: 100%、*Recall*: 100% HTTP-CSIC: *ACC*: 99.98%、*Precision*: 99.9%、*Recall*: 99.8%
[75]	SVM、GA	\	GA	KDD1999	多分类	*ACC:* 99.92%(all)、98.69%(Probe)、99.43%(DoS)、100%(U2R)、98.85%(R2L)
[76]	SVM、LMDRT	\	LMDRT	NSl-KDD Kyoto 2006+ Gure-KDD	二分类	NSL-KDD: *ACC*: 99.31%、*DR*: 99.20%、*FAR*: 0.6% Kyoto 2006+: *ACC*: 98.33%、*DR*: 99.85%、*FAR*: 3.25% Gure-KDD: *ACC*: 99.18%
[77]	SVM、k-means	\	\	KDD Corrected、NSL-KDD、Gure-KDD	二分类	KDD Corrected: *ACC*: 100%、*DR*: 100%、*FAR*: 0 NSL-KDD: *ACC*: 99.7%、*DR*: 99.8%、*FAR*: 0.3% Gure-KDD: *ACC*: 99.5%、*DR*: 96.7%、*FAR*: 0.7%
[82]	LR	\	稀疏逻辑回归	KDD1999	二分类	1074985 条训练数据, 67688 条测试数据 *DR*: 97.65%、训练时间: 11.6s
[85]	k-means、SVM、ELM	对数缩放(以 10 为底)	\	10%KDD1999	多分类	*ACC*: 95.75%、*DR*: 95.17%、*FAR*: 1.87% *DR*: 87.22%(Probe)、99.54%(DoS)、21.93%(U2R)、31.39%(R2L)
[91]	层次聚类、SVM	除以对应属性最大值	\	KDD1999	多分类	*ACC*: 95.72% *DR*: 97.55%(Probe)、99.53%(DoS)、19.73%(U2R)、28.81%(R2L)、99.29%(Normal)
[95]	GMM、DT	\	\	NSl-KDD	二分类	Dataset 1: *ACC*: 94.28%、*Precision*: 97.21%、*FAR*: 8.59% Dataset 2: *ACC*: 94.10%、*Precision*: 96.72%、*FAR*: 9.37%
[99]	PCA、IG、SVM、IBK、MLP	Min-Max 正则化	PCA、IG	ISCX2012, NSL-KDD、 Kyoto 2006+	二分类	ISCX2012: *ACC*: 99.01%、*DR*: 99.1%、*FAR*: 0.01%、*F-measure*: 99.2% NSL-KDD: *ACC*: 98.24%、*DR*: 98.2%、*FAR*: 0.017%、*F-measure*: 98.1% Kyoto 2006+: *ACC*: 98.95%、*DR*: 99.8%、*FAR*: 0.021%、*F-measure*: 99.1%

(注: “\”表示该方案无法参与该评分项。)

数据的质量和检测技术都会对入侵检测结果造成较大影响。在数据质量方面, 随着大数据时代的到来, 入侵数据量大幅增加, 其中存在着大量冗余和不相关数据, 而无监督机器学习技术能够有效地对

海量数据进行处理, 因此聚类和降维技术被广泛用于数据预处理阶段。在检测技术方面, 一些学者使用单一的传统机器学习算法来构建入侵检测系统, 然而单一分类器难以完全概括特定数据集的特征, 从而导致检测率较低。使用两种或多种传统机器学习技术来进行入侵检测通常能够表现出更好的检测性能, 因此大量研究通过集成多种分类器, 来提升入侵检测的检测准确率, 降低总体误差。

虽然没有哪一种传统机器学习方法能够很好地解决所有问题, 需要具体情况具体分析。但是对引用的论文中传统机器学习模型在入侵检测领域中检测能力的对比进行总结和梳理, 仍可以发现一些规律: 1)贝叶斯网络检测能力优于朴素贝叶斯, 因为贝叶斯网络能够处理存在相关关系的特征; 2)基于树的方法(例如决策树)检测能力优于基于概率的方法(例如贝叶斯网络和朴素贝叶斯), 因为基于树的方法计算复杂度低, 能够处理不相关特征的数据; 3)集成学习方法的检测能力优于单个分类器, 因为集成学习方法能够捕获到更多的信息。

表 2 总结了传统机器学习方法解决的问题对比。由表 2 可知, 目前大部分基于传统机器学习方法的 IDS 解决的问题有: 1)误报率高、检测率低; 2)数据特征维度高、数据量大导致检测困难、存在冗余无关数据、需要耗费大量时间; 3)算法模型自身存在的问题, 对算法进行改善。对于上述的三个问题, 已有工作解决的程度总结如下: 1)NB、HMM、KNN、DT、k-means、SVM、GMM 等算法均已用于解决第一个问题, 大部分研究均提升了检测率并降低了误报率, 但是有少量研究的误报率处于 10%以上, 仍有待改进, 并且很多研究评估指标较单一、未列出评估指标具体数值, 难以与其他研究进行比较。2)解决此类问题最常用的方式就是使用降维、聚类等方法先对海量数据进行预处理, 再结合模型进行分类。目前, 已经有较多研究来解决这一问题, 并取得了较好的结果, 但是部分研究只考虑了二分类的情况, 并未对多分类情况进行讨论, 并且随着数据量的增大, 部分研究的模型检测效果逐步降低。因此, 这一问题的研究仍存在提升空间。3)已经有很多研究提出了新的方法来改善算法模型本身存在的一些缺陷, 从而提升模型的检测能力, 例如朴素贝叶斯方法难以解决特征相关性较大的问题、传统的 k-means 算法容易受到噪声和孤立点的影响, 且必须预先设置“k”的值、PCA 对噪声和离群值敏感等算法自身存在的问题, 这些问题均已得到了相应的改善, 但是部分研究检测率较低, 仍有待提升。

无监督机器学习方法和监督机器学习方法的结合使用以及集成学习使用多种分类器来构建入侵检测系统, 已经成为了未来入侵检测的研究趋势, 但是流量数据中的异常流量远远小于正常流量的数据不平衡问题仍是基于传统机器学习方法的入侵检测技术发展的一大瓶颈, 且文中引用的大部分基于传统机器学习方法的 IDS 研究工作均没有解决该问题。

4 基于深度学习的入侵检测

传统的机器学习方法通常需要人工选取特征, 且需要大量的领域专业知识, 是较为浅层的学习方法。然而, 随着网络中海量数据的增加, 网络带宽的提升, 数据的复杂性和特征的多样性也在不断提升, 浅层学习难以达到分析和预测的目的。2006 年, Hinton 教授提出了深度学习理论[102], 与传统机器学习不同, 深度学习方法学习的是样本数据的内在规律和表示层次, 构建多个隐藏层组建的非线性网络结构能够适应较高维度学习和预测的要求, 效率更高, 节省了大量特征提取的时间, 可以根据问题自动建立模型, 不局限于某个固定的问题, 在解决入侵检测问题中很有前景。目前, 深度学习在入侵检测[103]、图像识别[104]、自然语言处理[105]、语音识别[106]等领域取得了很好的效果, 解决了很多复杂模式识别难题, 极大地推动了人工智能技术的进步。本文将从生成方法、判别方法、生成对抗网络等三个方面对基于深度学习方法的入侵检测进行详细阐述。

4.1 生成方法

4.1.1 自动编码器

自动编码器(Auto-Encoder, AE)[103]由编码器和生成重构的解码器组成, 包含输入层、编码层、解码层。深度自动编码器(Deep Auto-Encoder, DAE)相对于 AE 而言, 增加了隐藏层深度, 学习到的特征表示的层次更高。随着海量复杂数据的增长, 数据存在信息的缺失, 对数据进行类别标注是一项费时费力的工作, 作为一种无监督算法, AE 能够表征线性变换和非线性变换, 被广泛用于入侵检测领域中的降维任务。

处理大规模数据是入侵检测系统的主要挑战之一, 传统的 PCA 等线性降维方法难以捕获数据中的非线性信息, 为了解决这一问题, 文献[103]使用具有七个隐藏层的自动编码器在 NSL-KDD 数据集上进行降维实验, 所提方法减少了入侵检测系统时间和空间的复杂性, 且优于 PCA 等传统降维方法。文献[107]提出一种结合堆叠稀疏去噪自动编码器和

表 2 基于传统机器学习方法的入侵检测解决问题对比

Table 2 Comparison of problems solved by intrusion detection based on traditional machine learning methods

类型	文献	传统机器学习技术	解决的问题	存在的缺陷
监督机器学习技术(生成方法)	[56]	NB	误报率高、检测率低	误报率仍较高, 处于 10%~13%之间; 模型评估指标单一, 只有检测率误报率
	[58]	HNB	数据特征维度高、及朴素贝叶斯方法难以解决特征相关性较大的问题	多分类准确率为 93.72%, 仍有待提升; 只使用了 10%的 KDD1999 数据集
	[59]	BN	贝叶斯网络分类器通常选择次优模型的启发式方法来对数据进行训练	评估指标较单一, 只有准确率和 AUC; 只考虑了二分类的情况
	[61]	HMM	误报率高、检测率低	没有考虑有效载荷的长度
	[64]	KNN、PCA、模糊聚类	误报率高、检测率低; 数据集维度较高, 存在冗余数据	随着数据量增加, 该算法检测效率和准确率逐步降低
	[65]	KNN	数据集维度较高	并未列出准确率具体数值
	[66]	KNN、k-means	数据集维度较高, 检测需耗费大量时间	评估指标较单一且不直观
	[67]	KNN	攻击训练数据不足	U2R(8.12%)和 R2L(9.98%)错误率较高; 并未列出检测率具体数值
	[69]	DT	KDD1999 数据集不包含目前的网络攻击现状	未知攻击类别的 *F-measure* 较低
监督机器学习技术(判别方法)	[71]	DT	误报率高、检测率低	模型训练时间较长, 为 159.5s
	[73]	SVM	数据量较大, 检测困难	并未跟别的算法检测情况进行对比
	[74]	SVM	数据量较大, 检测困难	并未跟别的算法检测情况进行对比; 只考虑了二分类的情况
	[75]	SVM、GA	数据量较大, 特征维度高, 检测困难	评估指标较单一, 且检测耗时太长
	[76]	SVM、LMDRT	提升训练数据质量	只考虑了二分类的情况
	[78]	SVM	解决核函数的构造问题	并未跟别的算法检测情况进行对比
	[79]	LR	还没有学者使用多项式逻辑回归算法来预测攻击; 仍不清楚与个别重大攻击相关的风险因素	误报率仍较高, 为 18.9%
	[82]	LR	数据特征维度高, 且存在冗余数据	实验中, 不同对比算法的测试数据集大小不同, 难以直接对评估结果进行比较
	[85]	k-means、SVM、ELM	误报率高、检测率低	并未跟别的算法检测情况进行对比
无监督机器学习技术	[89]	k-means	传统 k-means 算法存在对聚类中心初始值敏感、容易受到噪声和孤立点的影响等问题	最高检测率为 90.2%, 仍有待提升
	[90]	k-means	传统的 k-means 方法必须预先设置“k”的值	所提算法的准确性较低, 为 72.91%
	[91]	层次聚类、SVM	数据集中存在冗余特征数据, 训练时间较长	U2R 和 R2L 检测率较低, 分别为 19.7%和 28.8%
	[95]	GMM、DT	检测率低	误报率较高, 为 9.37%
	[99]	PCA、IG、SVM、IBK、MLP	庞大的数据量对 IDS 构成了持续的挑战	只考虑了二分类的情况
	[101]	模糊 PCA	PCA 对噪声和离群值敏感, 仅限于线性主成分	并未列出评估指标具体数值

softmax 分类器来检测高维网络数据的入侵行为的 IDS, 该方法能够有效检测拒绝服务攻击和探针攻击。对流量数据进行标记需要专业知识和大量时间, 为了减少对数据的标记操作, 文献[108]、文献[109]、文献[110]、文献[111] 分别在不同入侵检测数据集上应用了基于变分自编码器(Variational Auto-Encoder, VAE)的无监督深度神经网络, VAE 是将变分推理与深度学习相结合的概率图模型, 能够重构输入。

