自适应邻域学习的加密恶意流量检测方法

张丽娜，杨阳，鲁亦群，贾弘瑜，段爱卓伦
（西安科技大学人工智能与计算机学院，陕西 西安 710600）

摘要：针对现有加密恶意流量检测方法使用静态图结构、邻域无法自适应选择节点、复杂网络环境下流量交互特征不足等问题，提出一种基于自适应邻域学习的加密恶意流量检测方法。该方法首先构建流量交互图建模流量会话间的关联关系，并在此基础上引入邻域学习机制，通过结合节点特征学习与邻域结构优化机制，实现流量交互模式的节点自适应建模。所提方法在CTU-13与MCFP数据集上取得了较好的效果，准确率、精确率、召回率与F1分数分别达到99.59%、99.56%、98.78%和99.16%，有效提升了加密恶意流量检测的性能。

关键词：加密恶意流量；图神经网络；自适应邻域学习；流量检测
中图分类号：TP303
文献标志码：A

Encrypted Malicious Traffic Detection Method Based on Adaptive Neighborhood Learning

Zhang Lina, Yang Yang, Lu Yiqun, Jia Hongyu, Duan Aizhuolun
School of Artificial Intelligence and Computer Science, Xi'an University of Science and Technology, Xi'an 710600, Shaanxi, China

Abstract: To address the limitations of existing encrypted malicious traffic detection methods (e.g., static graph structures, lack of adaptive neighborhood selection, and insufficient modeling of interaction features), we propose an adaptive neighborhood learning-based detection method. Specifically, we construct a traffic interaction graph and introduce a neighborhood learning mechanism, enabling adaptive modeling of traffic interaction patterns by jointly optimizing node features and neighborhood structures. Experimental results on the CTU-13 and MCFP datasets demonstrate that the proposed method achieves accuracy, precision, recall, and F1-score of 99.59%, 99.56%, 98.78%, and 99.16%, respectively, and effectively improves encrypted malicious traffic detection performance.

Key words: Encrypted malicious traffic, Graph neural network, Adaptive neighborhood learning, Traffic detection

0 引言

对网络通信内容的加密已经成为当前时代的主流趋势，以TLS(Transport Layer Security)、HTTPS (Hypertext Transfer Protocol Secure)、QUIC(Quick UDP Internet Connections)等加密网络协议被广泛应用，有效提升了数据传输的机密性与完整性。然而，加密的出现在提升了网络通信安全性的同时，也为网络安全带来了新的挑战。攻击者利用加密隐藏流量的恶意行为，使传统依赖端口号、明文载荷或深度包检测的方法难以发挥作用。近年来，安全报告显示，加密恶意流量在网络攻击中的占比逐渐上升，成为威胁网络安全中的关键之一。

早期研究主要使用统计特征对流量分类[1-4]，但这类方法依赖人工设计特征，泛化能力弱。随着深度学习的发展，研究者开始利用原始流量的字节

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：张丽娜，zhangln@xust.edu.cn
基金项目：西安市科技计划项目（No.22GXFW0063）；陕西省科技厅青年项目（No.2021JQ575）
Foundation Items：The Xi'an Science and Technology Plan Project (No.22GXFW0063), The Youth Project of Shaanxi Provincial Department of Science and Technology (No.2021JQ575)

序列[5-11]，使用CNN、RNN、Transformer等深度学习模型提取特征，在一定程度上提升了分类精度。然而，这类方法多注重欧氏空间表示，难以反映流量间的交互关系，并且部分方法可能存在泄露用户隐私的风险。

随着图神经网络的出现，加密恶意流量检测有了新的思路[12-19]。通过将网络流量构建为图结构，从而反映流量之间的交互关系和行为模式。然而，现有的图构建方式通常依赖静态邻域结构，可能引入冗余连接或噪声邻居，影响模型准确率。

针对上述问题，本文提出一种基于自适应邻域学习的加密恶意流量检测方法。首先基于流量的特征构建初始输入图，用来反映流量间的相似性关系；在此基础上，引入邻域学习机制，使模型能够在训练过程中自适应地调整邻域结构，从而缓解固定图结构的不足。将该自适应邻域学习模块与经典图神经网络模型相结合，实现对加密恶意流量的端到端建模与识别分类。

与现有方法相比，本文的主要贡献包括：

(1)针对单一特征难以全面刻画加密流量行为的问题，构建了一种融合图像特征与统计特征的流量交互图（Traffic Structure Graph,TSG）。该方法通过结合流量的图像特征与向量化特征，在图结构中引入多元化信息，从而更充分地刻画流量样本之间潜在的关联关系。

(2)针对传统图模型中邻域构建依赖固定规则、难以适应复杂流量结构的问题，设计了一种自适应邻域学习模块（Adaptive Neighborhood Refinement, ANR）。该模块通过对节点间关系及邻域规模进行动态建模，实现邻域结构的自适应优化，从而增强模型对多样化流量模式的表达能力。

