计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.010109] [CSTR: 32024.14.csa.010109]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于扩散模型的小样本入侵检测①

王舒淋[1,2], 林宏刚[1,2], 李鹏亮[1]

[1](成都信息工程大学 网络空间安全学院, 成都 610225)
[2](先进微处理器技术国家工程研究中心 (工业控制与安全分中心), 成都 610225)
通信作者: 林宏刚, E-mail: linhg@cuit.edu.cn

摘　要: 针对新型网络攻击初始阶段样本稀缺导致入侵检测模型泛化能力不足、检测效果差等问题, 本文提出一种基于扩散模型的小样本入侵检测方法. 该方法在数据增强层面构建噪声感知的条件扩散模型, 采用余弦噪声调度平衡生成效率与样本质量, 并通过残差连接增强特征传播稳定性, 从而提升合成流量数据的分布保真度. 在特征度量层面, 设计动态原型网络结构, 利用多头注意力优化类原型表示, 缓解小样本特征稀疏问题; 同时采用交叉熵损失与正交正则项的联合优化策略, 增强类内聚合与类间区分度. 在两个公开数据集上的实验结果表明, 该模型在小样本场景下的准确率和泛化能力均优于其他检测方法, 为小样本入侵检测提供了新的解决思路.

关键词: 入侵检测; 小样本检测; 扩散模型; 度量学习

引用格式: 王舒淋,林宏刚,李鹏亮.基于扩散模型的小样本入侵检测.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/10109.html

Few-shot Intrusion Detection Based on Diffusion Model

WANG Shu-Lin[1,2], LIN Hong-Gang[1,2], LI Peng-Liang[1]

[1](School of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China)
[2](SUGON Industrial Control and Security Center, Chengdu 610225, China)

Abstract: To address the insufficient generalization and degraded detection performance in intrusion detection models caused by scarce samples during the early stages of novel network attacks, this study proposes a few-shot intrusion detection method based on diffusion models. At the data augmentation level, the proposed method introduces a noise-aware conditional diffusion model that employs cosine noise scheduling to balance generation efficiency and sample fidelity, while residual connections are incorporated to enhance feature propagation stability and improve the distribution fidelity of synthesized traffic data. At the feature metric level, a dynamic prototype network is designed, leveraging multi-head attention to optimize class prototype representations and mitigate feature sparsity in few-shot scenarios. Simultaneously, a joint optimization strategy combining cross-entropy loss with an orthogonal regularization term is adopted to enhance intra-class compactness and increase inter-class separability. Experimental results on two public datasets demonstrate that the proposed model outperforms other detection methods in terms of accuracy and generalization capability under few-shot scenarios, providing a novel solution approach for few-shot intrusion detection.

Key words: intrusion detection; few-shot detection; diffusion model; metric learning

随着全球数字化进程的加速, 网络安全威胁日益严峻, 入侵行为呈现出多样化趋势. 作为网络安全防御的核心环节, 入侵检测系统 (intrusion detection system, IDS) 能通过分析网络流量或系统行为及时发现潜在的恶意活动. 近年来, 深度学习因其强大的特征提取能力成为入侵检测研究的热点. 基于深度学习的入侵检测

① 基金项目: 国家 242 信息安全计划 (2021-037); 四川省自然科学基金 (2024NSFSC0515)
收稿时间: 2025-09-10; 修改时间: 2025-10-10; 采用时间: 2025-10-29; csa 在线出版时间: 2026-01-16

方法需要依赖大量的标注样本来训练检测模型, 以准确识别已知攻击. 然而, 由于新型攻击不断出现, 难以收集足够的样本进行训练, 导致深度学习模型在面对新的和少见的威胁时表现不佳[1]. 具体来说, 目前基于深度学习的入侵检测方法在小样本场景下容易过拟合. 在此背景下, 小样本入侵检测逐渐成为研究热点, 已有研究主要沿几个方向展开探索: 一是元学习, 通过任务式训练模拟测试阶段的小样本场景让模型学会学习; 二是度量学习, 通过优化特征的嵌入空间, 提升类别区分度; 三是数据增强, 合成多样化的攻击样本以缓解数据稀缺带来的性能瓶颈. 尽管已有研究取得一定进展, 但仍面临多重挑战: 其一, 特征表达不充分, 有限样本难以支撑模型学习具有强判别力的特征表示; 其二, 动态适应能力弱, 传统模型采取静态训练范式, 难以适应持续变化的网络威胁环境.

针对上述挑战, 本文提出了一种结合条件扩散模型与动态原型网络的入侵检测方法 CD-DPN (conditional diffusion dynamic prototype network), 实现在小样本场景下提升对攻击的识别效果. 该方法研究基于残差连接的条件扩散模型, 实现符合真实流量分布的高质量样本生成, 有效缓解样本稀缺问题; 采用基于任务驱动的学习策略, 使模型能够快速适应新任务, 降低对大规模标注样本的依赖; 设计动态原型网络, 融合多头注意力机制优化特征嵌入空间, 突出关键特征, 提升检测效果.

综上所述, 本文的贡献如下.

1) 提出了一种基于残差连接的条件扩散模型, 将其应用于小样本入侵检测, 通过噪声调度策略和条件生成机制实现高质量攻击样本合成, 有效缓解小样本数据稀缺问题.

