面向边缘物联网的双教师蒸馏式轻量化入侵检测

韩 刚[1],常 轩[1],王耀天[1],王云俐[1],张忠良[2*]
(1. 西安邮电大学,陕西西安 710000;2. 杭州电子科技大学,浙江杭州 310000)

摘 要: 面向动态性强、协议多样且算力受限的边缘/物联网(Internet of Things,IoT)环境,本文提出一种轻量级深度学习入侵检测框架,以端到端方式统一"特征-模型-蒸馏"三层设计,在保证检测性能的同时降低计算与部署开销。针对边缘侧网络流量异构性强、攻击类别分布不均衡以及少数类样本难以充分学习等问题,框架在数据层引入自适应特征工程与动态类别均衡机制,提取流量方向不对称度、协议一致性、会话交互模式等领域相关特征,以增强对复杂攻击行为的表征能力,并缓解长尾分布对模型训练的影响。在表示学习层,构建双教师协同架构:时间分支采用增强型双向长短期记忆(Bidirectional Long Short-Term Memory,BiLSTM)网络结合多头注意力机制,用于捕获流量序列中的周期性变化、突发依赖及上下文关联;空间分支采用多尺度残差一维卷积与轻量门控模块,用于学习跨字段的局部模式与共现结构,从而提升对复杂协议语义和攻击模式的识别能力。在决策层,采用以拼接集成为主的自适应集成策略,对两路教师输出进行互补融合,并通过温度缩放软标签与真实标签联合构成蒸馏损失,将教师模型知识迁移至超轻量学生模型,在压缩模型规模的同时尽可能保持判别能力。基于场景适配的CICIDS-2017数据集实验结果表明:教师集成模型准确率达到96.5%,宏平均F_1(macro-F_1)达到95.0%;学生模型在参数量减少超过70%的条件下,仍取得93.91%的准确率和90.34%的macro-F_1,并在Patator、PortScan、Heartbleed等类别上实现超过98%的准确率,在类分布不均衡的测试条件下仍保持稳定识别。不同集成策略对比结果显示,拼接集成的macro-F_1为90.34%,优于注意力集成的90.19%、温度加权软投票的89.79%和无集成方案的88.67%,验证了该融合方式的有效性。模型端侧单次推理时延约为68 ms,满足边缘环境实时检测需求。研究结果表明:所提方法能够较好地兼顾检测精度、模型轻量化与推理效率,为边缘/IoT入侵检测提供了一种兼顾准确性与实用性的技术路径。

关键词: 物联网;入侵检测;深度学习;知识蒸馏;轻量化模型

基金项目: 国家自然科学基金(No.62102312);综合业务网国家重点实验室基金(No.ISN24-13);国家自然科学基金面上项目(No.72571079);陕西省重点研究发展计划项目(No.2024GX-YBXM-079);陕西省高校科协青年人才基金(No.20210119);陕西省高校青年创新团队(No.23JP160);陕西省高校青年创新团队(No.23JP162);

中图分类号: TP393.0;TP183 **文献标识码:** A **文章编号:** 0372-2112(2026)04-1497-16

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20251022

Lightweight Intrusion Detection in the Edge Internet of Things with Dual-Teacher Distillation

HAN Gang[1], CHANG Xuan[1], WANG Yaotian[1], WANG Yunli[1], ZHANG Zhongliang[2*]
(1. *Xi'an University of Posts and Telecommunications*, *Xi'an*, *Shaanxi* 710000, *China*;
2. *Hangzhou Dianzi University*, *Hangzhou*, *Zhejiang* 310000, *China*)

Abstract: For dynamic, protocol-diverse, and resource-constrained edge/internet of things (IoT) environments, this paper proposes a lightweight deep learning intrusion detection framework that unifies feature engineering, model construction, and knowledge distillation in an end-to-end manner, aiming to achieve an effective balance between detection accuracy, computational efficiency, and deployment feasibility. To address the strong heterogeneity of edge traffic, the severe class imbalance of attack categories, and the limited learnability of minority classes, the framework introduces adaptive feature engineering and dynamic class balancing at the data level, extracting domain-relevant features such as traffic directional asymmetry, protocol consistency, and session interaction characteristics to strengthen the representation of complex malicious behaviors while alleviating the adverse impact of long-tail distributions on model training. At the representation level, a dual-teacher architecture is designed to jointly model temporal and spatial dependencies in network traffic. The temporal branch employs an enhanced bidirectional long short-term memory (BiLSTM) network combined with multi-head attention to capture periodic patterns, bursty dependencies, and contextual correlations in traffic sequences, enabling the model to better characterize the dynamic evolution of attack behaviors over time. The spatial branch adopts multi-scale residual one-di-

收稿日期:2026-01-10;录用日期:2026-03-07;责任编辑:宋甲英
*通信作者:张忠良

mensional convolutions together with a lightweight gating mechanism to learn local feature interactions and cross-field co-occurrence structures, thereby improving the modeling of protocol semantics and hidden attack patterns across heterogeneous traffic attributes. At the decision level, an adaptive ensemble strategy is introduced, with concatenation-based fusion serving as the primary integration mechanism to combine the complementary outputs of the two teacher branches and enhance final classification performance. On this basis, the knowledge learned by the teacher ensemble is transferred to an ultra-lightweight student model through a joint distillation objective composed of temperature-scaled soft labels and ground-truth hard labels, so that the student model can preserve strong discriminative capability while substantially reducing parameter size and computational burden. Experiments conducted on a scenario-adapted CICIDS-2017 dataset demonstrate that the teacher ensemble achieves an accuracy of 96.5% and a macro-F_1 score of 95.0%, while the student model still attains 93.91% accuracy and 90.34% macro-F_1 with more than 70% parameter reduction. In addition, the student model yields over 98% accuracy on attack categories such as Patator, PortScan, and Heartbleed, and maintains stable recognition performance under imbalanced testing conditions, indicating strong robustness to class skew. The comparison results of different integration strategies show that concatenation integration achieves a macro-F_1 score of 90.34%, which outperforms attention integration at 90.19%, temperature-weighted soft voting at 89.79% and the non-integrated scheme at 88.67%, verifying the effectiveness of this fusion method. Moreover, the model exhibits an edge-side single-inference latency of approximately 68 ms, satisfying the real-time requirements of practical edge deployment. Overall, the proposed framework reduces computational and deployment costs while maintaining high detection quality, and provides a practical technical solution for intrusion detection in edge/IoT scenarios where both accuracy and efficiency are critical.

Keywords: internet of things; intrusion detection; deep learning; knowledge distillation; lightweight models

Foundation Item(s): National Natural Science Foundation of China (No.62102312); Fund of the State Key Laboratory of Integrated Services Networks (No.ISN24-13); General Program of the National Natural Science Foundation of China (No. 72571079); Shaanxi Provincial Key Research and Development Program (No. 2024GX-YBXM-079); Youth Talent Fund of the Shaanxi Association for Science and Technology in Universities (No.20210119); Young Innovation Team of Shaanxi Universities (No.23JP160); Young Innovation Team of Shaanxi Universities (No.23JP162)

0 引言

近年来,随着物联网(Internet of Things,IoT)与边缘通信等技术的快速发展,分布式异构网络在环境监测、应急响应与智慧城市等领域得到广泛应用[1-2]。然而,此类网络具有带宽受限、节点高动态与协议异构等特征,相较于传统网络更易暴露于复杂安全威胁,包括拒绝服务(Denial-of-Service,DoS)通过突发流量瘫痪控制链路、端口扫描导致资源受限设备的敏感信息泄露、协议异常破坏异构节点间通信[3-4]三种典型情形。依据国际电信联盟(International Telecommunication Union,ITU)2024 年报告,针对这类网络的攻击年增长率已达47%,其中60%以上为面向资源受限设备的定向攻击。传统入侵检测系统(Intrusion Detection System,IDS)在此环境下因难以适配上述特性而表现受限。最新综述对基于机器学习的加密流量分析进行了系统梳理,归纳出采集、表征、方法与评测的端到端工作流,并按资产识别、网络刻画、隐私泄露检测与异常检测四个目标对研究进行分类,同时指出,分类粒度与时效性的权衡是加密与异构条件下的关键瓶颈[5]。

现有入侵检测研究主要面临三方面挑战:

(1)数据特征适配。此类网络常呈现非对称通信与突发小包等模式,传统特征工程难以有效表征。同时,攻击样本呈长尾分布,常见攻击(如DoS)占比高,而新型攻击(如畸形协议报文)样本稀缺,削弱模型对罕见威胁的检出能力[6-7]。在加密流量场景中,已有工作尝试“多视角特征+图表示”的联合建模,如MalDiscovery以会话的图像特征构造属性k近邻(k-Nearest Neighbor,KNN)图,将向量化特征作为节点属性,并使用GraphSAGE(Graph Sample and Aggregate)对相邻会话进行相关性聚合以丰富嵌入,最终完成图上二分类并在对比实验中报告约99.9%的准确率[8]。GraphDApp提出流量交互图(Traffic Interaction Graph,TIG)以保留双向客户端-服务端交互的多维特征,将去中心化应用(Decentralized Application,DApp)指纹识别转化为图分类问题,构造为包含1 300个DApp、16.9万条流的真实数据集,在封闭世界与开放世界两类评测中均优于现有方法,并可迁移至移动应用识别任务[9]。然而,上述方法在跨场景迁移、在线推理效率与端侧部署代价方面仍受制约,难以直接满足资源受限设备上的实时检测需求。

(2)特征学习受限。攻击行为兼具时间依赖(如周期性恶意流量)与空间关联(如异常字段组合)。单一模型架构仅采用长短期记忆(Long Short-Term Memory,LSTM)网络或仅采用卷积神经网络(Convolu-

tional Neural Network,CNN)难以同时捕获两类特征,导致检测效果受限[10-11]。同时,近期研究还表明深度模型易受对抗扰动影响。FaceAdv通过生成-转换两阶段的对抗贴纸在物理世界实现对人脸识别系统的躲避与冒充攻击,在ArcFace、CosFace与FaceNet上显著提升成功率,并在多环境条件下表现出更强鲁棒性[12],这进一步凸显入侵检测对模型稳健性与泛化能力的需求。

(3)算力与资源约束。边缘设备的计算与存储能力有限[13]。复杂模型虽可提升精度,但难以满足实时处理需求。相对地,轻量模型在应对多样且复杂的现代网络攻击时,往往容量不足[14-15]。