(3)在上述基础上，将多元化特征的TSG图构建方法与ANR模块相结合，构建了一种面向加密恶意流量检测的图表示学习框架。

1 相关工作

早期加密恶意流量检测主要依赖人工提取特征，通过流量以及流量之间的规律分类。Shekhawat等人[1]利用IP地址、端口、服务器名称和加密套件等对加密恶意流量检测。Taylor等人[2]使用数据包长度的统计特征，Shen等人[3]则采用累计数据包长度特征训练随机森林，验证了统计特征在一定场景下的有效性。Wang等人[4]进一步总结了加密流量的特征，提取113个与协议无关的特征对恶意流量分类，提升了方法的通用性。这类方法训练成本低、特征匹配效率高，但分类的准确率依赖人工选择，泛化能力弱；面对网络环境变化和攻击者的特征混淆手段时，特征易失效。

为解决统计特征方法的缺陷，研究人员利用原始流量数据，通过使用深度学习模型提取特征，减少人工干预。Wang等人[5]首次提出将原始流量预处理为IDX3文件，输入卷积神经网络实现端到端加密流量分类；Shapira等人[6]将流量数据转化为图像，利用深度学习技术完成加密流量分类；Lotfollahi等人[7]提出DeepPacket模型，通过多个自编码器和卷积神经网络，分析原始流量负载。Lin等人[8]结合CNN和双向LSTM，利用流量的空间特征分类；Yao等人[9]则使用结合注意力机制的LSTM模型完成VPN类型流量的识别。Lin等人[10]提出的ET-BERT模型将原始流量转化为令牌，通过训练Transformer实现加密流量和恶意流量分类；Barut等人[11]提出的R1DIT模型在处理原始流量时规避用户负载信息，提升分类性能的同时兼顾了隐私保护。这类方法捕捉流量的深层特征时，无需人工干预。但大多数方法在预处理阶段使用了数据包载荷；同时其特征表示多停留在单个数据包层面，难以反映恶意流量间的交互关系。

随着图神经网络在关系建模与结构特征学习方面优势的逐渐显现，研究人员将加密流量构建为图结构，提升对恶意流量的分类性能。Wang等人[12]通过分析正常节点与异常节点的域名访问行为构建输入图，实现了对僵尸网络的分类。Shen等人[13]基于服务器-客户端交互关系提出了流量交互图，将加密流量分类问题转化为图分类任务。此外，Li等人[14]通过分析数据包之间、流之间的关联关系，将恶意流量建模为端点流量图，利用图神经网络学习图结构特征，实现了对DDoS攻击模式的识别。

在模型设计方面，图卷积神经网络(Graph Convolutional Network,GCN)[15]、图采样与聚合神经网络(Graph SAmple and aggreGatE,GraphSAGE)[16]及图注意力神经网络(Graph Attention Network,GAT)[17]等经典图神经网络模型被广泛应用于加密流量分析任务中。

在此基础上，部分研究进一步结合多视角特征

与注意力机制来增强模型的分类性能。Yang 等人[18]提出了 MTSecurity 方法，将原始字节序列特征与基于图的流量交互特征进行融合，结合 Transformer 与图神经网络实现加密恶意流量检测。Li 等人[19]提出 SAT-Net 通过构建加密流量图结合多层注意力机制，实现了节点结构与特征的融合，有效提升了加密流量分类的准确性。

尽管上述方法在一定程度上提升了加密恶意流量的检测性能，但其图结构构建与邻域特征聚合方面缺少了节点的自适应能力。针对这一问题，Zhao 等人[20]提出了通用图结构学习框架以提升模型泛化能力；Zou 等人[21]则通过结构约束优化邻接关系，增强了对噪声数据的鲁棒性。此外，针对端到端训练的需求，Zou 等人[22]提出了可微分图结构学习网络用于动态更新邻接矩阵；Xie 等人[23]进一步针对异质图场景提出了鲁棒图结构优化方法，提升了模型在特定环境下的适应能力。

综上所述，现有方法在建模流量交互关系方面取得了一定进展，但在邻域结构构建与聚合方法上仍主要使用固定规则，难以适应复杂多变的真实网络环境。如何在复杂的网络环境下，实现邻域结构的自适应学习，提升模型的检测性能，仍是当前研究需要解决的关键问题。

2　方法描述

2.1　方法概述

本文提出一种基于节点自适应邻域选择图神经网络的加密恶意流量检测方法。该方法通过引入可微分节点自适应模块实现自适应邻域选择与恶意流量精确检测。该方法主要结构如图 1 所示。

该框架以加密恶意流量会话为处理样本，依次完成数据预处理、初始关系图构建、特征提取与分类三个大模块。数据处理阶段，将原始流量会话分别转换为灰度图像与向量化特征矩阵。初始关系图构建阶段，基于节点相似度构建会话间的关联关系，形成流量会话图（TSG 图）。特征提取与分类阶段，通过自适应邻域学习模块(Adaptive Neighborhood Refinement, ANR)，多层感知机（Multi-Layer Perceptron，MLP）与 softmax 完成流量的恶意检测。

2.2　多元化特征提取

为了充分描述加密流量在结构行为与统计属性上的差异，本文分别从流量结构与原始数据两个层面反映加密流量的本质特征，为后续图结构构建与模型训练提供有效的输入信息。

向量化特征采用 Zeek 工具对 pcap 文件进行处理，生成 conn.log、ssl.log 及 x509.log 三类日志文件。向量化特征集合包括会话连接特征、SSL/TLS 特征和证书特征等。所有向量化特征均经过标准化处理，构建维度为 d 的节点属性特征向量，后续将直接赋予图节点作为基础属性信息。