2) 构建动态原型网络, 采用注意力驱动的特征加权策略与原型正交化空间约束, 提升小样本条件下的特征表示质量和类别区分能力.

1 相关工作

入侵检测系统经历了从基于规则匹配、特征工程到引入机器学习与深度学习的演进. 早期方法依赖于人工制定的规则或签名库, 难以应对动态变化的复杂攻击. 为突破规则系统的局限, 研究者开始引入机器学习方法, 如支持向量机[2]、随机森林[3]、K 近邻[4]等对流量特征进行建模, 实现初步的自动化检测.

近年来, 深度学习因其强大的特征提取能力在入侵检测中表现出色. Laghrissi 等人[5]对时间序列特征进行建模, 实现了基于长短期记忆 (long short-term memory, LSTM) 的 IDS. Wei 等人[6]引入注意力机制来提升 LSTM 对关键特征的聚焦能力. Ho 等人[7]采用卷积神经网络 (convolutional neural network, CNN) 有效提取网络流量的空间特征. 周璨等人[8]结合 CNN 和门控循环单元, 通过轻量级模型, 缩短了模型的攻击检测时间, 提高了模型的攻击检测性能. 池彬等人[9]改进自编码器中的 LSTM 并与流特征结合, 增强了复杂攻击模式的识别效果. 然而, 这些模型通常依赖大量标注数据, 在面对新型攻击时往往表现出较低的准确率, 难以有效学习最新攻击模式.

为突破这一限制, 小样本学习方法近年来逐渐被引入入侵检测任务. 元学习通过构建任务驱动的学习范式, 提升模型在有限样本条件下的快速适应能力. Xu 等人[10]首次将元学习框架用于入侵检测, 设计了 FC-Net 网络, 实现了小样本流量分类. Lu 等人[11]首次将模型无关元学习 (model agnostic meta learning, MAML) 与 CNN 结合应用于物联网入侵检测, 通过少量训练调整模型的参数, 实现快速适应. Hu 等人[12]构建基于隐私保护的小样本流量检测框架 PFTD, 结合个性化模型与全局共享模型, 在边缘设备上保持检测效能. 尽管 MAML 展现了快速适应的能力, 但对初始化敏感且复杂度较高. 相比之下, 基于度量的小样本学习方法如孪生网络[13]、原型网络[14]等, 通过学习任务间的相似性度量进行分类, 计算效率更高. Hindy 等人[15]利用孪生网络实现基于相似性的流量分类. Yang 等人[16]提出特征提取模块和距离度量模块协同工作的 FS-IDS 框架, 在未知攻击存在的情况下仍能保持较高的准确率. 林同灿等人[17]通过内部和外部对齐来优化原型生成. Wu 等人[18]开发的 MASiNet 框架通过在孪生网络中融入注意力模块提升了小样本检测性能. Wang 等人[19]则针对工业互联网中标记样本缺乏和新型攻击频发的问题, 结合 CNN 与原型网络, 提升了在复杂环境下的检测效果. 尽管已有研究尝试采用原型网络实现快速类别适配, 但在网络流量特征高维、异构的背景下, 静态原型易受干扰, 难以准确表达类别中心, 限制了模型的最终检测性能.

数据增强是缓解小样本问题的另一重要手段. Yang 等人[20]使用条件变分自动编码器 (conditional variational autoencoder, CVAE) 重构网络流量样本并识

别未知攻击. Xu 等人[21]通过选择代表性样本来训练 CVAE 模型, 生成基于语义嵌入的样本, 缓解类偏差问题. Zekan 等人[22]改进针对图像分类的 EC-GAN 以处理表格数据的分类任务, 结合深度神经网络解决小样本问题. 然而, CVAE 生成样本过于平滑, GAN 训练过程不稳定且对超参数敏感, 生成数据质量受输入分布限制, 难以保证与真实流量的分布一致性.

近年来, 扩散模型[23]依靠渐进式去噪生成, 在图像生成领域展现出卓越性能, 其稳定的训练过程和强大的样本建模能力为高质量流量生成提供了新的技术路径. 在入侵检测领域, 扩散模型开始应用于小样本和不平衡数据场景, 解决数据稀缺和异常检测的生成问题. Zhang 等人[24]针对类不平衡问题提出了一种新的入侵检测系统, 将数据进行增强数据特征相关性增强处理, 通过扩散模型和卷积神经网络结合, 提升了检测结果. Yang 等人[25]设计了 Diff-IDS 轻量级模型, 提高模型在解析复杂网络流量特征时的效率, 从而显著提高其检测速度和训练效率, 适用于资源受限的工业网络入侵检测系统. 此外, Wang 等人[26]提出了一种基于扩散模型和 Transformer 的物联网入侵检测模型, 利用扩散模型学习样本特征模式的少数类别, 并生成平衡数据集, 提高了非平衡数据集的检测性能. 这些研究表明, 扩散模型具备在小样本入侵检测中生成高质量样本的潜力, 并提供了新思路.

2 模型设计

2.1 总体架构

CD-DPN 方法的总体架构由数据预处理、条件扩散、动态原型这 3 大模块构成, 结合生成式增强与度量学习, 实现小样本场景下的入侵检测. 整个检测流程如图 1 所示.