传统的日志异常检测忽略了日志的时间模式, 且存在信息丢失问题, 为了弥补这些不足, 文献[112]将多模块卷积自动编码器(Convolutional Auto- Encoder, CAE)配置为不同的并行模块在不同的时间间隔中进行扫描以分别提取信息, 从而捕获时间的依赖性。然后, 将这些潜在空间的特征作为最终输入, 使用 SVM 算法进行分类。所提方法在 Blue Gene/L 日志数据集上的实验准确率为 94.37%, 所提方法优于 SVM 方法(92.62%)。

4.1.2 深度玻尔兹曼机

受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)[113]是一种通过输入数据集学习概率分布的随机生成神经网络, 包含一层可视层和一层隐藏层。深度玻尔兹曼机(Deep Boltzmann Machine, DBM)由多层受限玻尔兹曼机叠加, 是一个完全无向的模型。深度玻尔兹曼机能够从大量无标签数据中学习出高阶

特征, 鲁棒性较好。由于 RBM 能够以无监督的方式从复杂原始数据中学习到有效的高级特征信息, 因此近年来, 被较多学者用于入侵检测领域。文献[113]使用判别受限玻尔兹曼机将生成模型的表达能力与良好的准确分类能力结合起来, 从不完整的训练数据中推断出部分知识。文献[114]探索了受限玻尔兹曼机器作为网络入侵检测的方法, 使用 ISCX2012 入侵检测评估数据集, 但是检测准确率较低。文献[115]使用对比散度(Contrastive Divergence, CD)和持久对比散度(Persistent Contrastive Divergence, PCD)作为训练算法, 并结合受限玻尔兹曼机技术来区分正常和异常的 NetFlow 流量, 特异性(*Specificity measures*, *SPEC*)为 89%~95%, 但是该文献只对两层 RBM 的性能进行了研究。文献[116]对多层 RBM 进行实验来获取最佳 DBM 模型, 结合前馈神经网络(Feed-Forward Neural Network, FFNN)、自动前馈神经网络(Automated Feed-Forward Neural Network, AFFNN)、随机森林(Random Forest, RF)、SVM 四种算法来识别攻击行为。

4.1.3 深度信念网络

深度信念网络(Deep Belief Network, DBN)[117]是一种由若干层 RBM 和一层 BP 组成的有向深层神经网络, 通过隐层提取特征使得后面层次的训练数据更具有代表性, 还可以解决复杂高维数据的检测问题, 已经被应用于入侵检测领域。文献[117]使用无监督贪婪学习算法对 DBN 进行预训练和微调, 以学习非线性高维输入数据的相似性表示。文献[118]使用基于对比散度算法的 RBM 对数据进行训练, 得到数据的低维表示, 通过 BP 算法对 DBN 模型进行微调, 在 NSL-KDD 数据集上的实验结果的分类准确率为 95.25%, 效果优于 SVM(91.36%)和 BP(89.07%), 提升了检测速度, 能够有效检测攻击行为。此外, 将 DBN 算法和其他技术结合也能够提取低维特征。文献[119]使用 DBN 将原始数据转换为低维数据, 采用粒子群优化算法优化每层隐层节点的数量, 并使用概率神经网络用于对低维数据进行分类。文献[120]使用综合少数过采样技术消除类别不平衡问题, 并使用 DBN 来提升入侵分类准确度。文献[121]提出了一种基于改进遗传算法和 DBN 的 IDS, 通过遗传算法的多次迭代, 自适应地生成最优隐藏层数和每层中神经元的最佳数量, 以适应不同攻击类型。文献[122]结合 DBN 和 SVM 来构建 IDS, 提升了检测速度。文献[123]提出了一种新的联合优化算法来优化 DBN 的网络结构, 首先利用基于鱼群思想的粒子群优化算法(Particle Swarm Optimization, PSO)寻找初始优化解, 再基于初始优化解, 使用具有自调整交叉概率和变异概率的遗传算子来优化 PSO 以搜索全局优化解。最后, 将上述联合优化算法构造的全局优化解决方案用于构建 IDS。

4.1.4 循环神经网络

循环神经网络(Recurrent Neural Network, RNN)[124]是一类以序列数据为输入, 在序列演进方向进行递归的神经网络, 具有挖掘数据中的时序信息和语义信息的深度表达能力, 被广泛用于序列相关的入侵检测问题中。文献[124]使用组合 RNN 和模糊 C 均值聚类的云环境 IDS, 在聚类模块中, 使用模糊 C 均值聚类将输入数据集聚类分组, 在分类模块中, 使用 RNN 进行入侵分类。文献[125]结合 RNN 和区域自适应合成过采样算法来提升低频攻击的检测率。

循环神经网络存在梯度消失的问题, 难以建模长时间依赖。长短期记忆网络(Long Short-Term Memory networks, LSTM)[126]通过设计“门”结构实现信息的保留和选择的功能。门控循环单元(Gated Recurrent unit, GRU)是 LSTM 的一种变体, 与 LSTM 网络相比, GRU 结构更加简单, 而且效果也很好。由于 LSTM 和 GRU 能够较好的处理梯度消失的问题, 已被用于检测网络入侵行为。文献[126, 127]分别在 10%KDD1999 数据集和 CIDDS-001 数据集上应用 LSTM, 均能识别攻击行为。文献[128]使用双向长短期记忆神经网络(Bi-directional Long Short-Term Memory, BLSTM)检测物联网网络内的攻击。还有一些研究将 LSTM 方法与其他深度学习技术结合, 来检测大规模数据中的攻击行为。文献[129]利用 AE 的维数降低和特征提取属性来执行重建过程, 使用 LSTM 网络来处理计算机网络数据的序列特征。文献[130]提出了一种基于机器学习和卷积 LSTM 网络(Convolutional-LSTM, Conv-LSTM)的可扩展混合 IDS, 能检测全局和本地潜在威胁攻击。

为了解决流量样本的不平衡问题, 同时考虑流量内的时序关系, 文献[131]使用一种改进的局部自适应合成少数过采样技术来处理不平衡流量数据, 利用 GRU 实现流量异常检测, 具体建模过程如图 4 所示。为了提升 GRU 模型的异常检测能力, 文献[132]使用 SVM 替代 GRU 模型最终输出层中的 softmax 函数, 并将交叉熵函数替换为基于边缘的函数, 所提模型对攻击的检测能力优于传统的 GRU-softmax 模型。

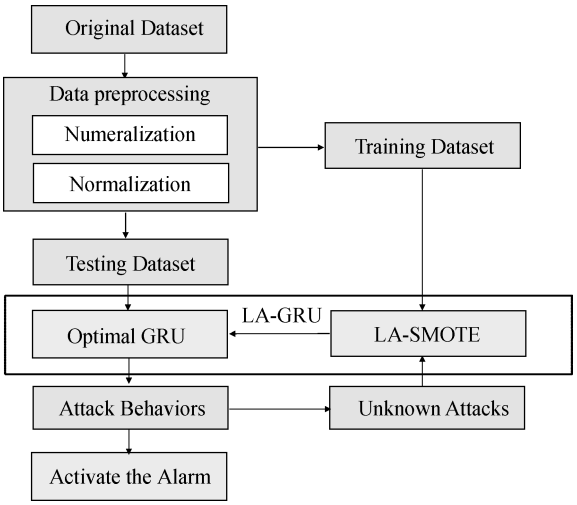


图 4　文献[131]的建模过程

Figure 4　The modeling process of [131]

4.2　判别方法

卷积神经网络(Convolutional Neural Networks, CNN)[133]是一种包含了卷积计算且具有深度结构的前馈神经网络, 能够更准确且高效地提取特征。文献[133]使用 PCA 和 AE 等降维方法去除流量数据中的冗余特征, 再使用 CNN 来提取特征, 将流量转化为图像进行检测。文献[134]通过数据填充将一维数据转换为二维图像数据, 使用 CNN 学习有效特征, 结合 softmax 识别攻击行为。文献[135]提出一种基于深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)的 IDS, 使用随机搜索方法对配置空间的模型参数进行调整。文献[136]使用 CNN 选择流量特征, 并根据数量设置每个类的成本函数权重系数, 来解决数据类别不平衡问题。为了提升检测性能, 一些学者将网络流量转化为序列数据进行处理。文献[137]提出一种基于 CNN 的字符级 IDS, 将网络流量记录视为字符序列, 记录中字符向量被聚合成矩阵, 作为 CNN 的输入。考虑到网络流量的有效载荷是类似于文本的序列数据这一事实, 文献[138]使用词向量和文本卷积神经网络从有效载荷中提取有效信息, 再在统计特征和有效载荷特征的组合上执行随机森林算法进行分类。该文献在 ISCX2012 数据集上进行实验, 所提模型准确率为 99.13%, 误报率为 1.18%, 优于 SVM(86.16%)、随机森林(97.21%)等传统方法, 能有效检测分布式拒绝服务攻击和渗透攻击。针对工业控制系统的入侵检测, 文献[139]提出一种基于 CNN 的过程状态转换的工业控制系统入侵检测算法, 实现了两级异常检测框架。

4.3　生成对抗网络

生成式对抗网络(Generative Adversarial Networks, GAN)[140]是近年来最具前景的无监督方法之一, 通过生成模型和判别模型的相互博弈学习产生高质量输出。网络中异常数据数量远远少于正常数据数量, 如果使用类别比率不平衡的数据集来训练模型, 会严重影响检测结果准确率, GAN 能够处理数据类别不平衡问题。文献[140]使用 Cycle-GAN 先将 ADFA-LD 数据集转换为图像, 再使用 Cycle-GAN 学习正常数据的图像来创建异常数据的图像, 将生成的综合异常数据与原始数据一起用于模型的训练。实验结果表明, 所提方法优于综合采样技术, 显示了生成对抗网络在异常生成中的潜力。

基于机器学习的模型容易受到对抗性的干扰, 导致 IDS 检测率降低, 为了解决这一问题, 文献[141]提出了基于 GAN 的新型对抗网络框架来生成对抗攻击, 该攻击可以欺骗和逃避入侵检测系统。考虑到攻击者不知道检测系统的内部结构, 对抗性攻击示例对检测系统执行了黑盒攻击。为了提升入侵检测系统抵御对抗攻击的健壮性, 文献[142]采用 GAN 在网络数据中引入战略对抗性干扰, 以损害 IDS 性能, 为了应对这种对抗性干扰, 作者在 IDS 机器学习模型中引入 GAN 模型以确保鲁棒性, 具体建模过程如图 5 所示。实验结果表明, GAN 技术不仅可以用来攻击 IDS, 还可以用来增强 IDS 对抗干扰的鲁棒性。

