

电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

# 专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE



论文题目    基于图神经网络的恶意流量检测技术  
研究

学科专业                      网络空间安全

学      号                      5806

作者姓名                      李海阳

指导老师                      张小松 教授

学      院                      计算机科学与工程学院

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

UDC 注 1 \_\_\_\_\_

# 学 位 论 文

基于图神经网络的恶意流量检测技术研究

(题名和副题名)

李海阳

(作者姓名)

指导老师

张小松 教授

电子科技大学 成都

(姓名、职称、单位名称)

申请学位级别 **专业硕士** 学科专业 **网络空间安全**

提交论文日期 \_\_\_\_\_ 论文答辩日期 \_\_\_\_\_

学位授予单位和日期 **电子科技大学** 年 月

答辩委员会主席 \_\_\_\_\_

评阅人 \_\_\_\_\_

注 1: 注明《国际十进分类法 UDC》的类号。

# **The Time Marching Scheme of Time Domain Integral Equation and Corresponding Fast Algorithm**

A Master Thesis for Professional Degree Submitted to  
University of Electronic Science and Technology of China

Discipline:	<b>Computer Science and Engineering</b>
Author:	<b>Li Haiyang</b>
Student ID:	<b>5806</b>
Supervisor:	<b>Dr. Xiaosong Zhang</b>
School:	<b>School of Computer Science and Engineering</b>

## 摘 要

关键词：



## **ABSTRACT**



## 目 录

第一章 绪论.....	1
1.1 研究工作的背景与意义 .....	1
1.2 国内外研究现状.....	2
1.2.1 数据不平衡处理 .....	3
1.2.2 恶意流量检测 .....	4
1.3 本文的主要贡献与创新 .....	10
1.4 本论文的结构安排 .....	11
第二章 相关理论及技术介绍 .....	12
2.1 恶意网络流量 .....	12
2.2 去噪扩散模型 .....	13
2.3 半监督学习 .....	15
2.4 图神经网络 (GNN).....	16
2.5 图卷积网络 (GCN) .....	17
2.6 图注意力机制 .....	18
2.7 本章小结.....	19
第三章 基于 LDNFDDPM 的数据增强方法 .....	20
3.1 问题描述和研究思路 .....	20
3.2 框架设计.....	23
3.2.1 数据预处理 .....	23
3.2.2 时序图构建 .....	27
3.2.3 基于 LDNFDDPM 恶意流量生成.....	31
3.3 本章小结.....	35
第四章 基于多头自适应图注意力网络的线图训练方法 .....	36
4.1 问题描述和研究思路 .....	36
4.2 框架设计.....	38
4.2.1 基于 LineGraph 的特征结合方法 .....	38
4.2.2 基于多头自适应注意力机制的图卷积方法 .....	40
4.3 本章小结.....	42
第五章 实验结果分析 .....	43
5.1 实验数据及环境.....	43



5.1.1 实验环境.....	43
5.1.2 实验数据.....	43
5.2 实验设计.....	45
5.2.1 评估指标.....	45
5.2.2 相关对比方法.....	46
5.3 数据增强方法测试.....	48
5.4 恶意流量检测方法测试.....	51
5.5 本章小结.....	56
<b>第六章 全文总结与展望.....</b>	<b>59</b>
6.1 全文总结.....	59
6.2 后续工作展望.....	59
<b>致 谢.....</b>	<b>60</b>
<b>参考文献.....</b>	<b>61</b>

## 第一章 绪论

### 1.1 研究工作的背景与意义

20 世纪以来,根据《第 54 次中国互联网络发展状况统计报告》的相关数据和分析,我国互联网发展与网络安全形势呈现出复杂且多层次的态势<sup>[1]</sup>。

首先,互联网的普及程度持续提高,截至 2024 年 6 月,我国网民总数已接近 11 亿人,互联网普及率达到 78.0%,相较 2023 年底增长了 0.5 个百分点。尤其是移动互联网的普及,手机网民规模已达 10.96 亿,占整体网民的 99.7%,几乎所有网民都依赖移动设备上网。此外,互联网应用逐渐多元化,即时通信、网络支付、网络购物、网络视频和直播等应用的用户规模持续增长,尤其是即时通信和网络支付,分别占到网民总数的 98.0% 和 88.1%。在数字消费方面,智能穿戴设备、人工智能产品以及数字服务如在线文旅、在线餐饮等都呈现出强劲增长势头,推动了我国数字消费市场的繁荣。同时,基础设施建设不断加强,IPv6 地址、域名数量、移动基站等互联网资源的供给持续增长,为我国数字经济的高质量发展提供了稳固的支撑。

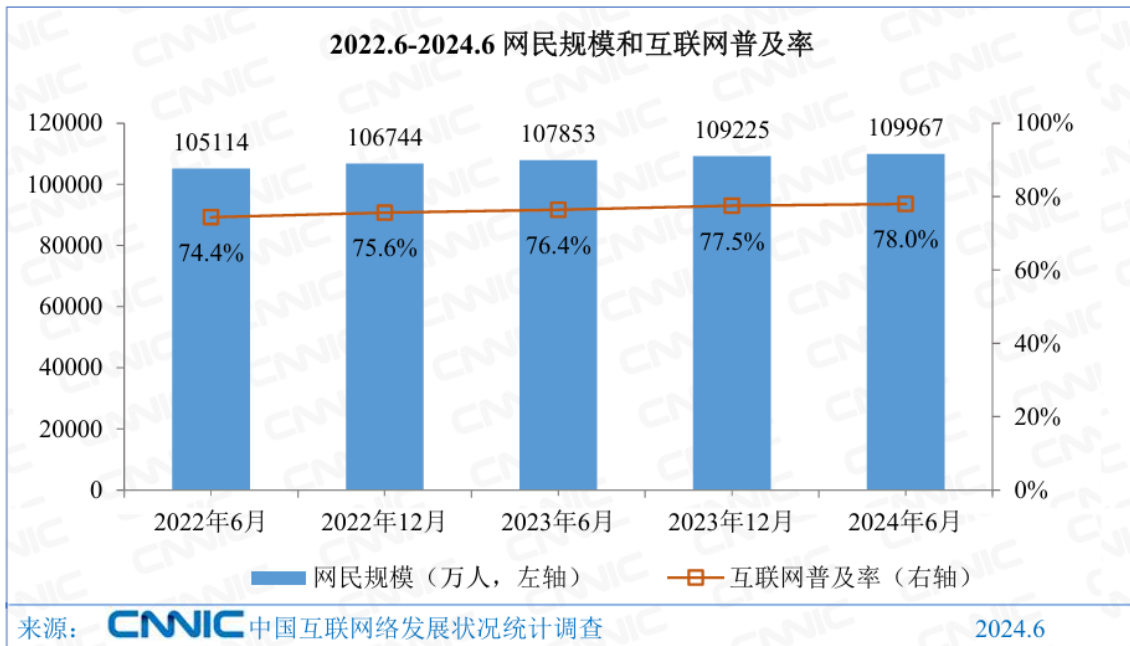


图 1-1 2022.6-2024.6 网民规模和互联网普及率

然而,随着互联网的蓬勃发展,网络安全形势日益严峻,网络攻击手段愈加复杂和多样,《中国网络安全产业分析报告(2024 年)》中指出,2023 年以来,网络攻击活动呈现组织化程度高、攻击目标明确、攻击数量增多、隐蔽性增强、攻击效

率提高等趋势，高级持续性威胁（APT）攻击、勒索软件攻击、数据窃取、零日攻击等网络攻击手段近年来已演化为结合社会工程学攻击、应用人工智能等前沿技术的综合体，网络空间安全威胁越来越严峻。据 Statista 网站统计显示，全球 72% 的企业成为勒索攻击受害者。2024 年上半年发生多起严重的网络攻击事件，例如，美国 IvantiVPN 设备零日漏洞受到攻击、美国医疗处方公司 ChangeHealthcare 遭受网络攻击、XZUtils 软件供应链攻击、美国国家环境保护局数据泄露 5 及 7·19 微软蓝屏事件造成全球近千万台设备宕机等，极大地暴露了网络安全防护的薄弱环节。

此外，随着互联网的快速发展，网络流量急剧增加，网络攻击的规模和手段也不断演变。恶意流量与正常流量存在显著差异，通常包含具有破坏性的行为，对网络安全构成严重威胁。因此，及时、准确地检测和防范恶意网络流量成为保障互联网健康发展的关键。网络恶意流量检测技术在入侵检测领域中发挥着至关重要的作用，它通过分析关键网络节点收集的信息，实时识别潜在的威胁，从而确保网络的安全性、可用性以及数据的保密性与完整性。

在当前的恶意流量检测研究中，快速准确地识别恶意流量仍然是一个亟待解决的问题。尽管机器学习和人工智能技术在此领域已取得一定进展，仍面临诸如流量数据处理、大规模数据适配、新型攻击检测以及网络环境适应性等挑战。为了克服这些难题，本研究结合了 DDPM 和图神经网络（GNN）的方法，提出了一种创新的恶意流量检测方案。DDPM（去噪扩散概率模型）在数据增强中的应用，可以有效地扩充训练数据集，提升模型的泛化能力，而图神经网络则通过其强大的图表示能力，能够在网络结构中有效地处理流量特征，适应不断变化的网络环境。

基于图神经网络的恶意流量检测方法能够结合网络拓扑结构与流量特征，提升对复杂网络环境下恶意流量的检测能力。通过对网络流量的深度分析，本方法能够实时识别潜在的攻击行为，从而有效提高网络的安全性。因此，结合 DDPM 和图神经网络的方法，在恶意流量检测中展现出巨大的应用潜力，成为网络安全领域的重要研究方向。

## 1.2 国内外研究现状

网络恶意流量检测本质上是对流量进行分类，将其分为正常流量和恶意流量，而在现实情况下，流量的占比往往存在着很多不同，尤其在分类任务中，这样就会引起数据不平衡问题，导致分类效果不佳。因此，本节将围绕数据不平衡处理和恶意流量检测两个方面介绍国内外研究现状。

### 1.2.1 数据不平衡处理

数据不平衡问题是机器学习中的一个重要课题，特别是在某一类别样本数量远少于其他类别时，传统的机器学习模型往往会偏向于多数类，导致对少数类的识别能力大幅下降。为了解决这一问题，研究者们提出了多种策略，包括数据采样、数据合成和加权等方法。

数据采样是一种直接应对数据不平衡问题的方法，主要包括欠采样和过采样两种方式。欠采样通过减少多数类样本的数量来平衡类别分布，这种方法简单易行，但可能会丢失多数类样本中的重要信息。过采样则通过增加少数类样本的数量来平衡数据集，尽管能够保留所有原始样本的信息，但也可能导致模型过拟合。为了克服这些缺点，研究者们提出了合成采样技术，如 SMOTE (Synthetic Minority Over-sampling Technique)。Nitesh V. Chawla 等人<sup>[2]</sup>提出的 SMOTE 技术通过在特征空间中生成合成的少数类样本，改善了少数类的分类性能，并在多个数据集上验证了其有效性。SMOTE 通过在少数类样本的邻域内生成新的合成样本，有效地增加了少数类的样本数量，并避免了简单复制样本所带来的过拟合问题。随后，Paria Soltanzadeh 和 Mahdi Hashemzadeh<sup>[3]</sup>提出了一种改进的范围控制合成少数过采样技术 (RCSMOTE)，通过样本分类和改进的生成过程，有效避免了过采样中噪声样本的引入，同时减少了类别重叠。

生成对抗网络 (GAN) 是一种强大的数据合成技术，能够生成逼真的样本数据。GAN 由生成器和判别器组成，生成器负责生成样本，判别器则负责区分生成样本和真实样本。近年来，基于 GAN 的技术在处理不平衡数据集时取得了显著的成果，尤其是在生成少数类样本方面。Ian J. Goodfellow 等人<sup>[4]</sup>提出的生成对抗网络 (GAN)，通过对抗过程同时训练生成模型和判别模型，以生成高质量的样本数据，尤其在图像生成领域显示出了巨大潜力。Ilyass Haloui 等人<sup>[1]</sup>提出了基于 Wasserstein 生成对抗网络 (WGAN) 的异常检测方法，用于时间序列数据集中的异常检测。该方法通过训练 WGAN 学习正常数据的分布表示，并结合编码器进行异常检测，从而提高了检测性能。Qiufeng Wu 等人<sup>[5]</sup>则提出了一种基于深度卷积生成对抗网络 (DCGAN) 的数据增强方法，用于番茄叶片疾病识别。该方法通过生成逼真的叶片图像，扩充了数据集，提高了识别模型的泛化能力和准确性。Sebastian Walter 等人<sup>[6]</sup>提出了渐进式生成对抗网络 (Progressive GAN) 用于 MIDI 音乐生成，通过逐步增加网络复杂度和数据分辨率，提升了音乐生成的质量和稳定性。Wei Yu 等人<sup>[7]</sup>则提出了一种改进的超分辨率生成对抗网络 (SRDAGAN) 应用于刨花板图像超分辨率重建，优化了网络结构和损失函数，从而有效提高了重建图像的质量和细节表现。

去噪扩散概率模型 (DDPMs) 是另一种在生成高质量样本方面表现出色的数据合成技术。DDPM 通过逐步去噪的过程生成样本, 其生成效果在图像生成和音频生成等领域中常常超越 GAN。DDPM 的训练过程相对稳定, 不容易出现模式崩溃等问题, 且可用于数据增强, 提高模型的泛化能力。最早由 Jonathan Ho 等人<sup>[8]</sup>提出, DDPM 通过训练一个参数化的马尔可夫链, 逆转逐步添加噪声的过程, 从而生成符合数据分布的样本。Xin Ding 等人<sup>[9]</sup>提出了连续条件扩散模型 (CCDM), 在给定标量连续变量 (回归标签) 的条件下生成高维数据 (如图像)。该模型通过改进的去噪 U-Net 架构和高效的条件采样程序, 克服了现有条件扩散模型的局限性, 显著提高了生成图像的质量。Hongxu Jiang 等人<sup>[10]</sup>提出了快速去噪扩散概率模型 (Fast-DDPM), 旨在加速训练和采样过程。通过仅使用 10 个时间步, Fast-DDPM 显著减少了训练和采样时间, 在医学图像生成任务中表现出优于传统方法的性能。Sang-gil Lee 等人<sup>[11]</sup>则提出了 PriorGrad 方法, 用于改进基于条件的去噪扩散模型, 尤其是在语音合成领域。PriorGrad 通过应用自适应先验, 提高了模型的去噪效率, 并在语音生成任务中显著提升了模型的感知质量和鲁棒性。

尽管在数据不平衡问题上取得了一定的进展, 但仍存在一些挑战。例如, SMOTE 及其变体对噪声和异常值较为敏感, 可能会生成具有异常特征的样本。基于 GAN 的方法则受到模型结构和超参数选择的影响较大, 需要大量的实验和调优。DDPM 虽然在生成质量上表现出色, 但其训练和采样过程较为复杂和耗时。

### 1.2.2 恶意流量检测

随着网络技术的不断发展, 网络安全问题日益严峻。恶意流量攻击 (如 DDoS 攻击、入侵行为、恶意扫描等) 已成为影响网络安全的主要威胁之一。为了有效检测和防御这些攻击, 恶意流量检测技术得到了广泛的研究与应用。传统的恶意流量检测方法主要依赖于规则匹配和统计特征分析, 但这些方法往往难以应对未知攻击或复杂的攻击行为。近年来, 随着机器学习、深度学习和图神经网络等技术的不断发展, 恶意流量检测技术逐渐向更加智能化、自动化的方向发展。

#### (1) 基于规则的恶意流量检测方法

基于规则的恶意流量检测方法依赖于人工构建的规则库, 将恶意流量的特征定义为签名, 匹配网络流量与规则库中的签名。当网络流量与规则匹配时, 系统判定其为恶意流量。Vikash Kumar 等人<sup>[12]</sup>提出一种基于规则的集成入侵检测系统 (IDS), 该系统在 UNSW-NB15 数据集和实时在线数据集 (RTNITP18) 上进行分析, 能够检测网络中的五类攻击: Exploit、DOS、Probe、Generic 和 Normal。该系统通过分析 UNSW-NB15 数据集, 设计了一个集成的分类模型, 并在实时数据集

上进行了性能评估。结果表明，该模型相较于其他决策树模型具有更高的检测性能，包括更高的准确率、攻击检测率和较低的误报率。此外，该系统还能够检测包括正常流量在内的五类攻击，具有较低的误报率和较高的检测率，为网络入侵检测提供了一种有效的解决方案。

Chaobo Zhang 等人<sup>[13]</sup>提出一种基于关联规则库的实时建筑能源系统异常操作模式检测方法。该方法结合了专家系统和关联规则挖掘的优势，首先从历史操作数据中提取关联规则，建立异常和正常操作模式的规则库。然后，利用这些规则库对实时操作数据进行检测，以识别异常操作模式。该方法在实际的制冷机房中进行了应用和评估，结果表明能够成功检测出已知和未知的异常操作模式，提高了建筑能源系统的运行效率和节能效果。Yuchen Yang 等人<sup>[14]</sup>提出一种名为 AnomalyRuler 的基于规则的推理框架，用于视频异常检测。该框架利用大型语言模型（LLMs），通过归纳和演绎两个阶段，从少量正常样本中总结规则，检测测试视频中的异常帧。在归纳阶段，LLM 从少量正常参考样本中总结出正常模式的规则；在演绎阶段，根据这些规则识别测试视频中的异常帧。此外，该框架还设计了规则聚合、感知平滑和鲁棒推理策略以增强检测的鲁棒性。实验结果表明，AnomalyRuler 在多个视频异常检测基准数据集上取得了优异的性能，展示了其在推理能力和领域适应性方面的优势。

Pengbin Feng 等人<sup>[15]</sup>提出一种名为 SCDFLOW 的方法，用于在 Android 应用中选择关键数据流以检测异常行为。该方法通过分析恶意应用与良性应用在敏感数据流上的差异，选择关键数据流作为特征。SCDFLOW 首先使用静态污点分析工具 FLOWDROID 提取应用中的敏感数据流，然后通过特征选择算法 CFlowSel 选择关键数据流。实验结果表明，SCDFLOW 能够有效减少数据流特征的维度，并在 MUDFLOW 和 DREBIN 数据集上提高了恶意软件检测率，同时对内存消耗的影响可以忽略不计。该方法为基于异常数据流的恶意软件检测提供了一种有效的解决方案。

这种方法在已知攻击的情况下能够快速检测并响应，处理速度较快。规则库的构建和维护相对简单，且技术门槛较低，适用于简单的恶意流量模式检测。然而，这种方法主要依赖于已知攻击的特征，对于未知或变种攻击的检测能力有限。攻击者可以通过绕过规则库中的签名或修改攻击方式来规避检测。此外，随着攻击手段的不断变化，规则库需要持续更新，这对于维护团队来说是一个巨大的挑战。

## （2）基于统计特征分析的恶意流量检测方法

基于统计特征的检测方法通过对网络流量的统计特性进行分析（如流量大小、

协议字段分布等), 利用统计学理论(如信息熵、概率论)区分正常流量和异常流量。Liu 等人提出的波形统计方法, 通过将数据包交换过程中的帧长信息转换为波形特征, 从中提取关键的统计特性进行分类 [13]。这种方法实时性较好, 且计算复杂度较低, 但由于攻击者可以模仿正常流量的统计特性, 因此容易发生误报或漏报, 尤其是在面临新型或变种攻击时。

Ankit Thakkar 和 Ritika Lohiya<sup>[16]</sup> 提出一种基于统计重要性融合的特征选择方法, 用于深度神经网络 (DNN) 基础的入侵检测系统 (IDS)。该方法通过计算特征的标准差、均值与中位数的差异来评估特征的重要性, 并根据这些统计量的融合结果对特征进行排序和选择, 旨在提取出具有高可辨识度和偏差的相关特征, 从而提高 DNN-IDS 在 NSL-KDD、UNSW\_NB-15 和 CIC-IDS-2017 等数据集上的性能, 包括准确率、精确率、召回率、F1 分数和误报率等指标。

Jingyu Liu 等人<sup>[17]</sup> 提出一种基于粒子群优化 (PSO) 的入侵检测模型, 用于物联网 (IoT) 环境中的安全防护。该模型首先利用 PSO-LightGBM 算法对数据进行特征提取和降维, 通过 PSO 优化 LightGBM 的参数来实现数据的双维度约简和特征提取, 以解决大规模数据集的不平衡分布问题。然后将处理后的特征数据输入单类支持向量机 (OCSVM) 进行建模, 以检测和识别正常数据以及各种异常数据, 包括低频攻击数据如后门、Shellcode 和蠕虫等。实验结果表明, 该模型在准确性、检测率、误报率和时间成本等方面均优于其他入侵检测技术, 具有在实时 IoT 环境中部署的潜力。

Hao Zhang 等人<sup>[18]</sup> 提出一种多维特征融合和堆叠集成机制 (MFFSEM), 用于网络入侵检测。该方法首先根据网络流量数据的特性, 如时间、空间和负载等维度, 提取多个基本特征数据集。然后, 考虑这些基本特征数据集之间的关联和相关性, 通过排列组合的方式构建多个综合特征数据集, 以满足现实世界异常行为检测的需求。接着, 在多个综合特征数据集上应用堆叠集成学习策略, 训练多个基本分类器, 并将它们的预测概率作为元分类器的输入, 从而实现多维全局异常检测模型。实验结果表明, MFFSEM 在 KDD Cup 99、NSL-KDD、UNSW-NB15 和 CIC-IDS2017 等数据集上的检测性能显著优于基本分类器和元分类器, 以及其他知名的集成方法, 具有较高的准确率、召回率、精确率和 F1 分数等指标。