为应对上述挑战,本文提出一种基于双教师模型集成与知识蒸馏(Knowledge Distillation,KD)的轻量级入侵检测框架:首先,通过自适应数据预处理构建面向场景的特征(如流量方向不对称度、协议一致性评分),并引入动态类别均衡策略,依据各类别分布自适应选择过采样算法并配合按类别评估机制,使罕见攻击的可检测性可度量;其次,通过构建多模态教师模型集成架构,设计两个互补教师——EnhancedTimeSensitiveModelV2(侧重时序依赖)与EnhancedFeatureSpaceModelV2(侧重字段与通道关联),通过加权融合与注意力融合联合建模,刻画复杂攻击中的时间与空间耦合模式;最后,提出面向轻量化部署的高效KD方法,将教师集成的判别能力迁移至超轻量学生模型,并采用温度缩放的混合损失(软标签与硬标签联合)与批级优化(logits缓存)以减少冗余计算,在保持较高检测性能的同时显著降低参数规模与推理时延,从而实现资源受限设备上的实时运行。

本文的主要贡献如下:

(1)提出自适应特征工程与类别均衡方案,面向高动态、异构网络环境,缓解特征不可迁移与样本不均衡问题。

(2)提出双教师自适应集成策略,联合学习时序与空间特征,提高对多类型攻击的检测准确度。

(3)提出资源约束场景的KD框架,在实验中,学生模型参数量相较教师侧整体减少超过70%,检测准确率下降小于3%,满足实时检测需求。

(4)面向少数类与罕见模式的鲁棒检测,结合动态重加权与类别级评测并借助自适应集成,使低频类别的识别显著提升,并同时保持常见类别的高精度与高召回。

1　相关工作

1.1　数据处理与特征工程

面向IoT IDS的加密流量与异构协议环境,数据表征与评测是基础环节。综述性研究对“采集-表征-方法-评测”的端到端流程进行了系统梳理,Xu等人[16]指出在加密与异构条件下,表征适配与时效性仍是当前难点。在高维稀疏与特征提取困难方面,离散哈希方法通过监督离散哈希学习等长二进制编码,并以汉明距离实现快速检索与判别,在CIC-IDS2017、NSL-KDD、UNSW-NB15三个数据集上取得了超过97%的准确率并显著降低误报。然而,该类方法对时序依赖与类别长尾的覆盖仍然有限。薛胤等人[17]针对类不均衡与类重叠,提出NaNDS数据集联合密度估计与自然近邻方案,从几何层面剔除不利的多数类重叠样本,并以密度与熵自适应确定少数类过采样权重,在KEEL数据集上优于多种先进算法,但缺乏按类别与未知类的一体化评测与端到端在线验证。Li等人[18]受傅里叶变换启发提出TFTKD模型,将时域灰度图映射至频域以缓解稀疏,并结合KD得到轻量学生网络,在多数据集上表现优良,但“图像化”编码与频域增强对协议语义与跨字段结构的保持仍需量化评估。此外,Wang等人[19]基于空天地一体化网络(Space-Air-Ground Integrated Network,SAGIN)方向的研究通过半监督与端到端异常流量检测降低人工标注与特征工程成本,验证了可行性与先进性,但多基于仿真环境,真实部署与跨场景泛化仍有不足。本文对此的改进是通过构建场景自适应特征工程,以流量不对称度与协议一致性评分为核心,兼顾可迁移性与实现成本。同时,引入动态类别均衡与类别级评测(含Unknown占位),在“训练-评测”闭环中量化稀有与未知威胁可检出性,弥补既有方法在长尾、未知类与在线性方面的不足。

1.2　模型架构与特征融合

在高动态流量条件下同时捕获时间依赖与特征空间结构,是提升检出率的关键。曹磊等人[20]提出的卷积神经网络-双向简单循环单元(Convolutional Neural Network-Bidirectional Simple Recurrent Unit,CNN-BiSRU)模型以卷积提取局部模式并以双向简单循环单元替代双向长短期记忆(Bidirectional Long Short-Term Memory,BiLSTM),在CICIDS2018上兼顾检测精度与推理速度,但对跨字段共现模式与复杂融合策略的系统性比较仍显不足。Cao等人[21]面向表征学习,提出面向基于Transformer的双向编码器表示(Bidirectional Encoder Representations from Transformers,BERT)的KD-BERT方案,其通过“报文头-载荷对”分词驱动BERT,并以KD将135 M参数压缩至约9 M,在TON-IoT、Edge-IIoTset及IoMT2024等数据集上取得领先性能。然而,对类别不均衡与未知类占位的细粒度评测不够充分。Xie等人[22]在双教师蒸馏方面,提出

双教师知识蒸馏入侵检测系统(Dual-Teacher Knowledge Distillation-Intrusion Detection System, DTKD-IDS)方法,通过原型蒸馏与互补蒸馏增强学生模型吸收效率,在 X-IIoTID、NSL-KDD 和 CICDDoS2019 等数据集上优于多种深度模型,但对融合策略选择与"时序-空间互补"机制的可解释性披露有限。此外,李波等人[23]在无监督图像异常检测任务中提出基于双向约束蒸馏的教师-学生框架,通过双向蒸馏分支对正常特征进行高效对齐,并借助多级过滤模块抑制异常干扰特征,从而提升异常检测与定位性能。该研究表明,蒸馏过程不仅可用于模型压缩,还可作为增强正常模式表征与异常分离能力的重要机制,但其研究对象主要面向视觉场景,在 IoT 加密流量与异构协议环境下的适用性仍有待验证。Chai 等人[24]在分布式与跨端依赖场景,提出融合时间卷积网络(Temporal Convolutional Network, TCN)与 Transformer 架构的深度学习模型(TCN-Transformer),结合时间卷积网络的局部建模与 Transformer 的全局依赖,并以联邦学习的精度加权聚合提升鲁棒性。然而,对非独立同分布、类不均衡与未知类评测的系统支持仍待加强。Ahmad 等人[25]在灾害态车联网场景中,使用离散时间马尔可夫链-机器学习方案将离散时间马尔可夫链与机器学习联合建模,以对数似然与受试者工作特征(Receiver Operating Characteristic, ROC)评估实现异常检测,但特征空间可迁移性与端到端实时性指标披露不充分。本文对此的改进是提出双教师自适应集成架构,以增强型 BiLSTM 关注时间依赖、以多尺度残差一维卷积关注字段共现。在可学习加权、注意力门控、拼接多层感知器(Multi-Layer Perceptron, MLP)与投票四类候选中,以验证集宏平均 F_1(macro-F_1)进行择优,系统对比了不同融合策略对多类别平衡与稳健性的影响。

1.3 资源感知与轻量化

边缘设备算力受限,需在精度、时延与能耗之间取得平衡。Medhi 等人[26]提出的无人机分布式隐私保护网络入侵检测(Unmanned Aerial Vehicle-Distributed and Privacy-preserving Network Intrusion Detection, UAV-DiPNID)方案将 KD 与剪枝用于入侵检测,在数据集 UAV-IDS 2020 上达到 99.61% 平均准确率,推理时间最多下降 80.70%,模型体积最多压缩 90%。但研究主要集中于单一数据域,对 macro-F_1、未知类与少数类召回的披露不足。Wang 等人[27]提出一种云-边协同的自适应蒸馏方案,并在 Jetson Xavier NX 上实现准确率提升 10.7%、推理速度提升 25.5%、内存占用降低 96.58%、处理速度提升 8.79 倍,展示了跨端知识迁移的部署价值,但其任务域为机械诊断,尚需在加密流量与协议语义场景验证可迁移性。Quyen 等人[28]在联邦学习方向,联邦知识蒸馏入侵检测系统(Federated Knowledge Distillation-Intrusion Detection System, FedKD-IDS)以蒸馏与投票替代参数聚合并引入抗投毒机制,在数据集 N-BaIoT 上显著优于半监督联邦学习(Semi-Supervised Federated Learning, SSFL)。面对 50% 标签翻转仍达 79% 准确率,并可在聚合阶段剔除超过 85% 的恶意协作者,但对通信开销、收敛轮次与更强威胁模型的量化与鲁棒性评估尚不完备。Wang 等人[29]的研究则是面向低时延业务的计算卸载,使数据在边缘节点间的传输与处理链路更加复杂,数据复制与异构资源协同会带来新的安全与隐私风险。因而,面向部署的轻量化 IDS 不仅要降低模型开销,还应在边缘侧提供稳定可靠的异常检测能力,以适应卸载场景下更动态的威胁形态。总体而言,深度学习在 IoT IDS 中较传统方法在高维表征与时序依赖建模方面更具优势,但仍受标注稀缺、边缘资源与动态攻击的制约,研究趋势正指向轻量化与跨模态融合[30]。本文对于以上问题的改进方法是设计超轻量学生模型与温度缩放的软硬混合蒸馏,配合 logits 批级缓存以降低训练与通信开销。在参数降低超过 70% 的前提下,将精度损失控制在 3% 以内,模型端侧单次推理时延约 68 ms,并以类别级评测确保轻量化不以牺牲稀有与未知威胁可检出性为代价。

1.4 小结与差异化定位

综上所述,现有研究在数据表征与评测、时空特征融合以及资源感知轻量化方面均取得进展,但仍存在三类共性不足:其一,加密与异构条件下的可迁移表征与按类别(含未知类)可检出性的量化仍不充分;其二,时间依赖与特征空间结构的互补建模与融合策略缺少系统比较与可解释选择;其三,面向边缘设备的蒸馏与压缩虽能降本增效,但在宏平均指标、少数类召回与端到端时延上的约束处理尚未完善。针对上述缺口,本文以场景自适应特征工程与动态类别均衡为数据底座,引入双教师自适应集成以显式分工"时序-空间"表征,并通过温度缩放的软硬混合蒸馏与批级 logits 缓存实现超轻量学生模型训练,在参数压缩超过 70% 的同时,将精度损失控制在 3% 以内并满足毫秒级推理时延。由此,本文为资源受限环境下的实时入侵检测提供了一条可复用、可部署且具可量化评测闭环的技术路径,为后续的大规模应用与跨场景迁移奠定基础。

2 轻量级双教师蒸馏入侵检测框架

本文提出的轻量级入侵检测框架,如图 1 所示。框架包含自适应数据预处理、多模态教师集成与高效

KD三部分。在数据表示阶段,将按时间到达的$N\times F$特征序列规整为长度为L的张量序列,并同步完成异常值处理、尺度对齐与协议一致性校验,得到可用于时空建模的规范化输入。随后进入教师端,特征空间教师(Feature-Space Teacher Model,FSTM)采用多尺度残差卷积与通道/空间门控以捕获字段共现与局部结构;时间敏感教师采用双向循环网络与注意力机制以建模周期、突发与顺序依赖。两路教师的输出通过可学习加权或注意力门控进行自适应融合,形成校准后的类别分布。为满足边缘侧的实时与资源约束,蒸馏阶段以轻量学生模型为目标,采用温度缩放的软标签与真实硬标签构成混合损失。通过对温度T与损失权重α的调度,使早期更多依赖软分布以学习细粒度类间关系,后期逐步强调硬标签以稳固决策边界。在实现上,学生网络使用小型化主干与全局平均聚合,并引入批级logits缓存以减少冗余前向,从而在保证精度的同时显著降低参数规模与推理时延。系统输出分类结果,并以混淆矩阵与ROC曲线进行评测,完成"采集-表示-教师集成-蒸馏-评测"的闭环。