图像特征可以完整保留加密流量原始数据的内在信息。首先按会话粒度对连续加密流量进行分割，得到独立的 pcap 文件；然后对分割后的会话

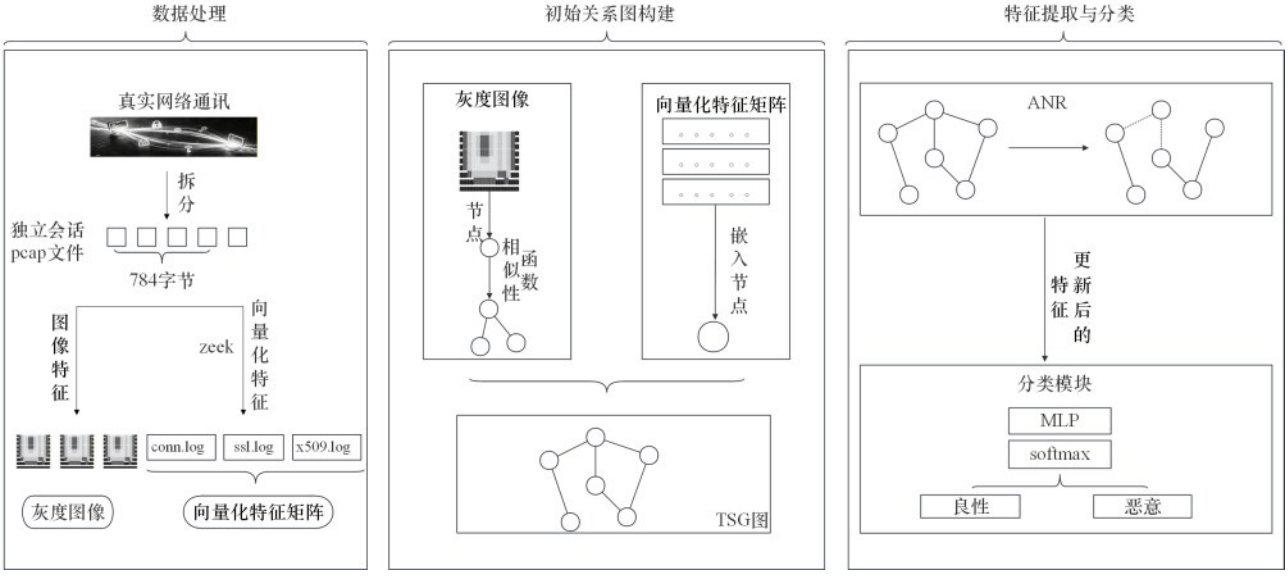

图 1　ANL-GNN 模型框架图

数据进行清洗、匿名化，同时删除空流与重复数据。最后截取会话前784字节，不足部分用0填充，将每个字节映射为[0,255]的灰度值，重塑为28×28的灰度图像。

尽管两类特征来源于同一加密流量数据，在信息层面可能存在一定程度的重合，但其表达侧重点存在明显差异。本文4.8节对两类特征做了详细的实验分析。

2.3　TSG图构建

构建TSG图G=(V,E,$\boldsymbol{X}$)，图像特征的相似性用来判断节点之间的关联，向量化特征作为节点的属性特征。

节点定义：每个加密会话对应图中的一个节点V_i∈V，节点总数为N；

边集构建：基于节点特征之间的距离关系构建节点连接,通过对节点特征向量之间的欧氏距离进行指数映射，得到节点间的相对关联强度，并在局部邻域内进行归一化处理，从而得到节点间相似度

$$S_{ij}=\frac{\exp\left(-\frac{\left\|\boldsymbol{X}_i-\boldsymbol{X}_j\right\|_2^2}{t^2}\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(-\frac{\left\|\boldsymbol{X}_i-\boldsymbol{X}_k\right\|_2^2}{t^2}\right)} \tag{1}$$

选其中，$\boldsymbol{X}_i$与$\boldsymbol{X}_j$分别表示节点i与节点j对应的流量灰度图像特征向量，$\|\cdot\|_2$表示二范数，t热核参数，用于控制特征距离对相似性衰减

选取每个节点在相似空间中的有限局部邻域作为初始连接关系，形成无向边集E；

特征矩阵：向量化特征作为节点属性，构建N×d的属性矩阵$\boldsymbol{X}$，实现图像特征与向量化特征的深度融合。

2.4　自适应邻域学习ANR模块

为了实现邻域结构的动态优化，本文引入一种可微分的自适应邻域学习模块（ANR），对邻接关系进行动态建模。其结构如图2所示。

该模块以节点特征为输入，主要由节点表示编码、边相似性估计、自适应领域大小估计以及可微分连续邻域筛选机制等组成。

2.4.1　节点表示编码

针对节点原始特征维度较高的问题，引入节点编码对节点特征进行映射，得到低维潜在表示。通过多层感知机将节点特征属性$\boldsymbol{X}$映射至latent特征空间，得到维度$R^{N\times d'}$的潜在表示$\hat{\boldsymbol{X}}$，其表达式为：