基于统计特征方法计算开销相对较小, 能够快速处理大规模网络流量, 适合实时监控, 并能够处理大部分常见的网络攻击, 尤其是在网络行为相对稳定的情况下。然而, 攻击者可以通过伪造正常流量的统计特征来绕过检测。对于高级攻击, 尤其是那些隐蔽性强或变种频繁的攻击, 基于统计特征的方法往往无法有效捕捉到其中的细节。同时, 网络波动或正常流量出现异常时, 可能会导致误报, 从

而降低检测系统的鲁棒性。

### （3）基于机器学习的恶意流量检测方法

基于机器学习的检测方法能够自动从网络流量数据中提取特征，利用机器学习模型（如 SVM、决策树、随机森林等）进行分类。Salman 等人<sup>[19]</sup>提出一种基于机器学习的框架，用于物联网（IoT）设备识别和异常流量检测。该框架旨在应对 IoT 设备的异构性、资源限制和大规模部署带来的安全挑战。框架在网络边缘提取网络流的特征，包括数据包方向、大小、时间戳和传输协议等，以识别设备类型、流量类型，并检测网络攻击。通过比较不同的机器学习算法，发现随机森林在设备类型识别、流量类型分类和异常流量检测方面表现最佳，分别达到了 94.5%、93.5% 和 97% 的准确率。该框架具有实时性和低开销的特点，能够在不进行数据包内容检查的情况下，仅通过统计特征实现对 IoT 流量的准确分类和异常检测。Waskle 等人<sup>[20]</sup>提出一种结合主成分分析（PCA）和随机森林算法的入侵检测系统（IDS）。该系统旨在提高入侵检测的准确性和效率。首先，通过 PCA 对数据集进行降维处理，以减少数据的维度并提取主要特征，从而提高数据集的质量和效率。PCA 通过计算数据集的协方差矩阵和特征向量，将高维数据转换为低维的主成分，保留了数据的主要信息。接着，利用随机森林算法进行分类，随机森林是一种集成学习方法，通过构建多个决策树并综合它们的预测结果来提高分类的准确性和鲁棒性。在实验中，该方法在 KDD 数据集上进行了测试，结果显示其性能优于传统的支持向量机（SVM）、朴素贝叶斯和决策树等方法。具体而言，该方法的性能时间为 3.24 分钟，准确率达到 96.78%，错误率仅为 0.21%。这表明该系统在检测网络入侵方面具有较高的准确性和较低的误报率，能够有效地识别和防御各种网络攻击。

相比于传统方法，机器学习能够自动从数据中学习到有用的特征，避免了人工特征选择的繁琐，具有较强的适应性，能够对未知的攻击类型进行有效学习。此外，机器学习方法可以结合多种算法进行集成学习，进一步提升分类效果。然而，在流量数据存在较大不平衡时，模型可能偏向于多数类，导致少数类的识别能力下降。机器学习模型通常需要大量的标注数据进行训练，对于恶意流量标注的获取和质量控制是一个挑战。与传统方法相比，机器学习模型的训练过程较为耗时，且对计算资源的需求较高。

### （4）基于深度学习的恶意流量检测方法

深度学习方法能够自动从原始数据中提取复杂的特征表示，适用于处理时序性强的流量数据，尤其是在面对动态网络流量时具有明显优势。Fu 等人<sup>[21]</sup>提出一种基于深度学习的网络入侵检测模型（DLNID），该模型旨在解决网络流



量数据不平衡和检测准确率低的问题。模型结合了注意力机制和双向长短期记忆 (Bi-LSTM) 网络, 首先通过卷积神经网络 (CNN) 提取数据流量的序列特征, 然后利用注意力机制重新分配每个通道的权重, 以突出重要特征, 最后使用 Bi-LSTM 学习序列特征之间的关系, 从而提高入侵检测的准确率和 F1 分数。在 NSL-KDD 数据集上的实验结果表明, 该模型的准确率和 F1 分数分别达到了 90.73% 和 89.65%, 优于其他比较方法。

Jay Sinha 和 Manollas M<sup>[22]</sup> 提出一种高效的深度 CNN-BiLSTM 模型, 用于网络入侵检测。该模型充分利用了卷积神经网络 (CNN) 在学习数据空间特征方面的优势, 以及双向长短期记忆 (Bi-LSTM) 在捕捉时间序列数据长期时间特征方面的能力。模型通过结合这两种网络结构, 能够更准确地识别网络流量中的异常行为, 提高检测率和降低误报率。实验在 NSL-KDD 和 UNSW-NB15 数据集上进行, 结果表明该模型在二元分类和多类分类任务中均表现出色, 具有较高的检测率和较低的误报率, 优于许多现有的网络入侵检测系统。

Ge 等人<sup>[23]</sup> 提出一种基于深度学习的入侵检测方法, 专门针对物联网 (IoT) 环境中的网络安全问题。该方法采用定制的深度学习技术, 利用包含 IoT 流量和真实攻击流量的先进 IoT 数据集, 构建了一个前馈神经网络模型, 并引入嵌入层来处理高维分类特征, 实现对拒绝服务 (DoS)、分布式拒绝服务 (DDoS)、数据收集和数据盗窃等攻击类型的多类分类。此外, 该方法还应用迁移学习概念, 将从多类分类模型中学习到的高维分类特征编码应用于二元分类器的构建, 以进一步提高分类性能。实验结果表明, 该方法在二元和多类分类任务中均取得了较高的分类准确率, 分别为 99.99% 和 99.79%, 显示出良好的入侵检测性能。

深度学习能够自动从原始流量数据中提取关键特征, 减少了人工特征工程的依赖, 能够捕捉流量数据中的复杂模式和深层次关系, 适应新型攻击模式。对于时序性强、长时间依赖关系明显的数据 (如网络流量), 深度学习能够更好地捕捉数据的动态特性。然而, 深度学习模型训练需要大量的计算资源, 且训练过程耗时较长, 这在一些资源受限的环境中可能不可行。此外, 深度学习模型的解释性较差, 缺乏透明度, 难以追踪决策过程, 这可能在实际部署时带来一定的不确定性。深度学习模型也需要大量的标注数据, 对于数据质量和数据集的多样性要求较高。

#### (5) 基于图神经网络 (GNN) 的恶意流量检测方法

图神经网络 (GNN) 作为一种新兴的深度学习技术, 能够处理图结构的数据。通过建模网络流量中的节点 (如 IP 地址、端口等) 及其之间的连接关系, GNN 能够有效捕捉复杂的流量模式, 提升检测的准确性。

刘杰等人<sup>[24]</sup> 提出了一种基于图神经网络的工控网络异常检测算法, 旨在提高

对联合异常攻击和恶意软件的检测能力。该算法首先为每个网络节点获取包含连接节点特征信息及节点间交互信息的状态向量，通过聚合节点自身的特征和邻居节点的特征来生成节点的新表示。然后，利用不动点理论对网络进行迭代更新，以求解每个节点状态向量的唯一解。接着，结合节点自身信息及邻域节点信息，通过神经网络提取更高层次的特征作为节点表示。最后，采用 K-means 聚类方法对节点特征进行聚类，判断网络节点是否是异常节点。实验结果表明，该算法在保持较高检测率的同时，也具有较高的鲁棒性，能够有效应对工控网络中的多点连接性，弥补了以往单点网络异常检测方法的不足。

Wang 等人<sup>[25]</sup>提出了一种基于深度强化学习（DRL）的种子选择方法，用于解决社交网络中的影响力最大化问题。该方法构建了一个端到端训练的双耦合图神经网络（DGN），结合了深度学习的表示能力和强化学习的决策能力。具体来说，DGN 利用双耦合图神经网络对节点进行嵌入，捕捉社交网络的拓扑结构和节点属性信息，生成丰富的节点向量表示。同时，采用深度 Q 网络（DQN）来探索社交网络的拓扑和节点属性信息，通过学习 Q 函数来近似当前状态和动作下的最优策略，从而自适应地选择种子节点，最大化影响力传播。实验结果表明，DGN 在合成网络和真实网络上的性能接近或超越了当前最先进的模型，如 ToupleGDD，并且在大规模、复杂和密集的社交网络中具有更好的鲁棒性。

Xu 等人<sup>[26]</sup>提出了一种基于层次聚类的图自编码器（HC-GAE），用于图表示学习。HC-GAE 在编码过程中，首先利用硬节点分配将样本图分解为多个分离的子图，然后对每个子图进行图卷积操作以进一步提取节点特征，并将属于每个子图的节点压缩成粗化节点，从而将原始图转换为粗化图。由于分离的子图之间没有连接，卷积操作无法在不同子图之间传播节点信息，从而显著减少了经典基于卷积的图自编码器中出现的过平滑问题。在解码过程中，采用软节点分配来重构原始图结构，通过扩展粗化节点。通过在解码过程中层次化地执行压缩过程以及在解码过程中执行扩展过程，HC-GAE 能够有效提取原始样本图的双向层次结构特征，生成有效的图表示用于节点分类或图分类。此外，文章还重新设计了损失函数，整合了来自编码器和解码器的信息，以平衡训练过程中的局部和全局信息，进一步提升模型的表达能力。

Busch 等人<sup>[27]</sup>提出了一种基于网络流图的恶意软件检测和分类方法，利用图神经网络模型来分析网络流量数据。该方法首先从网络流量中提取流图，其中节点对应于网络中的端点，边表示端点之间的通信。与传统的将网络流视为独立实体的方法不同，该方法通过构建流图来捕捉网络中丰富的通信模式。文章介绍了三种模型变体，分别用于监督学习和无监督学习设置下的恶意软件检测和分类。

监督学习模型 (NF-GNN-CLF) 通过在表示学习模块后添加池化层和预测层来实现图分类, 无监督学习模型包括图自编码器 (NF-GNN-AE) 和单类图神经网络 (NF-GNN-OC), 分别通过重建损失和单类损失来进行异常检测。实验结果表明, 该方法在不同的预测任务中均优于现有方法, 能够显著提高检测性能。该方法不仅能够处理标记数据, 还能在无标记数据和少量训练数据的情况下取得良好的性能, 展示了其在恶意软件检测领域的潜力。

Zhao 等人<sup>[28]</sup>提出了一种时间图卷积网络 (T-GCN) 模型, 用于交通预测。T-GCN 模型结合了图卷积网络 (GCN) 和门控循环单元 (GRU), 以同时捕捉交通数据中的空间和时间依赖性。具体来说, GCN 用于学习城市道路网络的复杂拓扑结构, 从而捕捉空间依赖性, 而 GRU 用于学习交通数据的动态变化, 捕捉时间依赖性。T-GCN 模型首先将历史交通信息作为输入, 通过 GCN 获取道路网络的空间特征, 然后将带有空间特征的时间序列输入到 GRU 模型中, 通过信息在单元之间的传递来获取时间特征。最后, 通过全连接层得到预测结果。实验结果表明, T-GCN 模型在不同预测时间范围内的预测性能优于现有的最先进基线方法, 如历史平均模型 (HA)、自回归积分滑动平均模型 (ARIMA)、支持向量回归模型 (SVR)、单独的 GCN 模型和 GRU 模型。此外, T-GCN 模型在面对数据噪声时表现出良好的鲁棒性, 能够处理高噪声问题, 适用于短期和长期交通预测任务。

### 1.3 本文的主要贡献与创新

本文针对物联网恶意流量检测中存在的模型鲁棒性问题, 提出了一种高鲁棒性的恶意流量检测方法。该方法的核心创新在于结合去噪扩散概率模型与图神经网络技术, 旨在提升模型在不平衡数据环境下的识别能力, 并增强模型在面对各种复杂网络流量时的适应性和稳定性。具体设计过程中, 本文提出了两项关键技术: 基于非功能性特征和低度 IP 的去噪扩散概率模型, 以及基于自适应注意力网络的线图训练方法。两者相辅相成, 有效地提升了恶意流量检测的准确性、鲁棒性和泛化能力。

(1) 基于非功能性特征和低度 IP 的去噪扩散概率模型为了解决网络流量数据中的不平衡问题, 尤其是低度 IP 的识别难题, 本文提出了去噪扩散概率模型。首先, 采用清洗、标签转换和标准化的方式确保数据质量, 并通过正则表达式提取功能性特征和非功能性特征, 处理低度 IP 数据。通过结合特征分类优化模型, 提高了低度 IP 的识别能力。进一步地, 为提升模型的表现, 本文引入了时序图结构来捕捉数据中的时间依赖性, 并增强了模型的上下文感知能力。同时, 结合低度 IP 掩码和数据增强技术, 模型通过添加噪声并逐步去噪生成符合真实分布的样本,

从而扩充了训练集并提升了模型的鲁棒性。

(2) 基于自适应注意力网络的线图训练方法为了进一步提升模型的表现, 本文提出了一种基于自适应注意力网络的线图训练方法。首先, 利用前述去噪扩散概率模型对训练数据进行增强, 并采用线图结构来更有效地捕捉边的特征, 从而提升模型的表达能力。接着, 结合自适应多头注意力机制, 动态调整节点之间的注意力权重, 实现更精确的消息传递和特征聚合。这一机制帮助模型更好地捕捉节点间的复杂关系, 增强了整体表现能力。最后, 本文使用 MAD (中位数绝对偏差) 指标来评估图神经网络在不同层次上的特征分布变化, 特别是邻居节点与远程节点之间的特征差异, 为图神经网络在流量数据中的应用提供了新的分析视角。

## 1.4 本论文的结构安排

围绕上述研究内容, 本文各章节内容安排如下:

第一章为绪论。首先, 说明了本论文的研究的背景和意义; 然后, 对恶意流量检测、流量不平衡处理进行国内外研究现状分析; 最后, 明确本文的主要研究内容和组织结构。

第二章为相关理论及技术介绍。相关基础理论的介绍, 涵盖了半监督学习、去噪扩散模型、恶意网络流量、神经网络和所使用的注意力机制相关介绍和原理。

第三章为基于 LDNFDDPM 的数据增强方法。首先介绍基于生成对抗网络的数据增强方法的不足, 并给出基于 DDPM 的数据增强方法; 然后按照数据预处理、时序图构建、基于 LDNFDDPM 恶意流量生成的顺序进行详细阐述。

第四章为基于多头自适应注意力网络的线图训练方法。首先介绍了传统图神经网络的不足和问题, 然后详细阐述了如何利用线图来重构图, 从而更好地捕捉边缘特征, 接着描述了利用多头注意力网络来进行图卷积操作。

第五章为实验结果分析。首先介绍了实验所使用的数据集和实验环境, 然后通过实验, 对比分析本文方法与 G-IDS、SIGMA、CWGAN-GP、SEMRes-DDPM、E-GraphSAGE、Extra Trees、Random Forest、Deep Feedforward Neural Network、Transductive Support Vector Machine 等方法在开源数据集上的表现, 验证本文所提方法的有效性。

第六章为总结与展望。总结了本文的工作内容, 并对下一步改进工作进行了展望。

## 第二章 相关理论及技术介绍

### 2.1 恶意网络流量

本节主要讨论恶意网络流量的概念和常见的网络攻击。网络流量一般分为正常和恶意流量，正常流量是按照网络协议规定传输的数据，用于正常通信行为。而恶意网络流量具有恶意意图或目的，通常包括网络攻击、病毒传播、网络钓鱼等恶意行为。检测恶意流量有助于更准确地发现和应对问题，保障网络正常运行和数据安全传输。恶意网络流量的攻击种类也多种多样，下面列举几种常见的网络攻击：

DDoS（分布式拒绝服务攻击）<sup>[29]</sup> 是一种通过大量分布在全球的受感染计算机（通常是“僵尸网络”）同时向目标服务器发送大量请求，超负荷目标服务器或网络资源，导致其无法正常服务的攻击方式。通过分布式源发起，DDoS 攻击难以追踪和防御，常常造成网站、服务器或网络基础设施的停机或服务不可用。

DoS（拒绝服务攻击）<sup>[30]</sup> 是一种通过发送大量无效请求，导致目标系统资源耗尽，使得合法用户无法访问目标服务的攻击。通常由单一源发起，攻击者通过消耗系统资源（如内存、带宽、CPU 等）使目标无法处理请求，造成系统崩溃或长时间不可用。

Backdoor（后门攻击）<sup>[31]</sup> 指攻击者通过某种手段在目标系统中植入隐藏的入口（“后门”），从而绕过正常的认证机制，随时访问目标系统。通过恶意软件（如病毒、木马等）创建的后门，攻击者可以在系统内外随时控制受害设备，窃取敏感信息或执行恶意操作。

Injection（注入攻击）<sup>[32]</sup> 是指攻击者将恶意代码（如 SQL 代码、脚本等）插入到输入字段、请求参数或 URL 中，通过漏洞让目标系统执行恶意命令。常见的注入攻击包括 SQL 注入、命令注入和 XSS（跨站脚本攻击）。这些攻击可以窃取数据、控制系统或破坏网络服务。

Password（密码攻击）<sup>[33]</sup> 是指攻击者通过破解目标用户的密码，获取非法访问权限的攻击方式。暴力破解、字典攻击和社交工程攻击是常见的密码攻击方法。成功的密码攻击可能导致系统或账户的控制权被窃取，进而造成数据泄露或系统滥用。

Scanning（扫描攻击）<sup>[34]</sup> 是攻击者利用扫描工具探测目标系统的开放端口、服务或漏洞，寻找潜在的攻击入口。通过端口扫描和漏洞扫描，攻击者可以识别目标系统的弱点，为后续攻击提供依据。

XSS（跨站脚本攻击）<sup>[35]</sup> 是攻击者向网站注入恶意脚本，使得用户在访问网站时执行这些脚本，从而泄露敏感信息（如 Cookie、凭证等）或执行其他恶意操作。XSS 攻击包括存储型 XSS 和反射型 XSS，分别通过存储在服务器上的脚本和通过 URL 传递的脚本来实现。

Theft（窃取攻击）<sup>[36]</sup> 指攻击者通过各种手段非法获取受害者的敏感数据，如用户名、密码、信用卡信息、机密文件等。数据窃取和身份盗窃是窃取攻击的常见形式，攻击者通过盗用信息进行非法活动，给受害者带来金融损失或法律问题。

Reconnaissance（侦察攻击）<sup>[37]</sup> 是攻击者在实际发起攻击之前，对目标系统和网络进行调查和侦察，以收集尽可能多的信息。攻击者可能通过社交工程、网络嗅探和域名查询等手段收集有关目标的信息，并通过网络映射来识别系统架构、端口、设备等，为后续攻击提供基础。

Brute Force（暴力破解攻击）<sup>[38]</sup> 是一种通过反复猜测用户名和密码组合来获取系统访问权限的攻击方式。攻击者通过尝试尽可能多的密码组合，直到找到正确的密码。暴力破解常用于破解简单的密码或账号，如果目标密码较弱，攻击者可能轻易获得访问权限。

## 2.2 去噪扩散模型

本节主要介绍去噪扩散模型。去噪扩散模型 (Denoising Diffusion Probabilistic Models, 简称 DDPM), 最初由 Jonathan Ho 等人<sup>[8]</sup> 提出, 旨在解决图像生成相关问题, 通过逐步添加噪声来训练模型, 生出数据则通过逐步去噪来实现。近年来, 去噪扩散模型也被用在恶意流量数据增强领域。Xingyu Gong、Siyu Chen 和 Na Li<sup>[39]</sup> 提出了一种基于 AOT-DDPM (Adaptive ODE Solver Transformer Network Diffusion Model) 的网络流量数据生成模型, 用于异常流量检测。该模型通过自适应采样策略、Transformer 网络结构和 ODE 求解器的结合, 有效解决了网络流量数据不平衡的问题, 提高了异常流量检测的准确率和效率。

DDPM 是一类基于扩散过程的生成模型, 其核心思想是通过一个渐进的过程, 将数据从清晰的状态变为噪声, 然后通过反向过程逐步去噪, 恢复出原始数据。

正向扩散过程的目标是将原始数据样本  $x_0$  逐步加入噪声, 直至将其转变为噪声。这个过程是一个马尔科夫链, 可以通过以下公式描述:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I) \quad (2-1)$$

其中,  $\mathcal{N}$  表示高斯分布,  $\beta_t$  是每个时间步的噪声调度参数, 控制每一步加入

的噪声强度。通过递归计算，我们可以得到每个时间步的样本：

$$x_t = \sqrt{1 - \beta_t} x_{t-1} + \sqrt{\beta_t} \varepsilon_t \quad (2-2)$$

其中  $\varepsilon_t \sim \mathcal{N}(0, \mathbf{I})$  是均值为 0，方差为 1 的高斯噪声。

最终，正向过程会将原始样本  $x_0$  转变为标准正态分布  $x_T$ ，即：

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (2-3)$$

其中， $\alpha_t = 1 - \beta_t$  且  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 。