针对 GAN 模型不适合处理具有离散值的数据的问题, 文献[143]提出了一种基于 GAN 损失函数的高效模型, 使用 GAN 学习正常数据的分布, 从学习到的分布中找到最相似的样本, 并通过基于测试样本与找到样本之间差异的强度所定义的异常评分, 来对测试样本的异常程度进行判别。在 10%KDD1999 数据集上的实验结果表明, 所提模型减少了开销, 可基于无监督训练有效识别未知数据中的异常行为。针对物联网中的网络入侵检测, 文献[144]提出了一种分布式生成对抗网络架构。在这种架构中, 每个物联网设备都可以监视自己的数据以及相邻的物联网设备来检测内部和外部攻击。车辆网络中的已知攻击特征很少, 任何攻击都会严重影响驾驶员安全, 文献[145]提出一种使用生成对抗网络的车载网络入侵检测模型, 使用简单的独热向量对控制器局域网的 ID 进行编码, 在训练过程中使用的是随机的假数据, 而不是真实的攻击数据。

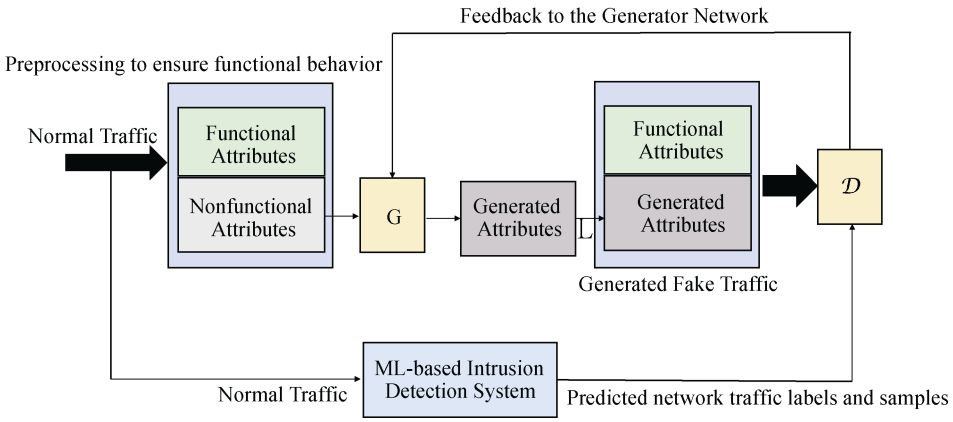


图 5　文献[142]的建模过程

Figure 5　The modeling process of [142]

4.4　总结与讨论

深度学习技术的发展促进了入侵检测研究的进步, 能够利用分层结构对数据进行无监督特征学习和模式分类, 可以将特征提取器和分类器集成到一个框架中, 不需要安全专家手动提取特征。深度学习能够有效处理大规模网络流量数据, 相比于浅层的传统机器学习方法, 具有更高的效率和检测率, 但是其训练过程较复杂, 模型可解释性较差。表 3 总结了深度学习代表性工作及方法特点的对应关系。由于常用的入侵检测数据集中存在 3.1 节中所述的问题, 因此在使用深度学习方法对其进行检测之前, 也需要对数据进行预处理。表 4 总结了深度学习方法解决的问题对比。由表可知, 目前基于深度学习的 IDS 解决的问题有: 1)网络中异常流量数据远远少于正常流量数据的类别不平衡问题; 2)网络数据量较大、特征维度增加、浅层机器学习技术难以对海量高维数据进行检测、难以提取数据中非线性特征信息; 3)对算法模型本身进行改善。

对于上述的三个问题, 已有工作解决的程度总结如下: 1)如果使用不平衡数据集来构建入侵检测模型, 则会对检测准确率造成严重影响。通常用于解决数据中的不平衡问题有两种方法, 分别是在算法级别上的解决方案(成本函数方法)和在数据级别上的解决方案(欠采样和过采样方法)。当前已经有一些研究使用这两种方法来处理 IDS 中的数据类别不平衡问题, 但是大部分研究工作都没有考虑到数据不平衡问题。随着近几年生成能力较强的 GAN 的出现, 为解决数据类别不平衡问题提供了新思路, GAN 能够通过生成新的数据来解决该问题, 因此, 数据类别不平衡问题仍然是目前研究的热点问题。2)大部分深度学习算法例如 AE、DBM、DBN、LSTM、CNN 等都已经用于解决这一问题, 在文中所引用的深度学习对比实验中, 可以发现, 通常深度学习方法的检测准确率优于传统机器学习技术, 例如 SVM、决策树、朴素贝叶斯、随机森林等, 体现了深度学习检测方法的有效性。但是, 一些基于深度学习 IDS 研究的二分类和多分类问题的检测准确率等评估指标的值仍有待提升。例如, 基于 NSL-KDD 数据集中 KDDTest+测试集上的多分类问题准确率大致处于 79%~85%的范围内, 而 KDDTest-21 测试集上的多分类问题准确率大致处于 60%~69%的范围内, 仍存在提升空间。3)一些研究对深度学习算法进行了改善, 从而提升模型检测能力, 例如使用 SVM 替换 softmax 函数提升 GRU 模型的检测能力、使用遗传算法优化 DBN 的网络结构等。还有针对深度学习模型存在的问题进行改善的研究, 例如文献[143]研究的是现有的基于 GAN 的模型不适合处理具有离散值的数据, 导致检测性能较差。但是部分研究并未列出具体评估数值, 难以与已有其他方法进行对比。

基于深度学习方法的入侵检测分为生成方法和判别方法, 其中大部分深度学习方法均为生成方法, 表明深度学习非常适合用于完成特征工程的任务。例如 AE, 能够在不需要类别标注的情况下, 对海量数据进行特征学习并降维, 相比于传统的降维技术有很大的优势。具体而言, 传统的线性降维技术例如 PCA, 难以提取数据中的非线性结构, 而传统的非线性降维技术例如基于核函数的 PCA, 具有固定的模型参数, 难以应用于大规模数据集中。而 AE、DBM 等深度学习方法能够较好地提取大规模高维数据中的非线性结构信息, 获得更多的隐藏信息。基于深度学习方法的入侵检测还能有效处理网络中的数据集不平衡问题, 例如具有强大生成能力的 GAN 方法, 能够通过生成器和判别器的博弈对抗来生成异常数据。除此之外, GAN 还能有效处理机器学习模型的对

表 3　深度学习代表性工作

Table 3　Deep learning representative work

文献	技术	数据预处理特征转换方法	特征选择/提取技术	数据集	任务类别	性能评价
[103]	AE	\	AE	NSL-KDD	多分类	训练数据: *ACC*: 91.49%、测试数据: *ACC*: 91.46%
[110]	VAE	\	VAE	10%KDD1999	二分类	*F-meaure*: 81.2%、*AUC*: 95.1%
[112]	CAE、SVM	\	CAE	Blue Gene/L	二分类	*ACC*: 94.37%、*Precision* : 71.81%、*Recall* : 75.81%、*F-meaure* : 73.76%
[115]	RBM、CD、PCD	连续特征离散化过滤器	\	ISCX2012	二分类	CD: *ACC*: 89.2%、*DR*: 89.2%、*SPEC*: 89.9% PCD: *ACC*: 89.7%、*DR*: 85.6%、*SPEC*: 94.8%
[116]	DBM、SVM、RF、FFNN、AFFNN	\	RBM	智能配水设备数据集	二分类	FFNN: *F-meaure*: 97.83%、AFFNN: *F-meaure*: 97.83% RF: *F-meaure*: 99.48%、SVM: *F-meaure*: 97.83%
[118]	DBN、CD	独热编码 Min-Max 正则化	\	NSL-KDD	二分类	*ACC*: 95.25%、分类时间: 11.65s
[121]	DBN、GA	Min-Max 正则化	GA	NSL-KDD	多分类	*ACC*: 99.37%(Probe)、99.45%(DoS)、98.68%(U2R)、97.78%(R2L) *DR*: 99.4%(Probe)、99.7%(DoS)、98.2%(U2R)、93.4%(R2L) *FAR*: 0.7%(Probe)、0.8%(DoS)、1.8%(U2R)、7.3%(R2L)
[123]	DBN、PSO、GA	独热编码、Min-Max 正则化	GA	NSL-KDD	二分类+多分类	二分类: KDDTrain+: *ACC*: 99.85%、KDDTest+: *ACC*: 83.86%、KDDTest-21: *ACC*: 68.75% 多分类: KDDTrain+: *ACC*: 98.55%、KDDTest+: *ACC*: 82.36%、KDDTest-21: *ACC*: 66.25%
[126]	LSTM	\	\	10%KDD1999	多分类	*ACC*: 96.93%、*DR*: 98.88%、*FAR*: 10.04%
[127]	LSTM	\	\	CIDDS-001	多分类	*ACC*: 84.83%、*Precision*: 85.14%、*Recall*: 88.34%
[128]	BLSTM	\	\	UNSW-NB15	二分类	*ACC*: 95.71%、*F-meaure*: 98%、*Recall* : 96%
[129]	LSTM、AE	\	AE	ISCX2012	二分类	*F-meaure*: 85.83%、*AUC*: 95.12%
[130]	Conv-LSTM	\	\	ISCX2012	二分类	*Precision* : 97.25%、*Recall* : 97.5%、*DR*: 97%、*FAR*: 0.71%、*F-meaure*: 97.29%
[131]	GRU	独热编码、Min-Max 正则化	\	NSL-KDD	多分类	训练数据集大小: 73906 条。 *ACC*: 99.04%、*DR*: 98.92%、*FAR*: 0.134% *DR*: 99.20%(Probe)、99.16%(DoS)、98.61%(U2R)、98.34%(R2L)、99.21%(Normal)
[133]	CNN、AE	独热编码、Min-Max 正则化	PCA、AE、CNN	10%KDD1999	多分类	*ACC*: 94%、*DR*: 93%、*FAR*: 0.5%
[135]	DCNN	\	\	NSL-KDD	二分类	KDDTest+: *ACC*: 85.22%、*AUC*: 96.5% KDDTest-21: *ACC*: 69.56%、*AUC*: 92.6%
[136]	CNN	独热编码、Min-Max 正则化	\	NSL-KDD	多分类	KDDTrain+: *ACC*: 99.46%、KDDTest+: *ACC*: 79.48%、KDDTest-21: *ACC*: 60.71%、*DR*: 68.66%、*FAR* : 27.90% *DR*: 81.87%(Probe)、83.21%(DoS)、13.00%(U2R)、21.68%(R2L) *FAR*: 2.09%(Probe)、2.35%(DoS)、0.06%(U2R)、0.69%(R2L)
[138]	CNN	\	\	ISCX2012	多分类	*ACC*: 99.13%、*DR*: 99.26%、*FAR*: 1.18% *DR*: 99.77%(Infiltration)、99.95%(BFSSH)、95.93%(DDoS)、99.70%(HttpDoS) *FAR*: 0.06%(Infiltration)、0.00%(BFSSH)、0.40%(DDoS)、0.07%(HttpDoS)
[140]	GAN	\	\	ADFA-LD	二分类	*F-meaure*: 41.64%、*AUC*: 71.30%
[143]	GAN	独热编码或虚拟编码	\	10%KDD1999	二分类	*Precision*: 93.24%、*Recall*: 94.73%、*F-meaure*: 93.98%

(注: “\”表示该方案无法参与该评分项。)

表 4 基于深度学习方法的入侵检测解决问题对比

Table 4 Comparison of problems solved by intrusion detection based on deep learning methods