2.1　自适应数据预处理机制

本文选用CICIDS-2017数据集作为研究基础[31-32]。该数据集由加拿大网络安全研究所(Canadian Institute for Cybersecurity,CIC)构建,涵盖丰富的正常流量与攻击场景,包括DoS(如Hulk、GoldenEye)、分布式拒绝服务(Distributed Denial-of-Service,DDoS)、端口扫描(PortScan)、Web攻击如结构化查询语言注入(SQL Injection,SQLi)、跨站脚本(Cross-Site Scripting,XSS)以及Patator暴力破解(brute-force)等。数据集同时提供流持续时长、报文大小、协议类型、标志位计数等细粒度特征,可较全面地模拟真实网络环境中的流量分布,因此被广泛用于入侵检测研究。然而,直接使用原始CICIDS-2017在资源受限、协议多样的边缘网络入侵检测任务中适配性不足:其一,流量类别分布高度不均衡,正常样本数量远超攻击样本[33];其二,不同攻击类型之间样本规模差异显著,易导致模型偏置与检出能力不稳定。针对带宽受限、协议差异化与样本不均衡等特征,本文设计了一条多级数据预处理流水线[34],作为后续模型学习的高质量输入底座。该流水线包含三类核心模块。

(1)特征构建(Feature Construction):基于场景相关流量模式构造流量不对称性指标与协议特征评分等可迁移特征,并与原始统计特征(时长、字节数、标志位等)按字段规范化后联合使用。

(2)样本均衡(Sample Balancing):引入动态类别均衡策略,依据类别占比与难例密度自适应选择过/欠采样与难例挖掘组合,以缓解长尾分布对训练与验证的一致性冲击。

(3)评测校准(Evaluation Calibration):建立类别级(class-wise)评测与可检出性(detectability)量化机制,将稀有攻击在验证/测试阶段的检出能力显式度量并纳入阈值与损失调度闭环,避免仅凭总体指标掩盖少数类退化。

上述预处理机制与后续的双教师集成、KD训练协同运行,前者提供可分可衡量的数据表征与评测信号,后者在此基础上完成时空互补表征学习与资源感知轻量化,共同支撑边缘环境中的实时入侵检测。

2.1.1　场景自适应特征工程

边缘/IoT网络通常呈现终端资源受限、传输带宽有限、链路相对不稳定等特征[35]。围绕这些特性开展特征工程,可有效提升IDS在此类环境中的识别

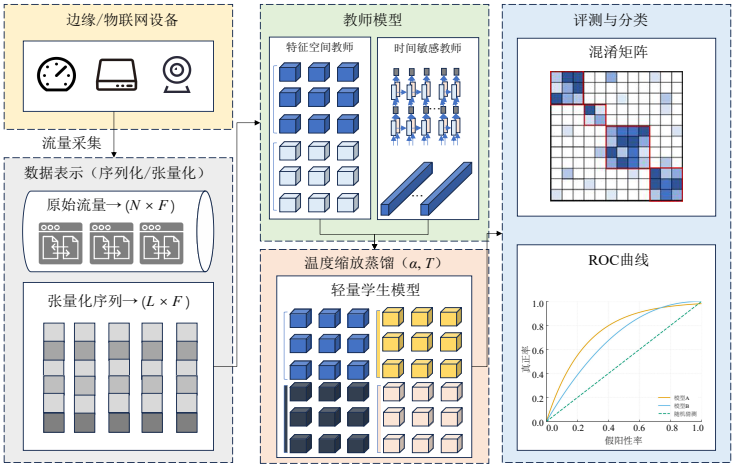

图1　系统总体框架与数据流

Figure 1　Overall architecture and data flow of the system

能力。

(1)原始网络特征的场景化改造

该部分包括协议特征和流持续时长(flow duration)两种特征类型。

协议特征指鉴于用户数据报协议(User Datagram Protocol,UDP)在此类环境中因其无连接特性被广泛采用,而传输控制协议(Transmission Control Protocol,TCP)的使用相对受限,需对协议类型进行有偏筛选:优先保留与UDP/TCP相关的关键字段,并剔除低频/非常见协议,以降低无关信息干扰;而流持续时长是指由于边缘场景的数据传输通常跨越时间短、波动性强,原始持续时长分布可能出现尺度差异与高方差。为稳定方差并减少极值影响,需要进行对数变换。其变换定义为

$$\text{Flow Duration} = \ln\left(\text{Flow Duration} + 1\right) \tag{1}$$

该变换可压缩取值范围,使该特征与网络的时间特性更为一致。就报文大小特征而言,数据包尺寸通常受设备能力与传输链路约束,异常偏大的数据包可能指示异常。因此,对前向报文长度均值(Fwd Packet Length Mean)与后向报文长度均值(Bwd Packet Length Mean)等特征实施截断,对相关特征进行基于第95百分位的上截断处理,将超过该分位点的取值统一裁剪为第95百分位对应的阈值,以减弱极端值对模型训练与统计量估计的干扰。截断方式定义为

$$\text{Feature} = \min\left(\text{Feature}, \text{Quantile_95}(\text{Feature})\right) \tag{2}$$

同时,诸如Init_Win_bytes_forward与Init_Win_bytes_backward等窗口大小特征限定在65 535以内,即Feature = min(Feature, 65 535)以满足窗口尺寸约束并抑制异常极值对统计的影响。

(2)场景自适应特征构建

流量不对称性特征用于度量双向数据流的不均衡程度。在部分边缘/IoT中,由于设备通信具有较强的单向性,该特征尤为关键。其计算式为

$$\text{Flow_Asymmetry} = \frac{\text{Total LFP} - \text{Total LBP}}{\text{Total LFP} + \text{Total LBP} + 1 \times 10^{-6}} \tag{3}$$

为避免分母为零,在分母处加入一个微小常数(1×10^{-6})。报文大小变异特征用于刻画报文大小的稳定性,其计算式为

$$\text{Packet_Size_Variation} = \frac{\text{FPL Std} + \text{BPL Std}}{\text{FPL Mean} + \text{BPL Mean} + 1\times10^{-6}} \tag{4}$$

该特征能够有效识别报文大小的异常波动,此类波动可能指示边缘/IoT网络中的攻击行为。同时,提出场景化协议评分特征(Protocol Score,PS),将协议类型与标志位信息进行融合,其计算式为

$$\begin{aligned}&\text{Low_Altitude_Protocol_Score}\\&= \left(\text{Protocol} == 17\right) \times 0.6\\&\quad + \left(\text{FIN Flag Count} + \text{SYN Flag Count}\right) \times 0.2\\&\quad + \left(\text{ACK Flag Count}\right) \times 0.2\end{aligned} \tag{5}$$

其中,Protocol == 17表示UDP协议,权重取0.6,用于体现UDP的重要性。标志位(Flag)信息用于刻画连接的建立与终止状态。到达间隔时间(Inter-Arrival Time,IAT)的变异性特征用于衡量时间间隔的稳定性,其计算式为

$$\text{IAT_Variability} = \frac{\text{Flow IAT Std} + \text{Fwd IAT Std} + \text{Bwd IAT Std}}{3} \tag{6}$$

该特征能够捕捉由网络链路不稳定引起的时间间隔波动,有助于识别异常通信模式。针对无线通信的特性,本文构建三类无线侧特征,其一为信噪比(Signal-to-Noise Ratio,SNR)的滑动窗口变异系数,计算式为

$$\text{SNR_Variation} = \frac{\text{std}(\text{SNR}_{\text{window}})}{\text{mean}(\text{SNR}_{\text{window}}) + 1\times10^{-6}} \tag{7}$$

该特征用于刻画无线信道的稳定性,突发异常可能指示蓄意干扰(jamming)攻击。进一步地,提取802.11p/4G协议的重传率,其定义为

$$\text{Frame_Retry_Rate} = \frac{\text{Retry_Frame_Count}}{\text{Total_Frame_Count} + 1\times10^{-6}} \tag{8}$$

恶意节点可能通过伪造帧首部或频繁重传发起攻击。为进一步刻画无线侧异常行为,本文同时引入首部异常率、重传率、到达时间差(Time Difference of Arrival,TDoA)和频偏等特征。其中,TDoA用于表征信号传播时延特性;频偏则反映节点相对运动引起的多普勒效应,其绝对值可表示为

$$\text{Doppler_Shift} = \left|\frac{f_c \cdot v \cdot \cos\theta}{c}\right| \tag{9}$$

其中,f_c表示载波频率;c表示光速;v表示节点运动速度;θ表示传播角。异常的频偏可能对应未授权节点伪装基站实施的欺骗攻击。通过对原始特征的自适应调整与上述新特征的构建,形成了一套契合目标网络环境特性的特征集合,为后续入侵检测模型提供更具判别力的输入,从而提升对各类攻击的检测准确性。

2.1.2 动态类别均衡策略

在网络入侵检测场景中,流量数据的类别分布往往明显不均衡[36]:正常样本数量远高于各类攻击样本,且不同攻击类别之间的规模差异显著。此类不均衡会使模型在训练中偏向多数类,从而削弱对少数类攻击的检出能力。因此,有必要引入动态类别均衡以优化数据分布。

该策略的核心在类别粒度上,依据各类样本规模

自适应地调整采样方式与目标样本量。一方面避免对少数类的过度合成而引入低质量样本,另一方面防止对多数类的过度下采样造成关键信息丢失。具体而言:先统计每一类别的原始样本数,记为count;随后基于类别规模动态确定该类别的目标样本量,记为target。对于多数类,若count > target,则执行下采样,随机保留target个样本以保持代表性,其过程可定义为

$$\text{Sampled_samples}=\text{Random_sample}\left(\text{Class_samples},\text{target}\right) \tag{10}$$

其中,Class_samples指该类别的全部样本。对于少数类,若样本数count小于目标值target,则根据样本规模选择合适的过采样算法:

(1)当count < 100(极少)时,采用ADASYN过采样算法[37],并设置较小的k_neighbors,利用少数类的局部分布生成合成样本,k_neighbors = min(3, count − 1)。

(2)当100 ≤ count ≤ 500(较少)时,采用边界线合成少数类过采样技术(Borderline Synthetic Minority Over-sampling TEchnique, Borderline-SMOTE)算法[38],设置中等k_neighbors,侧重在决策边界处合成样本以增强类间可分性,k_neighbors = min(5, count − 1)。

(3)当count > 500时,采用常规合成少数类过采样技术(Synthetic Minority Over-sampling Technique, SMOTE)算法[39],设置较大的k_neighbors进行标准过采样,k_neighbors = min(7, count − 1)。

而对于目标样本量target的确定应遵循以下原则:对于多数类,目标样本量不应超过预设的最大阈值(例如900 000),且不应小于5 000。其公式为

$$\text{target}=\min\left(\text{max_samples},\ \max\left(5\,000,\ \text{int}\left(\text{count}\times 2\right)\right)\right) \tag{11}$$