在反向扩散过程中，我们通过训练一个神经网络  $p_\theta(x_{t-1}|x_t)$  来近似学习如何从噪声中恢复原始数据。反向过程的公式为：

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_t) \quad (2-4)$$

其中， $\mu_\theta(x_t, t)$  是由神经网络预测的去噪值， $\Sigma_t$  是噪声方差（通常为常数或学习得到的调度）。

DDPM 的训练目标是最大化数据的对数似然  $\log p_\theta(x_0)$ ，即最小化正向过程和反向过程之间的 KL 散度。训练中的损失函数通常表示为去噪误差，公式为：

$$\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{q(x_0, x_1, \dots, x_T)} \left[ \sum_{t=1}^T \|\varepsilon_\theta(x_t, t) - \varepsilon\|^2 \right] \quad (2-5)$$

其中， $\varepsilon_\theta(x_t, t)$  是网络预测的噪声， $\varepsilon$  是实际的噪声， $T$  是扩散过程的步数。

在生成过程中，我们从一个标准正态分布的噪声样本  $x_T \sim \mathcal{N}(0, \mathbf{I})$  开始，通过反向扩散过程逐步去噪，生成样本。每一步的去噪操作可以表示为：

$$x_{t-1} = \mu_\theta(x_t, t) + \sigma_t \cdot \varepsilon_t \quad (2-6)$$

其中， $\varepsilon_t \sim \mathcal{N}(0, \mathbf{I})$  是用于增加噪声的高斯噪声， $\sigma_t$  是标准差，控制噪声的强度。

## 2.3 半监督学习

本节主要介绍半监督学习。在许多实际机器学习应用中，容易获取大量的无标签数据，但要通过特殊设备或昂贵且耗时的实验过程才能标注数据，导致有标签数据稀缺。为了充分利用无标签数据，半监督学习<sup>[40]</sup>应运而生，如图 2-1 所示，它将大量的无标签样本与少量的有标签样本结合起来，进行训练，从而改善学习性能。

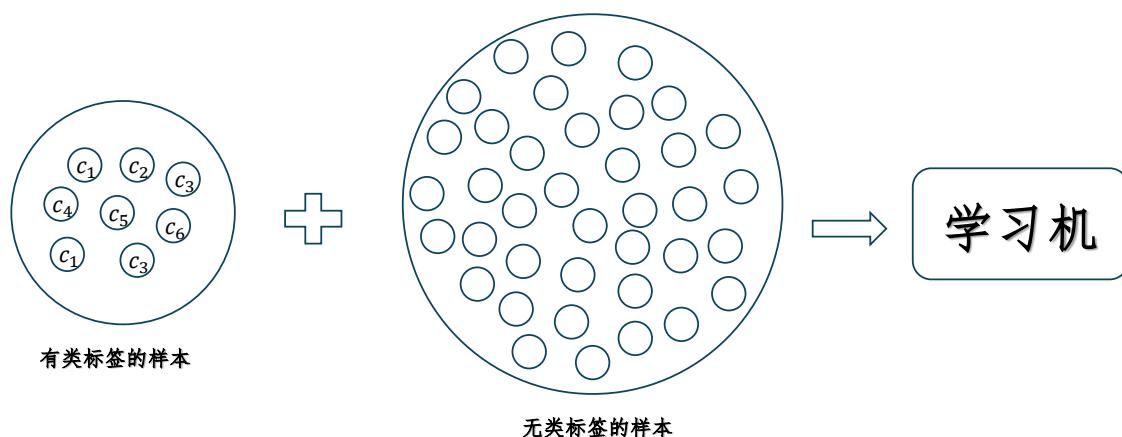


图 2-1 半监督学习

半监督学习的主要目标是避免数据和资源的浪费，同时解决传统监督学习模型泛化能力差和无监督学习模型精度不足的问题。半监督学习的常见应用场景如下：

1. 无标签数据远多于有标签数据
2. 直推半监督学习：直推半监督学习只考虑已知的训练数据，利用有标签数据和无标签数据进行训练，并预测无标签样本的标签。预测的目标是尽量让模型对训练集中的无标签数据进行准确预测。目标是提高模型的泛化能力，特别是在无标签数据的帮助下。

$$\hat{y} = f(x; \theta) \quad (2-7)$$

其中， $\hat{y}$  是对样本  $x$  的预测， $f(x; \theta)$  是模型， $\theta$  是参数。

3. 归纳半监督学习：归纳半监督学习不仅利用已标记和未标记的样本，还处理整个样本空间，包括未标记的数据和测试数据。通过归纳半监督学习，模型不仅可以对无标签的训练样本进行预测，还能预测未知的测试样本的标签。



$$\mathcal{L}_{\text{inductive}} = \sum_{i=1}^N \mathcal{L}(y_i, f(x_i; \theta)) \quad (2-8)$$

其中,  $\mathcal{L}$  是损失函数,  $y_i$  是标签,  $f(x_i; \theta)$  是模型预测。

在归纳半监督学习中, 假设训练数据中的无标签样本并不直接对应测试数据, 而是作为模型泛化能力提升的辅助数据。与此不同, 直推半监督学习假设未标记的样本恰恰是我们关注的待预测数据。

通过这种方式, 半监督学习能够在少量有标签数据的情况下, 通过利用大量无标签数据来提升模型的性能和泛化能力。

## 2.4 图神经网络 (GNN)

本节主要介绍图神经网络。图神经网络 (GNN) <sup>[41]</sup> 是一类专门设计来处理图结构数据的神经网络模型。图结构数据广泛存在于许多实际应用中, 如社交网络、分子结构、知识图谱等。在这些图中, 节点代表个体或物体, 边则表示个体之间的关系。GNN 的核心思想是通过在图结构上进行信息传播和聚合, 来学习节点、边或子图的低维表示。GNN 通过节点的邻域信息聚合来更新每个节点的特征表示, 从而捕捉图中的复杂连接模式和高阶关联信息。

GNN 的基本操作是基于节点邻域的消息传递。每个节点会接收其邻域节点的信息, 然后通过一个神经网络层对接收到的信息进行聚合、变换, 从而更新自己的特征表示。通过多层的堆叠, 信息会逐渐从局部传播到全局, 最终使得每个节点的特征包含了整个图的高阶信息。

GNN 的模型通常分为两大类:

**谱模型 (Spectral Models):** 通过图信号的谱变换 (例如图傅里叶变换) 进行卷积操作, 利用图的拉普拉斯算子或其变体进行处理。谱模型的方法通常依赖于图的全局性质, 计算复杂度较高。

**空间模型 (Spatial Models):** 直接在图的结构上进行卷积操作, 通过加权聚合邻居节点的信息来提取局部特征。这类模型通常计算更为高效, 且能较好地处理大规模图数据。

GNN 在许多领域得到了广泛的应用。例如, 在推荐系统中, GNN 能够处理用户-物品交互数据、用户社交关系等图结构数据, 从而提高推荐的准确性和多样性。在社交网络分析中, GNN 能够挖掘用户间的隐性关系, 提高社交推荐或广告精准度。此外, GNN 还被应用于生物信息学 (如分子结构预测) 和自然语言处理 (如图卷积网络在图像标注和图像描述中的应用)。

与传统的矩阵分解或浅层学习模型相比，GNN 具有显著优势：它能够利用图结构中的高阶连接信息，捕捉节点之间复杂的关系，从而增强模型的表现力和鲁棒性。

## 2.5 图卷积网络 (GCN)

本节主要介绍图卷积网络。图卷积网络 (GCN) [41] 是图神经网络 (GNN) 中的一种重要模型，属于谱模型。GCN 的核心思想是将图信号转换到谱域，通过图傅里叶变换来进行滤波，然后再将处理后的信号转换回空间域。GCN 利用图的邻接矩阵和度矩阵来定义卷积操作，从而实现对图信号的平滑和特征提取。

GCN 的卷积操作基于图的谱理论。首先，GCN 通过图的傅里叶变换将图信号映射到谱域，然后在谱域上应用滤波器进行卷积操作，最后再通过逆变换回到空间域。在图卷积的具体实现中，邻接矩阵和度矩阵被用来设计滤波器，进而对节点特征进行聚合和更新。

GCN 的基本公式如下：

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (2-9)$$

其中：-  $H^{(l)}$  是第  $l$  层的节点特征矩阵，表示每个节点在该层的特征表示；-  $\tilde{A} = A + I$  是带有自环的邻接矩阵 ( $A$  为原始邻接矩阵， $I$  为单位矩阵)；-  $\tilde{D}$  是度矩阵， $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ ；-  $W^{(l)}$  是第  $l$  层的可学习权重矩阵；-  $\sigma$  是激活函数 (如 ReLU)。

GCN 通常由多个卷积层堆叠而成，每一层的输出作为下一层的输入。在每一层，节点的特征向量通过卷积操作进行更新，更新后的特征向量包含了来自邻居节点的信息，完成了信息的传播与聚合。

GCN 的优势之一是其能够通过邻接矩阵和度矩阵的规范化，有效地捕捉图中的高阶邻域信息。相比传统基于邻域的图处理方法，GCN 能在更高的层次上提取图的特征，从而提升性能和泛化能力。

GCN 广泛应用于节点分类、图分类、链接预测等任务。例如，在推荐系统中，GCN 可以用于学习用户与物品之间的图嵌入，通过捕捉用户和物品之间复杂的关系来提高推荐系统的准确性和多样性。在社交网络中，GCN 可以用于社交关系的分析和社区发现。在图像处理和计算机视觉中，GCN 能够处理图像中的结构信息，提高模型的表现力。

## 2.6 图注意力机制

本节主要介绍图注意力机制。图注意力机制（Graph Attention Mechanism）<sup>[42]</sup>是一种用于处理图结构数据的注意力机制，它允许模型在聚合节点信息时为不同邻居节点分配不同的权重，从而更好地捕捉节点之间的复杂关系和图的结构特征。图注意力机制通过自注意力策略使得每个节点可以根据邻居节点的特征计算自身的隐藏表示，进而捕捉图中的复杂结构信息。图注意力机制的操作具有很强的并行性，适应不同度节点的特性，并且不需要全局图结构信息，使得模型能够高效地处理大规模图数据。

对于每个节点  $i$ ，首先对其邻居节点  $j$  的特征进行线性变换，然后通过共享的注意力机制  $a$  来计算注意力系数  $e_{ij}$ ，表示节点  $j$  的特征对节点  $i$  的重要性。计算公式为：

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j) \quad (2-10)$$

其中， $W$  是权重矩阵， $\vec{h}_i$  和  $\vec{h}_j$  分别是节点  $i$  和  $j$  的特征向量。为了使不同节点的注意力系数可比较，模型使用 softmax 函数对注意力系数进行归一化，得到最终的注意力权重  $\alpha_{ij}$ ：

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})} \quad (2-11)$$

其中， $\mathcal{N}_i$  是节点  $i$  的邻居节点集合。利用归一化后的注意力权重，模型对邻居节点的特征进行加权求和，得到节点  $i$  的最终特征表示：

$$\vec{h}'_i = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij} W \vec{h}_j \right) \quad (2-12)$$

其中， $\sigma$  是非线性激活函数，用于引入非线性特性。

为了进一步提升模型的表达能力，图注意力机制通常采用多头注意力（multi-head attention）。具体来说，模型将多个独立的注意力机制的输出进行拼接或平均，得到最终的节点特征表示。例如，对于  $K$  个注意力头，节点  $i$  的最终特征表示为：

$$\vec{h}'_i = \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right) \quad (2-13)$$

其中,  $\alpha_{ij}^k$  和  $W^k$  分别是第  $k$  个注意力头的注意力权重和权重矩阵。

图注意力机制通过引入动态的邻居加权策略, 使得每个节点能够根据邻居节点的特征来自动分配权重, 提升了模型在图结构数据中的表达能力。它不仅能有效捕捉图的局部结构信息, 还能在处理图数据时具有较高的计算效率, 尤其适合处理大规模图数据。图注意力机制在图分类、节点分类、链接预测等任务中展现了出色的性能, 并且可以与其他图神经网络模型结合, 进一步提升任务的准确性和鲁棒性。

## 2.7 本章小结

本章介绍了恶意网络流量、去噪扩散模型、半监督学习、图神经网络、图卷积网络和图注意力机制六个方面的知识。首先, 介绍了恶意网络流量的基本概念, 并介绍了几种常见的恶意网络流量攻击类型。然后, 介绍了本文工作所涉及的半监督学习、去噪扩散模型、图神经网络和图卷积网络的基本概念和原理。最后, 介绍了本文工作中所涉及的图注意力机制的基本概念和原理。

### 第三章 基于 LDNFDDPM 的数据增强方法

本章提出了一种基于非功能性特征和低度 IP 的去噪扩散概率模型，旨在解决流量数据不平衡问题。首先，对原始数据进行了清洗、标签转换和标准化，确保数据质量。然后，通过正则表达式提取功能性特征（如网络协议、端口、ICMP 类型）和非功能性特征（如流量统计、主机信息）。对训练数据中的低度 IP 进行处理，并结合特征分类优化模型。为了增强模型的表现，时序图被引入，捕捉数据的时间依赖性，提供更丰富的上下文信息。数据集被划分为训练集和测试集。训练集经过功能性和非功能性特征划分，并生成低度 IP 掩码，以便进行数据增强。

该模型通过向数据中添加噪声并逐步去噪，生成符合真实分布的样本，从而扩充训练数据集并提高模型的鲁棒性。该模型不仅改善了低度 IP 的识别能力，还增强了整体模型的泛化能力。

#### 3.1 问题描述和研究思路

恶意流量检测方法是确保网络安全的重要手段，它通过对网络流量进行实时监测和分析，以识别和防御潜在的恶意行为。然而，在实际物联网环境中，正常流量往往占据压倒性多数，而恶意流量则相对稀少，并且在复杂的交互网络中，游离 IP 更容易受到攻击，这导致了数据集中正常流量与恶意流量之间的显著类别不平衡问题。这种不平衡会使基于人工智能模型在训练过程中过于关注正常流量，而忽略掉恶意流量的特征，从而导致模型在识别恶意流量时的性能下降，甚至可能出现大量的误报和漏报。

去噪扩散概率模型（Denoising Diffusion Probabilistic Models，简称 DDPM）<sup>[8]</sup>是一种基于扩散过程的生成模型。它由 Sohl-Dickstein 等人在 2015 年首次提出，并在近年来成为生成模型领域的重要研究方向。DDPM 的核心思想是在数据生成过程中引入噪声，并通过逐步去噪来恢复原始数据，从而进行有效的样本生成。DDPM 广泛应用于图像生成、视频生成、语音生成等领域。近年来，DDPM 在多个生成任务上取得了超越 GAN 和 VAEs（变分自编码器）的表现。在图像生成领域，DDPM 通过逐步去噪的方式能够生成质量非常高的样本，特别是在高分辨率图像生成上表现尤为突出。

DDPM（去噪扩散概率模型）在处理流量图数据时，虽然具备生成高质量样本的能力，但也存在一些局限性，尤其是在流量数据的特征复杂性和数据不平衡问题上。流量图数据通常包括多种类型的节点和边特征，这些特征的多样性和异

质性使得 DDPM 在直接应用时面临挑战。特别是在流量图中的节点，如 IP 地址，可能会涉及到不同等级的 IP，其中部分低度 IP 节点（即与其他节点连接较少的节点）在生成过程中常常被忽略。低度 IP 节点通常代表着稀有或攻击类型的特征，但由于其在真实流量中出现的频率较低，传统 DDPM 模型可能无法有效学习到这些特征，从而影响生成模型对稀有攻击类型的生成能力。

此外，流量图中的特征通常分为功能性特征和非功能性特征两类。功能性特征包括协议类型、端口信息、ICMP 类型等，反映了网络流量的基础结构和通信方式。而非功能性特征则包括流量统计信息、主机信息、时延等，主要描述了流量的性质和行为模式。DDPM 在进行生成任务时，通常未针对这些不同类型的特征进行特殊处理，可能会对非功能性特征的噪声过度增强，进而导致生成数据的质量下降。

另外，传统的 DDPM 方法也没有充分考虑到数据的不平衡问题，特别是对于攻击类型之间的差异。攻击流量（如 DDoS、Bruteforce 等）相较于正常流量（如 Benign）通常较少，这种数据不平衡会影响 DDPM 的学习效果，导致模型无法很好地生成较为罕见的攻击流量，从而影响生成数据的多样性和完整性。

针对这些不足，本研究提出了一种改进的 DDPM 方法——基于低度 IP 与非功能性特征的 DDPM（LDNFDDPM）。该方法在传统 DDPM 的基础上引入了低度 IP 节点的计算和功能性、非功能性特征的划分，以更好地处理流量图数据中的特征复杂性和不平衡问题。通过计算低度 IP 节点，并专门为其设计增强策略，LDNFDDPM 能够有效地学习稀有攻击类型的特征，确保生成的流量样本包含了更多真实世界中可能遇到的攻击场景。同时，通过划分功能性特征和非功能性特征，并对其分别进行建模和处理，模型能够在生成过程中保留非功能性特征的重要信息，避免对这些特征的噪声过度增强，从而提高了生成数据的质量和多样性。

这种改进的 DDPM 方法不仅解决了传统模型在处理流量图数据时的不足，还增强了模型对稀有事件、低度 IP 节点和多样化特征的捕捉能力，从而为网络流量检测和生成任务提供了更为准确和鲁棒的数据增强方案。

LDNFDDPM 的总体结构如图 3-1 所示，包含三个主要模块：数据预处理模块、时序图构建模块和基于 LDNFDDPM 的恶意流量生成模块。

数据预处理模块：数据预处理是整个流程的第一步，主要包括数据集特征处理、数据标准化和任务集构造。首先，原始流量数据需要进行清洗，处理缺失值和异常值，并对标签进行转换，确保数据的有效性。接着，特征选择和标准化步骤确保了不同特征的尺度一致，为后续的建模和训练提供了合适的输入。此外，任务集构造根据实验需求构建训练集和测试集，为后续模型的训练与验证奠定基础。

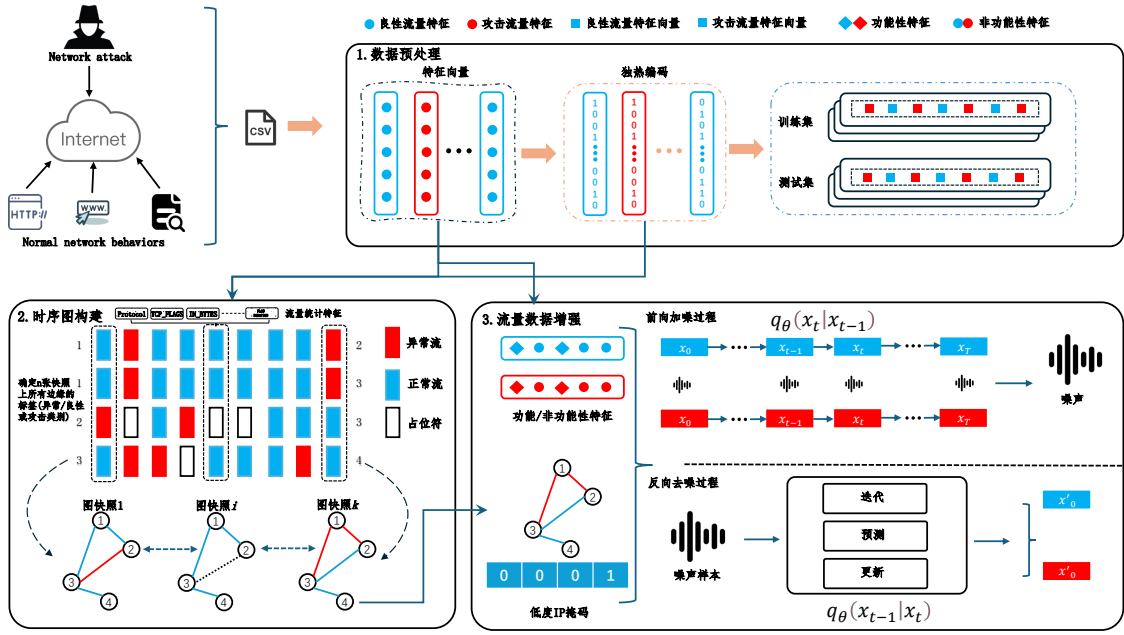


图 3-1 基于 LDNFDDPM 的恶意流量数据增强

时序图构建模块：该模块将主机之间的时序关系抽象为时序图。在流量图中，每个主机（即 IP 节点）之间的通信可以视为图中的边，时序图的构建则关注通信的顺序和时延等信息。通过捕获流量图中节点之间的时序依赖关系，时序图构建模块能够有效地表达网络流量的动态特性，为后续流量生成和异常检测提供支持。

基于 LDNFDDPM 的恶意流量生成模块：在时序图的基础上，LDNFDDPM 恶意流量生成模块根据物联网流量数据集的不平衡性进行流量增强。首先，基于特征类型与作用，将原始物联网流量特征向量分为功能性特征和非功能性特征。功能性特征主要包括网络协议类型、端口号等，而非功能性特征则涵盖流量统计、主机信息等。其次，统计低度 IP（即与其他节点连接较少的 IP）并生成相应的掩码。低度 IP 通常与异常或攻击行为相关，因此对这些 IP 的关注有助于检测潜在的攻击流量。

最后，LDNFDDPM 将非功能性特征和低度 IP 作为噪声关注点，结合去噪过程生成新的特征向量。此过程不仅提高了生成样本的多样性，还显著增强了流量生成的真实性，特别是在生成稀有攻击流量时。通过这一改进，LDNFDDPM 能够有效应对数据不平衡问题，生成更加符合实际网络流量分布的恶意流量样本，从而提升模型的鲁棒性和准确性。