$$\hat{\boldsymbol{X}}=N_\phi(\boldsymbol{X}) \tag{2}$$

其中N_ϕ为带可学习参数ϕ的MLP。该特征同时用于后续的边排序与度估计，并且作为后续图卷积的输入，避免梯度消失。

2.4.2　边相似性估计

基于节点的潜在表示，对节点之间的关联强度进行建模，通过计算节点间的相似性，对边进行排序，从而为邻域选择提供依据。

首先将节点潜在特征$\hat{X}_i$与$\hat{X}_j$拼接后输入MLPl_ϕ，生成包含局部信息的边嵌入$\hat{c_{ij}}\in R^{d'}$，表达式为：

$$\hat{c}_{ij}=l_\phi\left(\hat{\boldsymbol{X}}_i,\hat{\boldsymbol{X}}_j\right) \tag{3}$$

然后通过MLPm_θ将边嵌入$\hat{c_{ij}}$转换为边概率向量$p_i\in R^N$，表达式为：

$$p_i=\{\,m_\theta\left(\hat{c}_{ij}\right)|\forall j\in V\,\} \tag{4}$$

最后通过Gumbel-Softmax[24]生成可微分边权重e_i，平衡梯度稳定性，表达式为：

$$e_{ij}=\left\{\frac{\exp\left(\frac{\log\left(p_{ij}\right)+g_i}{\tau}\right)}{\sum_j\exp\left(\frac{\log\left(p_{ij}\right)+g_i}{\tau}\right)}|\forall j\in N\right\} \tag{5}$$

其中g~Gumbel(0,1)为Gumbel噪声，τ为温度参数，用于控制采样离散程度，τ越大，输出e_{ij}越平滑，

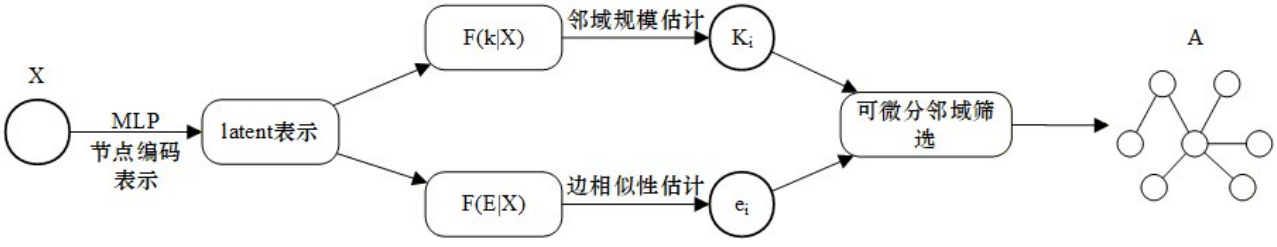

图2　ANR模块图

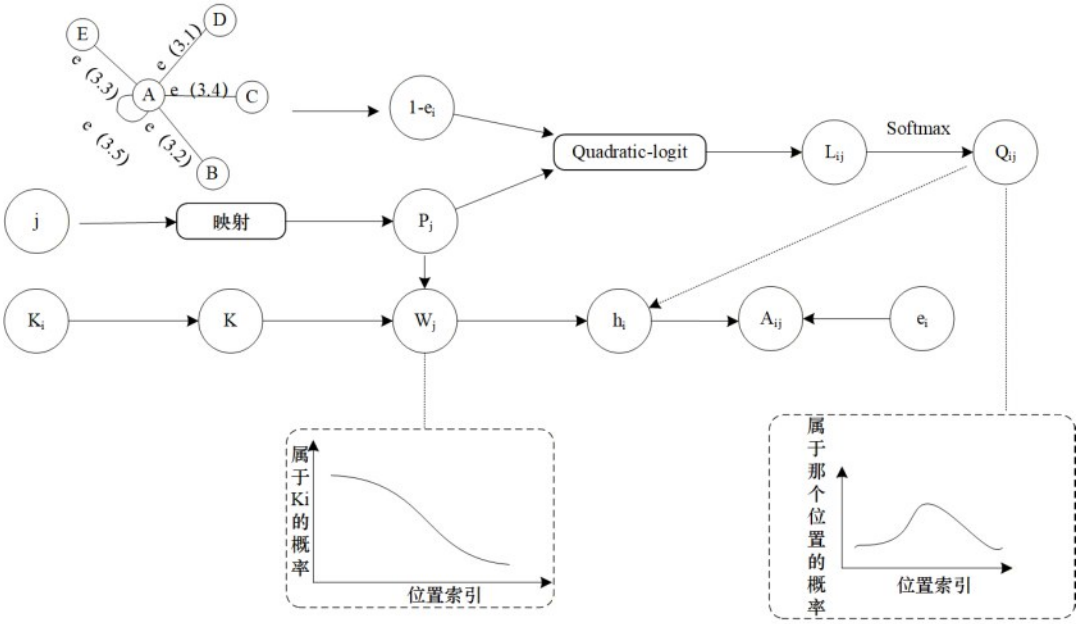

图3　可微分邻域筛选图

越接近连续分布，梯度越稳定；τ越小，输出出越接近离散的one-hot向量，梯度方差会增大。

2.4.3　自适应邻域大小估计

考虑到不同节点在结构与语义上的差异性，引入邻域规模自适应估计机制，对每个节点的邻域大小进行动态预测，避免采用统一邻域规模所带来的模型偏差。

首先基于类变分自编码器(VAE-like)框架[25]，将潜在特征$\hat{\boldsymbol{X}}$，通过MLPM_ρ和S_ρ分别编码为均值μ_i与方差σ_i，结合重参数化技巧，生成潜在变量z_i，表达式为：