原始网络流量数据包含正常流量和多种攻击类型, 包含统计特征、协议字段、时序信息等异构性特征. 为确保模型输入的统一性和质量, 需要对数据集进行预处理. 首先对数据进行清洗, 去除错误与重复样本; 随后将离散字符型字段转换为数值向量, 并对连续数值型特征进行归一化, 以保证训练过程的稳定性. 预处理后的数据将作为输入, 交由条件扩散模型进行数据增强. 该模块基于类别条件引导, 通过噪声注入与逐步去噪过程, 生成与真实分布高度一致的合成样本, 增强支持集的多样性. 最后, 动态原型度量模块基于增强后的支持集构造类别原型, 并结合多头注意力机制进行原型表示优化. 模型通过度量查询样本与各类原型之间的相似度实现精确分类.

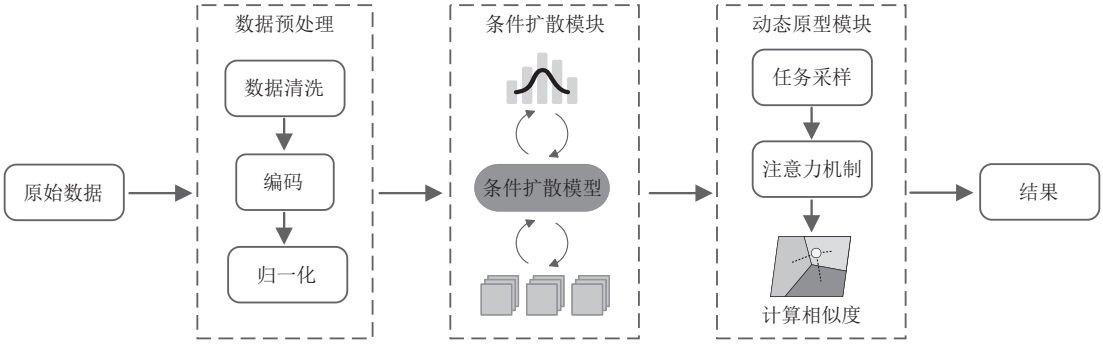

图 1 CD-DPN 整体流程

2.2 条件扩散模块

为解决小样本入侵检测中样本数量有限、类别分布不均的问题, 本文设计了一种基于去噪扩散概率模型的条件扩散生成器, 旨在通过合成高质量的攻击样本增强支持集多样性. 针对网络流量数据的特殊性, 本文提出以下改进: 采用余弦噪声调度与范围约束, 生成扩散系数序列并限制其范围, 实现平滑、稳定的扩散过程; 残差多层感知去噪网络和双重条件嵌入机制结合, 实现类别可控的条件生成过程并提升训练稳定性.

该模型通过模拟数据分布的逐步演化过程, 在隐空间中生成语义可控的目标类别样本, 从而有效扩充训练集. 与传统图像生成扩散模型不同, 本文将条件扩散机制应用于结构化的网络流量数据, 并结合攻击类型标签, 实现类别可控的条件生成过程, 从而有效提升样本的代表性与多样性, 缓解小样本带来的过拟合风险.

本模块核心流程包括前向扩散过程、条件嵌入机制、噪声调度策略、去噪重构网络这 4 个部分, 整体结构如图 2 所示. 在训练阶段, 模型学习如何从带噪样

本恢复出原始数据; 在推理阶段, 模型从纯噪声出发, 结合类别条件逐步去噪生成流量样本, 确保生成样本与指定类别在语义上的一致性.

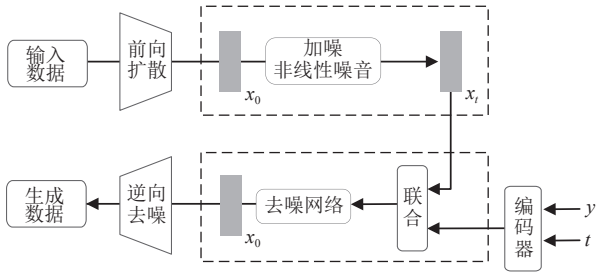

图 2 条件扩散模型结构示意

2.2.1 前向扩散过程

前向扩散阶段主要实现数据扰动, 用于模拟从原始数据到纯噪声的演化路径. 即在时间步t下对初始数据x_0添加高斯分布的随机噪声ϵ, 生成加噪样本x_t, 如式 (1) 所示:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon,\ \epsilon \sim \mathcal{N}(0,I) \tag{1}$$

其中, $\bar{\alpha}_t$是α_t的累积乘积, 控制数据在扩散过程中的保持程度. $\alpha_t = 1-\beta_t$表示每个时间步中残留原始信息的比例.

对于噪声的处理, 本文采用余弦调度策略, 决定添加噪声的强度β_t, 可以用式 (2) 表示:

$$\beta_t = 1 - \frac{\cos^2\left(\frac{t/T+s}{1+s}\cdot\frac{\pi}{2}\right)}{\cos^2\left(\frac{s}{1+s}\cdot\frac{\pi}{2}\right)} \tag{2}$$

其中, T是扩散总步数, s是平滑常数, 用于避免分母为空并控制曲线形状.

相较于线性调度, 余弦调度曲线具备良好的平滑性与收敛性, 能够更自然地模拟数据退化过程, 避免突变噪声对生成质量的冲击, 有效提升模型的稳定性与生成效果. 其前期噪声注入缓慢, 有利于模型学习数据结构, 后期则保留足够信息提升重建质量, 特别适用于网络流量等高维特征数据的建模.