类型	文献	深度学习技术	解决的问题	存在的缺陷
生成方法	[103]	AE	传统的降维方法难以捕获数据中的非线性信息	模型评估指标较单一, 只有准确率
	[109]	VAE	对流量数据进行标记需要专业知识和大量时间	NSL-KDD 数据集上 *Recall* 值较低为 85.99%
	[110]	VAE	挖掘数据的潜在信息并构造出简练的数据表示	*F-measure* 较低为 81.2%, 仍有待提升
	[112]	CAE	传统的日志异常检测忽略了日志的时间模式, 且存在向量表示造成的信息丢失问题	*F-measure* 较低为 73.76%, 仍有待提升
	[113]	RBM	攻击技术和方式的不断变化, 入侵检测较困难	并未列出评估指标具体数值
	[115]	RBM、CD、PCD	数据维度较高, 导致入侵检测变得困难	准确率为 89.7%, 仍有待提升; 文献只对两层 RBM 的性能进行研究
	[116]	DBM、SVM、RF、FFNN、AFFNN	分布式拒绝服务攻击是智能城市基础设施面临的最广泛的威胁之一	模型评估指标较单一, 只有 *F-measure*
	[117]	DBN	网络流量的不断增长, IDS 受到高维度入侵数据和复杂攻击类型的困扰	并未列出检测率、误报率具体数值
	[118]	DBN、CD	随着特征数量的增加, 数据变得越发复杂, 难以有效提取更好的低维特征	模型评估指标较单一, 只有准确率和时间
	[121]	DBN、GA	需要设计一个自适应模型来改变网络结构以适应不同的攻击类型	并未列出总体检测准确率、检测率和误报率
	[123]	DBN、PSO、GA	优化 DBN 的网络结构	KDDTest+和 KDDTest-21 的多分类准确率分别是 82.36%和 66.25%, 仍有待提升
	[126]	LSTM	传统机器学习技术无法利用海量数据解决入侵分类	误报率较高为 10.04%
	[127]	LSTM	网络数据量较大	多分类准确率较低为 84.83%
	[128]	BLSTM	传统机器学习技术无法利用海量数据解决入侵分类	只考虑了二分类的情况; 未考虑流量数据类别不平衡问题
	[129]	LSTM、AE	网络数据量较大	二分类 *F-measure* 较低为 85.38%
	[131]	GRU	数据集不平衡问题; 忽略了流量内的时序关系	只使用了 NSL-KDD 数据集, 并未在别的数据集上验证所提模型效果
	[132]	GRU、SVM	优化 GRU 的网络结构	评估指标单一, 只用了混淆矩阵 4 个值
	[133]	CNN、AE	传统机器学习模型无法确定数据特征之间的关系	多分类检测率为 93%, 仍有待提升
判别方法	[135]	DCNN	网络数据量巨大	KDDTest+和 KDDTest-21 的二分类准确率分别是 85.22%和 69.56%, 有待提升
	[136]	CNN	数据集不平衡问题	KDDTest+和 KDDTest-21 的多分类准确率分别是 79.48%和 60.71%, 有待提升
	[138]	CNN	将 NLP 中的现代深度学习技术应用于网络有效载荷的特征提取, 来提升入侵检测性能	未考虑流量数据类别不平衡问题
生成对抗网络	[140]	GAN	数据集不平衡问题	二分类 *F-measure* 较低为 41.64%
	[143]	GAN	基于 GAN 的模型不适合处理具有离散值的数据	只使用了 10%的 KDD1999 数据集

抗性干扰问题。

对于深度学习而言, 将生成模型与判别模型结合使用已经成为了未来的发展方向。然而, 基于深度学习方法的入侵检测也面临一些挑战: 1)训练速度、计算存储问题。深度学习的训练过程非常复杂, 负荷较大, CPU 已经难以满足深度学习所需要的强大计算量, 现在通常需要多个 GPU 并行来完成对大规模数据的处理任务; 2)模型调参问题。针对不同的问题, 采用多少层网络结构以及具体使用多少个神经元也是需要考虑的问题; 3)模型优化问题。在对深度学习模型进行训练时, 会遇到梯度爆炸、梯度消失、局部最优点的问题; 4)实时检测问题。网络中高维海量数据不断增加, 且复杂性较高, 将深度学习方法用于实时入侵检测也面临着一定的挑战。

5 基于强化学习的入侵检测

强化学习(Reinforcement Learning, RL)是用于描述和解决代理在与动态环境的交互过程中通过对策略的学习, 达到回报最大化或者达到特定目标的问题。不同于监督学习, 强化学习不需要明确的指导信号, 并且可以为复杂的随机任务自动构建顺序最优策略。目前, 强化学习已经被广泛用于机器人控制[146]、工业制造[147]、游戏博弈[148]等领域。强化学习是一个通用的框架, 奖励机制灵活, 一旦完成训练, 产生的策略功能通常是一个简单快速的神经网

络; 使用简单的奖励机制, 可以快速对网络状态的变化作出响应, 适合用于在线训练。

文献[149]将入侵检测问题转换为预测马尔可夫奖励过程的价值函数, 使用线性基函数的时间差异算法来进行值预测, 从而准确地预测主机过程的异常时间行为。文献[150]使用自动学习机方法, 通过与随机环境的交互, 能够从一组动作中选出最佳动作, 从而解决降维问题。与传统的 SVM(86.13%)相比, 所提方案实现了更高的准确率, 能达到 96.13%, 能够有效识别拒绝服务攻击。为了检测复杂的分布式攻击, 文献[151]在网络传感器代理的分层架构中提出了分布式强化学习方法, 但是所提模型精准率较差。因此, 文献[152]提出了一种分布式传感器和决策代理的体系结构来改善这一问题。文献[153]研究了分布式强化学习对入侵响应的适用性, 但是系统无法仅通过考虑流量来区分合法流量和攻击流量。

表 5 基于强化学习方法的入侵检测解决问题对比

Table 5 Comparison of problems solved by intrusion detection based on reinforcement learning methods

类型	文献	技术	解决的问题	存在的缺陷
强化学习	[149]	马尔可夫奖励过程模型	已有方法对于攻击检测的准确率和计算效率较低	模型评估指标较单一, 只有误报率和训练时间
	[150]	自动学习机	入侵检测数据集中存在冗余特征	模型评估指标较单一, 只有准确率和测试时间
	[151]	分布式强化学习	已有技术难以检测新的和复杂的分布式攻击	精准率处于 37%~92%, 有待提升
	[152]	分布式强化学习	检测分布式拒绝服务攻击的复杂性较高, 精准率差	并未列出检测结果具体数值
	[153]	分布式强化学习	攻击的分布式性质还未得到较好的解决	无法仅通过考虑流量来区分正常和攻击
深度强化学习	[155]	深度强化学习	强化学习中存在一些难以解决的决策问题	训练时间非常长, 且准确率较低为 80.16%

深度学习的兴起也推动了强化学习的进步, 形成了深度强化学习(Deep Reinforcement Learning, DRL)[154], 基本框架如图 6 所示。文献[155]提出一种基于神经网络的强化学习对抗性环境算法, 首次将对抗强化学习应用于入侵检测, 并将环境行为融入到改进的强化学习算法的学习过程中。将入侵检测中的术语与深度强化学习中的术语一一对应。该模型集成了强化和监督框架, 产生的环境能够与通过网络特征和相关入侵标签形成的预先记录的样本数据集进行交互, 并且选择具有优化策略的样本以实现最佳分类效果。

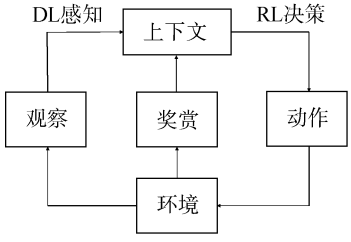


图 6 DRL 框架原理图[155]

Figure 6 DRL frame principle diagram [155]

如表 5 所示, 基于强化学习的入侵检测研究才刚刚起步。一方面, 深度强化学习结合了深度学习的感知能力和强化学习的决策能力, 能够通过端对端的学习方式来实现从原始输入到输出的控制。随着深度学习领域的不断发展, 深度强化学习方法会朝着模块复杂化、结构多样化的方向发展。基于深度强化学习的入侵检测方法能够具有类似高度非线性模型的预测性能, 预测耗费时间更少, 且能够处理高度不平衡的数据, 因此非常具有发展前景。另一方面, 深度强化学习自身也存在采样效率较低、奖励函数的设置困难、目标局部最优等问题。尤其是在入侵检测场景中, 当前的强化学习技术难以检测新的和复杂的分布式攻击行为, 且模型检测的性能不佳。因此, 基于强化学习的入侵检测研究仍有待进一步的探索。

6 基于可视化分析的入侵检测

随着网络攻击数量的增多、攻击规模的扩大、攻击复杂性的增加, 入侵检测系统通常会产生大量的告警信息, 而其中误报信息较多, 人工遍历警报日志需要耗费大量人力和时间。网络安全可视化技术通过将各种网络安全数据、警报信息等进行可视化, 将抽象的信息转换为便于直观理解的图像信息, 能够帮助安全管理人员快速识别潜在的异常事件和攻击行为[156-158]。目前, 相关研究人员也开发了很多可视化工具[159, 160]来辅助检测异常, 增强安全态势感知能力。网络安全可视化在全局性、智能性、高交互性等方面具有极大的优势, 因此成为了入侵检测领域的热点研究问题之一。

数据源是网络安全可视化的基础, 基于可视化分析方法的入侵检测领域中常用的网络数据源包含

网络流量数据和日志数据。本文根据可视化所处理的网络数据类型, 对入侵检测相关工作进行总结。

6.1 网络流量数据

文献[161]提出了一种基于四角星的可视化特征生成方法, 用于评估五分类问题中样本之间的距离, 使人们能够识别数据所属的类别, 所提方法能够较好地解决高维特征空间的多分类问题。为了解决机器学习技术在入侵检测中经常出现错误分类的问题, 文献[162]在常用的 NSL-KDD 和 UNSW-NB15 数据集上进行可视化展示。所提方法促进了安全管理人员对基于网络的入侵检测数据集的理解, 并反映了各种类型的网络流量之间的几何关系。文献[163]提出使用可视化技术来促进人们对机器学习复杂技术的理解, 调查了网络入侵数据集中二维降维技术的性能, 演示了在三维空间中可视化数据的结果。文献[164]介绍了一种基于多变量异常检测的可视化工具, 结合了 PCA 方法。该文献将这些方法与交互式可视化功能相结合, 生成的页面工具允许用户浏览网络中收集的大量数据, 便于检测攻击行为。

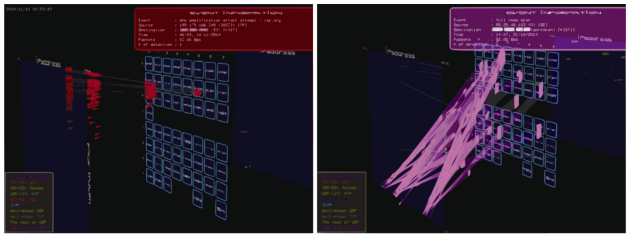
表 6 基于可视化分析方法的入侵检测解决问题对比

Table 6 Comparison of problems solved by intrusion detection based on visual analysis methods

数据类型	文献	技术	解决的问题	存在的缺陷
网络流量数据	[161]	基于四角星的可视化特征生成方法	数据的数量和复杂性显著增加, 人类对于数据分析的感知能力存在较大限制	仍然存在提高分类器准确性和改进视觉图形特征生成规则的空间
	[162]	3D 可视化网络数据	机器学习技术经常存在错误分类的情况	未实现实时网络流量可视化
	[163]	三维空间中可视化数据	降维技术、机器学习技术的复杂内在机制难以直接被人们理解	无法确定分类器在 NSL-KDD 的验证和测试集中产生相对较好结果的特定维数
	[164]	用于全网入侵检测的新型可视化工具	还没有工具能够识别网络安全/流量数据中复杂混合信息, 难以诊断异常	并未将用户界面与特征连接起来
日志数据	[165]	实时三维可视化	基于文本的用户界面难以适应增加的安全事件的数量, 超出了安全运营商的控制范围	未对检测结果进行评估说明
	[166]	IDS 警报的新型径向可视化工具	网络管理员手动遍历基于文本的警报日志来检测威胁非常麻烦	可扩展性较差, 对于大量数据的预处理效率需要提升, 缺少流量等安全数据