## 3.2 框架设计

### 3.2.1 数据预处理

数据预处理是恶意流量分类任务中的关键步骤。在实际的网络流量数据中，流量特征常常包含噪声、缺失值和冗余信息，这些问题若不加以处理，会影响模型训练的效果。因此，进行有效的数据预处理至关重要。本研究在数据预处理阶段，依照以下几个步骤进行操作，以确保数据质量和特征的有效性。具体的流程图如 3-2 所示。为了让模型能够更好的进行训练，首先，对原始数据集进行清洗。

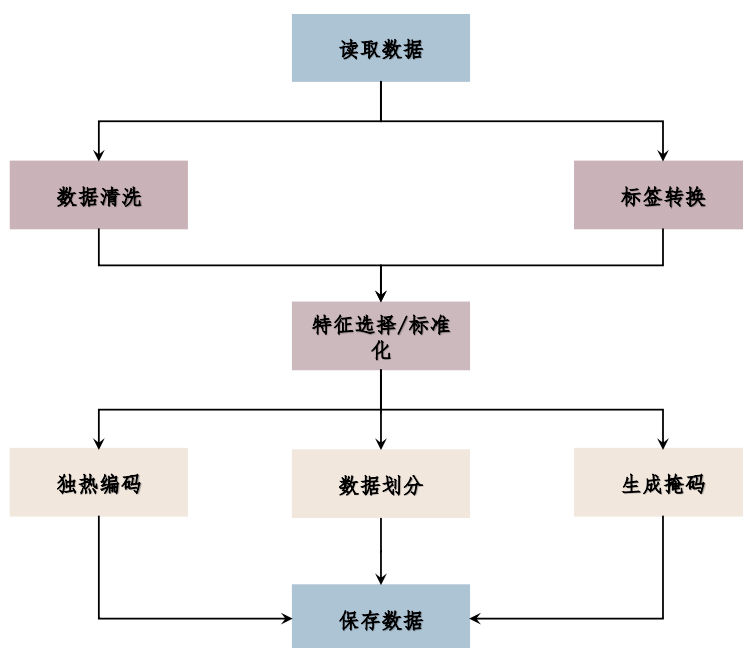


图 3-2 数据预处理

由于流量数据的多样性，可能会存在缺失值和异常值，这些不完整或错误的数​​据会严重影响后续分析的准确性。因此，通过删除或填补缺失值、去除异常值等方法，确保数据集的完整性与准确性。

针对攻击类型标签，本研究对原始标签进行了转换。网络攻击类型包括：DoS（拒绝服务攻击）、DDoS（分布式拒绝服务攻击）、Reconnaissance（侦察攻击）、Theft（盗窃攻击）以及正常流量（Benign）。为了便于后续模型训练，这些文本标签被转换为数值型标签。具体来说，DoS 被标记为 1，DDoS 为 2，Reconnaissance 为 3，Theft 为 4，正常流量为 0。

其次，针对特征选择，考虑到对模型训练的有效性，我们选择了约 60 个特征，它们分别是流量特征：这些特征反映了流量的传输量，包括输入和输出的字节数和流量的第二级别单位字节数（如源到目标和目标到源的字节数）。包数特征：这类特征记录了网络通信中的包的数量，包括传输的包数以及重新传输的包数，这



对于识别重复传输流量有很大帮助。流持续时间：表示网络连接持续的时间，能够帮助判断流量的持续性，尤其对 DDoS 类攻击的检测非常重要。时间相关特征：这些特征（如 TTL 值）有助于揭示数据包的生命周期和传输路径，特别是在网络攻击中，TTL 值的变化常常反映异常行为。流量大小特征：涉及流量包的最大最小尺寸，有助于揭示攻击流量的异常模式，如流量包过大或过小。吞吐量特征：衡量源和目标之间的数据传输速率，突发流量和持续性攻击（如 DDoS）通常在吞吐量上表现出异常特征。TCP 窗口特征：TCP 窗口大小影响数据流的传输速率，分析窗口大小的最大值可以揭示网络的流量控制机制，以及潜在的攻击行为。DNS 特征：这些特征与 DNS 查询相关，能够揭示网络中的域名解析活动，有助于检测 DNS 放大攻击等特定类型的攻击。部分选取的特征如表 3-1 所示。我们进一步对

表 3-1 部分网络流量特征

特征	描述
IN_BYTES	输入流量字节数
OUT_BYTES	输出流量字节数
FLOW_DURATION_MILLISECONDS	流持续时间
DURATION_IN	入站流持续时间
DURATION_OUT	出站流持续时间
MIN_TTL	最小 TTL（生存时间）
MAX_TTL	最大 TTL
LONGEST_FLOW_PKT	流最大数据包长度
SHORTEST_FLOW_PKT	流最小数据包长度
SRC_TO_DST_SECOND_BYTES	源到目标的第二级别字节数
DST_TO_SRC_SECOND_BYTES	目标到源的第二级别字节数
RETRANSMITTED_IN_BYTES	重新传输的入站字节数
RETRANSMITTED_OUT_BYTES	重新传输的出站字节数
SRC_TO_DST_AVG_THROUGHPUT	源到目标的平均吞吐量
DST_TO_SRC_AVG_THROUGHPUT	目标到源的平均吞吐量

这些特征进行处理，在流量数据分析中，许多类别特征如协议类型、端口号、标志位等通常是非数值型的，直接将其用于模型训练可能导致模型对特征关系产生误解。因此，必须将这些类别特征转化为数值特征，使得机器学习模型能够有效地进行训练。独热编码（One-Hot Encoding）是处理类别特征的常用方法，它将每个类别值转换为一个新的二元特征，确保模型不会误解类别之间的顺序或数值关系。在本研究中，对多种关键类别特征进行了独热编码，以便模型能够准确理解不同类别特征所代表的信息，提升模型在流量分类任务中的表现。

具体而言，研究中选择了包括 PROTOCOL（协议类型）、TCP\_FLAGS（TCP 标志）、CLIENT\_TCP\_FLAGS（客户端 TCP 标志）、SERVER\_TCP\_FLAGS（服务器端

TCP 标志)、DNS\_QUERY\_TYPE (DNS 查询类型)、FTP\_COMMAND\_RET\_CODE (FTP 命令返回代码) 等类别特征进行独热编码。这些特征直接影响流量的网络行为模式,且在恶意流量检测中具有重要作用。例如,协议类型(如 TCP、UDP、ICMP 等)对流量的传输和攻击行为有重要指示作用,TCP 标志和客户端/服务器端的 TCP 标志位反映了连接的状态信息,这些信息对流量分类尤其是恶意流量的识别至关重要。同样,DNS 查询类型和 FTP 命令返回代码也常常用于表征特定的攻击行为模式,如 DNS 放大攻击或特定的网络扫描行为。因此,独热编码能够帮助模型区分这些不同的网络行为模式,进一步提升对恶意流量的识别精度。

为了确保编码后的特征能被有效处理,本文采取了两种独热编码方式,如图 3-3 所示。对于类别取值较少且相对固定的特征,如 PROTOCOL、TCP\_FLAGS、DNS\_QUERY\_TYPE 等,直接采用标准的独热编码方法。即,为每个类别的不同取值创建独立的二元特征列,表示该类别是否属于某一特定值。这种方法能够明确表达各类别之间的不同,有助于模型更精确地捕捉流量中的模式。对于一些取值众多且频率不均的特征,如 PORT、L4\_SRC\_PORT、L4\_DST\_PORT、L7\_PROTO、ICMP\_TYPE 等,直接进行独热编码可能会导致特征维度过大,进而增加计算复杂度。因此,采用间接独热编码方法。具体来说,首先统计这些特征中出现频率较高的前  $n$  个取值,并为这些取值创建新的二元特征列,而将剩余不常见的取值归为一个“Other”类别。这样不仅减少了特征维度,还确保了编码后的特征能够代表大部分常见流量模式,同时避免了高维度带来的计算开销。最后,为了确保不

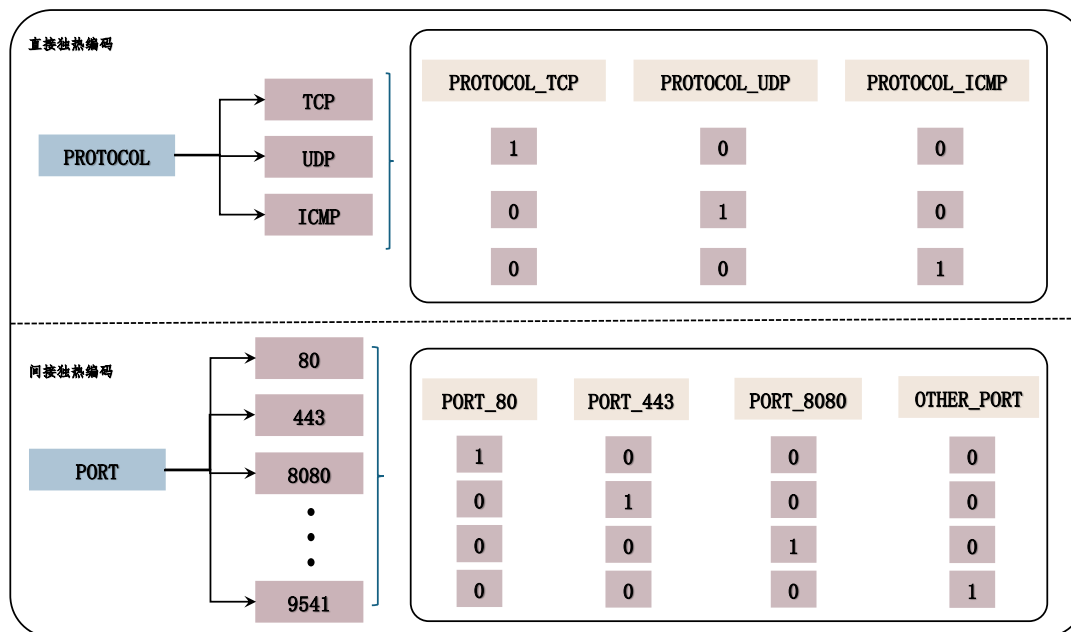


图 3-3 两种独热编码方式

同特征具有相同的尺度，从而避免某些特征因为数值范围大而在模型训练过程中占据主导地位。例如网络流量数据中，“输入字节数 (IN\_BYTES)” 可能在几千到几百万之间，而“TTL 值 (MIN\_TTL)” 则在 1 到 255 之间。如果不进行归一化处理，数值范围较大的特征可能会对模型的训练产生更大的影响，从而导致模型对某些特征的偏倚。本文采用 QuantileTransformer 对连续性特征进行归一化操作，QuantileTransformer 的过程可以分为以下几个步骤：排序数据并计算分位数，如公式 3-1 所示；将数据点的分位数映射到目标分布（如均匀分布或正态分布），如公式 3-2 所示；逆变化，将数据恢复到原始分布，如公式 3-3 所示。

$$p_i = \frac{i}{n} \quad \text{for } i = 1 \dots n \quad (3-1)$$

$$x_i = F_{\text{normal}}^{-1}(p_i) = \Phi^{-1}(p_i) \quad (3-2)$$

$$x'_i = F_{\text{normal}}(x_i) = \Phi(x_i) \quad (3-3)$$

此外，在网络流量分析和恶意流量检测中，标签数据的稀缺性和人工标注成本的高昂，往往限制了传统监督学习方法的应用。为了解决这个问题，半监督学习 (Semi-supervised Learning) 方法成为了一种有效的解决方案。半监督学习通过同时利用少量的标记数据和大量的未标记数据进行训练，从而能够在标签稀缺的情况下，获得较为准确的分类结果。这种方法特别适用于标签数据不平衡或数据难以标注的场景。具体来说，半监督学习方法在训练过程中，首先利用已标记数据进行监督学习，然后通过推测未标记数据的标签来增强学习过程。通过这种方式，未标记数据中的结构信息被有效地利用，进而提升了模型的分类性能。

在本研究中，数据集被划分为 70% 用于训练，30% 用于测试。这样的划分方式确保了训练集中的标记数据能够充分用于训练模型，同时也保证了测试集的独立性，可以有效评估模型在未见数据上的表现。为了实现半监督学习，采用了基于图的半监督学习方法。在这种方法中，数据被表示为图的形式，其中每个节点代表网络中的设备或 IP 地址，而边则表示设备之间的通信关系。图神经网络 (GNN) 通过对图结构的学习，能够对节点之间的复杂关系进行建模，并在传播信息的过程中逐步推断未标记节点的标签。这一过程不仅能够学习到每个节点的特征，还能够通过图中的拓扑结构捕捉到未标记数据的潜在标签信息，从而提高分类性能。

基于图的半监督学习方法相较于传统的机器学习方法具有明显优势，尤其在面对具有复杂关系和时序性的数据时。网络流量数据通常包含了丰富的时序信息和设备间复杂的通信拓扑，传统的机器学习方法往往无法充分利用这些信息。而

图神经网络通过传播节点间的信息，能够有效地捕捉到这些复杂的关系。此外，图神经网络还能够处理稀疏数据和不规则结构，这对于网络流量分析尤为重要，因为流量数据中往往存在不完全的标注或异常数据，图神经网络可以利用图的结构关系来补充这些缺失的信息，如图 3-4 所示。

另外，半监督学习方法的引入解决了标签数据不足的问题。在传统的监督学习方法中，模型训练依赖于大量的标记数据，而标记数据的获取通常非常昂贵且耗时。通过半监督学习，模型可以在大量未标记数据的基础上进行训练，从而减少了对标记数据的依赖，并能够提高数据的利用率。基于图的半监督学习方法，不仅提升了流量分类的准确性，还通过信息传播机制，改善了对少量标记数据的依赖，使得模型能够在标签稀缺的情况下有效地训练。

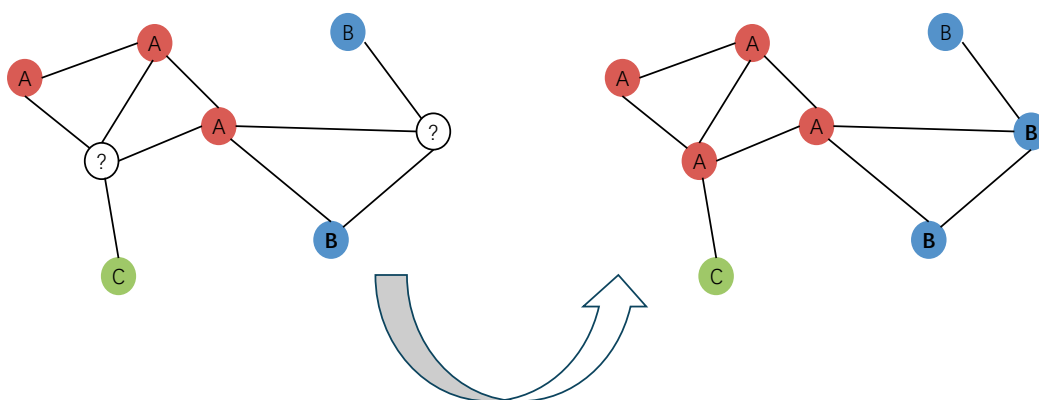


图 3-4 图半监督学习

### 3.2.2 时序图构建

为了应对日益复杂的网络攻击行为，传统的入侵检测方法已无法满足对长期、持续性攻击的有效识别需求。尤其是针对持续时间较长的攻击，如 DDoS 攻击或暴力破解等，网络流量的特征通常会随时间发生动态变化，因此，需要一种能够捕捉到这些时间演化特征的检测方法。本文采用时序图的构建与分析方法，旨在通过时序图捕捉网络流量的演化过程，进而实现对潜在攻击行为的精准检测。

时序图的构建首先依赖于网络流量数据的转换，如图 3-5 所示。具体来说，网络流量是由成对的 IP 地址之间的通信构成的双向流量，而每一对 IP 地址之间的

通信可以视为图中的一条边，IP 地址则对应图中的节点。在此框架下，网络流量数据按照时间戳被切割成多个图快照，构成时序图序列。每一个图快照代表某一时刻的网络拓扑结构，通过时间顺序连接这些图快照，可以表达网络流量的动态变化。类似的图结构在网络分析中已经得到了广泛应用 [?]

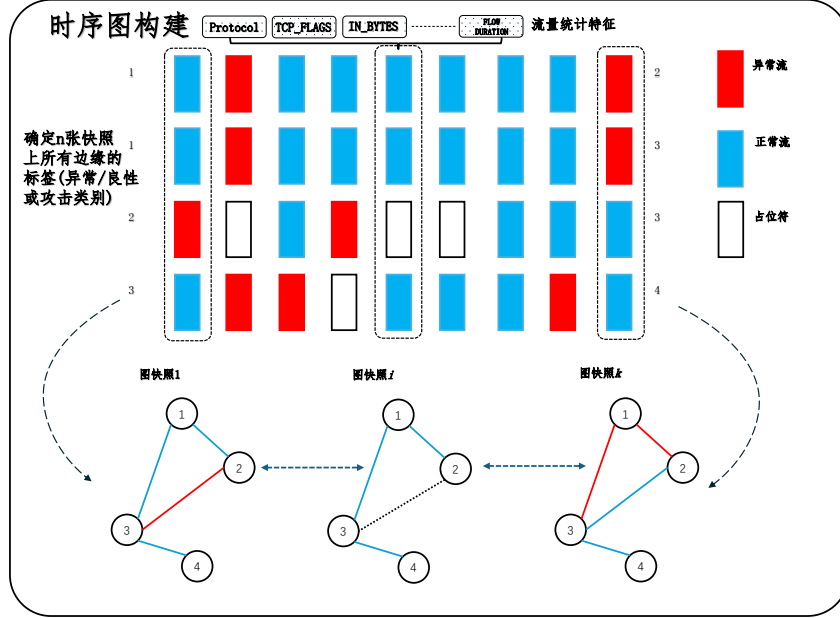


图 3-5 时序图构建

每个图快照由节点集合  $V_f$  和边集合  $E_f$  组成。节点集合  $V_f$  代表在该时刻参与通信的所有 IP 地址，而边集合  $E_f$  则表示在该时刻存在流量的 IP 地址对之间的连接，每条边的特征包括流量的协议类型、持续时间、入站流量字节数、每包字节数、TCP 标志位等。这些特征不仅描述了网络流量的属性，而且可以用来识别流量的正常性与异常性。例如，Zhang 等人提出了一种基于图模型的入侵检测方法，强调了流量特征在攻击检测中的重要性 [?]

$$A_{ij}^f = \begin{cases} 1 & \text{如果 } (v_i^f, v_j^f) \in E_f \\ 0 & \text{否则} \end{cases} \quad (3-4)$$

为了便于后续的图分析和模型训练，图的拓扑结构通过邻接矩阵表示。对于第  $f$  个快照  $G_f = (V_f, E_f)$ ，其邻接矩阵  $A_f$  是一个  $N \times N$  的矩阵，其中  $N$  是网络中 IP 地址的数量。如公式 3-4 所示，其中  $A_{ij}^f = 1$  表示节点  $v_i^f$  与节点  $v_j^f$  之间存在边，反之  $A_{ij}^f = 0$  表示无边。如果在第  $f$  个图快照中，这两个节点之间有通信流量，则  $A_{ij}^f = 1$ ，否则为 0。这种邻接矩阵的表示方式简洁而有效，能够直接反映网络中各个 IP 地址之间的连接关系，并为图分析模型提供输入 [?]

时序图的核心优势在于能够反映网络流量随时间的演化。现实中，攻击行为如 DDoS 攻击往往是一个持续的过程，攻击流量会随着时间的推移逐渐增加，这导致网络拓扑和流量特征发生变化。通过将每个时刻的网络状态表示为一个静态图，并将这些静态图按时间顺序连接成时序图，能够有效追踪网络流量的动态变化。具体来说，假设我们有  $K$  个图快照  $G_1 G_2 \dots G_K$ ，这些快照反映了从  $t_1$  到  $t_K$  时刻的网络状态。通过分析这些图快照的演化过程，可以揭示网络流量中的长期依赖关系，进而更好地捕捉到潜在的攻击行为。类似的研究已经表明，时序图能够有效建模流量的演变过程，进而实现对网络攻击的更精确检测 [?]。在恶意流量检测任务中，流量特征随着时间的推移会发生变化，特别是在攻击期间。为了捕捉这些时序依赖关系，图中的边特征会随时间演化，如公式 3-5 所示。假设第  $f$  到第  $f+t$  个图快照中的边特征分别为  $\{e_{ij}^f e_{ij}^{f+1} \dots e_{ij}^{f+t}\}$ ，则流量的变化可以表示为这些边特征的演变。图神经网络能够通过学习这些时序依赖关系，进一步提升流量检测的精度。

$$e_{ij}^{f+t} = f(e_{ij}^f \dots e_{ij}^{f+t}) \quad (3-5)$$

其中， $f$  表示流量特征随时间变化的函数。通过图的演化，时序图能够捕捉到攻击的长期演变特征，帮助识别持续性攻击。

时序图还能够处理网络流量中的稀疏性问题。在实际网络环境中，流量往往不是均匀分布的，设备可能会因故障或其他原因暂时断开连接，或是某些时间段内没有数据传输。为了应对这种稀疏性，通常会在缺失的图快照中插入占位符。占位符的作用是填补网络流量的空白时刻，以确保图的时序性和完整性。具体而言，占位符边的特征被设置为 0，并通过特殊标签标记这些边。通过这种方式，可以保持图的结构不受缺失数据的影响，从而保证时序图的完整性 [?]