$$\mu_i,\sigma_i = M_\rho\left(\hat{\boldsymbol{X}}_i\right),S_\rho\left(\hat{\boldsymbol{X}}_i\right) \tag{6}$$

$$z_i = \mu_i + \sigma_i \odot \epsilon,\epsilon \sim N(0,1) \tag{7}$$

后续将z_i通过MLPD_ρ解码后与边权重e_i的L_1范数拼接，得到连续值$k_i \in \mathrm{R}$，作为节点i的自适应邻域大小，实现节点级别的邻域规模优化，表达式为：

$$k_i = D_\rho\left(z_i\right) + /\!/\ e_i\ /\!/_1 \tag{8}$$

2.4.4　可微分连续邻域筛选机制

结合可微分的连续阈值筛选机制，对排序后的边进行筛选，生成连续可优化的邻接矩阵，从而在训练过程中实现邻域结构的自适应更新，具体步骤如图3所示。

1. 使用可微邻域掩码函数实现边选择，通过将离散排序位置映射为连续坐标，从而避免不可导排序操作对模型训练的影响[26]。

$$p_j = \frac{j-1}{n-1}\left(j = 1,2,\cdots,n\right) \tag{9}$$

2. 通过可学习函数$\phi(s)$将边权重e_i（score）映射至位置空间，其中

$$\phi(s) = 1 - s \tag{10}$$

确保分数与位置方向一致，提升筛选准确率；

3. 构建Quadratic-logit，计算logit值L_{ij}，形成以$\phi(e_i)$为中心的单峰二次函数，表达式为：

$$L_{ij} = -\frac{\left(p_j - \phi\left(e_i\right)\right)^2}{\tau} \tag{11}$$

其中τ为温度参数，τ越小，曲线越陡峭，τ越大，曲线越平缓;

后续L_{ij}进行softmax操作，得到单峰行随机矩阵$P \in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$，每行表示节点属于对应排序位置的概率，表达式为：

$$Q_{ij} = softmax\left(L_{ij}\right) \tag{12}$$

4. 构造位置权重函数，通过Sigmoid函数实现对邻域边的平滑阈值控制，从而在保持邻域选择连续性的同时避免硬阈值带来的梯度不稳定问题。表达式为：

$$w_j = \sigma\left(\frac{k - p_j}{\lambda}\right) \tag{13}$$

其中 $k=k_i/N$，σ 为 sigmoid 函数，λ 为平滑参数。当 $p_j<k$ 时，$w_j\approx1$；当 $p_j>k$ 时，$w_j\approx0$。保障邻域筛选的平滑性。

5. 生成邻域选择权重函数，通过加权求和得到节点 i 属于所选邻域的概率掩码 h_i，表达式为：

$$h_i=\sum_j Q_{ij}w_j \tag{14}$$

基于 h_i 与 e_i 生成优化后的邻接矩阵 $\boldsymbol{A}\in\mathrm{R}^{\mathrm{N}\times\mathrm{N}}$，，其中 A_{ij} 表示节点 i 与 j 的连接权重，表达式为：

$$A_{ij}=h_{ij}\cdot e_{ij} \tag{15}$$

为直观展示自适应邻域选择相比固定邻域结构的优势，图 4 给出了两种邻域构建方式的对比示意图。固定邻域结构对所有节点采用统一的邻域规模，如图左所示，节点均选择 2 个邻居，忽略节点特征差异，易引入噪声邻居；自适应邻域选择根据节点特征与边相似性动态调整邻域构成，如图右所示，节点根据自身特征分别选择 2 个、3 个高相关性邻居，仅保留有效连接，抑制冗余边，适配不同节点的结构与语义差异。

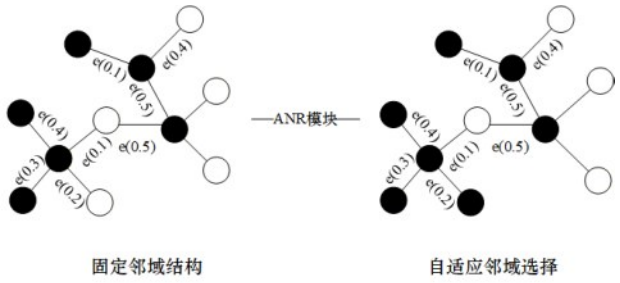

图 4　ANR 效果图

2.5　图卷积分类与模型优化

在特征传播阶段，本文使用 GraphSAGE[16] 进行节点表示学习。GraphSAGE[16] 方法采用固定邻域数量的随机采样，该方法难以适配不同节点在结构与语义上的差异性。因此，将传统 GraphSAGE[16] 中的固定邻域采样机制替换为所提出的 ANR 模块，通过动态学习的自适应邻域结构实现特征聚合，使模型能够根据节点特征优化邻域。