其中,max_samples表示预设的最大样本阈值。该设置可避免因多数类样本过多而引发的模型偏置,同时确保各类别均具有足够样本以支撑模型学习。对需要过采样的类别,先计算应合成的样本数n_samples = target − count。当n_samples为正时,按所选过采样算法生成相应数量的样本,并尽可能保持与原始数据的分布一致性。

当过采样过程中出现异常时,采用加噪复制(noise-added replication)作为回退策略:即先在原始样本上加入微小噪声(Noise = Class_samples + Normal)以增加多样性,再重复复制,直至达到目标样本量。其过程定义为

$$\text{Noisy_samples}=\text{Concatenate}\left(\left[\text{Class_samples},\text{Noise}\right]\times\text{repeat}\right) \tag{12}$$

其中,repeat表示复制次数。若复制后的样本数超过目标值,则采用随机抽样将样本数量降至目标规模。在完成各类别的采样后,将各类别处理后的样本合并以构建均衡数据集;最终对合并后的数据集进行随机打乱,其过程定义为

$$\text{Balanced_dataset}=\text{Shuffle}\left(\text{Concatenate}\left(\text{Processed_classes}\right)\right) \tag{13}$$

该随机化可消除数据顺序对模型训练的影响。通过上述动态类别均衡策略,可将各类别的样本规模维持在合理区间,既保留多数类信息的完整性,又提升少数类样本的代表性,从而为后续模型训练提供更稳健的数据基础。

2.1.3　类别级综合评测机制

为克服传统评测仅关注“已见类别”的局限,本文设计了一套覆盖九类已知攻击的综合评测体系,如表1所示,即便某一攻击类型在数据集中未出现,也通过将缺失类别映射为“Unknown”的方式将其纳入混淆矩阵与分类报告,从而保证评测的完整性。

表1　系统总体框架与数据流

Table 1　Overall system framework and data flow

解释含义	统一后的攻击标签	原始攻击标签
正常	BENIGN	BENIGN
DoS	DoS	DoS Hul DDoS DoS GoldenEye DoS slowloris DoS Slowhttptest
口令爆破	Patator	FTP-Patator SSH-Patator
Web攻击	Web Attack	Web Attack Sql Injection Web Attack XSS Web Attack Brute Force
端口扫描	PortScan	PortScan
僵尸网络	Bot	Bot
渗透	Infiltration	Infiltration
心脏出血漏洞	Heartbleed	Heartbleed
未知	Unknown	Unknown

2.2　多模态教师模型与自适应集成策略

在2.1节完成数据预处理与特征构建后,面向资源受限的IoT场景,本文设计双教师模型以联合建模时间-空间模式,结构如图2所示。体系由时序敏感教师(Temporally-Sensitive Teacher Model, TSTM)与FSTM构成:TSTM侧重周期/突发等序列依赖,FSTM强调跨字段共现与多尺度结构。两类教师分别产生类别概率p_{T}与p_{F},并通过自适应集成(温度加权软投票、拼

接-MLP或注意力/门控加权)。在样本级融合,得到融合概率向量$\boldsymbol{p}_{\text{fuse}}$作为最终判别结果。

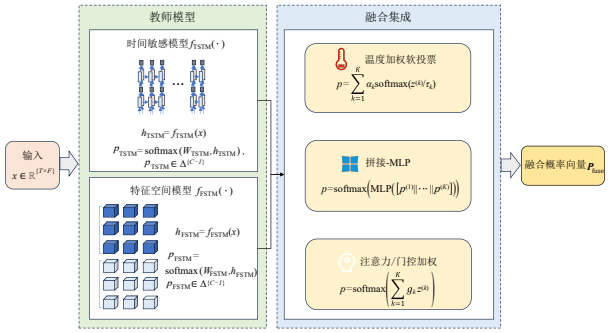

图2 双教师模型与自适应融合机制

Figure 2 The dual-teacher model and adaptive fusion mechanism

2.2.1 TSTM

为刻画入侵行为的周期性、突发性与顺序依赖,TSTM以“BiLSTM→自注意力→多尺度时间池化”的层次化结构获得判别性的时序表示。设输入为归一化后的特征序列$x\in\mathbb{R}^{T\times F}$(长度$T$、每步维度$F$),则TSTM的各子模块可形式化为:

(1)双向时序编码

前、后向LSTM分别产生隐藏状态,拼接后得到时间步表征:

$$\overrightarrow{h}_t=\text{LSTM}_f(\boldsymbol{x}_t,\overrightarrow{h_{t-1}}) \tag{14}$$

$$\overleftarrow{h}_t=\text{LSTM}_b(\boldsymbol{x}_t,\overleftarrow{h_{t+1}}) \tag{15}$$

$$\boldsymbol{h}_t=[\overrightarrow{h}_t;\ \overleftarrow{h}_t] \tag{16}$$

堆叠得$\boldsymbol{H}=[\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_T]^{\text{T}}\in\mathbb{R}^{T\times d_h}$。为稳定长序列训练,采用残差与层归一化:

$$\tilde{\boldsymbol{H}}=\text{LayerNorm}\left(\boldsymbol{H}+\phi(\boldsymbol{H})\right) \tag{17}$$

其中,$\phi(\cdot)$为线性投影与丢弃的组合。

(2)自注意力聚合

用可学习向量$\boldsymbol{u}$产生时间权重α_t,实现关键时刻的加权汇聚:

$$\alpha_t=\frac{\exp(\boldsymbol{u}^{\text{T}}\widetilde{\boldsymbol{h}}_t)}{\sum_{j=1}^{T}\exp(\boldsymbol{u}^{\text{T}}\widetilde{\boldsymbol{h}}_j)} \tag{18}$$

$$\boldsymbol{c}=\sum_{t=1}^{T}\alpha_t\widetilde{\boldsymbol{h}}_t\in\mathbb{R}^{d_h} \tag{19}$$

(3)多尺度时间池化

本文分别对短、中、长三种尺度窗口w_s,w_m,w_l内的特征表示进行均值池化,并将各尺度池化结果拼接得到多粒度统计量$\boldsymbol{m}$,其定义为

$$\boldsymbol{m}=\|_{w\in\boldsymbol{W}}\text{AvgPool}_w(\tilde{\boldsymbol{H}})\in\mathbb{R}^{d_m} \tag{20}$$

其中,$\|$表示拼接。最终的时序表示为

$$\boldsymbol{h}_{\text{TSTM}}=f_{\text{TSTM}}(\boldsymbol{x})\triangleq\psi\left(\left[\boldsymbol{c}\|\boldsymbol{m}\right]\right)\in\mathbb{R}^{d_{\text{TSTM}}} \tag{21}$$

其中,$\psi(\cdot)$为线性映射与激活的轻量组合。

(4)分类输出

通过线性分类头与softmax给出类别分布:

$$\boldsymbol{p}_{\text{TSTM}}=\text{softmax}(\boldsymbol{W}_{\text{TSTM}}\boldsymbol{h}_{\text{TSTM}}),\quad\boldsymbol{p}_{\text{TSTM}}\in\Delta^{C-1} \tag{22}$$

其中,Δ^{C-1}为C类概率单纯形。

该设计在保证计算开销可控的同时,能集中捕获短期突发与长期周期两类关键时序线索,为后续自适应融合提供稳健的时序证据。

2.2.2 FSTM

为学习协议字段、标志位、长度统计等在特征空间的局部结构与共现关系,FSTM采用“多尺度深度可分离卷积→残差瓶颈→通道/空间门控→金字塔池化”的层次化设计,在保证判别力的同时显著降低参数量与时延。设输入为经预处理后的序列特征矩阵$\boldsymbol{x}\in\mathbb{R}^{T\times F}$(长度$T$、每步特征维度$F$),其关键计算可形式化为。

(1)多尺度深度可分离卷积+残差瓶颈

在特征轴上采用不同感受野$\mathcal{K}=\{k_1,k_2,k_3\}$的一维深度可分离卷积(Depthwise+Pointwise),得到多分辨率响应并经瓶颈整合:

$$\boldsymbol{y}_k=\text{PW}\left(\delta\left(\text{DW}_{k,d}(\boldsymbol{x})\right)\right),\quad k\in\kappa \tag{23}$$

$$\boldsymbol{z}=\text{LayerNorm}\left(\boldsymbol{x}+\text{Conv}_{1\times1}\left(\|k\in\kappa\boldsymbol{y}_k\right)\right) \tag{24}$$

其中,$\delta(\cdot)$为非线性激活;d为可选空洞率;$\|$表示拼接。该模块以较低开销覆盖短/中/长感受野并通过残差稳定训练。

(2)轻量通道/空间门控

采用SE型通道注意力与卷积式空间注意力,强化判别性字段组合与跨特征交互:

$$\boldsymbol{a}_c=\sigma\left(\boldsymbol{W}_2\delta\left(\boldsymbol{W}_1\text{GAP}(\boldsymbol{z})\right)\right) \tag{25}$$

$$\boldsymbol{a}_s=\sigma\left(\text{Conv}_{1\times k}\left([\text{Avg}(\boldsymbol{z});\ \text{Max}(\boldsymbol{z})]\right)\right) \tag{26}$$

$$\boldsymbol{z}'=(\boldsymbol{a}_c\odot\boldsymbol{z})\odot\boldsymbol{a}_s \tag{27}$$

其中,$\sigma(\cdot)$为Sigmoid;GAP为全局平均池化(Global Average Pooling,GAP);$\odot$为逐元素乘。

(3)金字塔池化与表示汇聚

对尺度集合$\mathcal{R}=\{r_1,r_2,r_3,r_4\}$进行金字塔池化并拼接,获得多粒度统计量:

$$\boldsymbol{h}_{\text{FSTM}}=\psi\left(\text{GAP}(\boldsymbol{z}')\|\left\{\text{PP}_r(\boldsymbol{z}')\right\}_{r\in\mathbb{R}}\right)\in\mathbb{R}^{d_{\text{FSTM}}} \tag{28}$$

其中,$\psi(\cdot)$为线性投影与归一化的轻量组合。

(4)分类输出

$$\boldsymbol{h}_{\text{FSTM}}=f_{\text{FSTM}}(\boldsymbol{x}) \tag{29}$$

$$\boldsymbol{p}_{\text{FSTM}}=\text{softmax}(\boldsymbol{W}_{\text{FSTM}}\boldsymbol{h}_{\text{FSTM}}),\quad\boldsymbol{p}_{\text{FSTM}}\in\Delta^{C-1} \tag{30}$$