此外，时序图还能够有效捕捉到网络攻击的演化特征。网络攻击通常不是突然发生的，而是一个逐步演变的过程。例如，在 DDoS 攻击期间，攻击流量通常从低谷逐渐上升，直到达到峰值。这种攻击模式在时序图中表现为边特征的逐步变化，时序图能够准确捕捉到这种演化过程。通过分析图的演化，能够有效区分正常流量与攻击流量，识别出攻击的发生及其持续时间。已有研究表明，时序图能够显著提高对这种持续性攻击的检测能力 [?]

时序图还为半监督学习提供了有效的支持。在实际应用中，由于标签数据的匮乏，半监督学习方法能够在少量标记数据的基础上进行训练，并在未标记数据上进行验证。在时序图的训练过程中，通过标记少量的边，可以对整个时序图中的

所有边进行检测，识别出潜在的异常行为。由于时序图能够捕捉到网络流量的长期依赖性和时序演化，半监督学习方法能够利用这些信息有效提高检测的准确性，特别是在面对持续性攻击时，时序图提供了比传统方法更为精细的分析能力 [?]。

时序图通过其时间维度上的演化特点，能够精准地捕捉到攻击行为的变化过程，尤其在面对 DDoS 攻击、暴力破解等持续性攻击时表现出色。通过构建时序图并结合半监督学习模型，可以有效地识别出攻击行为，提升入侵检测系统的性能和准确性。为了成功构建出时序图，本文采用了算法 3-1 生成时序图集，首先，算法接受多个图快照  $G_1 G_2 \dots G_K$ ，每个图快照由节点集合  $V_f$  和边集合  $E_f$  组成，表示在不同时间点上网络的拓扑结构和流量信息。其次，对于每一对 IP 地址  $(v_i v_j)$ ，算法计算它们之间的双向流量特征，例如协议类型、流量持续时间、字节数等。若该对 IP 地址之间存在流量连接，则在相应的图快照中创建边，并记录相关的流量特征。然后，算法为每个图快照构建邻接矩阵  $A_f$ ，该矩阵表示图中各个节点（IP 地址）之间的连接关系。如果两个节点之间存在流量连接，则矩阵中对应元素设为 1，否则设为 0。最后，由于实际网络中流量分布可能存在不均匀性，某些时刻可能缺失图快照。为了处理这种稀疏性，算法在缺失位置插入占位符边，并将这些边的特征设为 0，从而保持时序图的完整性。

**算法 3-1** 时序图构建

**Input:**  $K$  个图快照  $G_1 G_2 \dots G_K$ , 每个图快照包含节点集合  $V_f$  和边集合  $E_f$

**Output:** 时序图  $T = \{G_1 G_2 \dots G_K\}$

```

1  for  $f = 1$  to  $K$  do
2      foreach 每对 IP 地址  $(v_i v_j)$  do
3          计算该对 IP 地址之间的双向流量特征;
4          记录流量的协议类型、持续时间、字节数、TCP 标志等特征;
5          if 存在流量连接 then
6              在图快照  $G_f$  中创建边  $(v_i v_j)$  并记录其流量特征;
7          else
8              跳过此对 IP 地址;
9          end
10     end
11     foreach 每个图快照  $G_f$  do
12         构建邻接矩阵  $A_f$ , 其中  $A_{ij}^f$  表示节点  $v_i$  和节点  $v_j$  是否连接;
13         if 流量连接存在 then
14             设置  $A_{ij}^f = 1$ ;
15         else
16             设置  $A_{ij}^f = 0$ ;
17         end
18     end
19 end
20 for  $f = 1$  to  $K$  do
21     foreach 图快照中的空缺位置 do
22         插入占位符边, 所有特征值设为 0;
23         使用特殊标签标记该边为占位符;
24     end
25 end

```

## 3.2.3 基于 LDNFDDPM 恶意流量生成

为了确保生成的恶意攻击数据尽可能逼近真实数据, 时序图的构建不仅仅依赖于图的拓扑结构和时序关系, 还需要特别关注流量数据的特征选择。根据恶意



攻击流量的特点，特征选择过程被细化为保留功能性特征 ( $x_f$ ) 并生成非功能性特征 ( $x_{nf}$ )，最后再增加低度 IP 的权重，如图 3-6 所示。这种方法能够确保生成的数据在保留关键特征的基础上，增强数据的多样性与真实性，进而提高模型的鲁棒性和准确性。功能性特征是指那些对恶意攻击流量分类有显著影响的特征，这些

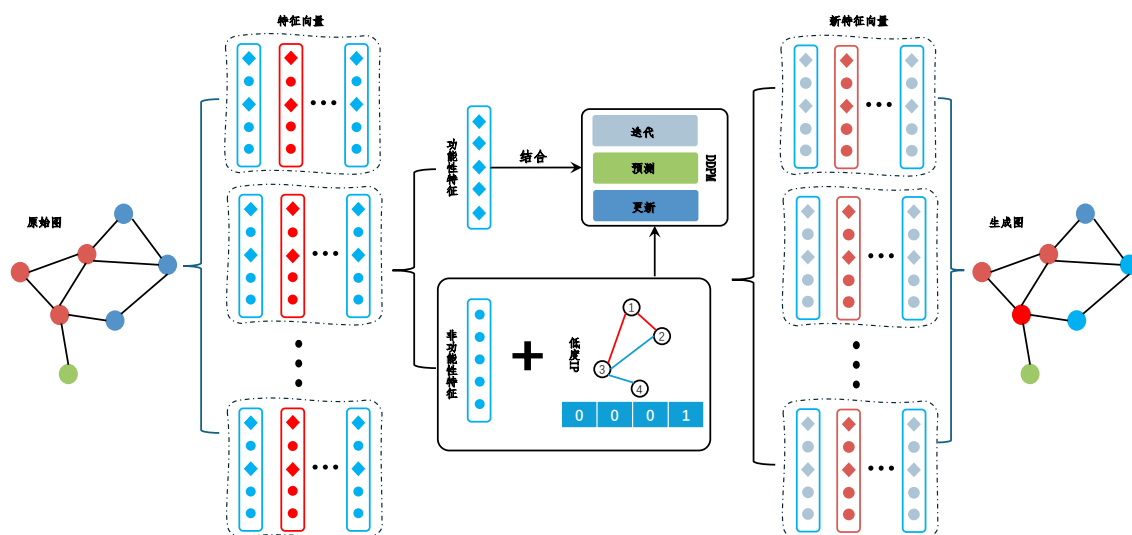


图 3-6 LDNFDDPM

特征通常能够反映攻击流量的核心特征和模式。例如，**源 IP (Src IP)**、**目标 IP (Dst IP)**、**协议类型 (Protocol)**、**目标端口 (Dst Port)** 等都被认为是功能性特征。这些特征在流量分析中具有明确的指示作用，通过它们可以快速区分不同类型的攻击和正常流量。因此，在数据增强过程中，这些特征通常被保留下来，确保生成的流量样本能够尽可能地模拟真实的攻击模式。

另一方面，非功能性特征是指那些对攻击类型的分类影响较小，或者可以通过生成模型推测的特征。例如，流量的**字节数**、**数据包数量**、**流的持续时间**等。虽然这些特征对流量的细节描述很重要，但它们更多地与网络状态和时序行为相关，不一定直接影响攻击的识别。因此，非功能性特征通常会通过生成模型来生成，以增加数据的多样性，尤其是在数据稀缺和标签不平衡的情况下。

为了确保生成的攻击流量数据逼近真实数据，首先需要从实际攻击流量数据中选择出那些最具代表性的功能性特征。为了量化每个特征的贡献，可以采用随机森林算法计算每个特征的重要性。随机森林通过训练多个决策树并综合各个树的输出，能够有效评估每个特征对攻击类型分类的贡献度。随机森林的训练过程可以表示为：

$$\hat{y} = \frac{1}{N} \sum_{n=1}^N h_n(x) \quad (3-6)$$

其中,  $\hat{y}$  是预测值,  $N$  是决策树的数量,  $h_n(x)$  是第  $n$  棵树的预测函数。通过计算每棵树中各特征的贡献度, 可以确定每个特征的重要性。

通过这种方式, 随机森林能够自动识别出在不同攻击类型下最有助于分类的特征。例如, 在 DDoS 攻击中, 源 IP 和目标 IP 可能具有较高的重要性, 而在网络嗅探攻击 (如 Reconnaissance) 中, 协议类型和目标端口可能会更加重要。因此, 在特征选择过程中, 随机森林算法能够帮助选择出最重要的功能性特征, 确保生成的流量数据更具代表性和准确性。

对于非功能性特征, 由于其在流量分类中的作用较小, 可以通过 DDPM 等生成模型进行数据增强。DDPM 通过在多个步骤中逐渐去噪, 能够生成与原数据相似的样本, 特别是在数据稀缺和标签不平衡的情况下, DDPM 能够有效扩展训练数据集。具体而言, DDPM 模型在每个时间步骤  $t$  上加入噪声, 并学习如何从噪声中恢复原始数据:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon \quad (3-7)$$

其中,  $x_0$  是原始数据,  $x_t$  是加噪后的数据,  $\varepsilon$  是从标准正态分布中采样的噪声,  $\bar{\alpha}_t$  是与噪声扩散相关的预定义时间表。

在训练过程中, DDPM 通过优化以下目标函数来学习去噪过程:

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0, t} [\|\varepsilon_\theta(x_t, t) - \varepsilon\|^2] \quad (3-8)$$

其中,  $\varepsilon_\theta(x_t, t)$  是去噪网络的输出,  $\varepsilon$  是真实的噪声。通过这种方式, DDPM 能够生成逼近真实攻击流量的非功能性特征, 并扩充数据集。

在生成过程中, 功能性特征和非功能性特征需要分别进行处理。对于功能性特征, 保留原始的数值, 而对于非功能性特征, 利用 DDPM 进行生成。具体来说, 每个攻击类型会对应一组特征, 其中功能性特征被保留下来, 非功能性特征则通过 DDPM 生成并填补。以 DoS 攻击为例, 假设其功能性特征为:

$$x_f = [\text{Src IP} \text{ Dst IP} \text{ Protocol} \text{ Dst Port}] \quad (3-9)$$

而非功能性特征可以通过 DDPM 生成：

$$x_{nf} = [\text{Flow Duration} \text{Packet Count} \text{Byte Count} \dots] \quad (3-10)$$

DDPM 模型通过采样得到的非功能性特征将与保留下来的功能性特征结合，形成最终的流量特征向量。这样，生成的恶意流量不仅具备攻击流量的核心特征，还在非功能性特征上具有多样性，从而增强了生成数据的真实感和多样性。

在流量图中，低度 IP 通常表示与其他 IP 的通信较少，可能代表某些异常或攻击行为。为了更好地处理这部分流量数据，可以根据 IP 的度数（连接数）来识别低度 IP。具体而言，对于某一 IP  $i$ ，其度数  $d_i$  可以通过以下公式计算：

$$d_i = \sum_j \mathbb{I}(e_{ij}^f) \quad (3-11)$$

其中， $\mathbb{I}(e_{ij}^f)$  为指示函数，表示在图快照  $f$  中，节点  $i$  和节点  $j$  是否有边相连。当  $d_i$  小于某个阈值  $T$  时，该 IP 被视为低度 IP。

生成过程中，针对低度 IP 的流量，特别加以关注。通过构建低度 IP 的掩码，并在生成过程中对这些掩码位置进行特征生成，能够增加模型对低度 IP 流量的敏感性，进一步增强数据的真实性。

通过特征的功能性与非功能性划分，结合 DDPM 生成非功能性特征，可以有效扩充恶意攻击流量的数据集，尤其是在面对数据不平衡和标签稀缺的情况下，数据增强方法能够显著提高模型的性能。功能性特征的保留确保了生成的数据具有代表性，而非功能性特征的生成则增加了数据的多样性，为恶意流量检测提供了更多的信息和支持。

**算法 3-2** 低度 IP 处理逻辑**Input:** 图数据集 graph**Output:** 处理后的图数据集 data

```

1 读取预处理的数据，构建图结构;
2 for 每个节点  $i$  do
3   | 计算节点  $i$  的度数;
4   | if 度数小于阈值  $T$  then
5   |   | 标记节点  $i$  为低度 IP;
6   | end
7 end
8 生成低度 IP 掩码;
9 for 每个低度 IP do
10  | 进行特征生成和加权平均;
11 end
12 更新图数据集;
13 return 处理后的图数据集 data;

```

### 3.3 本章小结

本章提出了一种基于 LDNFDDPM 的数据增强方法，旨在解决恶意流量检测中的数据不平衡问题。通过对流量数据进行清洗、标签转换、特征提取和标准化处理，结合低度 IP 节点的增强策略和功能性、非功能性特征的划分，提出了一个改进的去噪扩散概率模型。该模型通过引入时序图捕捉流量数据的时间依赖性，增强了对稀有攻击类型和低度 IP 节点的生成能力，从而提高了模型的鲁棒性和生成样本的质量，为网络流量检测提供了更为有效的数据增强方案。

## 第四章 基于多头自适应图注意力网络的线图训练方法

本章提出了一种基于自适应注意力网络的线图训练方法。首先，利用第三章提出的数据增强技术对离线训练数据进行处理，并采用线图结构来更有效地捕捉边的特征，从而提升模型的表达能力。接着，引入自适应多头注意力机制，在图神经网络（GNN）中通过动态调整节点间的注意力权重，实现更精确的消息传递和特征聚合。这一机制帮助模型更好地捕捉节点间的复杂关系，进一步增强其表达能力。最后，采用 MAD（中位数绝对偏差）指标，评估数据集中各个数据点与其均值之间的偏差，分析图神经网络在不同层次上的特征分布变化，尤其是邻居节点和远程节点之间的特征差异。

### 4.1 问题描述和研究思路

现如今，图神经网络（Graph Neural Networks, GNNs）已经成为处理图数据的强大工具，广泛应用于社交网络分析、推荐系统、化学分子建模、网络安全等领域。图结构数据的独特性使得传统的深度学习方法（如卷积神经网络、循环神经网络等）难以直接应用，因而推动了图神经网络的发展。图神经网络通过消息传递机制，使节点能够根据图的拓扑结构和节点特征与邻居节点交换信息，从而逐层更新节点的表示。这种机制对于大多数图数据任务，如节点分类、图分类和链接预测等，表现出了较强的能力。

然而，随着图数据的复杂性逐步增加，尤其是当图结构规模扩大或者图中的节点和边表现出更复杂的关系时，现有的图神经网络方法常常面临一系列挑战。首先，图中节点间的关系并非是静态且均匀的。在实际应用中，图结构往往包含多种类型的边，且不同的边在图中对节点的影响和重要性有所不同。许多传统的图神经网络方法通常通过固定的消息传递规则，假定所有边对节点的影响是相等的，这使得它们难以捕捉图中复杂的边的特征和节点之间的不同关系。

此外，节点间的关系不仅仅是邻近节点的作用，图中的远程节点之间也可能存在潜在的联系。这些远程节点的特征可能对于节点的表示至关重要，但在传统的图神经网络中，随着图的层次加深，远程节点的影响会逐渐被忽视，导致模型无法有效地捕捉长距离的依赖关系。为了克服这一问题，通常需要更为复杂的网络架构和更精细的信息传播策略。

再者，现有的图神经网络方法往往依赖于图的邻接矩阵或边列表来表示图结构，但这也意味着图的边的特征需要被简单地表示为固定的权重或标签，未能充

分利用边之间的高阶关系。传统的图神经网络方法通常聚焦于节点的邻居信息，并在每一层将节点特征与邻居节点的特征进行聚合，但这一过程忽略了边与边之间的潜在复杂关系，限制了模型的表达能力。

这些问题表明，现有的图神经网络方法存在一些局限性，需要新的方法来更好地捕捉节点和边之间的复杂关系，尤其是在节点间关系动态变化的情况下。因此，如何优化图神经网络中的消息传递机制，提高模型对边特征的敏感度，以及如何改进远程节点之间信息的传递与融合，成为了提升图神经网络性能的关键问题。本章提出了一种基于多头自适应注意力网络的线图训练方法（Training Method for Line Graphs Based on Multi-Head Adaptive Attention Mechanism, LG-MHAAM），旨在通过创新的结构与机制来克服现有方法的局限，从而优化图神经网络的性能。

LG-MHAAM 总体结构图如图 4-1 所示，首先，在扩充的图数据集基础上，我们构建了线图结构。线图结构能够更好地捕捉图中节点之间的关系，从而提高模型的表达能力。通过将原始图转换为线图，我们能够更细致地分析节点之间的交互关系，这对于复杂网络数据的处理尤为重要。

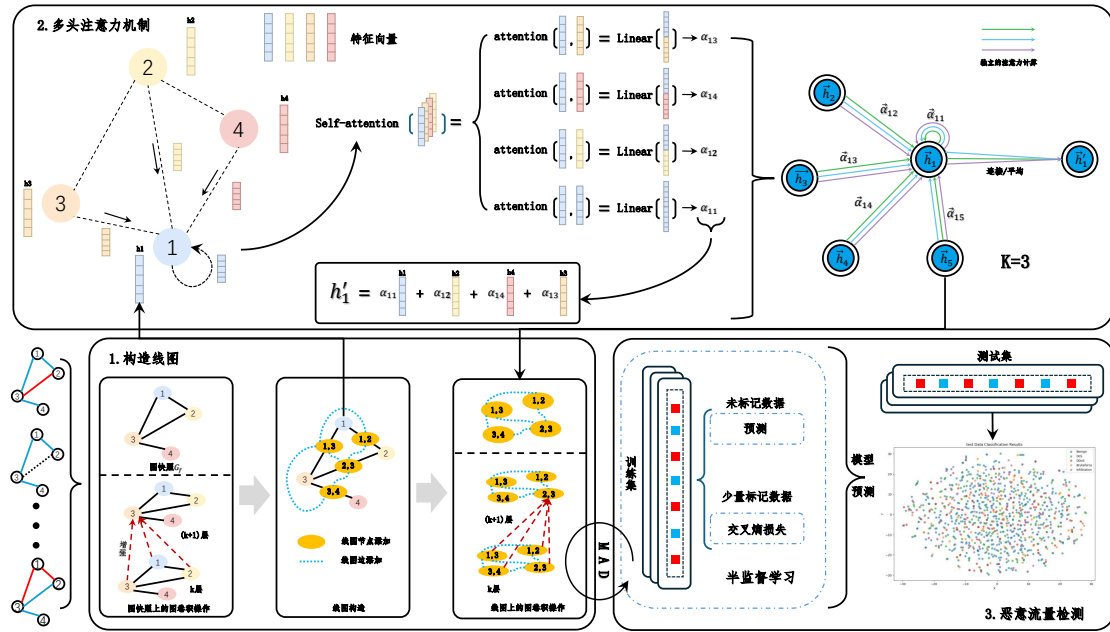


图 4-1 基于多头自适应注意力网络的线图训练模型

然后，我们引入了多头注意力网络。多头注意力机制通过同时关注图中不同部分的信息，能够更有效地捕捉复杂的图结构特征，提高模型的分类性能。具体来说，多头注意力机制允许模型在不同的子空间中独立地学习节点之间的关系，从而增强模型的表达能力和鲁棒性。

为了进一步提升模型的性能，我们还引入了 MAD-GAP 正则化器（MAD-GAP

Regularizer)。MAD-GAP 正则化器通过在训练过程中对注意力权重进行正则化，能够有效防止模型过拟合，并提高模型的泛化能力。具体来说，MAD-GAP 正则化器通过在注意力权重上施加约束，防止模型过度依赖某些特定的节点或边，从而增强模型的鲁棒性和稳定性。

在此基础上，我们构建了一个半监督学习模型。半监督学习能够利用少量标注数据和大量未标注数据进行训练，从而在标注数据有限的情况下仍能取得较好的分类效果。通过结合有监督和无监督学习的优势，半监督学习模型能够更好地利用数据中的信息，提高分类的准确性和泛化能力。

## 4.2 框架设计

### 4.2.1 基于 LineGraph 的特征结合方法

为了更有效地分析网络流量中的复杂交互关系，本研究采用线图 (Line Graph) 作为分析工具。线图能够将原始图中的边转换为新图中的节点，并在这些节点之间建立连接，以反映原始图中边的相互关系。这种方法在检测网络攻击和异常流量时具有显著优势 (Harary, F. (1969). Graph Theory. Addison-Wesley Publishing Company)。

线图在恶意流量检测中的作用体现在以下几个方面。首先，在网络攻击中，攻击流量往往通过多个源 IP 发起，边与边之间可能存在复杂的交互。线图将原图中的边转化为节点，捕捉边与边之间的共享节点关系。例如，DDoS 攻击中多个源 IP 向同一目标 IP 发起流量，线图能够有效捕捉这些攻击源之间的高阶依赖关系。

其次，线图将原图中的边特征映射为新节点特征，从而加强了对边特征的学习。举例来说，Brute Force 攻击可能通过特定端口发起，而正常流量则有不同的协议和端口分布，线图能够识别这些特征并提高检测精度。

此外，恶意流量可能跨越多个路径，导致攻击源与目标之间的关系较远。线图增强了对远程节点间关系的建模能力，能够识别例如 DDoS 攻击中分布式流量的协调行为。

线图还能够帮助识别潜在的攻击模式，尤其是当恶意流量依赖于多次交互模式时。例如，持续的流量模式可能通过多次交互形成攻击特征，线图帮助识别这些复杂模式。

最后，恶意流量常常表现为低频或隐蔽的异常行为，线图通过聚合边间的关系，提高了图神经网络在处理这些稀疏数据时的鲁棒性。线图可以帮助识别出在常规流量中难以察觉的低频攻击。