2.2.2 逆向去噪过程

逆向去噪阶段从纯噪声开始, 结合时间步t和类别标签y, 由神经网络ϵ_θ逐步预测并移除噪声, 恢复目标数据分布.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t,y)\right) + \sqrt{\beta_t}\epsilon \tag{3}$$

为增强生成条件控制能力, 本文设计双重嵌入机制: 时间步嵌入采用正弦位置编码构造周期性时序表示; 类别嵌入使用嵌入层将离散标签映射至连续向量空间. 两者拼接后形成条件向量, 作为去噪网络的条件输入.

去噪重构网络采用多层感知器作为主干, 并引入残差连接结构, 以简化训练目标. 该结构通过在输入和输出之间建立直通通道, 使网络仅需学习输入与输出之间的差值, 从而简化优化目标并增强训练稳定性. 相较于 U-Net 等复杂结构, 该轻量级模型以更少的参数, 在不牺牲性能的前提下大幅降低计算资源需求.

2.3 动态原型模块

为模拟小样本场景, 动态原型网络采用基于任务的学习策略, 通过动态任务采样器从数据集中构造多个 N-way K-shot 任务. 每个任务由支持集 S 和查询集 Q 组成, 分别用于原型构建和分类评估. 其中支持集包含 N 个类别, 每类 K 个样本; 查询集包含相同的 N 个类别, 每类 Q 个样本, 两者类别相同但样本互不重叠. 在这种任务构造下, 模型通过多轮任务训练不断积累跨任务的通用知识, 从而具备对新类别的快速适应能力.

在小样本场景中, 传统的度量学习方法难以处理不同攻击类型间边界模糊、特征重叠的问题. 为此, 本文在原型网络框架基础上, 引入任务式训练策略与多头注意力机制, 以增强对小样本数据的表征能力与新攻击类别的适应性. 网络架构如图 3 所示, 主要由 3 部分组成: 特征编码器、多头注意力机制和损失函数.

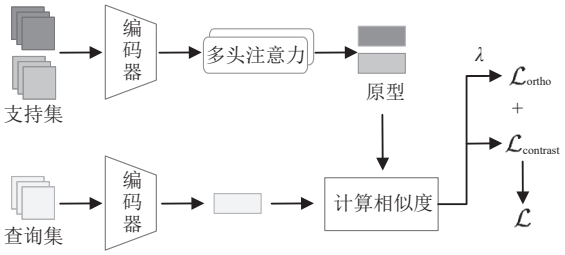

图 3 动态原型网络结构示意

为实现统一的特征表示, 动态原型网络对所有样本使用共享的嵌入编码器, 将预处理后的特征向量映射到高维隐藏空间, 生成嵌入向量. 在原型构造过程中, 传统原型网络直接对支持样本取均值, 忽略了不同样本对原型表示的贡献差异. 为此, 本文引入多头注意力机制对支持样本进行加权聚合, 构造动态类别原型. 通过计算查询矩阵Q、键矩阵K和值矩阵V, 生成注意力

权重, 公式如下:

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{4}$$

$$p_c=\frac{1}{K}\sum_{i=1}^{K}Attention(E_c) \tag{5}$$

其中, d_k为注意力头的维度大小, E_c为嵌入向量通过注意力机制, 对K个头的输出聚合得到原型p_c.

注意力多头机制从多个子空间提取差异性表示, 确保原型表达能力覆盖类别的主要特征方向. 相比静态的均值表示, 该机制能够动态关注关键样本, 弱化噪声样本影响, 从而提升类内一致性和类间区分度.

为了进一步提升原型间的判别能力, 本文设计了联合优化的损失函数, 由原型对比损失$\mathcal{L}_{\text{contrast}}$与原型正交正则项$\mathcal{L}_{\text{ortho}}$共同构成, 并通过超参数$\lambda$平衡两者的权重:

$$\mathcal{L}=\mathcal{L}_{\text{contrast}}+\lambda\mathcal{L}_{\text{ortho}} \tag{6}$$

其中, 原型对比损失通过温度参数τ控制查询样本与各类别原型之间相似度分布的陡峭程度, 引导模型聚焦于目标类别原型:

$$\mathcal{L}_{\text{contrast}}=-\frac{1}{Q}\sum_{j=1}^{Q}\log\frac{\exp(q_j\cdot p_{y_j}/\tau)}{\sum_{c=1}^{N}\exp(q_j\cdot p_c/\tau)} \tag{7}$$

其中, Q是查询集样本数, q是查询样本的嵌入, p_y是其对应的真实类别, N是原型的类别数, p_c对应的各类别原型.

正交正则项采用 Frobenius 范数计算原型矩阵的正交性偏差:

$$\mathcal{L}_{\text{ortho}}=\left\|P^{\mathrm{T}}P-I_N\right\|_F^2 \tag{8}$$

其中, P是所有类别原型的矩阵, I_N是单位矩阵.

该正则项鼓励类别原型在特征空间中保持最大化区分, 提升类别边界清晰度, 有效缓解攻击流量分布重叠导致的混淆问题, 使模型能够更准确地区分不同攻击类型.