6.2 日志数据

基于文本的用户界面难以适应增加的安全事件的数量, 为了对现有基于文本的用户界面进行补充, 文献[165]设计了针对安装在多个网络中的 IDS 的安全事件可视化的工具来实现实时三维可视化, 由三个平行的平面正方形组成, 分别代表全局源网络, 目标网络和全球目的网络。大规模事件的可视化表示如图 7 所示, a 图展示了 DNS 放大攻击的可视化图, 该攻击的源地址是土耳其, b 图展示了 Xmas 扫描信息, 该信息来自丹麦, 发往对应的目标组织。可视化结果表明, 该工具能够有效地显示网络安全运营中心中 IDS 收集的大量安全事件日志。文献[166]提出了新型径向可视化工具 IDSPlanet, 可以帮助管理员分析 IDS 警报日志, 分析攻击模式, 了解不断变化的网络情况。IDSPlanet 由计时环, 警报和交互式核心组成, 这些组件分别编码警报类型的时间特征, 主机中的行为模式以及警报类型, 攻击者和目标之间的相关性。实验采用了 Snort 生成的 IDS 日志进行案例研究, 实验证明 IDSPlanet 能够有效地检测误报和潜在威胁。



(a) DNS amplification attacks (b) Xmas scanning

图 7 大规模事件的可视化表示[165]

Figure 7 Visual representation of large-scale events[165]

6.3 总结与讨论

网络安全可视化将网络安全数据分析和可视化技术相结合, 把高维度不可见的抽象数据用图像的方式进行展现, 并通过提供图形交互工具, 提高了安全管理人员感知、分析、理解网络安全态势的能力[167]。表 6 总结了基于可视化分析方法解决的问题对比, 通过可视化不断增加的海量数据, 从而辅助

分析人员更好地识别攻击行为。网络安全可视化属于一个新型研究方向, 虽然目前已经取得了很多较好的成果, 但是随着网络环境日益复杂, 网络中攻击形式逐渐多样化、网络数据越发庞大, 基于可视分析方法的入侵检测仍面临着一些挑战: 1)如何实时显示并处理海量数据。随着互联网中带宽的增加, 数据复杂性的增加, 对海量数据的处理需要耗费大量时间和计算资源, 实时显示并处理大规模数据变得更加困难; 2)如何搭建入侵检测可视化协同工作环境。随着网络的复杂化, 一些复杂问题通常需要多视图、多数据源、多人协同分析, 已有的技术难以较好地用于检测分析, 在多视图多数据源多人协同方面还有较大的发展空间; 3)如何提升入侵检测可视化系统的易用性、交互性、扩展性。基于入侵检测的可视化需要在网络数据的不同属性和图形属性建立映射关系, 不同的研究者的设计思路和设计结果均会产生较大的差异, 交互设计也不尽相同, 因此, 用户对于不同的入侵检测可视化系统的切换使用变得非常困难, 即易用性、交互性较低, 导致安全管理人员工作效率较低。入侵检测可视化系统通常是在有限的区域内进行图像展示, 随着网络环境的变化, 会造成图像重叠等现象, 即扩展性较低; 4)如何建立一套规范统一的可视化评估体系。目前网络安全可视化缺乏统一的评估指标和理论基础, 不同研究者出于不同的需求, 来构建基于入侵检测的可视化系统, 当前评估方法主观性太强, 难以对不同的可视化系统进行客观地评估。

7 讨论与展望

随着网络环境的日益复杂化, 攻击方式不断变化、攻击种类不断增加, 入侵检测对于维护网络空间安全起着重要的作用。本文从以下几个方面对 IDS 的研究进行讨论:

1) 入侵检测数据的选择

入侵检测系统的构建常常需要用到有类别标注的数据集, 目前使用最广泛的公开数据集有 KDD1999、DARPA1998、DARPA1999, 但是这些数据集的年代久远, 无法反映当前最新的网络入侵行为模式, 因此相关研究的检测结果不具有代表性。近年来, 一些安全机构陆续发布了一些较新的数据集, 例如 UNSW-NB15、CICIDS2017、CIDDS-001 等。这些数据集包括了新出现的一些攻击类型, 例如后门攻击、蠕虫攻击、Shellcode 攻击等。在未来的研究中, 使用较新的数据集将会更具说服力, 更能有效地检测入侵行为。

2) 入侵检测面临的挑战

当前入侵检测研究的主要挑战包括海量高维数据、数据不平衡和实时检测等问题。

- 海量高维数据问题。网络数据量的成倍增长, 对 IDS 来说是首先需要解决的问题。已有的方法中, 常用的是基于传统机器学习方法的入侵检测和使用深度学习方法的入侵检测。基于传统机器学习方法通常先使用聚类或降维的方式对大规模数据进行处理, 再使用分类器对处理好的数据进行类别划分。但是传统机器学习是浅层的技术, 难以达到分析的目的, 使用较多的是深度学习方法, 深度学习方法具有强大的表征能力, 适合用于特征工程, 对海量高维数据进行处理, 但是深度学习的训练过程复杂, 甚至需要多个 GPU 设备并行处理, 并且已有的研究工作评估指标结果较差, 检测性能仍有较大的提升空间。因此, 在未来, 对于高维海量数据的处理仍是入侵检测领域面临的难点之一。
- 数据不平衡问题。由于异常流量数据远远小于正常流量数据, 数据不平衡问题严重影响入侵检测系统的检测准确率。在已有的研究中, 成本函数等算法级别上的解决方案在 IDS 数据不平衡问题上的研究较少, 部分研究是通过使用欠采样、过采样等方法来解决数据不平衡问题, 但是欠采样的方式缩小了整体的样本数量, 而过采样的方式又容易引发过拟合问题, 并不能较好地处理数据不平衡问题。深度学习方法中的 GAN 具有强大的生成能力, 能够通过生成异常数据来解决该问题, 但是目前这方面的研究还较少。因此在未来的研究中, 数据不平衡问题仍是需要解决的难点问题。
- 实时检测问题。目前大部分入侵检测系统的研究都是使用公开数据集, 研究离线的入侵检测。而随着网络中数据规模的扩大, 攻击种类和数量的增加, 对攻击行为进行实时检测变得越发重要, 实时的攻击检测是亟需解决的问题。

3) 检测技术的发展

基于传统机器学习的入侵检测是较浅层的方法, 难以捕获到一些重要信息, 深度学习方法能够有效地从大规模数据中提取出重要信息, 表征学习能力强, 基于深度学习方法的入侵检测系统通常具有更高的处理效率和更高的检测准确率, 因此, 未来在入侵检测领域中, 深度学习方法、强化学习方法、可视化方法会应用地越来越广泛。这些方法也可以进

行组合使用从而提升入侵检测系统的检测性能, 例如深度强化方法就是深度学习方法和强化学习方法相结合的产物, 目前前沿的深度强化学习研究方向有分层深度强化学习(将复杂任务分解成若干子任务)、多智能体深度强化学习(多智能体合作完成任务)、多任务迁移深度强化学习(训练单个模型来完成多个任务)等。同样地, 也可以将可视化分析方法与深度学习、强化学习结合使用, 来构建更强大的入侵检测系统。

4) 检测性能的评估

在入侵检测领域中, 通常准确率、检测率、精准率、*F-measure* 值越高, 误报率越低, 则说明该 IDS 检测性能较好, 但是部分文献中使用的是其他的评估指标, 难以和已有工作进行对比, 且不同的文献使用的数据集不同, 或者是同一数据集的不同子集, 也难以统一进行对比。虽然很多文献中所提方法的入侵检测准确率等评估指标较高, 但是通常是对二分类的检测结果, 并未对多分类的检测结果进行详细说明, 且未列出测试集评估指标的具体数值, 评估指标较单一。在未来的研究中, 统一使用常用的评估指标对测试数据集的检测结果进行详细展示, 将有利于增加说服力, 能够较好地体现所提方法的泛化性能。

5) 应用领域

目前, 入侵检测技术在军事、医疗、交通、物联网安全、工业控制等领域均有广泛的应用, 未来可能会在更多的领域出现入侵检测的身影, 对于跨网络的入侵检测研究也许会增加, 例如将交通领域和工业控制领域结合的入侵检测研究, 从而既保障了交通领域的安全又维护了工业控制领域的安全。随着物联网技术的兴起, 其他领域和物联网领域相结合的入侵检测研究也非常有价值。

8 结束语

随着互联网的快速发展和普及, 互联网面临着严峻的安全问题, 入侵检测技术作为一种主动的安全防御手段已在学术界和工业界获得了广泛的研究。本文对入侵检测技术和系统相关研究情况进行了总结, 从数据来源和检测技术对入侵检测系统进行划分, 其中, 数据来源分为基于主机和基于网络两部分内容, 检测技术分为基于签名和基于异常两部分内容。基于异常的入侵检测系统从统计学习、传统机器学习、深度学习、强化学习和可视化分析五个方面进行详细梳理与总结。本文旨在对基于异常的入侵检测系统中新出现的大数据技术进行概述, 为目前入侵检测领域的研究提供框架及总结, 并为相关安全研究人员提供各类研究技术现状、面临的挑战以及有待改进的方向。

致 谢 感谢中国科学院网络测评技术重点实验室的各位老师和同学提出的有益建议。感谢审稿专家和编辑部老师对本文提出的有益建议及指导。

参考文献

[1] Wu J X, Li J H, Ji X S. Security for Cyberspace: Challenges and Opportunities[J]. *Frontiers of Information Technology & Electronic Engineering*, 2018, 19(12): 1459-1461.

[2] Internet Security Threat Report. Symantec, https://www. symantec.com/security-center/threat-report, Feb. 2019.

[3] J.P. Anderson, Computer security threat monitoring and surveillance[R]. Technical report, James P.Anderson Company, Fort Washington, Pennsylvania, 1980.

[4] Ghosh K, Neogy S, Das P K, et al. Intrusion Detection at International Borders and Large Military Barracks with Multi-sink Wireless Sensor Networks: An Energy Efficient Solution[J]. *Wireless Personal Communications*, 2018, 98(1): 1083-1101.

[5] P. Yi, F. Zou, V. Zou, et al. Performance analysis of mobile ad hoc networks under flooding attacks[J]. *J. Syst. Eng. Electron*, 2011, 22(2): 334-339.

[6] Gurung S, Chauhan S. Performance Analysis of Black-hole Attack Mitigation Protocols under Gray-hole Attacks in MANET[J]. *Wireless Networks*, 2019, 25(3): 975-988.

[7] Le Fessant F, Papadimitriou A, Viana A C, et al. A Sinkhole Resilient Protocol for Wireless Sensor Networks: Performance and Security Analysis[J]. *Computer Communications*, 2012, 35(2): 234-248.

[8] Arthur M P, Kannan K. Cross-layer Based Multiclass Intrusion Detection System for Secure Multicast Communication of MANET in Military Networks[J]. *Wireless Networks*, 2016, 22(3): 1035-1059.

[9] Mitchell R, Chen I R. Behavior Rule Specification-Based Intrusion Detection for Safety Critical Medical Cyber Physical Systems[J]. *IEEE Transactions on Dependable and Secure Computing*, 2015, 12(1): 16-30.