线图  $L(G)$  是从原始图  $G = (V, E)$  构建而来的，其中  $V$  是节点集合， $E$  是边集

合。在线图中，节点集合  $V'$  对应于原始图中的边集合  $E$ ，即：

$$V' = \{e_i \mid e_i \in E\} \quad (4-1)$$

对于任意两条边  $e_i, e_j \in E$ ，如果它们在原始图中共享一个公共节点，则在线图中添加一条边  $(e_i, e_j)$ 。这种边的添加可以通过以下方式表示：

$$E' = \{(e_i, e_j) \mid e_i \cap e_j \neq \emptyset\} \quad (4-2)$$

在该公式中， $e_i \cap e_j$  表示边  $e_i$  和  $e_j$  之间的公共节点。通过这种转换，线图能够将原图中的边之间的关系转化为节点之间的连接，从而揭示流量之间的复杂交互关系，转换过程如图 4-2 所示。

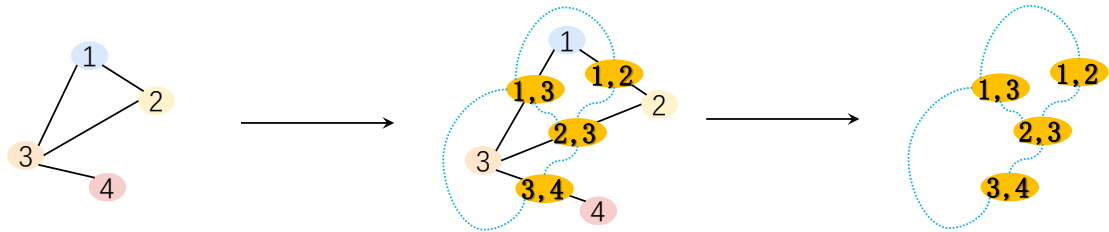


图 4-2 线图转换过程

在构建线图时，首先初始化一个新的边索引列表 `new_edge_index`，该列表包含两个空列表，用于存储线图上的边。同时，初始化一个字典 `ip_pair_newest_index_dict`，用于记录每对 IP 对应的最新流量信息。接下来，遍历原始图中的每个流量（边），获取源 IP (`src_ip`) 和目标 IP (`dst_ip`)。如果源 IP 等于目标 IP，则跳过该流量，因为不需要自连接边；如果源 IP 大于目标 IP，则交换这两个 IP，以确保源 IP 小于目标 IP，从而统一 IP 对的表示形式。然后，生成 IP 对的名称 `ip_pair_name`，格式为 `src_ip - dst_ip`。

在遍历过程中，如果当前流量不是第一个流量，则检查当前流量与之前的流量是否有共同的 IP。遍历 `ip_pair_newest_index_dict` 中的每个过去的 IP 对名称 `past_ip_pair_name`，如果过去的 IP 对名称与当前的 IP 对名称相同，则跳过；否则，获取过去流量的源 IP (`past_src_ip`) 和目标 IP (`past_dst_ip`)，并判断当前流量的源 IP 或目标 IP 是否与过去流量的源 IP 或目标 IP 相同。如果相同，则在 `new_edge_index` 中添加一条新边，连接当前流量和过去的流量。最后，更新 `ip_pair_newest_index_dict`，记录当前流量的最新索引，将当前 IP 对的名称作为键，源 IP、目标 IP 和索引作为值存储在字典中。



最终，返回包含新边索引的张量 `torch.tensor(new_edge_index)`。详细的线图转换算法如算法 4-1 所示。

**算法 4-1** 将原始图转换为线图

**Input:** 原始图  $G = (VE)$ ，其中  $V$  是节点集合， $E$  是边集合  
**Output:** 线图  $L(G)$

```

1  $L(G) \leftarrow$  空图;
2 foreach 边  $e = (uv) \in E$  do
3   | 在  $L(G)$  中添加一个节点  $e$ ;
4 end
5 foreach 一对边  $e_1 = (u_1v_1)$  和  $e_2 = (u_2v_2)$ ，满足  $v_1 = u_2$  do
6   | 在  $L(G)$  中添加一条边  $(e_1e_2)$ ;
7 end
8 return  $L(G)$ ;
    
```

#### 4.2.2 基于多头自适应注意力机制的图卷积方法

在线图转换后，由于节点和边的数量急剧增加，如图 4-3 所示。传统的图神经网络（GNN）面临着巨大的计算资源压力。在这种情况下，为了提高训练效率和模型的表达能力，结合多头注意力机制和均值绝对差（MAD）正则化提供了一个有效的解决方案。多头注意力机制能够有效捕捉节点间的复杂依赖关系，而 MAD 正则化则能够抑制特征间的噪声，提升信息传递的稳定性。两者的结合不仅缓解了计算负担，还增强了模型的鲁棒性和表达能力。

多头注意力机制通过并行计算多个注意力头，分别学习不同的特征子空间，并将这些子空间的结果进行拼接，最终生成更加丰富的节点表示。在图神经网络中，多头注意力可以用来动态调整节点之间的信息传递强度。每个注意力头能够独立地关注不同的邻居节点，学习到不同的依赖关系，这使得信息传递变得更加多样化，从而增强了模型的表达能力。

具体地，多头注意力的计算可以通过以下公式表示：

$$\text{MultiHead}(\mathbf{h}) = \text{Concat}(\text{head}_1 \dots \text{head}_K) \mathbf{W}^O \quad (4-3)$$

其中， $\text{head}_i = \text{Attention}(\mathbf{h})$  是第  $i$  个注意力头的输出， $K$  是注意力头的数量， $\mathbf{W}^O$  是输出的线性变换矩阵。每个注意力头通过自注意力机制计算节点间的相关性，然后将这些输出拼接起来，进一步通过线性变换生成最终的节点表示。

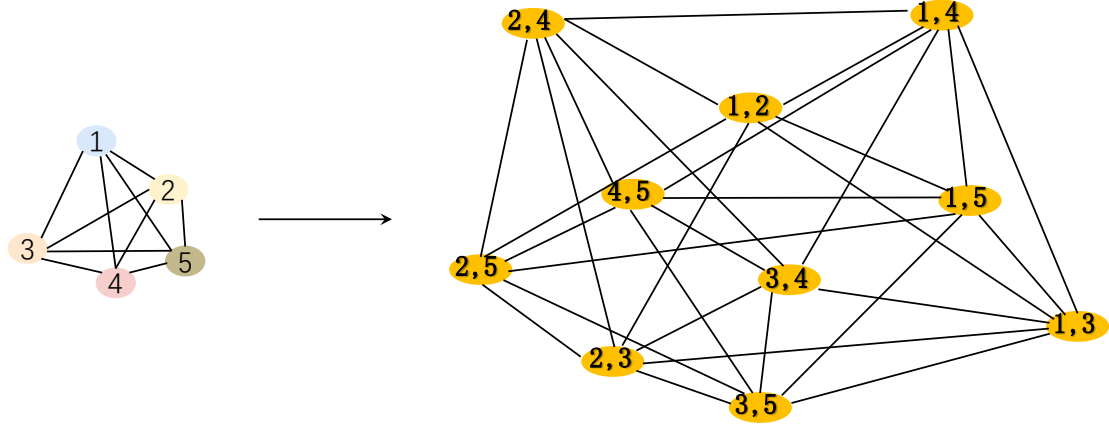


图 4-3 线图转换规模增加

多头注意力机制可以有效减少冗余计算，并且能够从多个角度捕捉图的结构信息。这在处理大规模线图时尤为重要，因为多个注意力头可以并行计算，减轻了计算负担并提高了模型的学习能力。

然而，尽管多头注意力机制能够有效捕捉到多样化的节点特征，在线图中，由于节点特征之间的差异较大，信息传递过程中可能会存在较大的噪声，尤其是在处理具有高复杂度的图时。这种噪声会导致节点特征的过度偏移，进而影响模型的收敛性和最终性能。为了减少这种影响，结合 MAD 正则化是一种有效的策略。

MAD（均值绝对差）是一种衡量数据离散程度的统计量，能够有效地约束节点特征在信息传播过程中的偏差。在节点特征的聚合过程中，MAD 通过计算特征的离散程度，动态调整每个节点的权重，确保特征传播更加稳定。具体地，MAD 的计算方式为：

$$\text{MAD}(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \mu| \quad (4-4)$$

其中， $x_i$  是节点  $i$  的特征， $\mu$  是该节点特征的均值， $n$  是节点数。通过引入 MAD 正则化，可以在特征聚合过程中减少不重要节点特征的影响，使得信息传递更加精确。

结合多头注意力机制与 MAD 正则化，可以在计算复杂图结构时有效减少计算负担并提高模型的鲁棒性。多头注意力机制通过并行处理多个不同的注意力视角，能够捕捉到节点特征之间的多样化依赖关系；而 MAD 正则化通过对特征的离

散程度进行约束，抑制了噪声的影响，确保信息传递过程中更加稳定。

具体来说，节点特征更新公式可以表示为：

$$h'_i = \sigma \left( \sum_{j \in \mathcal{N}(i)} \frac{A_{ij}h_j}{d_j} + \beta \cdot \text{MAD}(h_i) \right) \quad (4-5)$$

其中， $h_i$  是节点  $i$  的特征， $\mathcal{N}(i)$  是节点  $i$  的邻居节点集合， $A_{ij}$  表示节点  $i$  和节点  $j$  之间的邻接矩阵元素， $d_j$  是节点  $j$  的度数， $\beta$  是用于调节 MAD 正则化项影响的超参数。这个公式结合了邻接矩阵的加权聚合和 MAD 正则化项，确保了节点特征的稳定更新。

通过这种结合，多头注意力机制能够在处理大规模图时并行计算不同的注意力视角，而 MAD 正则化则帮助减少特征差异带来的噪声影响，增强了信息传递的稳定性。这种方法不仅提高了计算效率，还提升了模型对复杂图结构的表达能力。

在线图转换后的节点数和边数急剧增加，带来了计算资源的挑战。结合多头注意力机制和 MAD 正则化，不仅能够减轻计算负担，还能有效提升模型的鲁棒性和表达能力。多头注意力机制通过并行计算多个注意力头，捕捉不同的节点依赖关系，增强了图的表达能力；而 MAD 正则化则通过约束节点特征的离散度，抑制了噪声的影响，确保了信息传递的精度和稳定性。两者的结合提供了一种有效的解决方案，使得图神经网络在处理大规模、复杂图数据时能够高效地进行训练，并且具有较强的泛化能力。

### 4.3 本章小结

本章提出了一种基于多头自适应图注意力网络的线图训练方法（LG-MHAAM），通过利用自适应注意力机制和线图结构提升图神经网络（GNN）处理复杂图数据的能力。首先，结合第三章的数据增强技术，构建了线图结构，使得模型能够更有效地捕捉边的特征。应用自适应多头注意力机制，在图神经网络中动态调整节点间的注意力权重，实现更加精确的信息传递与特征聚合。引入 MAD（中位数绝对偏差）指标评估不同层次上节点特征的变化，分析模型在处理不同类型节点和边时的鲁棒性。通过该方法，能够克服现有图神经网络的局限性，特别是在节点间关系动态变化和远程节点特征传递方面。

## 第五章 实验结果分析

### 5.1 实验数据及环境

#### 5.1.1 实验环境

本章所提的 CLDA 方法在配备 3.50GHz 的 12 核 Intel(R)Core(TM)i9-10920X CPU、256 GB 内存和 Linux 内核 v.5.11.0 的 Ubuntu 20.04 LTS 操作系统上进行评估。选用 PyTorchv1.7.0 与 Jupyter notebook 一起实现相关实验，具体实验设置如表 5-1 所示。

表 5-1 硬件配置

名称	配置
CPU	i9-10920X
GPU	技嘉 RTX3080
内存	镁光 256GB 3200HZ 8
硬盘	三星 2TB 2

#### 5.1.2 实验数据

本文采用 NF-BoT-IoT-V2、NF-ToN-IoT-V2、NF-CSE-CIC-IDS2018-V2 三个数据集<sup>[43]</sup>，详细介绍如下：

NF-BoT-IoT-V2 数据集是基于原始 BoT-IoT 数据集生成的，BoT-IoT 数据集最初由澳大利亚国立计算机安全中心（ACCS）实验室创建。通过使用 nProbe 提取原始 PCAP 数据中的 43 个 NetFlow 特征，生成了该数据集。NF-BoT-IoT-V2 包含 37,763,497 条流量数据，其中良性流量比例仅为 0.36%，其余均为攻击流量。该数据集覆盖了 293 个 IP 地址和 65,536 个端口，平均流持续时间为  $3.999 \times 10^6$  毫秒。它的入站和出站流量字节数较低，分别为 546 和 204。其主要支持的协议类型包括 ICMP、TCP 和 UDP。NF-BoT-IoT-V2 的攻击类型涵盖了探测攻击、DDoS 攻击、DoS 攻击以及数据盗窃。作为物联网流量分析的一个重要数据集，NF-BoT-IoT-V2 在攻击类型和规模上具有显著的失衡性，适合用于异常检测和攻击分类的研究。

NF-ToN-IoT-V2 数据集是 ToN-IoT 数据集的扩展版本，最初由澳大利亚国防部数据 61 实验室开发，旨在模拟真实物联网环境中的网络流量。该数据集包含 16,940,495 条网络流量，其中良性流量占 36.01%。与其他物联网数据集相比，NF-ToN-IoT-V2 的流量数据覆盖范围较广，包括 29,245 个 IP 地址和 65,536 个端口，支持的协议类型包括 ICMP、TCP、UDP 和 IPv6-ICMP。平均流持续时间为

表 5-2 NF-BoT-IoT-V2 数据集主要属性

属性	描述
流量总数	37, 763, 497
良性流量比例	0.36%
IP 地址数	293
端口数量	65, 536
平均流持续时间	$3.999 \times 10^6$ 毫秒
平均输入字节数	546
平均输出字节数	204
TCP 标志数量	25
支持协议类型	ICMP、TCP、UDP
攻击类别	Reconnaissance, DDoS, DoS, Theft

$7.929 \times 10^5$  毫秒，入站和出站流量字节数分别为 726 和 837。NF-ToN-IoT-V2 中的攻击类型与 NF-BoT-IoT-V2 类似，主要包括探测攻击、DDoS 攻击、DoS 攻击和数据盗窃，如表 3-3 所示。该数据集具有更高的良性流量比例和丰富的流量特征，适合于物联网环境下的多协议分析和跨网络的异常检测研究。

表 5-3 NF-ToN-IoT-V2 数据集主要属性

属性	描述
流量总数	16, 940, 495
良性流量比例	36.01%
IP 地址数	29, 245
端口数量	65, 536
平均流持续时间	$7.929 \times 10^5$ 毫秒
平均输入字节数	726
平均输出字节数	837
TCP 标志数量	34
支持协议类型	ICMP、TCP、UDP、IPv6-ICMP
攻击类别	DDoS, DoS, Backdoor, Injection, Password, Scanning, XSS

NF-CSE-CIC-IDS2018-V2 数据集基于原始 CICIDS2018 数据集生成，原始数据集由加拿大网络安全实验室（CIC）和通信安全研究中心（CSE）联合创建，旨在模拟现代企业网络中的攻击行为。该数据集包含 18, 893, 708 条网络流量，其中良性流量比例为 11.95%。NF-CSE-CIC-IDS2018-V2 拥有 255, 042 个独立 IP 地址和 65, 492 个端口，协议类型覆盖范围广，包括 ICMP、TCP、UDP、IPv6-ICMP 和 GRE。平均流持续时间为  $5.746 \times 10^5$  毫秒，入站和出站字节数分别为 1, 812 和 6, 913。该数据集的攻击类别多样化，包括暴力破解、僵尸网络、DoS 攻击、DDoS 攻击和渗透攻击等，如表 3-4 所示。NF-CSE-CIC-IDS2018-V2 数据集的特点

在于其流量特征的多样性和复杂性，是研究企业网络安全威胁检测的重要资源，尤其适用于多种攻击类型的分类和流量分析任务。

表 5-4 NF-CSE-CIC-IDS2018-V2 数据集主要属性

属性	描述
流量总数	18, 893, 708
良性流量比例	11.95%
IP 地址数	255, 042
端口数量	65, 492
平均流持续时间	$5.746 \times 10^5$ 毫秒
平均输入字节数	1, 812
平均输出字节数	6, 913
TCP 标志数量	50
支持协议类型	ICMP、TCP、UDP、IPv6-ICMP、GRE
攻击类别	BruteForce, Bot, DoS, DDoS, Infiltration

## 5.2 实验设计

本章实验工分为两个部分：数据增强方法测试、恶意流量检测方法测试。采用 FID（Frechet Inception Distance）和 IS（Inception Score）等指标来评估数据增强效果，采用准确率、召回率和 F1-score 等指标来评估检测效果。

### 5.2.1 评估指标

混淆矩阵用于分类任务中，表示模型在不同类别上的预测结果。通过混淆矩阵可以计算多个关键指标，包括准确率、精确率、召回率和 F1-Score 等。二分类任务中的混淆矩阵结构如下：

	预测为正类	预测为负类
真实为正类	TP (True Positive)	FN (False Negative)
真实为负类	FP (False Positive)	TN (True Negative)

其中：TP：真正例，表示模型正确地将正类样本分类为正类。FP：假正例，表示模型错误地将负类样本分类为正类。TN：真负例，表示模型正确地将负类样本分类为负类。FN：假负例，表示模型错误地将正类样本分类为负类。

#### （1）准确率 (Accuracy)

准确率是模型正确分类的样本占总样本的比例，计算公式如 5-1 所示：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5-1)$$

(2) 精确率 (Precision)

精确率衡量的是所有被模型预测为正类的样本中，真正为正类的比例，计算公式如 5-2 所示：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5-2)$$

(3) 召回率 (Recall)

召回率衡量的是所有真实为正类的样本中，模型正确预测为正类的比例，计算公式如 5-3 所示：

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5-3)$$

(4) F1-Score

F1-Score 是精确率和召回率的调和平均数，计算公式如 5-4 所示：

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \times \frac{TP}{2TP + FP + FN} \quad (5-4)$$

(5) 误报率 (False Positive Rate)

误报率衡量的是所有负类样本中，模型错误预测为正类的比例，计算公式为：

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN} \quad (5-5)$$

(6) FID

FID 是评估生成模型（如生成对抗网络 GAN）的生成样本与真实样本之间的相似度的指标。FID 的原理是通过比较两个分布（真实数据和生成数据）在某一预训练网络（如 Inception v3）上的特征表示来衡量它们之间的差距。

## 5.2.2 相关对比方法

根据本文的研究方法和实验数据，针对数据增强和恶意流量检测部分，分别选择了近年来的优秀工作进行对比。为了评估 LDNFDDPM 在数据增强方面的有

效性，将其与现有的几种数据增强方法进行了比较，具体如下：

**G-IDS<sup>[44]</sup>**：G-IDS 是一种利用基本的 GAN 生成流量数据的方法，旨在解决网络物理系统中数据不平衡问题。通过生成合成流量数据，G-IDS 可以增加少数类（如攻击流量）的样本量，从而提高入侵检测系统（IDS）的检测性能。虽然其生成的数据能够缓解数据不平衡的问题，但由于使用的是基本的 GAN，生成的数据可能存在一定的偏差和不足。

**SIGMA<sup>[45]</sup>**：SIGMA 采用随机森林方法对特征进行分割，保留重要特征，然后使用基本的 GAN 生成对抗样本来训练 IDS。通过这种方式，SIGMA 能够有效地提高 IDS 的检测能力，尤其是在数据不平衡的情况下。尽管生成的数据能够增强模型的训练集，但由于 GAN 本身的局限性，生成的数据可能不能完全捕捉到所有攻击流量的特征。

**CWGAN-GP<sup>[46]</sup>**：CWGAN-GP（条件 Wasserstein 生成对抗网络与梯度惩罚）是一种先进的深度生成对抗网络模型，专门用于处理不平衡数据问题。它通过引入条件信息来指导生成器生成特定类别的样本，从而在生成过程中更好地控制样本的类别分布。CWGAN-GP 采用 Wasserstein 距离作为生成器和判别器之间的距离度量，相较于传统的 GAN 模型，能够更稳定地训练并生成高质量的样本。此外，该模型还引入了梯度惩罚机制，以替代传统的权重裁剪方法，进一步提高了模型的训练稳定性和生成样本的多样性。CWGAN-GP 在生成合成数据以增强少数类样本方面表现出色，被广泛应用于图像生成、数据增强以及不平衡数据分类等领域，为解决数据不平衡问题提供了有效的解决方案。

**SEMRes-DDPM<sup>[47]</sup>**：SEMRes-DDPM 是一种创新的过采样方法，专门针对不平衡表格数据的分类问题。它基于去噪扩散概率模型（DDPM），通过引入一种新颖的神经网络结构 SEMST-ResNet 来实现高效的去噪和数据合成。SEMST-ResNet 结合了多头自注意力机制和软阈值处理，能够有效去除噪声并提取原始数据的特征，从而生成更接近真实数据分布的少数类样本。