3 实验分析

3.1 实验环境与数据

所有实验在 64 位 Windows 10 操作系统下进行, 使用 Intel Core i5-8300U CPU@2.3 GHz 处理器、NVIDIA GeForce GTX 1060 显卡和 16 GB 内存, 实验环境为 Python 3.10, CUDA 11.8 和 PyTorch 2.3.1.

为验证所提方法在小样本入侵检测任务中的有效性与通用性, 本文选用由加拿大网络安全研究所 (CIC) 所提供的 CIC-IDS2017、CSE-CIC-IDS2018[27]公开数据集. 具备攻击种类丰富、特征维度高、数据可扩展等特点, 被广泛用于入侵检测研究. CIC-IDS2017 数据集在真实网络环境中模拟攻击行为, 生成了大规模、结构化的流量记录以供分析. CSE-CIC-IDS2018 数据集进一步增强了数据的多样性与复杂性. 其模拟环境覆盖多个网络拓扑与平台配置, 包含更多种类的攻击模式与精细的流量特征, 能够更全面地反映现代网络环境中的安全威胁, 为入侵检测研究提供了更具挑战性的测试平台.

为保证模型训练与评估的公平性与有效性, 本文对原始数据进行了预处理和筛选. 为了确保数据质量, 由于某些流量类型的样本量低于 20 条, 不足以支持训练和测试, 则予以剔除. 最终从 CIC-IDS2017 数据集和 CSE-CIC-IDS2018 数据集分别保留 14 和 15 个流量类型作为实验对象, 覆盖正常流量及常见的攻击场景. 具体流量类型如表 1 所示.

表 1 数据集基本情况

数据集	类型数量	流量类型
CIC-IDS2017	14	Benign, FTP-Patator, SSH-Patator, DoS Hulk, DoS GoldenEye, DoS Slowloris, DoS Slowhttptest, DDoS, Web Attack-Brute Force, Web Attack-XSS, Web Attack-Sql Injection, PortScan, Infiltration, Bot
CSE-CIC-IDS2018	15	Benign, FTP-BruteForce, SSH-Bruteforce, DoS attacks-Hulk, DoS attacks-GoldenEye, DoS attacks-Slowloris, DoS attacks-SlowHTTPTest, DDoS attacks-LOIC-HTTP, DDOS attack-LOIC-UDP, DDOS attack-HOIC, Brute Force-Web, Brute Force-XSS, SQL Injection, Infiltration, Bot

3.2 实验设置

为模拟真实入侵检测中样本有限、类型未知的任务环境, 采用典型的 N-way K-shot 小样本设置进行实验, 对 CD-DPN 模型的学习能力与泛化能力进行综合评估. 具体而言, 本文采用 5-way K-shot 的任务配置, 在每轮任务中随机选取 5 个目标类型, 即正常流量和 4 个攻击类型作为当前任务的目标类. 从每个目标类别中分别抽取 K 个样本构建支持集, 以模拟小样本场景下的分类挑战. 在训练阶段, 模型仅接触一部分攻击类型, 测试阶段则评估其在未见类型上的识别能力, 从而

模拟真实环境中新型攻击的检测任务, 考验模型在小样本下的学习能力.

为了确保所提出的方法能够有效运行, 需要对模型的训练和推理过程进行参数设置. 本文模型的超参数设置详见表 2.

表 2 参数设置

超参数	设置
扩散步数T	500
噪音参数β_t	[0.0001, 0.02]
学习率	0.0001
优化器	Adam
Dropout	0.8
温度参数τ	0.1
训练轮数	200

3.3 评估指标

为了全面评价所提方法的检测效果并确保研究结果的可比性, 本文采用以下常见的分类指标: 准确率、召回率、精确率和 $F1$ 值. 上述指标均采用宏平均计算, 即对每个类别分别计算指标后取平均, 以避免数据不平衡的影响. 这些指标能够直观体现模型在均衡数据集中的综合判别能力, 提供全局性能基准, 确保与经典研究的纵向可比性. 从整体分类效果和类别平衡两个维度构建多层次评估体系. 为减少实验随机性的影响, 提高实验结果的可信度, 最终指标取多次独立实验的平均值.

3.4 实验结果与分析

3.4.1 K 值变化实验

为探究支持样本数量对模型性能的影响, 本文在两个数据集上开展了不同 K 值设置下的对比实验. K 分别取值为 1、5、10 和 15, 固定 N 值为 5, 保持每类查询样本数量不变. 评估指标选用准确率和 $F1$ 值, 以衡量模型在不同支持集规模下的检测能力. 实验结果如图 4 所示.

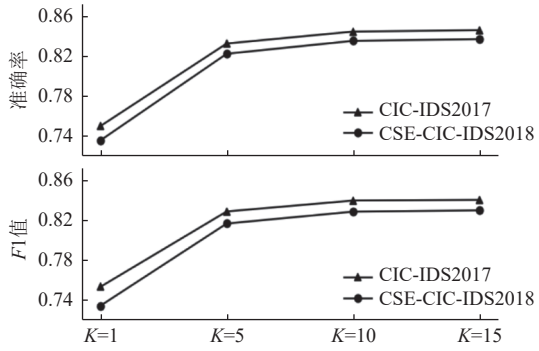

图 4 K 值变化实验结果 (N=5)

从图 4 中可以看出, K 从 1 增大至 5 时准确率和 $F1$ 值显著提升, 表明样本变多提供了更丰富的类内特征, 增强了对类别的表征能力. K 从 5 增大至 15 准确率和 $F1$ 值的提升减缓, 从 10 增大至 15 时明显平缓. 反映了类内特征的多样性已基本被捕捉, 增加样本的边际收益递减. 由于实验结果表明, 模型在 K=10 时已基本捕捉到类别的主要特征分布, 且在准确率与 $F1$ 值上的提升较为平稳, 进一步增大 K 值带来的性能增益有限. 因此, 为在保证检测性能的同时避免训练成本过高, 本文后续的小样本方法对比实验统一采用 10-shot 设置, 作为典型的小样本场景进行性能评估.