相较于单尺度或纯MLP表征,FSTM能够同时捕

获局部结构与跨字段共现,并在深度可分离与瓶颈设计的加持下保持轻量化与低时延,为后续自适应融合提供稳定的特征空间证据。

2.2.3　自适应集成

为在样本级动态权衡两路教师的贡献,本文引入门控融合。设 TSTM 与 FSTM 的隐藏表示分别为 $\boldsymbol{h}_{\mathrm{TSTM}}\in\mathbb{R}^{d_{\mathrm{TSTM}}}$、$\boldsymbol{h}_{\mathrm{FSTM}}\in\mathbb{R}^{d_{\mathrm{FSTM}}}$,对应的对数几率与温度缩放概率为 $\boldsymbol{z}^{(t)}\in\mathbb{R}^{C}$, $\boldsymbol{p}^{(t)}=\mathrm{softmax}(\boldsymbol{z}^{(t)}/\tau_{\mathrm{TSTM}})$ 与 $\boldsymbol{z}^{(s)}\in\mathbb{R}^{C}$, $\boldsymbol{p}^{(s)}=\mathrm{softmax}(\boldsymbol{z}^{(s)}/\tau_{\mathrm{FSTM}})$。且为避免固定融合权重在不同样本与不同类别上的不适配,采用样本级自适应加权,使融合权重随输入变化而动态调整:当某一路教师对当前样本给出更清晰、更稳定的表征时,融合结果更偏向该分支;当两者信息互补时,则通过平衡加权获得更稳健的综合判别。基于上述动机,门控权重的计算如下。

(1)门控权重

以拼接向量 $\boldsymbol{c}=[\boldsymbol{h}_{\mathrm{TSTM}};\ \boldsymbol{h}_{\mathrm{FSTM}}]\in\mathbb{R}^{d_{\mathrm{TSTM}}+d_{\mathrm{FSTM}}}$ 为输入,计算两路分支的原始门值:

$$\boldsymbol{s}=[s_{\mathrm{TSTM}},s_{\mathrm{FSTM}}]=\boldsymbol{W}_{\mathrm{g}}\boldsymbol{c}+\boldsymbol{b}_{\mathrm{g}}\in\mathbb{R}^{2}\tag{31}$$

并归一化得到样本级权重:

$$\boldsymbol{\alpha}=\mathrm{softmax}(\boldsymbol{s})=[\alpha_{\mathrm{TSTM}},\alpha_{\mathrm{FSTM}}]\tag{32}$$

$$\alpha_{\mathrm{TSTM}}+\alpha_{\mathrm{FSTM}}=1,\quad \alpha_{\mathrm{TSTM}},\alpha_{\mathrm{FSTM}}\in[0,1]\tag{33}$$

式(31)的意义在于先用两路教师的隐藏表示拼接得到 $\boldsymbol{c}=[\boldsymbol{h}_{\mathrm{TSTM}};\ \boldsymbol{h}_{\mathrm{FSTM}}]$,再经一层线性映射得到二维打分向量 $\boldsymbol{s}=[s_{\mathrm{TSTM}},s_{\mathrm{FSTM}}]$。这里的 s_{TSTM} 与 s_{FSTM} 可理解为当前样本上两路分支的相对可信度打分,并不直接等价于类别概率,而是对分支贡献的粗粒度评价。式(32)则对 $\boldsymbol{s}$ 进行 softmax 归一化得到 $\boldsymbol{\alpha}=[\alpha_{\mathrm{TSTM}},\alpha_{\mathrm{FSTM}}]$,从而保证两路权重非负且和为 1,如式(33)所示。由于 $\boldsymbol{\alpha}$ 由样本的联合表征 $\boldsymbol{c}$ 决定,因此它实现了样本级动态加权,可在不同输入上自适应调整两路教师的融合比例。

(2)融合输出

本文在同一门控框架下考察三类候选融合,并在验证集上以 Macro-F_1 选优:

温度加权软投票:

$$\boldsymbol{p}_{\mathrm{fuse}}=\alpha_{\mathrm{TSTM}}\mathrm{softmax}(\boldsymbol{z}^{(t)}/\tau_{\mathrm{TSTM}})+\alpha_{\mathrm{FSTM}}\mathrm{softmax}(\boldsymbol{z}^{(s)}/\tau_{\mathrm{FSTM}})\in\Delta^{C-1}\tag{34}$$

拼接-MLP:

$$\boldsymbol{p}_{\mathrm{fuse}}=\mathrm{softmax}\Big(\mathrm{MLP}\big(\big[\boldsymbol{p}^{(t)}\|\boldsymbol{p}^{(s)}\big]\big)\Big)\tag{35}$$

注意力/门控加权:

$$\boldsymbol{p}_{\mathrm{fuse}}=\mathrm{softmax}(\alpha_{\mathrm{TSTM}}\boldsymbol{z}^{(t)}+\alpha_{\mathrm{FSTM}}\boldsymbol{z}^{(s)})\tag{36}$$

其中,Δ^{C-1} 为 C 类概率单纯形;$\|$ 表示向量拼接。三种实现分别对应概率级融合、概率后非线性重构与 logit 级加权,涵盖从“后验集成”到“证据合成”的不同假设。式(34)~(36)给出了三种候选融合方式,分别对应不同的融合假设:温度加权软投票在概率空间进行加权求和,形式最简单、可解释性强,适合在两路输出较为一致时获得平滑而稳定的融合结果;拼接-MLP 先将两路概率向量拼接,再由 MLP 学习非线性重构,能够显示建模两路输出之间的交互关系,适合处理互补性更强或存在系统性偏差的情形;注意力/门控加权在 logit 空间完成加权再 softmax,相当于在决策边界形成前进行融合,能够在保持类别间相对偏好结构的同时引入门控权重的自适应调节。本文在后续实验中采用拼接-MLP 作为默认融合方式,以获得更稳定的融合表示与更强的表达能力。其余两种形式可视为同一框架下的替代实现,便于在不同部署约束下进行取舍。

(3)训练与稳定化

在训练阶段,融合模块的参数 $\{\boldsymbol{W}_{\mathrm{g}},\boldsymbol{b}_{\mathrm{g}}\}$ 以及对应的 MLP 参数与两路教师网络端到端联合优化,主目标为融合输出的交叉熵损失 $L_{\mathrm{CE}}(y,\boldsymbol{p}_{\mathrm{fuse}})$。为避免门控权重在训练早期过早退化为单一路分支长期占优,本文引入式(37)的门控熵正则 R_{gate},其计算式为

$$R_{\mathrm{gate}}=-\gamma(\alpha_{\mathrm{TSTM}}\ln\alpha_{\mathrm{TSTM}}+\alpha_{\mathrm{FSTM}}\ln\alpha_{\mathrm{FSTM}})\tag{37}$$

以鼓励 α_{TSTM} 与 α_{FSTM} 保持必要的多样性,从而提升融合的稳定性与泛化性,其中 γ 控制约束强度。与此同时,温度缩放参数 $\tau_{\mathrm{TSTM}},\tau_{\mathrm{FSTM}}>0$ 用于平滑教师输出分布,缓解过度置信带来的梯度不稳定,使融合模块更容易学习到合理的权重分配。最终,本文在验证集上以 macro-F_1 为主指标(并兼顾少数类召回)选择融合策略,并在测试与蒸馏阶段固定该策略,以保证推理行为的一致性。

2.3　面向资源受限设备的高效 KD 机制

为在边缘/IoT 场景下在精度-时延-参数量之间取得平衡,本文在双教师模型(TSTM/FSTM)上构建轻量学生模型 $f_s(\cdot)$,并采用温度缩放的 KD 进行训练,推理阶段仅部署学生模型。总体流程如图 3 所示。

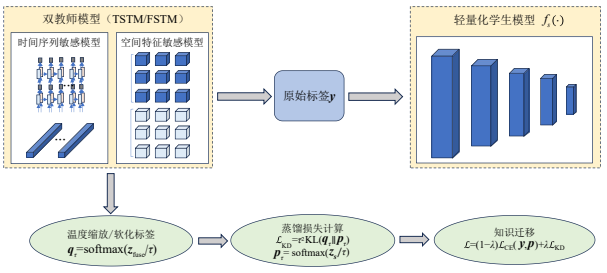

图 3　KD 训练流程

Figure 3　The knowledge distillation training process

2.3.1 超轻量学生模型

学生端主干以“低通道+深度可分离卷积+GAP+精简分类头”为设计原则，显著压缩容量与FLOPs，同时保留对关键判别信息的表达能力：

（1）特征提取轻量化：首层通道由32降至16，并在各阶段采用深度可分离卷积与残差瓶颈的最小化配置。

（2）注意力稀化：去除复杂注意力，仅保留GAP以聚合通道信息。

（3）分类头精简：全连接层维度由128降至32修正线性单元（Rectified Linear Unit，ReLU），输出与类别数C对齐，得到学生logits $\boldsymbol{z}_s$与概率向量$\boldsymbol{p}=\mathrm{softmax}(\boldsymbol{z}_s)$。

经后续实验数据验证，该结构在保证精度的同时，将参数量与推理时延显著降低。

2.3.2 优化的蒸馏过程

（1）软标签与温度缩放

由2.2.3节的自适应融合得到教师端融合logits $\boldsymbol{z}_{\mathrm{fuse}}$。为暴露“暗知识”，对教师与学生均使用统一温度$\tau>0$进行软化：

$$\boldsymbol{q}_\tau=\mathrm{softmax}\left(\frac{\boldsymbol{z}_{\mathrm{fuse}}}{\tau}\right) \tag{38}$$

$$\boldsymbol{p}_\tau=\mathrm{softmax}\left(\frac{\boldsymbol{z}_s}{\tau}\right) \tag{39}$$

（2）蒸馏与总目标函数

软标签蒸馏损失采用带τ^2缩放的KL（Kullback-Leibler），与交叉熵等价：

$$\mathcal{L}_{\mathrm{KD}}=r^2\mathrm{KL}(\boldsymbol{q}_\tau\|\boldsymbol{p}_\tau)=r^2\mathrm{CE}(\boldsymbol{q}_\tau,\boldsymbol{p}_\tau) \tag{40}$$

结合硬标签交叉熵$\mathcal{L}_{\mathrm{CE}}(\boldsymbol{y},\boldsymbol{p})=-\sum_{k=1}^{C}\boldsymbol{y}_k\ln\boldsymbol{p}_k$，得到混合目标：

$$\mathcal{L}=(1-\lambda)\mathcal{L}_{\mathrm{CE}}(\boldsymbol{y},\boldsymbol{p})+\lambda\mathcal{L}_{\mathrm{KD}},\lambda\in[0,1] \tag{41}$$

其中，λ控制硬/软监督的权衡。训练早期取较大的τ,λ以侧重“暗知识”学习，后期退火至较小的λ以贴合真实标签分布。

（3）训练效率与优化策略

为降低教师端重复前向的计算开销，采用批级logits缓存与周期性刷新机制：对每个mini-batch生成的融合logits $\boldsymbol{z}_{\mathrm{fuse}}$进行缓存，并以固定间隔$K$（按迭代数或按epoch）统一刷新。该策略在不破坏数值稳定与收敛性的前提下，显著减少蒸馏阶段的计算量与I/O开销。蒸馏训练过程中教师参数保持冻结，仅优化学生端与融合相关的可学习标量/门控。