为了评估 LG-MHAAM 在恶意流量方面的检测性能，选取了现有的几种物联网恶意流量检测方法进行比较，具体如下：

**E-GraphSAGE<sup>[48]</sup>**：E-GraphSAGE 是一种基于图神经网络（GNN）的网络入侵检测方法。它通过同时捕捉流量数据中的边特征和拓扑信息来进行入侵检测。与传统的基于节点特征的 GNN 方法不同，E-GraphSAGE 强调了图中节点间的关系和数据流动方式，适用于具有复杂网络结构的入侵检测任务。通过从邻居节点聚合信息，E-GraphSAGE 能够有效捕捉到流量中的时序依赖和复杂的拓扑特征，在处理图结构数据时，表现出较强的性能。



Extra Trees<sup>[49]</sup>: Extra Trees 是一种基于决策树的集成学习方法，它通过构建多个极端随机树来进行分类。与随机森林类似，Extra Trees 使用了多个决策树的投票结果来进行决策，但它在训练过程中对特征的选择和切分点的选择引入了更多的随机性。这种方法在数据特征较多时，能够提高训练速度，并且具有较强的鲁棒性和较低的过拟合风险。Extra Trees 通常在处理大规模数据时表现出色，尤其是在数据存在噪声时。

Random Forest (RF)<sup>[50]</sup>: 随机森林是一种广泛使用的集成学习方法，通过建立多棵决策树来提高模型的准确性。每棵决策树都是在数据的不同子集和特征子集上训练的，因此它能够有效减少过拟合现象。随机森林的优点包括对异常值和噪声的鲁棒性强，能够自动评估特征的重要性，因此适用于处理具有高维度特征的数据。它在分类任务中，尤其是流量数据分类中表现良好，能够快速高效地处理大规模数据集。

Deep Feedforward Neural Network (DNN)<sup>[50]</sup>: 深度前馈神经网络（DNN）是一种基础的神经网络模型，由多层全连接层组成，每一层通过激活函数将输入数据映射到新的空间。DNN 能够捕捉数据中的复杂非线性关系，尤其适用于处理高维度的结构化数据。在入侵检测任务中，DNN 能够从原始特征中自动学习出潜在的模式，从而提高分类性能。尽管 DNN 对计算资源的需求较高，但在数据量充足时，它能够通过深度学习方式处理更加复杂的特征关系。

Transductive Support Vector Machine (TSVM)<sup>[51]</sup>: TSVM 是一种半监督学习方法，旨在通过最大化标签数据和未标签数据之间的间隔来提高分类性能。它通过引入正则化项，约束分类超平面的选择，使得分类边界尽量远离未标记的数据，从而在数据稀缺的情况下实现更好的泛化能力。TSVM 在流量分类中的优势在于其能够利用少量的标记数据，同时结合大量的未标记数据进行训练，提高模型的鲁棒性和准确性，尤其在标签数据有限的情况下，TSVM 展现出了较强的优势。

### 5.3 数据增强方法测试

数据增强方法的实验评估共分为两个部分：对比实验和消融实验。在对比实验中，为了评估 LDNFDDPM 和几种不同的数据增强方法的性能，统计每种数据增强方法的 FID 值，具体来书每次随机抽取每类攻击的 10 份生成样本计算 FID 值，一共进行十次，统计 FID 平均值，最后利用每类生成样本的 FID 值来得到该数据增强方法生成样本的凭据 FID 值，此外，为了进一步探究生成数据的真实可用性，选用一个学习原始数据得到 MLP 模型来测试生成的样本数据。。在消融实验中，测试划分功能性特征于非功能性特征以及关注游离 IP 策略对于性能的影响。

## 1. 对比实验

图 5-1 显示了不同的数据增强方法在不同的数据集上的 FID 值，而表 5-5、表 5-6、5-7 和 5-8 显示了不同数据增强算法生成数据在原始数据训练得到的模型的性能表现。

(1) 从整体来看，对于三个数据集上的 FID 表现来说，NF-BoT-IoT-V2 略高于 NF-CSE-CIC-IDS2018-V2，NF-CSE-CIC-IDS2018-V2 略高于 NF-ToN-IoT-V2。这是因为由于其攻击流量比例极高（99.64% 是攻击流量），且攻击类型较为简单，生成模型可能较容易生成攻击流量，但难以生成正常流量。虽然生成模型会聚焦于攻击流量的特征，但可能忽略了正常流量的多样性，这导致生成数据质量较低，FID 值较高。而 NF-ToN-IoT-V2 良性流量比例相对较高（36.01%），且该数据集包含多种协议（包括 IPv6-ICMP）。生成模型在生成攻击流量时的难度相对较大，但正常流量的特征较为明显，有助于生成模型较好地捕捉这些特征，因此 FID 值较低。对于 NF-CSE-CIC-IDS2018-V2，该数据集复杂度最高，攻击类型多样，且协议种类丰富。生成模型在处理此类数据时需要学习到更复杂的特征，因此其 FID 处于两者之间。

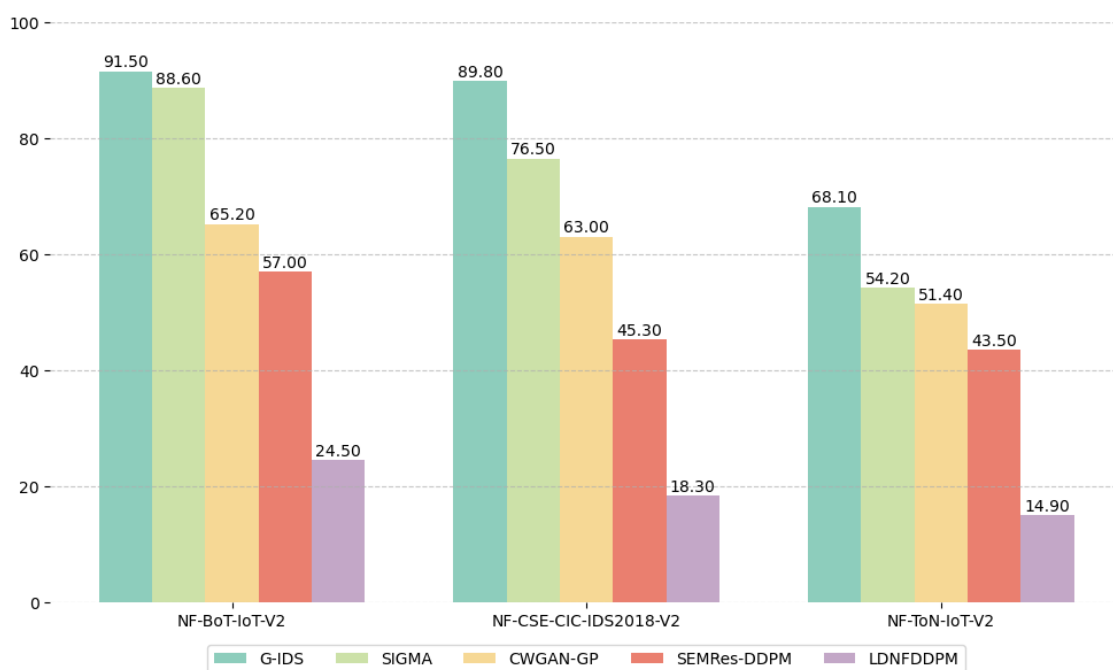


图 5-1 不同方法数据增强的效果表现

(2) 从图 5-1 很容易得出结论，使用传统 GAN（G-IDS、SIGMA）生成的数据，都表现出了较大的 FID 值，并且在表 5-5、5-6、5-7、5-8 中的各类别检测中表现出较低的 Rec 值，说明其生成的数据于真实数据偏差较大。这是因为传统 GAN

在训练阶段只能从有限的样本中学习到有限的样本特征，并且模型在有限数据下难以收敛。优化后的 GAN（CWGAN-GP）在表 5-7 和表 5-8 中生成数据的检测的表现要优于 G-IDS，但图 5-1 中 FID 值仍然较高。这表明，在训练 GAN 时引入条件信息指导生成器生成特定类别的样本在一定程度上可以提高 GAN 性能。

（3）与其他类型的数据增强方法相比，基于 DDPM 的方法（SEMRes-DDPM、LDFDDPM）的 FID 值明显更低。这是因为 DDPM 通过逐步去噪的过程生成数据，避免了 GAN 中的模式崩溃和训练不稳定的问题，同时能够生成更加多样且高质量的样本。DDPM 不依赖于生成器和判别器的对抗博弈，而是通过最大化数据似然进行训练，这使得它在数据稀缺或复杂分布的情况下表现得更加鲁棒和可靠。此外，与 SEMRes-DDPM 相比，LDFDDPM 具有更低的 FID 值和同类型数据集下更高的 Rec。这是因为 LDFDDPM 只生成样本中的非功能特征部分，并且将游离 IP 增加权重，不仅进一步提升了样本的真实性，还降低了数据训练的难度。

表 5-5 NF-BoT-IoT-V2 数据集评估结果

Method	DDoS		DoS		Reconnaissance		Theft	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
G-IDS <sup>[44]</sup>	60.25	64.57	68.32	61.21	55.12	78.16	48.56	42.12
SIGMA <sup>[45]</sup>	72.14	67.54	80.34	74.13	67.10	71.24	52.12	45.79
CWGAN-GP <sup>[46]</sup>	81.46	86.42	89.42	82.78	86.34	79.84	50.78	44.31
SEMRes-DDPM <sup>[47]</sup>	90.24	81.91	81.42	85.54	88.02	72.63	53.45	47.46
<b>LDFDDPM</b>	<b>93.68</b>	<b>89.47</b>	<b>92.01</b>	<b>86.45</b>	<b>88.56</b>	<b>83.56</b>	<b>59.12</b>	<b>60.34</b>

表 5-6 NF-CSE-CIC-IDS2018-V2 数据集评估结果

Method	DDos		Dos		Bot		BruteForce		Infiltration	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
G-IDS <sup>[44]</sup>	66.42	62.18	54.72	59.01	68.05	55.14	77.98	64.98	63.56	66.23
SIGMA <sup>[45]</sup>	77.58	74.63	76.12	62.28	68.78	67.11	78.64	66.67	74.98	79.48
CWGAN-GP <sup>[46]</sup>	77.10	73.67	75.68	71.42	78.42	76.32	78.28	75.84	64.45	68.56
SEMRes-DDPM <sup>[47]</sup>	88.05	93.56	91.78	93.54	92.14	88.24	89.02	90.59	94.68	91.24
<b>LDFDDPM</b>	<b>98.48</b>	<b>96.32</b>	<b>97.10</b>	<b>94.17</b>	<b>99.56</b>	<b>98.96</b>	<b>99.32</b>	<b>98.31</b>	<b>96.02</b>	<b>91.95</b>

## 2. 消融实验

此次实验的主要目的是测试划分功能/非功能特征以及关注游离 IP 行为对于数据增强效果的影响。本节基于此设计了两种变体：LDDPM 和 NFDDPM。

LDDPM：不对特征进行功能性、非功能性划分，直接进行加噪去噪过程。

NFDDPM：不对游离 IP 做额外处理。

各变体在不同数据集上的实验结果如表 5-5 所示，可以看出，LDFDDPM

表 5-7 NF-ToN-IoT-V2 数据集评估结果

Method	DDoS		DoS		Backdoor		XSS	
	Pre	Rec	Pre	Rec	Pre	Rec	Pre	Rec
G-IDS <sup>[44]</sup>	60.15	67.54	57.32	32.34	65.86	64.23	70.23	45.12
SIGMA <sup>[45]</sup>	72.30	79.73	70.42	66.37	67.08	65.54	63.47	78.21
CWGAN-GP <sup>[46]</sup>	81.42	88.56	89.18	84.52	76.32	84.72	82.15	76.98
SEMRes-DDPM <sup>[47]</sup>	93.75	90.59	91.56	87.74	98.12	96.75	84.96	79.87
<b>LDNFDDPM</b>	<b>94.68</b>	<b>91.47</b>	<b>92.12</b>	<b>88.42</b>	<b>98.56</b>	<b>97.28</b>	<b>86.34</b>	<b>81.23</b>

表 5-8 NF-ToN-IoT-V2 数据集评估结果（续表）

Method	Scanning		Injection		Password	
	Pre	Rec	Pre	Rec	Pre	Rec
G-IDS <sup>[44]</sup>	63.45	59.17	69.56	53.52	60.58	63.12
SIGMA <sup>[45]</sup>	65.02	61.28	61.25	65.56	73.85	66.14
CWGAN-GP <sup>[46]</sup>	74.12	79.94	80.34	84.35	72.42	64.71
SEMRes-DDPM <sup>[47]</sup>	96.23	92.85	92.45	87.08	75.12	67.85
<b>LDNFDDPM</b>	<b>97.01</b>	<b>93.56</b>	<b>85.01</b>	<b>87.65</b>	<b>76.43</b>	<b>69.12</b>

的 FID 值要低于 LDDDP, 这表明, LDNFDDPM 通过保存数据功能性特征, 有助于生成模拟数据。同样, LDNFDDPM 的 FID 值要低于 NFDDPM, 这表明, 关注游离 IP, 对生成数据的真实性有一定提升。此外, 在数据集 NF-ToN-IoT-V2 上 NFDDPM 的 FID 值要小于 LDDDP, 这是因为 NF-ToN-IoT-V2 数据集的游离 IP 偏多, 这也恰恰说明针对游离 IP 进行权重提升能够提高生成数据的真实性。

表 5-9 不同消融设置下的数据增强表现

Method	FID		
	NF-BoT-IoT-V2	NF-CSE-CIC-IDS2018-V2	NF-ToN-IoT-V2
LDDDP	47.15	41.54	37.32
NFDDPM	46.28	42.56	34.17
<b>LDNFDDPM</b>	<b>24.50</b>	<b>18.30</b>	<b>14.90</b>

#### 5.4 恶意流量检测方法测试

在本实验中, 将针对模型的性能和实际应用进行三个系列的实验分析: 分类效果实验、不同监督比例实验和消融实验。分类效果实验, 以评估模型在不同数据集上的分类精度、召回率、精确度和 F1-Score 等关键指标。不同监督比率的样本对模型性能的影响, 通过调整标签数据的比例, 观察模型在标注数据稀缺情况

下的表现。消融实验，逐步移除模型的各个组成部分，评估每一部分对模型整体性能的贡献。通过这些实验，旨在全面评估模型在不同条件下的鲁棒性和有效性。

在本实验中，使用了一些关键的超参数来调整模型的训练过程。以下表格列出了模型的默认超参数值及其相应的描述：

表 5-10 模型的默认超参数值

超参数	描述	默认值
<b>Learning rate</b>	每次迭代时朝着损失函数最小值前进的步长	0.001
<b>Weight decay</b>	损失函数的 L2 惩罚项	0.001
<b>Sequence length</b>	每个批次中的网络快照数量	100
<b>Sliding window overlap</b>	连续快照之间的重叠部分	50%
<b>Early stop patience</b>	在最后一次验证损失改进后等待的轮数	20

### 1. 分类效果分析

在三个数据集的分析中，LG-MHAAM 表现出了显著的优势。

在 NF-BoT-IoT-V2 数据集上，LG-MHAAM 的准确率为 99.78%，远高于其他对比方法。尽管 RandomForest 和 DNN 的准确率分别为 100% 和 99.54%，它们的误报率（FAR）较高，尤其是 DNN 方法的误报率为 0.20%，而 LG-MHAAM 的误报率为 0.01%，显示出其更强的鲁棒性。在召回率方面，LG-MHAAM 的值为 0.9998，几乎达到了完美，且精确度（0.9979）也优于其他方法。总体来看，LG-MHAAM 在综合性能（准确率、召回率、精确度和 F1-Score）上都表现得非常优秀，尤其是在低误报率方面，显示了更高的检测可靠性。

在 NF-ToN-IoT-V2 数据集上，LG-MHAAM 的准确率为 98.69%，略低于 E-GraphSAGE（99.69%）和 ExtraTrees（99.66%）。尽管如此，LG-MHAAM 在召回率（0.9794）和精确度（0.9982）方面仍保持了较高水平，显示出其在减少假阳性（精确度）和提高检测率（召回率）方面的平衡能力。与 RandomForest 和 DNN 相比，LG-MHAAM 在误报率（FAR）上表现更为优秀，尤其在减少假阳性方面表现突出。

在 NF-CSE-CIC-IDS2018-V2 数据集上，LG-MHAAM 的表现更为显著，其准确率为 99.80%，大幅领先于其他方法。ExtraTrees 的准确率为 95.33%，而 TSVM 的准确率仅为 91.22%。在召回率方面，LG-MHAAM 为 0.9986，几乎完美，远超 ExtraTrees 和 TSVM。在精确度上，LG-MHAAM 的值为 0.9988，接近完美，而 ExtraTrees 的精确度较低，为 0.7420。F1-Score 指标也表明，LG-MHAAM（0.9987）明显优于 ExtraTrees（0.83）和 TSVM（0.9127）。最为关键的是，LG-MHAAM 的误报率为 0.13%，在所有方法中最低，这在流量分类中非常重要，因为较低的误报

率有助于减少系统负担。

综合来看, LG-MHAAM 在所有三个数据集中的表现都非常优秀, 尤其在准确率、召回率、精确度和 F1-Score 方面均表现突出。与 RandomForest 和 DNN 等方法相比, LG-MHAAM 在误报率方面更具优势。特别是在 NF-CSE-CIC-IDS2018-V2 数据集上, LG-MHAAM 展现了极强的稳定性和高效的流量分类能力, 显示了其在处理复杂网络流量分类任务时的优势。

表 5-11 NF-BoT-IoT-V2 数据集方法对比

Method	Accuracy	Recall	Precision	F1-Score	FAR
<b>LG-MHAAM</b>	99.78%	0.9998	0.9979	0.9989	0.01%
<b>E-GraphSAGE</b> <sup>[48]</sup>	93.57%	0.9343	1.0	0.97	0.38%
<b>ExtraTrees</b> <sup>[49]</sup>	93.82%	1.0	0.9417	0.97	1.13%
<b>RandomForest</b> <sup>[50]</sup>	100.0%	1.0	1.0	1.0	0.25%
<b>DNN</b> <sup>[50]</sup>	99.54%	0.9954	1.0	1.0	0.20%
<b>TSVM</b> <sup>[51]</sup>	92.0%	0.91	0.92	0.915	7.5%

表 5-12 NF-ToN-IoT-V2 数据集方法对比

Method	Accuracy	Recall	Precision	F1-Score	FAR
<b>LG-MHAAM</b>	98.69%	0.9794	0.9982	0.9887	2.05%
<b>E-GraphSAGE</b> <sup>[48]</sup>	99.69%	0.9985	1.0	1.0	0.15%
<b>ExtraTrees</b> <sup>[49]</sup>	99.66%	0.9967	0.9995	1.0	0.37%
<b>RandomForest</b> <sup>[50]</sup>	99.66%	0.9980	0.9991	1.0	0.58%
<b>DNN</b> <sup>[50]</sup>	94.74%	0.9527	0.9674	0.96	6.08%
<b>TSVM</b> <sup>[51]</sup>	91.0%	0.89	0.91	0.90	7.5%

表 5-13 NF-CSE-CIC-IDS2018-V2 数据集方法对比

Method	Accuracy	Recall	Precision	F1-Score	FAR
<b>LG-MHAAM</b>	99.80%	0.9986	0.9988	0.9987	0.13%
<b>E-GraphSAGE</b> <sup>[48]</sup>	93.0%	0.92	0.95	0.935	0.50%
<b>ExtraTrees</b> <sup>[49]</sup>	95.33%	0.9471	0.7420	0.83	4.59%
<b>RandomForest</b> <sup>[50]</sup>	99.47%	0.9682	0.9921	0.98	0.17%
<b>DNN</b> <sup>[50]</sup>	99.24%	0.9467	0.9944	0.97	0.14%
<b>TSVM</b> <sup>[51]</sup>	91.22%	0.9121	0.9133	0.9127	8.06%

另一方面, 在三个不同的数据集上, 所提出的 LG-MHAAM 方法显示出了较强的分类能力, 特别是在 NF-CSE-CIC-IDS2018-V2 数据集上, 几乎所有类别的准确率和 F1-Score 都达到了 100%, 并且加权平均的 F1-Score 高达 0.9996。这表明该模型在处理大多数类别时表现出了非常优越的分类性能, 尤其是在 Bot、BruteForce、

DDoS 和 Infiltration 类别上, 均获得了完美的性能。

在 NF-ToN-IoT-V2 数据集上, LG-MHAAM 的表现较好, 准确率分别为 92.68%、99.89% 和 98.51%, 且 F1-Score 也分别为 0.8919、0.9983 和 0.9778。尽管 Injection 和 Password 类别的表现相对较弱, 准确率分别为 89.53% 和 82.85%, F1-Score 也偏低, 但其他攻击类别的高准确率和 F1-Score 表明该模型在大多数类别上仍然表现良好。整体加权平均准确率为 93.37%, F1-Score 为 0.9340, 反映了该模型在大部分类别上的较为均衡的表现。

在 NF-CSE-CIC-IDS2018-V2 数据集上, Bot 和 Infiltration 类别的准确率和 F1-Score 达到 100%, 表明该模型在这两个类别上的分类表现极为优越。BruteForce 类别也表现出色, 准确率为 100%, F1-Score 为 1.0, 表明该模型能够很好地捕捉这一攻击类别的特征。虽然 DoS 类别的准确率为 99.50%, F1-Score 为 0.9924, 但依然保持了很高的性能。整体加权平均准确率为 99.96%, F1-Score 为 0.9996, 显示出该模型在此数据集上的极高表现。