具体而言，对于节点 v_i，通过聚合其自适应邻域内节点的表示，并与自身特征进行组合，得到新的节点嵌入表示:

$$h_i^{(l|1)}=\sigma\left(W^{(l)}\cdot \mathrm{AGG}\left(h_i^{(l)},\{h_j^{(l)}|j\in\mathcal{N}_i\}\right)\right) \tag{16}$$

其中，N_i 表示由 ANR 模块动态学习得到的节点 i 的邻域集合，$AGG(\cdot)$ 表示邻域聚合函数，$W^{(l)}$ 为可学习参数，$\sigma(\cdot)$ 为非线性激活函数。该聚合方式避免了固定归一化邻接矩阵的假设，使模型能够更灵活地适应自适应邻域结构。

在分类与损失函数的选择上，使用 MLP 分类器，将节点嵌入后，经 softmax 函数输出分类概率

$$Z_C=softmax\left(\mathrm{MLP}\left(h_v^l\right)\right) \tag{17}$$

采用交叉熵损失函数优化模型参数，最小化分类误差：

$$Loss=-\frac{1}{N}\sum_{n=1}^{N}\sum_{C=0}^{1}y_C\log\left(Z_C\right) \tag{18}$$

其中 y_C 为真实标签，N 为节点总数。训练过程中采用随机梯度下降优化所有参数，包括 ANR 模块、图卷积权重与分类器参数。

虽然本文主要主要优化分类损失，但自适应邻域学习模块在训练过程中隐式实现了对图结构复杂度的正则化。

3　数据集与预处理

本文实验采用 CTU-13[28] 与 MCFP 两个真实网络流量数据集。CTU-13 侧重复杂僵尸网络场景，MCFP 则覆盖多类型恶意通信且规模较大。本文采用基于会话的数据划分策略，按 8:1:1 比例随机划分为训练集、验证集和测试集。最终合并数据集包含恶意加密会话 63,720 条、良性加密会话 69,358 条，共计 133,078 条。为 TSG 图构建与 ANR 模型训练提供了数据基础。

4　实验设置

4.1　实验环境

实验基于 PyTorch 深度学习框架实现，具体软硬件参数信息如表 1 所示。模型训练中，TSG 图构建的热传导时间参数设为 t=0.5；优化器选用 Adam (始学习率 0.001，权重衰减 5e-4)。在核心的 ANR 模块中，Gumbel-Softmax 温度参数 τ 与 Sigmoid 平滑参数 λ 分别设定为 0.5 和 0.2；分类任务采用交叉熵损失函数。

4.2　评价指标

为验证所提方法的有效性，分类指标采用准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1 分数（F1-score），全面衡量模型的分类性能。其中准确率反映整体分类性能，精确率与召

表1　环境配置信息

组件	型号
操作系统	Ubuntu 22.04 64位
深度学习框架	PyTorch 2.1.2 on Python 3.10
CPU	12 vCPU Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz
GPU	RTX 3080 Ti(12GB) * 1
CUDA	CUDA 11.8
内存	90GB

回率衡量模型在恶意流量识别中的误报情况，F1分数反映模型在不平衡数据场景下的检测能力，各评价指标的具体计算公式如(19)-(22)所示：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{20}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

其中，TP、TN、FP、FN分别表示真阳性、真阴性、假阳性及假阴性，即正样本预测正确的数量、负样本预测正确的数量、正样本预测错误的数量、负样本预测错误的数量。

4.3　实验结果

为了验证本文方法的可行性，在合并数据集上开展实验，同时为减小结果随机性带来的影响，对实验重复运行5次，并取其平均值作为最终结果，实验结果如图5所示。

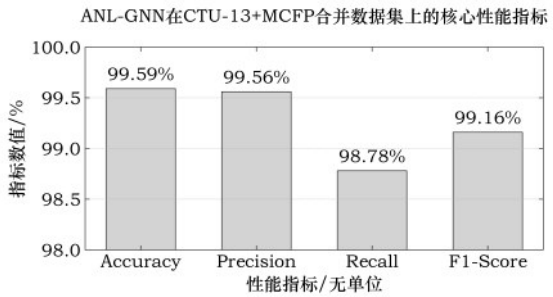

图5　合并数据集实验结果

由图5可知，本文方法在四大指标上均表现优异：Accuracy达99.59%，Precision为99.56%，Recall为98.78%，F1-Score为99.16%，充分验证了自适应邻域学习与多元化特征融合的有效性。

4.3　模型超参数分析

4.4.1　Gumbel-Softmax中温度参数τ敏感性分析

温度参数τ用于平衡Gumbel-Softmax输出的离散程度与梯度稳定性，τ越小输出越接近离散one-hot向量，稀疏性强但梯度方差大；τ越大输出越平滑，梯度稳定但稀疏性弱。实验τ取值范围为[0.1-0.9]，结果如图6所示