3 实验设计与结果分析

本研究实验采用真实网络流量数据集，通过构建双教师模型集成与超轻量学生模型验证方法有效性。实验流程以多阶段数据预处理为起点，包括特征工程、类别均衡与序列化处理，为模型训练提供数据基础。随后分别构建时序敏感模型与特征空间模型作为教师模型，并在多种融合策略下形成最优教师集成。在此基础上，采用KD与动态温度调整训练轻量学生模型。其中，后续的FeatureSpaceModel和TimeSensitiveModel分别指仅使用FSTM和仅使用TSTM来进行分类，TeacherEnsemble_concatenation是指结合FSTM和TSTM两个模型集成进行分类。

实验结果表明，在参数量降至教师模型的1/5条件下，所提方法保持较高检测性能。在预处理阶段，优化的类别均衡策略将原始不均衡分布调整为相对平衡。为保证增广样本保真与质量，各类别样本规模被控制在5 000~900 000。在特征工程方面，引入流量不对称（Flow Asymmetry）与报文大小变异（Packet Size Variation）等场景自适应特征。经随机森林评估，这些特征的平均重要性提升超过15%。在教师模型集成方面，软投票融合在验证集上取得macro-F_1为95.0%，较单一教师模型提升约1.4%。

在KD阶段，本文对温度参数T与软硬标签混合系数α采用动态调度，以兼顾训练稳定性与最终判别精度。具体地，T初始设为2.0，并随训练轮次线性衰减至1.0。较高的初始温度用于平滑教师分布、增强类间相对信息，从而在训练早期为学生提供更稳定的“暗知识”监督。随着训练推进逐步降低温度，使学生输出分布由“软”向“硬”过渡，避免长期过度平滑影响边界收敛。在混合系数方面，α由0.7逐步增至0.9，用于提高软标签损失在总损失中的占比。该设置基于验证集对α的敏感性观察，较低的初始权重可缓解学生初期对噪声/不确定样本的过拟合，而在学生已具备基本判别能力后提升软监督比重，有助于更充分地对齐集成教师的决策结构并提升长尾类别的稳定性。采用上述调度策略后，学生模型参数量约为教师集成的1/5，在测试集上仍取得94.8%的准确率与90.34%的macro-F_1。

3.1 消融实验

如图4~6所示，本文进行了消融实验，以两条教师单分支与基线集成模型（TeacherEnsemble_concatenation）在相同设置下输出的混淆矩阵进行对比。结果显示，TeacherEnsemble_concatenation在整体与各类均衡性上最优：在Patator、PortScan、Web Attack、Bot与Unknown等典型类别上，其混淆矩阵主对角线对应类别的识别比例均稳定在98%以上，如Patator 99.8%、PortScan 99.6%、Web Attack 98.6%、Bot 98.8%等，非对角误判质量与范围均低于两条单分支。相比之下，单分支仅在个别类别呈现局部优势，如FeatureSpaceModel在DoS类上达到78.8%，虽高于集成模型的67.3%，但同时带来更多交叉误判，并暴露出类别不

均衡问题:TimeSensitiveModel 将 Web Attack 误判为 PortScan 的比例上升至 1.7%,而集成模型为 1.0%;Unknown 类被正确识别的比例降至 99.2%,集成模型则为 99.8%;FeatureSpaceModel 在 Unknown 上也略低,相比集成模型低近 0.1%。综合混淆矩阵中对角项与非对角项的占比,以及不同类别之间的误差分布可以看出,集成模型在保持高准确率的同时显著提升了类间平衡与整体稳健性,更适合作为多场景下的统一检测器。

图 4　FeatureSpaceModel 混淆矩阵

Figure 4　Confusion matrix of the FeatureSpaceModel

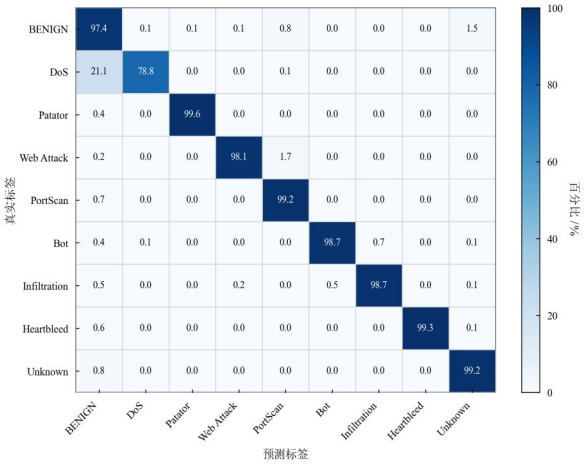

图 5　TimeSensitiveModel 混淆矩阵

Figure 5　Confusion matrix of the TimeSensitiveModel

为进一步验证集成方案的优势,图 7~9 给出了三种模型的 P-R 曲线与各类平均精度。基线 TeacherEnsemble_concatenation 的曲线整体更贴近坐标顶边:在 BENIGN、Patator、PortScan、Web Attack、Heartbleed 上 AP 等于 1.00,在 Bot 与 Infiltration 上 AP 约为 0.99;在 DoS 上 AP 为 0.92,显著高于 TimeSensitiveModel 的 0.84,并与 FeatureSpaceModel 持平,但其曲线在高精度区间维持更长,召回提升前精度下降更缓。与前述混淆矩阵相呼应,基线在默认阈值下对 BENIGN、Patator、PortScan、

图 6　TeacherEnsemble_concatenation 混淆矩阵

Figure 6　Confusion matrix of the teacherEnsemble_concatenation

Web Attack、Bot、Infiltration、Heartbleed、Unknown 的召回率分别为 97.7%、99.8%、99.6%、98.6%、98.8%、98.7%、99.7%、99.8%,DoS 的主要误差来自将约三成样本误报为 BENIGN。两条教师分支的曲线印证了其短板:FeatureSpaceModel 在 DoS 的曲线较为平稳且 AP 为 0.92,但 Unknown 的 AP 仅为 0.58,阈值敏感性更高;TimeSensitiveModel 的 DoS 与 Unknown AP 分别为 0.84 与 0.34,曲线更早坍塌,说明对复杂攻击与未知类别的分离能力不足。需要强调的是,Unknown 在 P-R 度量上对阈值极为敏感,虽然基线在该类的 AP 为 0.35,但其在默认阈值下的准确率最高,这一点已由混淆矩阵的主对角占比给出。综合 P-R 与混淆矩阵两方面证据,TeacherEnsemble_concatenation 在多数类别实现最稳的精度与召回折中,并在部署所关注的默认阈值下取得最佳总体识别,因此本实验选择其作为后续蒸馏的教师模型。

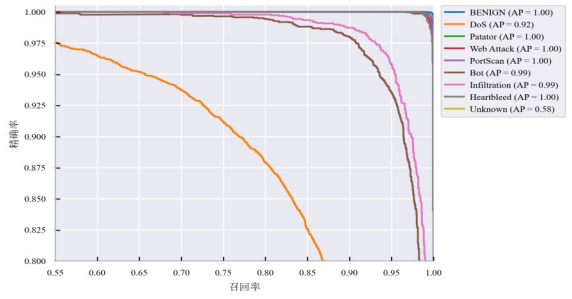

图 7　FeatureSpaceModel 的 P-R 曲线

Figure 7　P-R curve of the FeatureSpaceModel

由图 10 的测试集数据分布可知,数据呈显著长尾:Unknown 样本数有 181 898 个、BENIGN 样本数有 146 572 个、PortScan 样本数有 73 167 个、Patator 样本数有 44 251 个,数据集以这些类别为主体;而反观 DoS 样本数有 758 个、Infiltration 样本数有 987 个、Bot 样本数

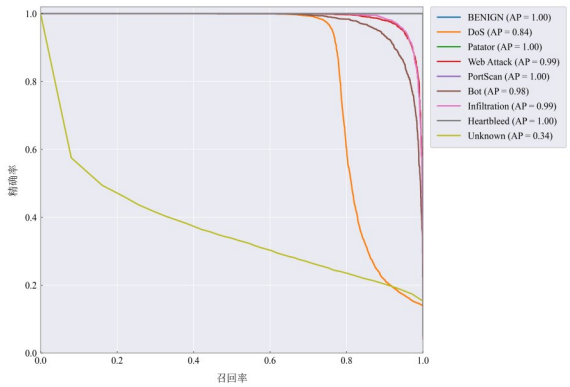

图8 TimeSensitiveModel的P-R曲线

Figure 8 P-R curve of the TimeSensitiveModel

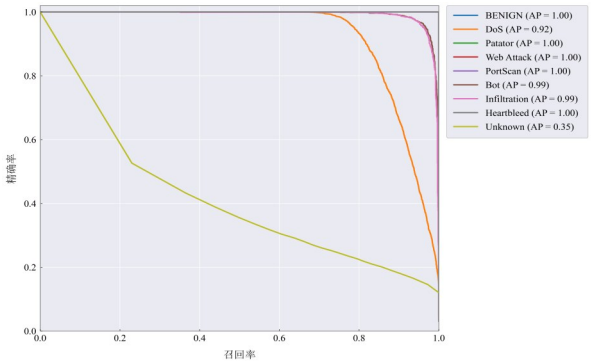

图9 TeacherEnsemble_concatenation的P-R曲线

Figure 9 P-R curve of the teacherensemble_concatenation

有1 352个、Web Attack样本数有2 414个、Heartbleed样本数有3 175个，这些类别明显属于少数类。消融结果与图4~6一致，表明TeacherEnsemble_concatenation不仅总体最优，也对少数类检测更有增益。以表2和表3中分别对少数类Infiltration和Bot的模型比较结果为例，可以观察到在Infiltration类别上，教师集成的精确率、召回率、F_1分数值分别为99.08%、98.68%、98.88%，均高于两条单教师的98.78%、98.58%、98.68%与98.58%、98.68%、98.63%。在Bot类别上，教师集成达到97.02%、98.83%、97.92%，同样领先于单教师的97.02%、98.82%、97.91%与96.88%、98.74%、97.80%。这说明拼接集成能够在样本层面更充分整合两路教师的互补证据，显著提升少数类的召回与F_1分数值，减少低样本场景中的偏置与漏检，在高度类不均衡的网络行为分类任务中具有明确优势。在表2和表3中，F_1分数表示对应类别的单类F_1。