3.4.2 数据增强方法对比

为评估 CD-DPN 中条件扩散模型在小样本场景下的有效性, 实验在 CIC-IDS2017 和 CSE-CIC-IDS2018 数据集上, 分别对比了 4 种不同的数据增强方式: 原始支持集 (无增强)、CVAE[21]、EC-GAN[22]和本文提出的条件扩散模型, 比较不同生成方法在对分类结果的影响. 为公平比较, 均使用本文的动态原型网络作为分类器, 评估生成方法对分类效果的影响. 实验采用准确率和 $F1$ 值作为评价指标, 结果如表 3、表 4 所示.

表 3 CIC-IDS2017 数据集上的对比实验结果 (%)

方法	准确率	精确率	召回率	$F1$值
无增强	81.58	80.42	80.70	81.16
CVAE	82.48	81.55	82.13	82.05
EC-GAN	83.61	81.97	82.36	82.92
本文	84.27	84.54	82.62	83.97

表 4 CSE-CIC-IDS2018 数据集上的对比实验结果 (%)

方法	准确率	精确率	召回率	$F1$值
无增强	80.10	79.11	78.90	80.96
CVAE	81.92	80.05	81.59	81.73
EC-GAN	82.37	81.56	81.87	82.52
本文	83.34	82.54	83.12	82.85

表 3、表 4 显示, 本文提出的 CD-DPN 方法在两个数据集上均取得最优表现. 在采用不同的增强方法后检测效果均有所增加, 验证了增强对小样本入侵检测任务的有效性.

具体而言, CVAE 通过条件建模生成流量样本, 但其本质为潜变量重构机制, 在异构特征场景下难以精准拟合攻击样本的高阶结构, 提升幅度有限. EC-GAN 在生成对抗机制下可生成更贴近真实分布的样本, 但在 CSE-CIC-IDS2018 上性能下降说明其处理高维复杂网络环境时适应性有限. 相比之下, CD-DPN 所采用的

条件扩散模型通过逐步去噪过程建模攻击样本分布, 在余弦噪声调度与残差连接的协同作用下, 有效提升了样本生成的质量与多样性. 此外, 扩散过程通过类别标签进行条件控制, 使得合成样本在语义上更贴近目标类别, 进一步提升了增强样本的判别性. 实验结果验证了本文提出的方法有效缓解了小样本问题带来的过拟合风险, 提升了整体检测性能.

3.4.3 小样本方法对比

为了全面评估本文所提出的 CD-DPN 方法在小样本入侵检测任务中的性能优势, 本文选取了 5 种具有代表性的小样本学习方法作为对比: 原型网络[14]、孪生网络[15]、Meta-Baseline[28]、PFTD[12]和 FS-CT[29]. 所有方法在 10-shot 设置下进行评估, 分别在 CIC-IDS2017 和 CSE-CIC-IDS2018 数据集上测试其分类准确率、召回率、精确率和 $F1$ 值, 结果如表 5 和表 6 所示.

表 5 CIC-IDS2017 数据集上的对比实验结果 (%)

方法	准确率	精确率	召回率	$F1$值
原型网络	76.67	75.68	76.67	75.87
孪生网络	79.44	77.96	79.55	78.11
Meta-Baseline	80.89	79.88	81.02	79.90
PFTD	81.07	75.57	78.09	75.78
FS-CT	82.11	83.29	81.58	81.44
本文	84.27	84.54	82.62	83.97

表 6 CSE-CIC-IDS2018 数据集上的对比实验结果 (%)

方法	准确率	精确率	召回率	$F1$值
原型网络	74.33	74.28	73.51	73.72
孪生网络	77.56	75.14	76.52	75.16
Meta-Baseline	79.67	78.67	79.38	78.89
PFTD	80.48	75.22	77.57	75.01
FS-CT	81.64	78.55	80.43	79.25
本文	83.34	82.54	83.12	82.85

从实验结果可以看出, CD-DPN 在两个数据集上均取得了最佳的检测结果. 在 CIC-IDS2017 数据集中, CD-DPN 的 $F1$ 值为 83.97%, 相较次优模型 FS-CT 提升 2.5 个百分点; 在更具挑战性的 CSE-CIC-IDS2018 数据集中, CD-DPN 的 $F1$ 值为 82.85%, 相比 FS-CT 提升了 3.6 个百分点, 且召回率与精确率较为均衡. 这一结果充分体现了 CD-DPN 在应对复杂异构流量环境时的强泛化能力与鲁棒性, 表明 CD-DPN 在小样本场景下能够更好地平衡类迁移性和任务泛化性.