[10] M.B. Mohamed, A. Meddeb-Makhlouf, A. Fakhfakh. Intrusion cancellation for anomaly detection in healthcare applications[C]. *International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019: 313-318.

[11] Al-Jarrah O Y, Maple C, Dianati M, et al. Intrusion Detection Systems for Intra-Vehicle Networks: A Review[J]. *IEEE Access*, 2019, 7: 21266-21289.

[12] Choi W, Joo K, Jo H J, et al. VoltageIDS: Low-Level Communica-

tion Characteristics for Automotive Intrusion Detection System[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(8): 2114-2129.

[13] S. Checkoway, D. McCoy, B. Kantor, et al. Comprehensive Experimental Analyses of Automotive Attack Surfaces[C]. *USENIX Security Symposium*, 2011:89-96.

[14] Gao L L, Li F, Xu X, et al. Intrusion Detection System Using SOEKS and Deep Learning for In-vehicle Security[J]. *Cluster Computing*, 2019, 22(S6): 14721-14729.

[15] Halder S, Ghosal A, Conti M. Efficient Physical Intrusion Detection in Internet of Things: A Node Deployment Approach[J]. *Computer Networks*, 2019, 154: 28-46.

[16] Chaabouni N, Mosbah M, Zemmari A, et al. Network Intrusion Detection for IoT Security Based on Learning Techniques[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2671-2701.

[17] Jan S U, Ahmed S, Shakhov V, et al. Toward a Lightweight Intrusion Detection System for the Internet of Things[J]. *IEEE Access*, 2019, 7: 42450-42471.

[18] Hu Y, Yang A, Li H, et al. A Survey of Intrusion Detection on Industrial Control Systems[J]. *International Journal of Distributed Sensor Networks*, 2018, 14(8): 155014771879461.

[19] Langner R. Stuxnet: Dissecting a Cyberwarfare Weapon[J]. *IEEE Security & Privacy Magazine*, 2011, 9(3): 49-51.

[20] R.M. Lee, M.J. Assante, T. Conway. Analysis of the cyber attack on the Ukrainian power grid[C]. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016:96-102.

[21] Shen C, Liu C, Tan H L, et al. Hybrid-Augmented Device Fingerprinting for Intrusion Detection in Industrial Control System Networks[J]. *IEEE Wireless Communications*, 2018, 25(6): 26-31.

[22] S.H. Amer, J.A. Hamilton. Intrusion Detection Systems (IDS) Taxonomy-A Short Review[J]. *Journal of Software Technology*, 2010,23(2):102-110.

[23] S. Maza, M. Touahria. Feature Selection Algorithms in Intrusion Detection System: A Survey[J]. *TIIS*, 2018, 12(10): 5079-5099.

[24] Mishra P, Varadharajan V, Tupakula U, et al. A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(1): 686-728.

[25] 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.

[26] Ring M, Wunderlich S, Scheuring D, et al. A Survey of Network-based Intrusion Detection Data Sets[J]. *Computers & Security*, 2019, 86: 147-167.

[27] T.R. Glass-Vanderlan, M.D. Iannacone, M.S. Vincent, et al. A Survey of Intrusion Detection Systems Leveraging Host Data[J]. *CoRR abs*, 2018, 4(3):157-168.

[28] H. Hindy, D. Brosset, E. Bayne, et al. A Taxonomy and Survey of Intrusion Detection System Design Techniques, Network Threats and Datasets[J]. *CoRR abs*, 2018, 8(5):65-68.

[29] B. Rebecca. An Introduction to Intrusion Detection & Assessment[C]. *International Conference on Software Architecture(ICSA)*, 1998: 89-96.

[30] Rahul-Vigneswaran K, Poornachandran P, Soman K P. A Compendium on Network and Host Based Intrusion Detection Systems[M]. Lecture Notes in Electrical Engineering. Singapore: Springer Singapore, 2020: 23-30.

[31] Computational Intelligence in Intrusion Detection System. Heba Fathy Ahmed Mohamed Eid, http://scholar.cu.edu.eg/sites/default/files/abo/files/phd_thesis_computational_intelligence_in_intrusion_detection_system_2013.pdf, Feb. 2016.

[32] Mallissery S, Prabhu J, Ganiga R. Survey on Intrusion Detection Methods[C]. *3rd International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom 2011)*, 2011: 224-228.

[33] G. Creech ad J. Hu. Generation of a new IDS test dataset: Time to retire the KDD collection[C]. *Wireless Communications and Networking Conference (WCNC)*, 2013: 4487-4492.

[34] A.J. Oliner, J. Stearley. What Supercomputers Say: A Study of Five System Logs[C]. *Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2007: 575-584.

[35] Hamed T, Dara R, Kremer S C. Network Intrusion Detection System Based on Recursive Feature Addition and Bigram Technique[J]. *Computers & Security*, 2018, 73: 137-155.

[36] Chawla A, Lee B, Fallon S, et al. Host Based Intrusion Detection System with Combined CNN/RNN Model[M]. ECML PKDD 2018 Workshops. Cham: Springer International Publishing, 2019: 149-158.

[37] R. Lippmann, J.W. Haines, D.J. Fried, et al. Evaluating intrusion detection systems: The 1998 DARPA offline intrusion detection evaluation[C]. *IEEE DARPA Inf. Surviv. Conf. Expo.*,2000: 12–26.

[38] Lippmann R, Haines J W, Fried D J, et al. The 1999 DARPA Off-line Intrusion Detection Evaluation[J]. *Computer Networks*, 2000, 34(4): 579-595.

[39] University of California Irvine, KDD repository. S. J. Stolfo, KDD Cup 1999 Data Set, http://kdd.ics.uci.edu, Jun. 2014.

[40] M. Tavallaee, E. Bagheri, W. Lu, et al. A detailed analysis of the KDD CUP 99 data set[C]. *IEEE International Conference on Computational Intelligence for Security & Defense Applications (CISDA)*, 2009: 1-6.

[41] S.K. Sahu, S. Sarangi, S.K. Jena. A detail analysis on intrusion detection datasets[C]. *International Advance Computing Conference(IAdCC)*, 2014: 1348–1353.

[42] M. Ring, S. Wunderlich, D. Grüdl, et al. Flow-based benchmark

data sets for intrusion detection[C]. *European Conference on Cyber Warfare and Security (ECCWS)*, 2017: 361-369.

[43] Sharafaldin I, Habibi Lashkari A, Ghorbani A A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization[C]. *The 4th International Conference on Information Systems Security and Privacy*, 2018: 108-116.

[44] 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.

[45] Song J, Takakura H, Okabe Y, et al. Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation[C]. *The First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, 2011: 29-36.

[46] N. Moustafa, J. Slay. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]. *Military Communications and Information Systems Conference (MilCIS)*, 2015: 1-6.

[47] Agarwal R, Joshi M V. PNrule: A New Framework for Learning Classifier Models in Data Mining (a Case-Study in Network Intrusion Detection)[C]. *The 2001 SIAM International Conference on Data Mining*, 2001: 1-17.

[48] P.A. Porras, R.A. Kemmerer. Penetration state transition analysis: A rule-based intrusion detection approach[C]. *Annual Computer Security Application Conference(ACSAC)*, 1992: 220-229.

[49] T.F. Sheu, N.F. Huang, H.P. Lee. NIS04-6: A Time- and Memory-Efficient String Matching Algorithm for Intrusion Detection Systems[C]. *The Global Telecommunications Conference (Globecom)*, 2006: 1-5.

[50] Z. Pan, H. Lian, G. Hu, et al. An Integrated Model of Intrusion Detection Based on Neural Network and Expert System[C]. *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2005: 671-672.

[51] Bronte R, Shahriar H, Haddad H M. A Signature-Based Intrusion Detection System for Web Applications Based on Genetic Algorithm[C]. *The 9th International Conference on Security of Information and Networks*, 2016: 32-39.

[52] Nikolova E, Jecheva V. Some Similarity Coefficients and Application of Data Mining Techniques to the Anomaly-based IDS[J]. *Telecommunication Systems*, 2012, 50(2): 127-135.

[53] Denning DE, Neumann PG. Requirements and model for IDES – a real-time intrusion detection system[OL]. Computer Science Laboratory SRI International, 1985.

[54] Ye N, Emran S M, Chen Q, et al. Multivariate Statistical Analysis of Audit Trails for Host-based Intrusion Detection[J]. *IEEE Transactions on Computers*, 2002, 51(7): 810-820.

[55] García-Teodoro P, Díaz-Verdejo J, Maciá-Fernández G, et al. Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges[J]. *Computers & Security*, 2009, 28(1/2): 18-28.

[56] Yu N. A Novel Selection Method of Network Intrusion Optimal Route Detection Based on Naive Bayesian[J]. *International Journal of Applied Decision Sciences*, 2018, 11(1): 1.

[57] X.K. Ren, W.B. Jiao, D. Zhou. Intrusion Detection Model of Weighted Navie Bayes Based on Particle Swarm Optimization Algorithm[J]. *Computer Engineering and Applications*, 2016, 52(7): 122-126. (任晓奎, 缴文斌, 周丹. 基于粒子群的加权朴素贝叶斯入侵检测模型[J]. *计算机工程与应用*, 2016, 52(7): 122-126.)

[58] Koc L, Mazzuchi T A, Sarkani S. A Network Intrusion Detection System Based on a Hidden Naïve Bayes Multiclass Classifier[J]. *Expert Systems With Applications*, 2012, 39(18): 13492-13500.

[59] L. Xiao, Y. Chen, and C.K. Chan, “Bayesian Model Averaging of Bayesian Network Classifiers for Intrusion Detection[C],” in *International Computer Software and Applications Conference Workshops (COMPSACW)*, pp. 128-133, 2014.

[60] J. Xu and C.R. Shelton, “Intrusion Detection using Continuous Time Bayesian Networks[J],” *Journal of Artificial Intelligence Research*, vol. 39, no. 4, pp. 745-774, 2010.

[61] Xu J, Shelton C R. Intrusion Detection Using Continuous Time Bayesian Networks[J]. *Journal of Artificial Intelligence Research*, 2010, 39: 745-774.

[62] D. Ariu, R. Tronci, and G. Giacinto, “HMMPayl: An intrusion detection system based on Hidden Markov Models[J],” *Computers & Security*, vol. 30, no. 4, pp. 221-241, 2011.

[63] Ariu D, Tronci R, Giacinto G. HMMPayl: An Intrusion Detection System Based on Hidden Markov Models[J]. *Computers & Security*, 2011, 30(4): 221-241.

[64] T. Hurley, J.E. Perdomo, and A. Perez-Pons, “HMM-Based Intrusion Detection System for Software Defined Networking[C],” in *IEEE International Conference on Machine Learning & Applications(ICMLA)*, 2017.

[65] J. Liang, M. Ma, M. Sadiq, and K. Yeung, “A filter model for intrusion detection system in Vehicle Ad Hoc Networks: A hidden Markov methodology[J],” *Knowledge-Based Systems*, 2019.

[66] Liang J W, Ma M D, Sadiq M, et al. A Filter Model for Intrusion Detection System in Vehicle Ad Hoc Networks: A Hidden Markov Methodology[J]. *Knowledge-Based Systems*, 2019, 163: 611-623.

[67] H. Benaddi, K. Ibrahimi, and A. Benslimane, “Improving the Intrusion Detection System for NSL-KDD Dataset based on PCA-Fuzzy Clustering-KNN[C],” in *International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pp. 1-6, 2018.