表4表明，拼接集成（concatenation ensemble）在统一数据与评测协议下取得最高macro-F_1值为90.34%，优于注意力/门控加权的90.19%、温度加权软投票的89.79%以及无集成的88.67%。这说明本文的集成设计不仅有效，而且还具自适应性：拼接集成把各教师分支的概率向量按通道拼接为$[\boldsymbol{p}^{(1)}\|\boldsymbol{p}^{(2)}]$，交由MLP学习样本级、类别级的非线性交互与隐式权重，从而根据当前样本的分布自动强调信息更充分的分支、抑制噪声分支，在精度与召回之间获得更优折中。相比之下，温度加权与注意力方法依赖显式（或近似线性）的权重，对类别差异与场景漂移的适配能力相对受限，不做集成的性能最低进一步佐证了多教师互补的必要性。基于上述结果，本文在后续蒸馏阶段采用拼接集成作为最终教师输出，以充分保留两路教师的互补信息并实现对不同网络行为场景的自适应识别。

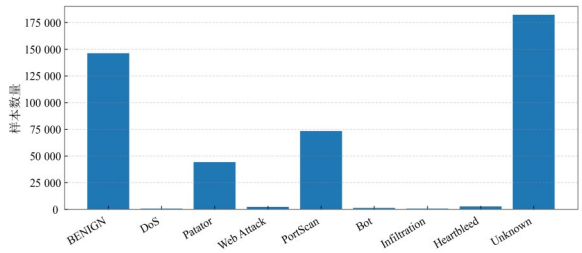

图10 测试集数据分布

Figure 10 Distribution of the test dataset

表2 Infiltration类别下的不同模型指标 单位：%

Table 2 Metrics of different models under the infiltration class unit：%

模型	精确率	召回率	F_1分数
FeatureSpaceModel	98.78	98.58	98.68
TimeSensitiveModel	98.58	98.68	98.63
TeacherEnsemble_concatenation	99.08	98.68	98.88

表3 Bot类别下的不同模型指标 单位：%

Table 3 Performance metrics of different models for the bot class unit：%

模型	精确率	召回率	F_1分数
FeatureSpaceModel	97.02	98.82	97.91
TimeSensitiveModel	96.88	98.74	97.80
TeacherEnsemble_concatenation	97.02	98.83	97.92

表4 集成策略比较 单位：%

Table 4 Comparison of integration strategies unit：%

集成策略	macro-F_1
温度加权软投票	89.79
拼接-MLP集成	90.34
注意力/门控加权集成	90.19
无集成方案	88.67

如表5所示，本实验在相同训练与评测设置下，教师集成模型达到96.5%的准确率与95.0%的macro-F_1值，给出了本任务的性能上限。基于双教师蒸馏的轻量化学生模型在大幅压缩模型规模的同时，仍取得93.91%准确率与90.34%的macro-F_1值，整体判别趋势与教师侧保持一致。该结果表明，本文的实验设计能够稳定、有效地迁移教师模型的判别能力，在“精度-效率”权衡上具有明显优势，验证了轻量化方案在网

络行为多分类任务中的可靠性与有效性。

表5　轻量化学生模型与教师集成模型对比　　单位:%

Table 5　Comparison between the lightweight student model and the teacher ensemble model　　unit:%

指标	准确率	macro-F_1
教师集成模型	96.50	95.00
轻量化学生模型	93.91	90.34

3.2　对比实验及鲁棒性测试

将本文的轻量化学生模型与同一数据集CICIDS-2017、同一环境下的代表性方法进行对比,结果汇总于表6。学生模型端侧单次检测时延约68 ms,并取得93.91%的准确率与90.34%的macro-F_1值。相较传统机器学习范式,LS-SVM的准确率与macro-F_1分别为82%与82%、端侧单次检测时延1.8 s;AdaBoost为80%与88%、推理时间7.8 s;LASSO+EnsDT为80%与83%、端侧单次检测时延3.9 s。在这些对比中,学生模型在准确率上分别提升11.9~13.9个百分点,在macro-F_1上提升2.3~7.3个百分点,且速度分别约为26倍、115倍、57倍,体现出在边缘约束场景中的显著效率优势。与深度序列模型LSTM+DAC相比,该模型以17.6倍的速度换取可用精度,后者准确率与macro-F_1均为97%、端侧单次检测时延1.2 s,显示出本方案在“精度-时延”折中的优势。与联合深度框架CNN+RNN/LSTM+SMOTE+FGSM相比,学生模型在准确率上领先5.9个百分点,macro-F_1略低于92%,但在实际部署中凭借更低时延与更小计算负载提供了更好的实时可用性与能效表现。综合上述结果,本文的轻量化学生模型在保证检测效能的同时,实现了对计算成本与响应时间的系统性压降,在同类方法中呈现出最优的效率优势。

为评估模型在极端类别不均衡条件下的鲁棒性,

表6　模型对比

Table 6　Model comparison

相关研究	方案	准确率/%	macro-F_1/%	端侧单次检测时延/s
本文	轻量化学生模型	93.91	90.34	0.068
文献[40],2025	LSTM+DAC	97.00	97.00	1.200
文献[41],2025	LS-SVM	82.00	82.00	1.800
	AdaBoost	80.00	88.00	7.800
	LASSO+EnsDT	80.00	83.00	3.900
文献[42],2025	CNN+RNN/LSTM+SMOTE+FGSM	88.00	92.00	—

本文进一步从风险视角统计了少数类样本被误判为正常流量的比例,如表7所示。与总体精度等指标相比,该统计更能刻画长尾场景下的关键安全风险:当少数攻击被判为正常时,尽管其占比很低,仍会在实际部署中造成不可忽视的漏报隐患。由表7可知,教师集成模型在多类少数攻击上的误判率整体更低且更稳定;相较于单一教师模型,集成教师模型在Patator、Web Attack、PortScan等少数类别上均进一步降低了被判为正常的比例,表明融合时序特征与特征空间信息能够在多数类占优的背景下更有效地保持对稀有攻击模式的区分能力,从而提升模型在长尾分布下的部署鲁棒性与风险控制能力。进一步地,表8以Heartbleed类为例开展鲁棒性测试:在训练样本数从原始规模逐步降至2 500、1 000乃至500的情况下,F_1分数由99.48%下降至94.98%,精确率由99.28%下降至95.52%,召回率由99.69%下降至94.44%,整体仍保持在较高水平,说明即便少数类样本进一步稀缺,模型对该类攻击的识别能力依然相对稳定,为长尾与低样本场景下的实际部署提供了更可靠的性能保障。此处F_1分数表示对应类别的单类F_1。

表7　三种模型少数类的误判率　　单位:%

Table 7　Misclassification rates of minority classes for the three models　　unit:%

类别	Patator	Web Attack	PortScan	Bot	Infiltration	Heartbleed
FeatureSpaceModel	0.6	0.7	0.6	0.5	0.7	0.8
TimeSensitiveModel	0.4	0.2	0.7	0.4	0.5	0.6
TeacherEnsemble_concatenation	0.2	0.3	0.4	0.5	0.6	0.2

表8　Heartbleed类别下的鲁棒性测试　　单位:%

Table 8　Robustness test results for the heartbleed class　unit:%

训练样本数	F_1分数	精确率	召回率
3 175	99.48	99.28	99.69
2 500	98.78	98.74	98.82
1 000	97.34	97.86	96.83
500	94.98	95.52	94.44

3.3　类别级评估及通信开销分析

在真实网络IDS的部署场景中,检测器所面对的流量分布往往无法完全由已知攻击类型覆盖。除数据集中标注明确的攻击行为外,系统还需应对未知攻击、攻击变种以及由网络环境或设备行为变化引发的分布漂移等情况。因此,将未能被可靠归入任何已知攻击类别的流量统一建模为Unknown类别,对于提升IDS的安全性与实用性具有重要意义。该类别在工程语义上并不对应某一种具体攻击类型,而是承担“异常兜底”的角色,其核心目标在于降低未知或异常流量被误判为BENIGN的风险。与常规已知类别

不同，Unknown类别在评估层面具有明显特殊性：一方面，其样本分布通常高度异质，内部可能包含多种不同模式，导致在基于阈值变化的P-R曲线下表现出较强的不稳定性；另一方面，在实际部署中，检测系统往往采用固定或近似固定的决策阈值，而非针对单一类别动态调整阈值。因此，仅依赖平均精度或曲线下面积等整体指标，难以全面反映模型在Unknown类别上的实际判别能力。

基于上述考虑，本文对Unknown类别的评估更侧重于部署导向的判别行为分析，具体包括在统一决策阈值下，其在混淆矩阵中的主对角占比，以及其被误判为 BENIGN 或其他已知攻击类别的比例。该评估方式能够更直接地反映模型在面对未知流量时的拒识能力和安全保守性。由图4~6所示的混淆矩阵结果可知，相较于单一教师模型，教师集成模型在Unknown类别上的主对角占比最高，同时对Unknown误判成BENIGN的比例也是最低。这表明自适应集成能够在样本级充分融合时序特征与特征空间信息，在保持整体检测性能的同时，更有效地区分未知流量与正常流量。尽管在P-R曲线评估中，Unknown类别的AP值相对较低，但在默认阈值设置下，集成模型仍表现出最稳定的判别行为，这与其在实际部署场景中的设计目标是一致的。随后，表9从训练与资源占用角度给出了教师侧与学生侧在通信开销方面的对比结果。可以看到，学生模型在保持较高检测性能的前提下显著降低了训练成本与运行开销，其训练累计耗时由61.9 s降至26.16 s，内存占用由0.66%降至0.22%。这说明蒸馏后的学生模型能够以更低的资源消耗承载教师集成所学习到的判别知识，为边缘与IoT环境中对实时性与资源约束更为敏感的应用场景提供了可参考性。

表9 教师与学生模型少数通信开销对比

Table 9 Communication overhead comparison between the teacher and student models for minority classes

指标	训练累计耗时/s	内存占用百分比/%
教师模型	61.90	0.66
学生模型	26.16	0.22

综上所述，本文在类别级评估中引入Unknown的工程语义，并结合更贴近部署需求的判别行为分析，对传统曲线指标难以充分刻画的部分进行了补充；同时，教师—学生的效率对比表明：该方案在保持整体检测能力的基础上具备一定的轻量化优势，更符合资源受限场景的应用诉求。总体而言，上述分析结果均能为后续面向真实网络环境的落地应用与持续优化提供了参考依据与方法支撑。

4 结论与展望

针对入侵检测中时空语义利用不足、冗余特征与噪声干扰、类别失衡等问题，本文提出一种能够提炼时间与特征空间语义关系并抑制冗余信息的轻量化检测模型。模型以层次化表征学习挖掘流量序列的周期、突发与跨字段共现等深层语义，增强对多源网络特征的判别力。为减少语义学习阶段的噪声影响，引入可学习的先验提示与原始序列共同构成初级特征，使模型在早期对异常模式进行筛选并提供更充分的上下文信息。为提升时间分支与空间分支的语义表达，设计跨分支交互机制，以互补证据促进难区分类别的识别与边界样本的稳健判决。同时，利用已知类别与潜在未知类别之间的语义一致性线索，将高置信度的教师表征与学生表征进行匹配，过滤仅靠监督信号难以消除的无关特征。在多模型融合阶段，采用自适应门控与以拼接为主的集成策略，动态整合两路教师的关键信息并抑制冗余输出。通过对比实验、消融研究与可视化分析，验证了所提出方法在宏观性能、少数类检出以及端侧效率三方面的有效性。

参考文献

[1] Hossein Motlagh N, Taleb T, Arouk O. Low-altitude unmanned aerial vehicles-based Internet of Things services: Comprehensive survey and future perspectives[J]. IEEE Internet of Things Journal, 2016, 3(6): 899-922.