在 NF-BoT-IoT-V2 数据集上, DDoS 和 DoS 类别的表现较好, 准确率分别为 99.05% 和 95.95%, F1-Score 为 0.9859 和 0.9399。Reconnaissance 类别的准确率和 F1-Score 稍微下降, 准确率为 94.21%, F1-Score 为 0.9146。Theft 类别的表现相对较弱, 准确率仅为 59.08%, F1-Score 为 0.4975, 表明该类别在样本分布或特征提取方面可能存在一定问题。整体加权平均准确率为 95.72%, F1-Score 为 0.9579, 说明该模型在大多数类别上表现良好, 但在少数类别上可能存在提升空间。

LG-MHAAM 模型在大多数类别上具有非常强的分类能力, 特别是在 NF-CSE-CIC-IDS2018-V2 和 NF-ToN-IoT-V2 数据集上, 表现出色。对于 Bot、BruteForce、DDoS 等类别, LG-MHAAM 提供了几乎完美的分类结果, F1-Score 达到 1.0。从表 5-16 可以看到, 对于 Theft 的表现略显不足, 这是由于这 Theft 这类攻击类型仅占 NF-BoT-IoT-V2 数据集的 0.21%, 样本稀疏带来不稳定检测。加权平均准确率和 F1-Score 表现出模型在各个类别之间的平衡能力, 尤其在 NF-CSE-CIC-IDS2018-V2 数据集上, 模型的综合表现尤为突出, 表明该模型在应对复杂场景下能够取得稳定的效果。

## 2. 不同监督率的样本的影响

为了进一步探索模型的表现能力, 在 **NF-CSE-CIC-IDS2018-V2** 和 **NF-ToN-IoT-V2** 数据集上进行了半监督训练, 使用了 30%、50% 和 70% 的标记样本。通过观察图 5-2 中的结果, 可以看出对于这两个数据集中的每个类别, 使用 30% 标记样本的效果与使用 70% 标记样本的效果并没有显著差距。

具体来说, 在 **NF-CSE-CIC-IDS2018-V2** 数据集上, 使用 30% 标记样本的模型

表 5-14 NF-ToN-IoT-V2 数据集多分类结果

ClassName	ACC	F1-Score
Backdoor <sup>[31]</sup>	97.66%	0.9653
DDoS <sup>[29]</sup>	92.68%	0.8919
DoS <sup>[30]</sup>	99.89%	0.9983
Injection <sup>[32]</sup>	89.53%	0.8459
Password <sup>[33]</sup>	82.85%	0.7517
Scanning <sup>[34]</sup>	98.51%	0.9778
XSS <sup>[35]</sup>	95.70%	0.9359
Weighted Average	93.37%	0.9340

表 5-15 NF-CSE-CIC-IDS2018-V2 数据集多分类结果

ClassName	ACC	F1-Score
Bot <sup>[52]</sup>	100.00%	1.0
BruteForce <sup>[38]</sup>	100.00%	1.0
DDoS <sup>[29]</sup>	100.00%	1.0
DoS <sup>[30]</sup>	99.50%	0.9924
Infiltration <sup>[53]</sup>	100.00%	1.0
Weighted Average	99.96%	0.9996

准确率与使用 70% 标记样本准确率基本持平甚至略高。同样地, 在 **NF-ToN-IoT-V2** 数据集上, 使用 30% 标记样本时, XSS 攻击检测模型准确率为 95.7%, 而使用 70% 标记样本时, 准确率为 97.0%, 仅下降了 1.3%。

这些结果表明, 即使只有 30% 的样本被标记, 模型仍然表现出非常强的表达能力和区分能力, 能够有效地识别不同类型的网络攻击。与传统的全监督训练方法相比, 这种半监督方法不仅减少了标注成本, 同时也证明了我们的模型在处理不完全标注的数据时仍然能够保持较高的准确性和稳定性。

### 3. 消融实验

第一项消融研究检查了时间演变相关性对最终入侵检测性能的影响。在这个实验中, 我们去除了融合时空特征的 GCNII GRU 层, 以切断时间演变信息在不同图快照之间的传播。也就是说, 我们只利用 GCNII 层从多个图中学习网络流量统计和网络拓扑信息, 并观察在没有时间演变特征参与的情况下实现的检测性能。如表 5-17 所示, 当仅使用网络的拓扑信息和统计特征时, 检测的 ACC 和 F1 分数值会下降。基于这些发现, 表明考虑网络流量的“演变信息”确实有助于理解和描述网络行为。

为了以有效和直观的方式了解折线图的作用及其广泛的适用性, 采用当前模



表 5-16 NF-BoT-IoT-V2 数据集多分类结果

ClassName	ACC	F1-Score
DDoS <sup>[29]</sup>	99.05%	0.9859
DoS <sup>[30]</sup>	95.95%	0.9399
Reconnaissance <sup>[37]</sup>	94.21%	0.9146
Theft <sup>[36]</sup>	59.08%	0.4975
Weighted Average	95.72%	0.9579

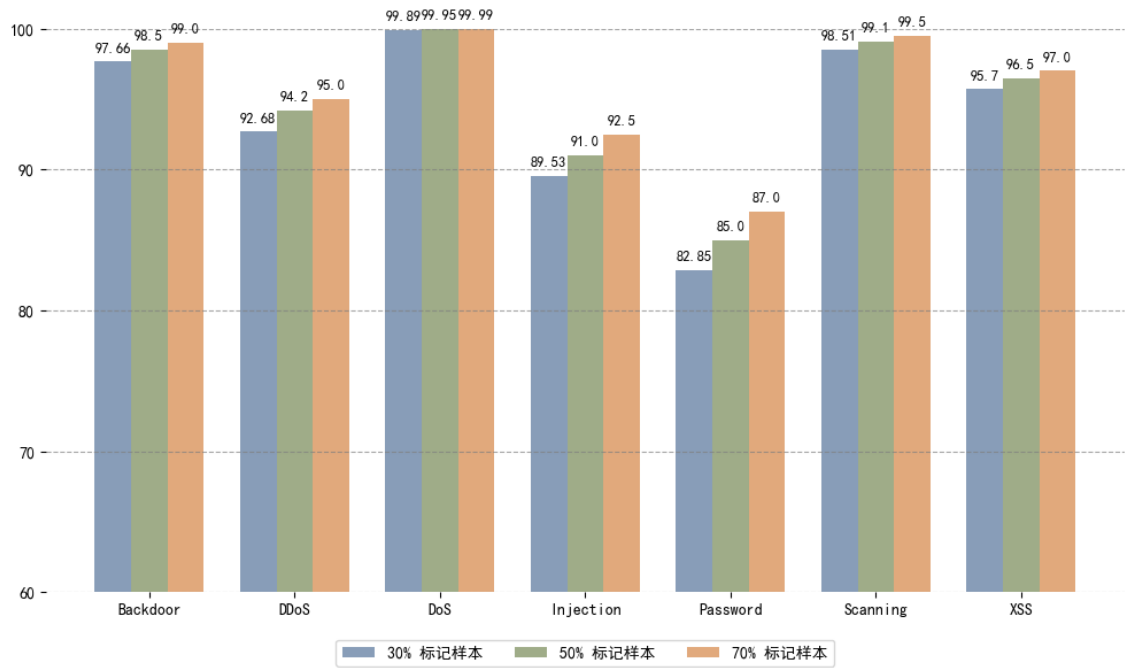


图 5-2 在 NF-ToN-IoT-V2 数据集上进行半监督学习时在不同标签比例下实现的准确率。

型和 E-GraphSAGE 模型的折线图版本进行了验证实验。如前所述，E-GraphSAGE 通过 GraphSAGE 卷积操作在图边缘部署消息传递机制。因此，我们通过采用原始版本（GraphSAGE）与我们的线图结构相结合进行了对比实验。表 5-17 中所示的比较实验结果表明，具有线形图结构的 GraphSAGE 比 E-GraphSAGE 实现了更高的检出率和 F1 分数值。这些结果表明，我们的线图结构可以增强图卷积运算，而且这种效果不仅限于特定的图卷积层。

## 5.5 本章小结

本章详细分析了实验环境、数据集和实验设计。在实验环境部分，介绍了硬件配置和操作系统，确保实验的重复性和结果的可靠性。数据集部分则描述了

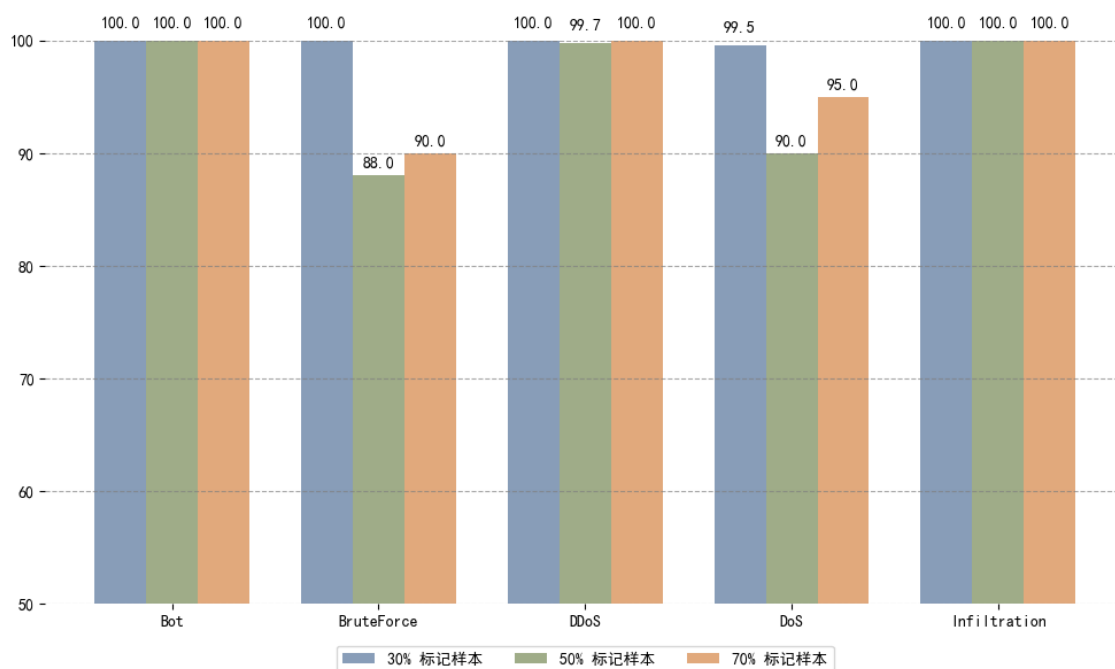


图 5-3 在 NF-CSE-CIC-IDS2018-V2 数据集上进行半监督学习时在不同标签比例下实现的准确率。

NF-BoT-IoT-V2、NF-ToN-IoT-V2 和 NF-CSE-CIC-IDS2018-V2 三个数据集的特征，为后续对比分析提供基础。

在实验设计中，定义了准确率、精确率、召回率、F1-Score 等评估指标，同时引入 FID 值来评估数据增强效果。通过与其他数据增强方法的对比，LDNFDDPM 在生成样本质量和恶意流量检测性能上表现出较大优势。消融实验进一步验证了特征划分和游离 IP 策略对性能的影响。

实验结果表明，LDNFDDPM 在数据增强和恶意流量检测中具有显著优势，证明其在物联网环境下的可行性和有效性。

表 5-17 消融实验结果

Model	Dataset	ACC	F1
LG-MHAAM	NF-BoT-IoT-V2	99.78%	0.9989
LG-MHAAM without time-evolving features	NF-BoT-IoT-V2	91.74%	0.9233
LG-MHAAM	NF-ToN-IoT-V2	98.69%	0.9887
LG-MHAAM without time-evolving features	NF-ToN-IoT-V2	89.43%	0.8945
LG-MHAAM	NF-BoT-IoT-V2	99.78%	0.9989
LG-MHAAM without lineGraph	NF-BoT-IoT-V2	98.16%	0.9100
LG-MHAAM	NF-ToN-IoT-V2	98.69%	0.9887
LG-MHAAM without lineGraph	NF-ToN-IoT-V2	97.16%	0.9300
GraphSAGE+line graph	NF-BoT-IoT	85.30%	0.8500
E-GraphSAGE	NF-BoT-IoT	80.56%	0.8250
GraphSAGE+line graph	NF-ToN-IoT	75.10%	0.7800
E-GraphSAGE	NF-ToN-IoT	72.22%	0.6900

## 第六章 全文总结与展望

### 6.1 全文总结

### 6.2 后续工作展望

## 致 谢

## 参考文献

- [1] Haloui I, Gupta J. S, Feuillard V. Anomaly detection with wasserstein gan[J]. arXiv preprint arXiv:1812.02463, 2018.
- [2] Chawla N. V, Bowyer K. W, Hall L. O, et al. Smote: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [3] Soltanzadeh P, Hashemzadeh M. RcsMOTE: Range-controlled synthetic minority over-sampling technique for handling the class imbalance problem[J]. Information Sciences, 2020, 542: 92-111.
- [4] Goodfellow I. J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Advances in Neural Information Processing Systems, 2014, 2672-2680.
- [5] Wu Q, Chen Y, Meng J. DcGAN-based data augmentation for tomato leaf disease identification[J]. IEEE Access, 2020, 8: 98716-98728.
- [6] Walter S, Mougeot G, Sun Y, et al. Midipgan: A progressive gan approach to midi generation[C]. Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design, 2021, 1166-1171.
- [7] Yu W, Zhou H, Liu Y, et al. Super-resolution reconstruction of particleboard images based on improved srgan[J]. Forests, 2023, 14: 1842.
- [8] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]. Advances in Neural Information Processing Systems, 2020, 6840-6851.
- [9] Ding X, Wang Y, Zhang K, et al. CcDM: Continuous conditional diffusion models for image generation[C]. Journal of L<sup>A</sup>T<sub>E</sub>X Class Files, 2021, 1-13.
- [10] Jiang H, Imran M, Ma L, et al. Fast-ddpm: Fast denoising diffusion probabilistic models for medical image-to-image generation[C]. arXiv preprint arXiv:2405.14802, 2024, .
- [11] Lee Sgil, Kim H, Shin C, et al. Priorgrad: Improving conditional denoising diffusion models with data-dependent adaptive prior[C]. Proceedings of the International Conference on Learning Representations, 2022, .
- [12] Kumar V, Sinha D, Das A. K, et al. An integrated rule based intrusion detection system: analysis on unsw-nb15 data set and the real time online dataset[J]. Cluster Computing, 2020, 23: 1397-1418.

- [13] Zhang C, Zhao Y, Zhou Y, et al. A real-time abnormal operation pattern detection method for building energy systems based on association rule bases[J]. Building Simulation, 2022, 15: 69-81.
- [14] Yang Y, Lee K, Dariush B, et al. Follow the rules: Reasoning for video anomaly detection with large language models[J]. arXiv preprint arXiv:2407.10299, 2024.
- [15] Feng P, Ma J, Sun C. Selecting critical data flows in android applications for abnormal behavior detection[J]. Mobile Information Systems, 2017, 2017: 1-16.
- [16] Thakkar A, Lohiya R. Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system[J]. Information Fusion, 2023, 90: 353-363.
- [17] Liu J, Yang D, Lian M, et al. Research on intrusion detection based on particle swarm optimization in iot[J]. IEEE Access, 2021, 9: 38254-38268.
- [18] Zhang H, Li J.-L, Liu X.-M, et al. Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection[J]. Future Generation Computer Systems, 2021, 122: 130-143.
- [19] Salman O, Elhajj I. H, Chehab A, et al. A machine learning based framework for iot device identification and abnormal traffic detection[J]. Transactions on Emerging Telecommunications Technologies, 2019.
- [20] Waskle S, Parashar L, Singh U. Intrusion detection system using pca with random forest approach[C]. Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020), 2020, 803-808.
- [21] Fu Y, Du Y, Cao Z, et al. A deep learning model for network intrusion detection with imbalanced data[J]. Electronics, 2022, 11: 898.
- [22] Sinha J, M M. Efficient deep cnn-bilstm model for network intrusion detection[J]. AIPR, 2020.
- [23] Ge M, Syed N. F, Fu X, et al. Towards a deep learning-driven intrusion detection approach for internet of things[J]. Computer Networks, 2021, 186: 107784.
- [24] Liu J, Li X.-W. Anomaly detection algorithm in industrial control network based on graph neural network[J]. Computer Systems Applications, 2020, 29(12): 234-238.
- [25] Wang J, Cao Z, Xie C, et al. Dgn: influence maximization based on deep reinforcement learning[J]. The Journal of Supercomputing, 2025, 81: 130.
- [26] Xu Z, Bai L, Cui L, et al. Hc-gae: The hierarchical cluster-based graph auto-encoder for graph representation learning[J]. arXiv preprint arXiv:2405.14742, 2024.

- [27] Busch J, Kocheturov A, Tresp V, et al. Nf-gnn: Network flow graph neural networks for malware detection and classification[C]. 33rd International Conference on Scientific and Statistical Database Management (SSDBM 2021), 2021, 12.
- [28] Zhao L, Song Y, Zhang C, et al. T-gcn: A temporal graph convolutional network for traffic prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(9): 3848-3858.
- [29] Mirkovic J, Reiher P. A taxonomy of ddos attack and ddos defense mechanisms[J]. ACM SIGCOMM Computer Communication Review, 2004, 34(2): 39-53.
- [30] Carl G, Brooks R. R, Rai S, et al. Denial-of-service attack detection techniques[J]. IEEE Internet Computing, 2006, 10(1): 82-89.
- [31] Gao Y, Doan B. G, Zhang Z, et al. Backdoor attacks and countermeasures on deep learning: A comprehensive review[J]. arXiv preprint arXiv:2007.10760v3, 2020.
- [32] Alghawazi M, Alghazzawi D, Alarifi S. Detection of sql injection attack using machine learning techniques: A systematic literature review[J]. Journal of Cybersecurity and Privacy, 2022, 2: 764-777.
- [33] Wang X, Yan Z, Zhang R, et al. Attacks and defenses in user authentication systems: A survey[J]. Journal of Network and Computer Applications, 2021, 188: 103080.
- [34] Lee C. B, Roedel C, Silenok E. Detection and characterization of port scan attacks[C]. Proceedings of the 2000 USENIX Workshop on Steps to Reducing Unwanted Traffic on the Internet, 2000, .
- [35] Gupta S, Gupta B. B. Cross-site scripting (xss) attacks and defense mechanisms: classification and state-of-the-art[J]. International Journal of System Assurance Engineering and Management, 2017, 8(Suppl. 1): S512-S530.
- [36] Leevy J. L, Hancock J, Khoshgoftaar T. M, et al. Detecting information theft attacks in the bot-iot dataset[C]. 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), 2021, 807-812.
- [37] Jafarian J. H, Al-Shaer E, Duan Q. An effective address mutation approach for disrupting reconnaissance attacks[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(12): 2562-2577.
- [38] Knudsen L. R, Robshaw M. The block cipher companion[M]. Springer-Verlag Berlin Heidelberg, 2011.
- [39] Gong X, Chen S, Li N. A network traffic data generation model based on aot-ddpm for abnormal traffic detection[J]. Evolving Systems, 2025, 16(1): 15.



- [40] Learning S.-S. Semi-supervised learning[J]. CSZ2006. html, 2006, 5: 2.
- [41] Gao C, Zheng Y, Li N, et al. A survey of graph neural networks for recommender systems: Challenges, methods, and directions[J]. ACM Transactions on Recommender Systems, 2023, 1(1): 1-51.
- [42] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017.
- [43] Sarhan M, Layeghy S, Portmann M. Towards a standard feature set for network intrusion detection system datasets[J]. Mobile networks and applications, 2022, 1-14.
- [44] Shahriar M. H, Haque N. I, Rahman M. A, et al. G-ids: Generative adversarial networks assisted intrusion detection system[C]. 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020, 376-385.
- [45] Msika S, Quintero A, Khomh F. Sigma: Strengthening ids with gan and metaheuristics attacks[J]. arXiv preprint arXiv:1912.09303, 2019.
- [46] Kang Y, Chen L, Jia N, et al. A cwgan-gp-based multi-task learning model for consumer credit scoring[J]. Expert Systems with Applications, 2022, 206: 117650.
- [47] Zheng M, Yang Y, Zhao Z.-H, et al. Semres-ddpm: Residual network based diffusion modelling applied to imbalanced data[J]. arXiv preprint arXiv:2403.05918, 2024.
- [48] 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.
- [49] Sarhan M, Layeghy S, Moustafa N, et al. Netflow datasets for machine learning-based network intrusion detection systems[C]. Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings 10, 2021, 117-135.
- [50] Sarhan M, Layeghy S, Portmann M. Evaluating standard feature sets towards increased generalisability and explainability of ml-based network intrusion detection. arxiv 2021[J]. arXiv preprint arXiv:2104.07183, 2022.
- [51] Abdel-Basset M, Hawash H, Chakraborty R. K, et al. Semi-supervised spatiotemporal deep learning for intrusions detection in iot networks[J]. IEEE Internet of Things Journal, 2021, 8(15): 12251-12265.

- [52] Zhang L, Yu S, Wu D, et al. A survey on latest botnet attack and defense[C]. 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications, 2011, 53-60.
- [53] 陈可, 鲁辉, 方滨兴, et al. 自动化渗透测试技术研究综述 [J]. 软件学报, 2023, 1-21.