[68] B. Senthilnayaki, K. Venkatalakshmi, and A. Kannan, “Intrusion detection system using fuzzy rough set feature selection and modified KNN classifier[J],” *Int. Arab J. Inf. Technol*, vol. 16, no. 4, pp. 746-753, 2019.

[69] Y.Y. Aung and M.M. Min, "Hybrid Intrusion Detection System using K-means and K-Nearest Neighbors Algorithms[C]," in *ACIS International Conference on Computer and Information Science (ICIS)*, pp. 34-38, 2018.

[70] X. Jing, Y. Bi, and H. Deng, "An innovative two-stage fuzzy kNN-DST classifier for unknown intrusion detection[J]," *Int. Arab J. Inf. Technol*, vol. 13, no. 4, pp. 359-366, 2016.

[71] P.I. Radoglou-Grammatikis, and P.G. Sarigiannidis, "An Anomaly-Based Intrusion Detection System for the Smart Grid Based on CART Decision Tree[C]," in *Global Information Infrastructure and Networking Symposium (GIIS)*, pp. 1-5, 2018.

[72] S. Sahu and B.M. Mehtre, "Network intrusion detection system using J48 Decision Tree[C]," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2023-2026, 2015.

[73] F. Jiang, C.P. Chun, H.F. Zeng. Relative Decision Entropy Based Decision Tree Algorithm and Its Application in Intrusion Detection[J]. *Computer Science*, 2012, 39(4): 223-226. (江峰, 王春平, 曾惠芬. 基于相对决策熵的决策树算法及其在入侵检测中的应用[J]. 计算机科学, 2012, 39(4): 223-226.)

[74] A. Ahmim, L.A. Maglaras, M.A. Ferrag, et al. A Novel Hierarchical Intrusion Detection System Based on Decision Tree and Rules-Based Models[C]. *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019: 228-233.

[75] L. Teng, S. Teng, F. Tang, et al. A Collaborative and Adaptive Intrusion Detection Based on SVMs and Decision Trees[C]. *IEEE International Conference on Data Mining Workshop (ICDM)*, 2014: 898-905.

[76] Chen S X, Peng M L, Xiong H L, et al. SVM Intrusion Detection Model Based on Compressed Sampling[J]. *Journal of Electrical and Computer Engineering*, 2016, 2016: 1-6.

[77] R.R. Reddy, Y. Ramadevi, K.V.N. Sunitha. Effective discriminant function for intrusion detection using SVM[C]. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016: 1148-1153.

[78] C.Y. Hou, G.W. Wang, C.J. Wang. Intrusion detection method based on improved genetic algorithm to optimize SVM[J]. *Journal of Ordnance Equipment Engineering*, 2019(6):15-24.
(侯春雨, 王戈文, 王崇峻. 一种改进遗传算法优化SVM的入侵检测方法[J]. *兵器装备工程学报*, 2019(6) :15-24.)

[79] Wang H W, Gu J, Wang S S. An Effective Intrusion Detection Framework Based on SVM with Feature Augmentation[J]. *Knowledge-Based Systems*, 2017, 136: 130-139.

[80] Sahu S K, Katiyar A, Kumari K M, et al. An SVM-Based Ensemble Approach for Intrusion Detection[J]. *International Journal of Information Technology and Web Engineering*, 2019, 14(1): 66-84.

[81] Y. Liu, D. Pi. A Novel Kernel SVM Algorithm with Game Theory for Network Intrusion Detection[J]. *TIIS*, 2017, 11(8): 4043-4060.

[82] Yun W. A Multinomial Logistic Regression Modeling Approach for Anomaly Intrusion Detection[J]. *Computers & Security*, 2005, 24(8): 662-674.

[83] M.H. Kamarudin, C. Maple, T. Watson, et al. Packet Header Intrusion Detection with Binary Logistic Regression Approach in Detecting R2L and U2R Attacks[C]. *International Conference on Cyber Security, Cyber Warfare, and Digital Forensic (CyberSec)*, 2015: 101-106.

[84] Ioannou C, Vassiliou V. An Intrusion Detection System for Constrained WSN and IoT Nodes Based on Binary Logistic Regression[C]. *The 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2018: 259-263.

[85] Shah R, Qian Y T, Kumar D, et al. Network Intrusion Detection through Discriminative Feature Selection by Using Sparse Logistic Regression[J]. *Future Internet*, 2017, 9(4): 81.

[86] Besharati E, Naderan M, Namjoo E. LR-HIDS: Logistic Regression Host-based Intrusion Detection System for Cloud Environments[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2019, 10(9): 3669-3692.

[87] M. Fritz, B. Leibe, B. Caputo, et al. Integrating Representative and Discriminant Models for Object Category Detection[C]. *Tenth IEEE International Conference on Computer Vision IEEE Computer Society(ICCV)*, 2005:25-36.

[88] Al-Yaseen W L, Othman Z A, Nazri M Z A. Multi-level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System[J]. *Expert Systems With Applications*, 2017, 67: 296-303.

[89] Y.Y. Aung, M.M. Min. A collaborative intrusion detection based on K-means and projective adaptive resonance theory[C]. *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2017: 1575-1579.

[90] Y.Y. Aung, M.M. Min. Hybrid Intrusion Detection System Using K-Means and Classification and Regression Trees Algorithms[C]. *International Conference on Software Engineering Research, Management and Applications (SERA)*, 2018: 195-199.

[91] Y. F. Yi, J. Yang. PSO-based K-means Algorithm and Its Application in Network Intrusion Detection System[J]. *Computer Science*, 2011, 38(5): 54-55, 73. (易云飞, 杨舰. 基于 PSO 的 k-means 算法及其在网络入侵检测中的应用[J]. *计算机科学*, 2011, 38(5): 54-55, 73.)

[92] Wang F L, Hezhou University School of Mathematics and Computer Hezhou Guangxi China. Research on Application of Improved K-means Algorithm in Network Intrusion Detection[J]. *Journal of Software*, 2018, 13(3): 192-200.

[93] J.V.A. Sukumar, I Pranav, M.M. Neetish, et al. Network Intrusion

Detection Using Improved Genetic k-means Algorithm[C]. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018: 2441-2446.

[94] Horng S J, Su M Y, Chen Y H, et al. A Novel Intrusion Detection System Based on Hierarchical Clustering and Support Vector Machines[J]. *Expert Systems With Applications*, 2011, 38(1): 306-313.

[95] Song J P, Zhu Z L, Price C. Feature Grouping for Intrusion Detection System Based on Hierarchical Clustering[M]. Advanced Information Systems Engineering. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 270-280.

[96] Butun I, Ra I H, Sankar R. An Intrusion Detection System Based on Multi-Level Clustering for Hierarchical Wireless Sensor Networks[J]. *Sensors*, 2015, 15(11): 28960-28978.

[97] L. Su, Y. Yao, N. Li, et al. Hierarchical Clustering Based Network Traffic Data Reduction for Improving Suspicious Flow Detection[C]. *IEEE International Conference On Trust, Security And Privacy In Computing And Communications (Trust-Com/BigDataSE)*, 2018: 744-753.

[98] M. Bitaab, S. Hashemi. Hybrid Intrusion Detection: Combining Decision Tree and Gaussian Mixture Model[C]. *International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC)*,2017: 8-12.

[99] Blanco R, Malagón P, Briongos S, et al. Anomaly Detection Using Gaussian Mixture Probability Model to Implement Intrusion Detection System[M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 648-659.

[100] Z. Wang, Y. Zhu. Intrusion Detection System Based on Gaussian Mixture Model Using Hadoop Framework[C]. *International Conference on Applied Computing and Information Technology (ACIT/CSII/BCD)*, 2017: 125-130.

[101] de la Hoz E, de la Hoz E, Ortiz A, et al. PCA Filtering and Probabilistic SOM for Network Intrusion Detection[J]. *Neurocomputing*, 2015, 164: 71-81.

[102] Salo F, Nassif A B, Essex A. Dimensionality Reduction with IG-PCA and Ensemble Classifier for Network Intrusion Detection[J]. *Computer Networks*, 2019, 148: 164-175.

[103] M.Y. Qi, M. Liu, Y.M. Fu. Research on PCA-based SVM Network Intrusion Detection[J]. *Information Network Security*, 2015(2): 15-18.
(戚名钰, 刘铭, 傅彦铭. 基于 PCA 的 SVM 网络入侵检测研究[J]. *信息网络安全*, 2015(2):15-18.)

[104] A. Hadri, K. Chougdali, R. Touahni. Intrusion detection system using PCA and Fuzzy PCA techniques[C]. *International Conference on Advanced Communication Systems and Information Security (ACOSIS)*, 2016: 1-7.

[105] LeCun Y, Bengio Y, Hinton G. Deep Learning[J]. *Nature*, 2015, 521(7553): 436-444.

[106] B. Abolhasanzadeh. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features[C]. *Conference on Information and Knowledge Technology (IKT)*, 2015:265-271.

[107] A. Krizhevsky, I. Sutskever, G.E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks[C]. *Annual Conference on Neural Information Processing Systems (NIPS)*, 2012: 1106-1114.

[108] A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is All you Need[C]. *Annual Conference on Neural Information Processing Systems (NIPS)*, 2017: 5998-6008.

[109] Kamath U, Liu J, Whitaker J. Deep Learning for NLP and Speech Recognition[M]. Cham: Springer International Publishing, 2019.

[110] X.D. Guo, X.M. Li, R.X. Jing, et al. Intrusion Detection Based on Improved Sparse Denoising Autoencoder[J]. *Journal of Computer Applications*, 2019, 39(3): 769-773. (郭旭东, 李小敏, 敬如雪, 等. 高玉琢基于改进的稀疏去噪自编码器的入侵检测[J]. 计算机应用, 2019, 39(3): 769-773.)

[111] R. Yao, C. Liu, L. Zhang, et al. Unsupervised Anomaly Detection Using Variational Auto-Encoder based Feature Extraction[C]. *International Conference on Prognostics and Health Management (ICPHM)*, 2019: 1-7.

[112] Osada G, Omote K, Nishide T. Network Intrusion Detection Based on Semi-supervised Variational Auto-Encoder[M]. Computer Security – ESORICS 2017. Cham: Springer International Publishing, 2017: 344-361.

[113] Sun J Y, Wang X Z, Xiong N X, et al. Learning Sparse Representation with Variational Auto-Encoder for Anomaly Detection[J]. *IEEE Access*, 2018, 6: 33353-33361.

[114] J. An, S. Cho. Variational Autoencoder based anomaly Detection using Reconstruction Probability[OL]. 2015.

[115] Y. Cui, Y. Sun, J. Hu, et al. A Convolutional Auto-Encoder Method for Anomaly Detection on System Logs[C]. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018: 3057-3062.

[116] U. Fiore, F. Palmieri, A. Castiglione, et al. Network anomaly detection with the restricted Boltzmann machine[J]. *Neurocomputing*, 2013, 122: 13-23.

[117] Gouveia A, Correia M. A Systematic Approach for the Application of Restricted Boltzmann Machines in Network Intrusion Detection[M]. Advances in Computational Intelligence. Cham: Springer International Publishing, 2017: 432-446.

[118] 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.

[119] Elsaeidy A, Munasinghe K S, Sharma D, et al. Intrusion Detection

in Smart Cities Using Restricted Boltzmann Machines[J]. *Journal of Network and Computer Applications*, 2019, 135: 76-83.

[120] N. Gao, L. Gao, Q. Gao, et al. An Intrusion Detection Model Based on Deep Belief Networks[C]. *Second International Conference on Advanced Cloud and Big Data (CBD)*, 2014: 247-252.