在所有方法中, 原型网络与孪生网络的性能相对较低, 主要原因在于这两类方法通常依赖静态的类原型构造方式或固定的相似性度量函数, 无法充分适应类别分布差异较大的流量数据. 而 CD-DPN 采用注意力机制和联合优化的原型分类器, 使得类原型更具表达能力和判别能力. Meta-Baseline 在元学习框架下具备一定的任务泛化能力, 但其训练依赖充分的任务抽样结构, 对于复杂网络流量分类任务难以完全拟合. 相比之下, CD-DPN 借助条件扩散模型在类内生成多样化样本, 有效缓解了过拟合和迁移困难的问题. 尽管 PFTD 引入了联邦训练策略, 但联邦学习中的分布式训练对边缘设备数量和通信效率敏感, 且受限于通信开销与模型聚合方式, 导致全局模型对某些类别的偏见从而影响整体检测性能. 而 CD-DPN 使用可控的生成机制和动态原型网络, 无需通信代价即可高效模拟多样任务, 有效降低了训练难度与资源开销. FS-CT 使用了余弦注意力用于增强支持-查询对之间的相关性建模, 但类别原型生成是一次性的, 面对网络流量数据时难以有效建模复杂类间分布. 而 CD-DPN 中的注意力模块作用于原型构建阶段, 能够在特征维度层面自适应捕捉任务相关性强的表示, 提升了原型的判别力, 联合优化策略进一步增强类原型的判别能力与表达鲁棒性. 最终在不同场景下均保持稳定的高性能表现.

综上, CD-DPN 在小样本入侵检测任务中展现出明显优于现有方法的性能, 验证了其结构设计的有效性和在现实复杂网络环境下的实用潜力.

3.4.4 消融实验

为了评估所提出 CD-DPN 方法中各组件的有效性, 对 CIC-IDS2017 和 CSE-CIC-IDS2018 数据集上的 10-shot 任务进行了消融研究, 以评估每个组件的有效性. 表 7 展示了比较结果.

表 7 消融实验结果 (%)

数据集	模型	准确率	$F1$值
CIC-IDS2017	Exp1	81.58	81.16
	Exp2	82.34	82.56
	Exp3	82.16	82.51
	本文	84.27	83.97
CSE-CIC-IDS2018	Exp1	80.10	80.96
	Exp2	81.89	81.86
	Exp3	81.32	81.24
	本文	83.34	82.85

实验设置包括 3 种变体: Exp1 去掉条件扩散模块, 直接使用原始数据; Exp2 移除多头注意力机制, 使用均值生成原型进行表示; Exp3 移除正交正则项与对比

损失, 使用经典原型损失作为训练目标.

实验结果显示, 完整 CD-DPN 模型在两项指标上均优于所有变体, 证明了其整体架构设计的合理性. 从 Exp1 中可以看出, 移除扩散模型后导致显著的性能下降, 表明扩散模型在增强模型的判别能力方面发挥了关键作用, 缓解了数据稀缺导致的模型过拟合问题. Exp2 移除注意力机制后, 准确率与 $F1$ 值均出现了一定程度降低. 这说明注意力机制能够更充分地捕捉样本的关键特征, 从而生成更具代表性的类原型, 而简单的均值表示无法实现对关键特征的差异化聚合, 导致类代表性下降. Exp3 中移除损失项检测效果有所下降, 说明联合优化策略能够有效优化空间结构, 拉近类内距离、区分类间特征. 综上所述, CD-DPN 的 3 个核心组件均对模型效果具有重要贡献. 它们分别从数据增强、特征聚合和特征优化这 3 个不同角度提升了小样本入侵检测的泛化能力与鲁棒性.

4 结论与展望

针对小样本场景下入侵检测面临的挑战, 本文提出了一种融合条件扩散生成与动态原型表示的检测框架 CD-DPN. 该方法通过条件扩散模型生成高质量网络流量样本, 从而提升数据表达能力; 设计了结合多头注意力机制的动态原型网络, 实现高效原型构造与度量推理, 实现了自适应加权聚合与类间对齐, 提升了模型对少量样本的判别能力.

在公开数据集上的实验结果表明, CD-DPN 在小样本场景下展现出更好的检测结果和泛化能力. 未来工作可进一步探索更高效的生成机制、跨域迁移能力及与联邦学习等新型隐私保护技术的融合, 以提升小样本入侵检测方法在实际场景中的应用价值.

参考文献

1 Xu CY, Zhan Y, Chen GH, *et al*. Elevated few-shot network intrusion detection via self-attention mechanisms and iterative refinement. PLoS One, 2025, 20(1): e0317713. [doi: 10.1371/journal.pone.0317713]

2 Du RZ, Li Y, Liang XY, *et al*. Support vector machine intrusion detection scheme based on cloud-fog collaboration. Mobile Networks and Applications, 2022, 27(1): 431–440. [doi: 10.1007/s11036-021-01838-x]

3 Gaber T, Awotunde JB, Folorunso SO, *et al*. Industrial internet of things intrusion detection method using machine learning and optimization techniques. Wireless Communications and Mobile Computing, 2023, 2023(1): 3939895. [doi: 10.1155/2023/3939895]

4 Singh KS, Singh KJ. Network intrusion detection system using decision tree and KNN algorithm. Proceedings of the 3rd International conference on Power Electronics and IoT Applications in Renewable Energy and its Control. Mathura: IEEE, 2024. 275–280.