[2] Huang C Q, Fang S F, Wu H, et al. Low-altitude intelligent transportation: System architecture, infrastructure, and key technologies[J]. Journal of Industrial Information Integration, 2024, 42: 100694.

[3] Anagnostis I, Kotzanikolaou P, Douligeris C. Understanding and securing unmanned aerial vehicle (UAV) services: a comprehensive tutorial[EB/OL]. (2024-03-06)[2026-03-13]. https://www.techrxiv.org/users/750957/articles/722052.

[4] Cai L Y, Wang J C, Zhang R C, et al. Secure physical layer communications for low-altitude economy networking: A survey[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 2497-2530.

[5] Shen M, Ye K, Liu X T, et al. Machine learning-powered encrypted network traffic analysis: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 791-824.

[6] Azaria A, Richardson A, Kraus S, et al. Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data[J]. IEEE Transactions on Computational Social Systems, 2014, 1(2): 135-155.

[7] Gupta I, Lycklama H, Opel E, et al. Fragile giants: understanding the susceptibility of models to subpopulation attacks [EB/OL]. (2024-10-11) [2026-03-13]. https://arxiv.

org/pdf/2410.08872.

[8] Hong Y P, Li Q, Yang Y Q, et al. Graph based encrypted malicious traffic detection with hybrid analysis of multi-view features[J]. Information Sciences, 2023, 644: 119229.

[9] Shen M, Zhang J P, Zhu L H, et al. Accurate decentralized application identification *via* encrypted traffic analysis using graph neural networks[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2367-2380.

[10] Sinha P, Sahu D, Prakash S, et al. A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning[J]. Scientific Reports, 2025, 15: 9684.

[11] Xie Y B, Gardi A, Sabatini R. Cybersecurity trends in low-altitude air traffic management[C]//2022 IEEE/AIAA 41st Digital Avionics Systems Conference. Piscataway: IEEE, 2022: 1-9.

[12] Wilson A N, Kumar A, Jha A, et al. Embedded sensors, communication technologies, computing platforms and machine learning for UAVs: A review[J]. IEEE Sensors Journal, 2022, 22(3): 1807-1826.

[13] Shen M, Yu H, Zhu L H, et al. Effective and robust physical-world attacks on deep learning face recognition systems[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4063-4077.

[14] Aslan Ö, Aktuğ S S, Ozkan-Okay M, et al. A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions[J]. Electronics, 2023, 12(6): 1333.

[15] Sarker I H. Multi-aspectsAI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview[J]. Security and Privacy, 2023, 6(5): e295.

[16] Xu H T, Han S Y, Li X H, et al. Anomaly traffic detection based on communication-efficient federated learning in space-air-ground integration network[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 9346-9360.

[17] 薛胤, 魏松杰. 基于深度监督离散哈希神经网络的网络入侵检测方法[J]. 信息安全学报, 2025, 10(4): 66-76.
Xue Yin, Wei Songjie. Network intrusion detection with deep neural network for supervised learning of discrete hash[J]. Journal of Cyber Security, 2025, 10(4): 66-76. (in Chinese)

[18] Li X Q, Liu Q C. A hybrid sampling algorithm for imbalanced and class-overlap data based on natural neighbors and density estimation[J]. Knowledge and Information Systems, 2025, 67(3): 2259-2290.

[19] Wang L H, Dai Q, Du T, et al. Lightweight intrusion detection model based on CNN and knowledge distillation[J]. Applied Soft Computing, 2024, 165: 112118.

[20] 曹磊, 温蜜, 何蔚. 基于深度学习的车联网的路网监测系统的DoS和DDoS攻击的入侵检测方法[J]. 计算机应用与软件, 2025, 42(1): 303-311.
Cao Lei, Wen Mi, He Wei. Deep learning based dos and ddos attack detection method in the highway monitoring system of iov[J]. Computer Applications and Software, 2025, 42(1): 303-311. (in Chinese)

[21] Cao Z, Liu X H, Zhou Z, et al. KD-BERT: A lightweight knowledge distillation bidirectional encoder representations from transformers for IoT network intrusion detection[J]. IEEE Transactions on Industrial Informatics, 2025, 21(11): 8475-8483.

[22] Xie B, Wang Z D, Zeng Z Y, et al. DTKD-IDS: A dual-teacher knowledge distillation intrusion detection model for the industrial Internet of Things[J]. Ad Hoc Networks, 2025, 174: 103869.

[23] 李波, 李泽超, 邢鹏, 等. 基于双向约束蒸馏的无监督图像异常检测[J]. 电子学报, 2025, 53(3): 895-909.
Li Bo, Li Zechao, Xing Peng, et al. Unsupervised image anomaly detection based on constrained BidiRectional distillation[J]. Acta Electronica Sinica, 2025, 53(3): 895-909. (in Chinese)

[24] Chai Y E, Liu M Q, Li M. Navigation spoofing and jamming signals identification of UAV based on federated learning[J]. IEEE Internet of Things Journal, 2025, 12(21): 44177-44188.

[25] Ahmad T, Hadi M U, Vassiliou V, et al. Real-time anomaly detection in smart vehicle-to-UAV networks for disaster management[J]. Transactions on Emerging Telecommunications Technologies, 2025, 36(5): e70162.

[26] Medhi J, Liu R, Wang Q, et al. A lightweight and efficient intrusion detection system (IDS) for unmanned aerial vehicles[J]. Neural Computing and Applications, 2025, 37(20): 15819-15836.

[27] Wang Y Z, Yu Z Y, Wu J H, et al. Adaptive knowledge distillation-based lightweight intelligent fault diagnosis framework in IoT edge computing[J]. IEEE Internet of Things Journal, 2024, 11(13): 23156-23169.

[28] Quyen N H, Duy P T, Nguyen N T, et al. FedKD-IDS: A robust intrusion detection system using knowledge distillation-based semi-supervised federated learning and anti-poisoning attack mechanism[J]. Information Fusion, 2025, 117: 102807.

[29] Wang F, Diao B Y, Sun T, et al. Data security and privacy challenges of computing offloading in FINs[J]. IEEE Network, 2020, 34(2): 14-20.

[30] 周品希, 沈岳, 李伟. 基于深度学习的物联网入侵检测系统综述[J]. 网络安全与数据治理, 2025, 44(6): 1-10.
Zhou Pinxi, Shen Yue, Li Wei. A review of IoT intrusion detection systems based on deep learning[J]. Cyber Security and Data Governance, 2025, 44(6): 1-10. (in Chinese)

[31] Kurniabudi, Stiawan D, Darmawijoyo, et al. CICIDS-2017 dataset feature analysis with information gain for anomaly detection[J]. IEEE Access, 2020, 8: 132911-132921.

[32] Panwar S S, Raiwani Y P, Panwar L S. An intrusion detection model for CICIDS-2017 dataset using machine learning algorithms[C]//2022 International Conference on Advances in Computing, Communication and Materials. Piscataway: IEEE, 2022: 1-10.
[33] Zuech R, Hancock J, Khoshgoftaar T M. Detecting web attacks in severely imbalanced network traffic data[C]//2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science. Piscataway: IEEE, 2021: 267-273.
[34] Wang Y X, Sun G, Sun Z M, et al. Toward realization of low-altitude economy networks: Core architecture, integrated technologies, and future directions[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 11(5): 2788-2820.
[35] 俞宁宁, 毛盛健, 周成伟, 等. DroneRFa: 用于侦测低空无人机的大规模无人机射频信号数据集[J]. 电子与信息学报, 2024, 46(4): 1147-1156.
Yu Ningning, Mao Shengjian, Zhou Chengwei, et al. DroneRFa: A large-scale dataset of drone radio frequency signals for detecting low-altitude drones[J]. Journal of Electronics & Information Technology, 2024, 46(4): 1147-1156. (in Chinese)
[36] Vu L, Bui C T, Nguyen Q U. A deep learning based method for handling imbalanced problem in network traffic classification[C]//Proceedings of the 8th International Symposium on Information and Communication Technology. New York: ACM, 2017: 333-339.
[37] Marlisa H, Satyahadewi N, Imro'ah N, et al. Application of adasyn oversampling technique on k-nearest neighbor algorithm[J]. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2024, 18(3): 1829-1838.
[38] Dey I, Pratap V. A comparative study of SMOTE, borderline-SMOTE, and ADASYN oversampling techniques using different classifiers[C]//2023 3rd International Conference on Smart Data Intelligence. Piscataway: IEEE, 2023: 294-302.
[39] Wang J J, Xu M T, Wang H, et al. Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding[C]//2006 8th international Conference on Signal Processing. Piscataway: IEEE, 2006
[40] Ayad A G, Sakr N A, Hikal N A. Fog-empowered anomaly detection in IoT networks using one-class asymmetric stacked autoencoder[J]. Cluster Computing, 2025, 28(8): 550.
[41] Waghmode P, Kanumuri M, El-Ocla H, et al. Intrusion detection system based on machine learning using least square support vector machine[J]. Scientific Reports, 2025, 15: 12066.
[42] Sinha P, Sahu D, Prakash S, et al. An efficient data driven framework for intrusion detection in wireless sensor networks using deep learning[J]. Scientific Reports, 2025, 15: 34046.

作者简介

韩 刚 男,1990年7月出生于陕西省西安市。现为西安邮电大学网络空间安全学院副教授。主要研究方向为物联网和区块链技术。
E-mail: hangang668866@163.com

常 轩 男,2002年5月出生于辽宁省营口市。现为西安邮电大学网络空间安全学院研究生,已参与多项人工智能领域相关研究工作。主要研究方向为物联网安全和入侵检测。
E-mail: 2923547941@qq.com

王耀天 男,2000年3月出生于山西省太原市。现为西安邮电大学网络空间安全学院研究生,已参与多项人工智能领域相关研究工作。主要研究方向为物联网安全和入侵检测。
E-mail: 1767178707@qq.com

王云俐 女,2001年5月出生于陕西省西安市。现为西安邮电大学网络空间安全学院研究生,已参与多项人工智能领域相关研究工作。主要研究方向为物联网安全和联邦学习。
E-mail: 2920500595@qq.com

张忠良 男,1986年2月出生于浙江省嘉兴市。现为中国杭州电子科技大学信息管理与信息系统系教授。主要研究方向为联邦学习、数据挖掘和进化算法。
E-mail: zlzhang2525@163.com