由图6可知，温度参数τ的取值对模型性能具有一定影响，其中当τ=0.5时模型综合性能最优，准确率、精确率与F1分数分别达到99.59%、99.56%和99.16%，仅召回率为98.78%，略低于τ=0.3时98.95%的峰值，但整体仍保持高水准。当τ<0.5时，精确率与F1分数呈上升趋势，这一现象可能源于过小的τ导致边权重分布过于集中，使得部分低相似度的邻居被忽略；而当τ>0.5时，所有评价指标均呈下降趋势，这是因为过大的τ使边权重过度平滑，难以有效区分关键邻域与噪声邻域，进而导致邻域选择精度降低，影响模型整体性能。

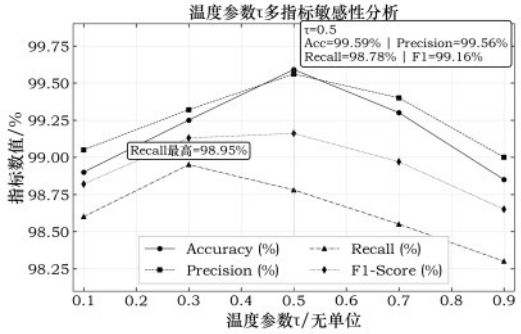

图6　温度参数τ敏感性分析

4.4.2　平滑参数λ敏感性分析

平滑参数λ控制Sigmoid阈值函数的陡峭度，λ越小曲线越陡，邻域筛选越接近硬阈值，梯度易不稳定；λ越大曲线越平缓邻域筛选模糊，冗余连接

增多。实验中λ取值范围为[0.1-0.9]，结果如图7所示。

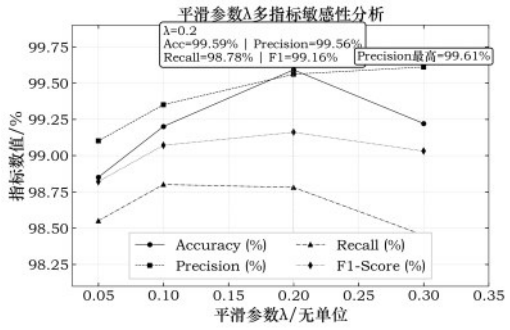

图7　平滑参数λ多指标敏感性分析

由图7可知，平滑参数λ对邻域筛选精度与模型整体性能具有一定影响，其中当λ=0.2时模型达到最优性能，准确率、召回率与F1分数均达到峰值，仅精确率略低于λ=0.3时99.61%的数值，充分说明该取值能够精准筛选邻域边，有效平衡筛选精度与梯度稳定性。当λ<0.2时，召回率从98.00%提升至98.78%，但精确率增长较为缓慢，这一现象可能源于过小的λ使Sigmoid阈值曲线过于陡峭，导致部分边缘有效邻域被误筛除；而当λ>0.2时，精确率虽有小幅提升，但准确率与F1分数均呈现明显下降趋势，推测是过大的λ让阈值筛选趋于平缓，使得邻域中冗余连接增多，引入噪声邻域干扰特征聚合过程，进而影响模型的判别能力。

4.4.3　图构建中参数敏感性分析

为分析在图构建过程中关键参数对模型性能的影响，本文在合并数据集上对热核参数t及邻域规模P进行实验分析。在实验过程中，采用控制变量法分别对两个参数进行评估：在分析t时固定P=15，在分析P时固定t=1。实验结果如表2所示。

由表2可知，模型在不同参数下性能波动较小。这主要因为ANR模块的动态邻域学习与噪声抑制机制能够有效削弱初始图构建中参数的影响；但合理的参数选择依然对模型性能具有重要作用。

4.4.4　超参数选择结论

结合上述敏感性分析的结果，最终选择t=1,P=15,τ=0.5、λ=0.2作为ANR模块的超参数。该组合下模型在核心指标Accuracy、F1-Score值最高，同时Precision与Recall值合理，充分适配本文需求。

4.5　与现有方法进行对比

为验证所提方法的有效性，选取3种传统机器学习方法、2种深度学习方法及3种经典的图神经网络方法进行对比。实验结果如表3所示。

表2　图构建中参数敏感性分析实验结果

参数配置		Accuracy(%)	Precision(%)	Recall (%)	F1-score(%)
t=1	P=10	99.48	99.45	98.68	99.07
	P=15	99.59	99.56	98.78	99.16
	P=20	99.52	99.50	98.72	99.11
P=15	t=0.5	99.50	99.48	98.70	99.09
	t=1	99.59	99.56	98.78	99.16
	t=1.5	99.53	99.51	98.73	99.12

从表3中可以看出：

(1)本文ANL-GNN方法性能表现：准确率99.59%、精确率99.56%、召回率98.78%、F1-Score99.16%。该结果主要源于自适应邻域学习模块能够动态优化图结构，根据数据分布筛除噪声边、强化强关联节点，同时多元化特征融合为模型提供了更多的特征表达，使模型在复杂加密流量场景中仍有较好的分类性能。

(2)SVM、Random Forest、XGBoost，3个传统学习方法的准确率低于96%，其中XGBoost准确率仅为93.65%。

(3)CNN与TSCRNN，2个深度学习方法通过提取原始流量特征，准确率分别为96.03%和98.21%二者仅聚焦欧氏空间，忽略了流量间的交互关系，导致性能提升受限。