[121] Qu F, Zhang J T, Shao Z T, et al. An Intrusion Detection Model Based on Deep Belief Network[C]. *The 2017 VI International Conference on Network*, 2017: 97-101.

[122] G. Zhao, C. Zhang, L. Zheng. Intrusion Detection Using Deep Belief Network and Probabilistic Neural Network[C]. *IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 2017: 639-642.

[123] S.H. Adil, S.S.A. Ali, K. Raza, et al. An Improved Intrusion Detection Approach using Synthetic Minority Over-Sampling Technique and Deep Belief Network[C]. *New Trends in Software Methodologies, Tools and Techniques (SoMeT)*, 2014: 94-102.

[124] Zhang Y, Li P S, Wang X H. Intrusion Detection for IoT Based on Improved Genetic Algorithm and Deep Belief Network[J]. *IEEE Access*, 2019, 7: 31711-31722.

[125] Y.Y. Zhang, Z.T.Liu, W.Zhou. An Intrusion Detection Model Based on Deep Belief Networks[J]. *Modern Computer*, 2015(1): 10-14. (张亚军, 刘宗田, 周文. 基于深度信念网络的入侵检测模型[J]. *现代计算机(普及版)*, 2015(1): 10-14.)

[126] Wei P, Li Y F, Zhang Z, et al. An Optimization Method for Intrusion Detection Classification Model Based on Deep Belief Network[J]. *IEEE Access*, 2019, 7: 87593-87605.

[127] Manickam M, Ramaraj N, Chellappan C. A Combined PFCM and Recurrent Neural Network-based Intrusion Detection System for Cloud Environment[J]. *International Journal of Business Intelligence and Data Mining*, 2019, 14(4): 504.

[128] B.H. Yan, G.D. Han. Combinatorial Intrusion Detection Model Based on Deep Recurrent Neural Network and Improved SMOTE Algorithm[J]. *Chinese Journal of Network and Information Security*, 2018, 4(7)): 48-59. (燕昺昊, 韩国栋基于深度循环神经网络和改进 SMOTE 算法的组合式入侵检测模型[J]. *网络与信息安全学报*, 2018, 4(7)): 48-59.)

[129] J. Kim, J. Kim, H.L.T. Thu, et al. Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection[C]. *International Conference on Platform Technology and Service (PlatCon)*, 2016:268-274.

[130] S.A. Althubiti, E.M. Jones, K. Roy. LSTM for Anomaly-Based Network Intrusion Detection[C]. *International Telecommunication Networks and Applications Conference (ITNAC)*, 2018: 1-3.

[131] B. Roy, H. Cheung. A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network[C]. *International Telecommunication Networks and Applications Conference(ITNAC)*, 2018: 1-6.

[132] A.H. Mirza, S. Cosan. Computer network intrusion detection using sequential LSTM Neural Networks autoencoders[C]. *Signal Processing and Communications Applications Conference (SIU)*, 2018: 1-4.

[133] Khan M A, Karim M R, Kim Y. A Scalable and Hybrid Intrusion Detection System Based on the Convolutional-LSTM Network[J]. *Symmetry*, 2019, 11(4): 583.

[134] Yan B H, Han G D. LA-GRU: Building Combined Intrusion Detection Model Based on Imbalanced Learning and Gated Recurrent Unit Neural Network[J]. *Security and Communication Networks*, 2018, 2018: 1-13.

[135] Agarap A F M. A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data[C]. *The 2018 10th International Conference on Machine Learning and Computing*, 2018: 26-30.

[136] 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.

[137] Zhang S C, Xie X Y, Xu Y. Intrusion Detection Method Based on a Deep Convolutional Neural Network[J]. *Journal of Tsinghua University (Science and Technology)*, 2019, 59(1): 44-52. (张思聪, 谢晓尧, 徐洋. 基于 dCNN 的入侵检测方法[J]. *清华大学学报(自然科学版)*, 2019, 59(1): 44-52.)

[138] S. Naseer, Y. Saleem. Enhanced Network Intrusion Detection Using Deep Convolutional Neural Networks[J]. *KSII Transactions on Internet and Information Systems*, 2018, 12(10):5159-5178.

[139] Wu K H, Chen Z G, Li W. A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks[J]. *IEEE Access*, 2018, 6: 50850-50859.

[140] S.Z. Lin, Y. Shi, Z. Xue. Character-Level Intrusion Detection Based On Convolutional Neural Networks[C]. *International Joint Conference on Neural Networks (IJCNN)*, 2018: 1-8.

[141] 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: 1-9.

[142] J. Liu, L. Yin, Y. Hu, et al. A Novel Intrusion Detection Algorithm for Industrial Control Systems Based on CNN and Process State Transition[C]. *IEEE International Performance Computing and Communications Conference, IPCCC (IPCCC)*, 2018: 1-8.

[143] M. Salem, S. Taheri, J. Yuan. Anomaly Generation Using Generative Adversarial Networks in Host-Based Intrusion Detection[C]. *IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2018: 683-687.

[144] Z. Lin, Y. Shi, Z. Xue. IDSGAN: Generative Adversarial Networks for Attack Generation against Intrusion Detection[J]. *CoRR abs*, 2018, 2(1):12-18.

[145] M. Usama, M. Asim, S. Latif, et al. Generative Adversarial Networks For Launching and Thwarting Adversarial Attacks on Network Intrusion Detection Systems[C]. *International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019: 78-83.

[146] H. Chen, L. Jiang. GAN-based method for cyber-intrusion detection[J]. *CoRR abs*, 2019, 4(2):56-68.

[147] A. Ferdowsi, W. Saad. Generative Adversarial Networks for Distributed Intrusion Detection in the Internet of Things[J]. *CoRR abs*, 2019, 2(3): 58-64.

[148] E. Seo, H.M. Song, H.K. Kim. GIDS: GAN based Intrusion Detection System for In-Vehicle Network[J]. *CoRR abs*, 2019, 5(3):25-36.

[149] Kober J, Bagnell J A, Peters J. Reinforcement Learning in Robotics: A Survey[J]. *The International Journal of Robotics Research*, 2013, 32(11): 1238-1274.

[150] B.Q.Huang, G.Y.Cao, Y.Q.Fei, et al. Study on an Average Reward Reinforcement Learning Algorithm[J]. *Chinese Journal of Computers*, 2007, 30(8): 1372-1378. (黄炳强, 曹广益, 费燕琼, 等. 平均奖赏强化学习算法研究[J]. *计算机学报*, 2007, 30(8): 1372-1378.)

[151] Tesauro G. TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play[J]. *Neural Computation*, 1994, 6(2): 215-219.

[152] Xu X, Xie T. A Reinforcement Learning Approach for Host-Based Intrusion Detection Using Sequences of System Calls[M]. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 995-1003.

[153] Di C, Su Y, Han Z R, et al. Learning Automata Based SVM for Intrusion Detection[M]. Lecture Notes in Electrical Engineering. Singapore: Springer Singapore, 2018: 2067-2074.

[154] Servin A, Kudenko D. Multi-agent Reinforcement Learning for Intrusion Detection[M]. Adaptive Agents and Multi-Agent Systems III. Adaptation and Multi-Agent Learning. Berlin, Heidelberg: Springer Berlin Heidelberg, : 211-223.

[155] Servin A, Kudenko D. Multi-Agent Reinforcement Learning for Intrusion Detection: A Case Study and Evaluation[M]. Multiagent System Technologies. Berlin, Heidelberg: Springer Berlin Heidelberg, : 159-170.

[156] K. Malialis. Distributed reinforcement learning for network intrusion response[M]. University of York, 2014.

[157] Q. Liu, J.W. Zhai, Z.Z. Zhang, et al. A review of deep reinforcement learning[J]. *Chinese Journal of Computers*, 2018(1): 1-27.(刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. *计算机学报*, 2018(1): 1-27.)

[158] Caminero G, Lopez-Martin M, Carro B. Adversarial Environment Reinforcement Learning Algorithm for Intrusion Detection[J]. *Computer Networks*, 2019, 159: 96-109.

[159] B. Yuan, D.Q. Zhou, H. Jin, et al. Network Security Visualization：A Survey[J]. *Journal of Cyber Security*, 2016, 1(3): 10-20. (袁斌, 邹德清, 金海, 等. 网络安全可视化综述[J]. *信息安全学报*, 2016, 1(3): 10-20.)

[160] Liu S X, Cui W W, Wu Y C, et al. A Survey on Information Visualization: Recent Advances and Challenges[J]. *The Visual Computer*, 2014, 30(12): 1373-1393.

[161] Shiravi H, Shiravi A, Ghorbani A A. A Survey of Visualization Systems for Network Security[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(8): 1313-1329.

[162] Koike H, Ohno K. SnortView: Visualization System of Snort Logs[C]. Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security - VizSEC/DMSEC '04, 2004: 143-147.

[163] K. Abdullah, C.P. Lee, G.J. Conti, et al. IDS RainStorm: Visualizing IDS Alarms[C]. *IEEE Symposium on Information Visualization's Workshop on Visualization for Computer Security (VizSEC)*, 2005: 25-36.

[164] Luo B, Xia J B. A Novel Intrusion Detection System Based on Feature Generation with Visualization Strategy[J]. *Expert Systems With Applications*, 2014, 41(9): 4139-4147.

[165] W. Zong, Y. Chow, W. Susilo. A 3D Approach for the Visualization of Network Intrusion Detection Data[C]. *International Conference on Cyberworlds (CW)*, 2018: 308-315.

[166] Zong W, Chow Y W, Susilo W. Dimensionality Reduction and Visualization of Network Intrusion Detection Data[M]. Information Security and Privacy. Cham: Springer International Publishing, 2019: 441-455.

[167] R. Therón, R. Magán-Carrión, J. Camacho, et al. Network-wide intrusion detection supported by multivariate analysis and interactive visualization[C] *IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2017: 1-8.

[168] Song B, Choi S S, Choi J, et al. Visualization of Intrusion Detection Alarms Collected from Multiple Networks[J]. *Information Security*, 2017: 437-454.

[169] Shi Y, Zhao Y, Zhou F F, et al. A Novel Radial Visualization of Intrusion Detection Alerts[J]. *IEEE Computer Graphics and Applications*, 2018, 38(6): 83-95.

[170] Y. Zhao, X.P. Fan, F.F. Zhou, et al. Overview of network security data visualization[J]. *Journal of computer aided design and graphics*, 2014,26(5): 687-697.(赵颖, 樊晓平, 周芳芳, 等. 网络安全数据可视化综述[J]. *计算机辅助设计与图形学学报*, 2014, 26(5): 687-697.)

蹇诗婕 于 2018 年在北京科技大学信息安全专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读硕士学位。研究领域为网络安全态势感知、入侵检测等。Email: jianshijie@iie.ac.cn

卢志刚 于 2010 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所高级工程师, 中国科学院网络空间安全学院副教授。研究领域为网络安全态势感知、网络攻击检测、移动终端安全等。Email: luzhigang@iie.ac.cn

杜丹 于 2016 年在中国科学院大学计算机技术专业获得硕士学位。现任中国科学院信息工程研究所工程师。研究领域为网络入侵检测、移动安全、电子取证等。Email: dudan@iie.ac.cn

姜波 于 2016 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为网络安全态势感知、知识图谱、数据挖掘等。Email: jiangbo@iie.ac.cn

刘宝旭 于 2002 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所研究员, 第六研究室主任。研究领域为网络安全攻防对抗、网络安全测评技术等。Email: liubaoxu@iie.ac.cn