5 Laghrissi F, Douzi S, Douzi K, *et al*. Intrusion detection systems using long short-term memory (LSTM). Journal of Big Data, 2021, 8(1): 65. [doi: 10.1186/s40537-021-00448-4]

6 Wei WT, Gu HX, Deng WS, *et al*. ABL-TC: A lightweight design for network traffic classification empowered by deep learning. Neurocomputing, 2022, 489: 333–344. [doi: 10.1016/j.neucom.2022.03.007]

7 Ho S, Al Jufout S, Dajani K, *et al*. A novel intrusion detection model for detecting known and innovative cyberattacks using convolutional neural network. IEEE Open Journal of the Computer Society, 2021, 2: 14–25. [doi: 10.1109/OJCS.2021.3050917]

8 周璨, 杨栋, 魏松杰. 融合 GRU 和 CNN 的轻量级网络入侵检测模型. 计算机系统应用, 2023, 32(8): 162–170. [doi: 10.15888/j.cnki.csa.009194]

9 池彬, 胡辉, 周天宇, 等. 一种改进自编码器与流特征结合的入侵检测方法. 重庆理工大学学报 (自然科学), 2025, 39(13): 119–126.

10 Xu CY, Shen JZ, Du X. A method of few-shot network intrusion detection based on meta-learning framework. IEEE Transactions on Information Forensics and Security, 2020, 15: 3540–3552. [doi: 10.1109/TIFS.2020.2991876]

11 Lu CM, Wang XF, Yang AM, *et al*. A few-shot-based model-agnostic meta-learning for intrusion detection in security of internet of things. IEEE Internet of Things Journal, 2023, 10(24): 21309–21321. [doi: 10.1109/JIOT.2023.3283408]

12 Hu YL, Wu J, Li GL, *et al*. Privacy-preserving few-shot traffic detection against advanced persistent threats via federated meta learning. IEEE Transactions on Network Science and Engineering, 2024, 11(3): 2549–2560. [doi: 10.1109/TNSE.2023.3304556]

13 Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. Proceedings of the 2015 ICML Deep Learning Workshop. Lille: ILMS, 2015. 1–30.

14 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Proceedings of the 31st International

Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 4080–4090.

15 Hindy H, Tachtatzis C, Atkinson R, *et al*. Developing a siamese network for intrusion detection systems. Proceedings of the 1st Workshop on Machine Learning and Systems. ACM, 2021. 120–126. [doi: 10.1145/3437984.3458842]

16 Yang JC, Li HW, Shao S, *et al*. FS-IDS: A framework for intrusion detection based on few-shot learning. Computers & Security, 2022, 122: 102899.

17 林同灿, 葛文翰, 王俊峰. 基于对齐原型网络的小样本异常流量分类. 四川大学学报 (自然科学版), 2024, 61(3): 3–14. [doi: 10.19907/j.0490-6756.2024.030001]

18 Wu YM, Lin GY, Liu LS, *et al*. MASiNet: Network intrusion detection for IoT security based on meta-learning framework. IEEE Internet of Things Journal, 2024, 11(14): 25136–25146. [doi: 10.1109/JIOT.2024.3395629]

19 Wang YH, Zhang ZY, Zhao KJ, *et al*. A few-shot learning based method for industrial internet intrusion detection. International Journal of Information Security, 2024, 23(5): 3241–3252. [doi: 10.1007/s10207-024-00889-x]

20 Yang J, Chen X, Chen SW, *et al*. Conditional variational auto-encoder and extreme value theory aided two-stage learning approach for intelligent fine-grained known/unknown intrusion detection. IEEE Transactions on Information Forensics and Security, 2021, 16: 3538–3553. [doi: 10.1109/TIFS.2021.3083422]

21 Xu JY, Le H. Generating representative samples for few-shot classification. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 8993–9003.

22 Zekan M, Tomičić I, Schatten M. Low-sample classification in NIDS using the EC-GAN method. Journal of Universal Computer Science, 2022, 28(12): 1330–1346. [doi: 10.3897/jucs.85703]

23 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 574.

24 Zhang WX, Chen ZJ, Chen DW, *et al*. DID-IDS: A novel diffusion-based imbalanced data intrusion detection system. Proceedings of the 11th IEEE International Conference on Information, Communication and Networks. Xi'an: IEEE, 2023. 364–369.

25 Yang Y, Tang XY, Liu ZW, *et al*. Diff-IDS: A network intrusion detection model based on diffusion model for imbalanced data samples. Computers, Materials & Continua, 2025, 82(3): 4389–4408. [doi: 10.32604/cmc.2025.060357]

26 Wang P, Song YF, Wang XD, *et al*. DIFT: A diffusion-Transformer for intrusion detection of IoT with imbalanced learning. Journal of Network and Systems Management, 2025, 33(3): 48. [doi: 10.1007/s10922-025-09926-z]

27 Sharafaldin I, Lashkari AH, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy. Funchal: INSTICC, 2018. 108–116.

28 Chen YB, Liu Z, Xu HJ, *et al*. Meta-baseline: Exploring simple meta-learning for few-shot learning. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9062–9071.

29 Nguyen QH, Nguyen CQ, Le DD, *et al*. Enhancing few-shot image classification with cosine Transformer. IEEE Access, 2023, 11: 79659–79672. [doi: 10.1109/ACCESS.2023.3298299]

(校对责编: 张重毅)