(4)所有图模型的准确率超过97%，优于前两类方法。其中，静态图模型GCN[15]、GraphSAGE[16]、GAT[17]受限于固定邻域结构，难以适配真实流量的复杂性，性能低于近年来的图学习模型；SAT-Net通过注意力机制强化关键关联，MT-Security融合多视角特征，准确率分别提升至98.97%和99.46%。

本文方法在Accuracy与Precision指标上略优于部分对比方法，而在Recall与F1指标上略低。

这源于ANR模块在邻域筛选过程中优先保留高置信度邻接节点，并抑制噪声干扰，有效降低了误报率，增强了模型判断的稳定性。然而，对于部分位于类别边界区域的样本，其关联特征在筛选机制下被弱化，所以Recall与F1指标呈现小幅下降趋势。

表3　对比实验

方法	图建模方式	Accuracy(%)	Precision(%)	Recall (%)	F1-score (%)
SVM	×	92.10	91.84	90.76	91.30
RF	×	94.71	93.73	93.54	93.63
XGBoost	×	93.65	92.51	92.42	92.61
CNN	×	96.03	95.74	95.21	95.47
TSCRNN	×	98.21	97.83	97.58	97.60
GCN	静态图	97.84	97.51	96.93	97.22
GAT	静态图	98.12	97.88	97.20	97.54
GraphSAGE	静态图	98.45	98.16	97.62	97.89
SAT-Net	注意力图	98.97	98.74	98.02	98.38
MT-Security	多视角图	99.46	99.54	**99.26**	**99.40**
ANL-GNN	自适应图结构	**99.59**	**99.56**	98.78	99.16

4.6　ANR机制在不同神经网络中的适用性分析

为验证所提出的ANR模块在3种经典图神经网络模型中的性能表现，本文将ANR分别嵌入GCN、GAT和GraphSAGE。在该实验中，各模型在相同条件下进行，仅替换基础图神经网络模型以保证对比的公平性，最终实验结果如表4所示。

表4　不同图神经网络性能对比

模型配置	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
GCN	98.74	98.52	97.91	98.21
GAT	99.12	98.96	98.43	98.69
GraphSAGE	99.59	99.56	98.78	99.16

由表4结果可知：ANR模块在GCN、GAT及GraphSAGE三种经典图神经网络中取得了较好的检测效果。其中GraphSAGE结果最优，这源于其邻域采样与聚合方式能够更好地适配ANR模块所学习到的自适应邻域结构，在保证建模能力的同时有效抑制了噪声传播。

4.7　消融实验

4.7.1　多元化特征与整体模块消融实验

本节实验围绕TSG图构建与自适应邻域模块ANR两大关键部分展开，分析图构建方式、核心模块对模型性能的影响。实验结果如表5所示。

由表5结果可知：

(1)图像特征与向量化特征分别从流量空间分布与统计行为两个角度描述流量信息，二者具有较好的互补性。仅使用单一特征时，模型对复杂流量模式的表达能力存在一定局限；而融合两类特征后，模型各项指标均得到提升，说明多元化特征融合能够有效增强模型检测能力。。

(2)移除ANR模块后，模型采用固定邻域的静态图结构进行训练，难以充分适应真实流量之间的差异性，导致模型性能下降。其中，Accuracy由99.59%下降至98.45%，F1-score由99.16%下降至97.89%。说明ANR模块能够通过动态优化邻域关系，有效提升模型对节点关系的建模能力。

表5　消融实验结果

模型配置	Accuracy(%)	Precision(%)	Recall (%)	F1-score (%)
仅向量化特征	98.62	98.45	97.83	98.14
仅图像化特征	98.87	98.71	98.05	98.38
ANL-GNN	99.59	99.56	98.78	99.16
w/o自适应邻域模块	98.45	98.16	97.62	97.89

4.7.2　ANR模块内部结构消融实验

为进一步分析所提出自适应邻域学习模块（ANR）中各组成部分的作用，本文在保持其他设置不变且相同实验条件下，从边相似性估计、自适应邻域大小估计以及可微分连续邻域筛选机制三个方面分别进行消融实验，包括(1)去除边相似性估计，(2)去除自适应邻域大小估计，(3)去除可微分连续邻域筛选机制，以验证各子模块对整体性能的影响。

需要说明的是，在涉及固定邻域规模与Top-k邻域选择的消融实验中，本文统一选择邻域大小K=10。实验结果如表6所示。

从表6可以看出，去除任一子模块后，模型性能均出现不同程度的下降。其中，去除边相似性估计后，模型难以有效区分关键邻居与非关键邻居，导致性能下降较为明显；在去除自适应邻域大小估计后，固定邻域规模难以同时适应不同流量样本的结构差异，模型表现有所下降；而将可微分连续邻域筛选机制替换为Top-k策略后，由于离散选择方式缺乏平滑性，容易引入不稳定的邻域结构，从而对模型性能产生一定影响。

综合来看，ANR模块中各组成部分均对模型

表6 ANR模块内部消融实验结果

模型配置	Similarity	Size	Mask	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
ANR	√	√	√	99.59	99.56	98.78	99.16
边相似性估计	×	√	√	98.42	98.30	97.95	98.10
自适应邻域大小估计	√	×	√	98.63	98.52	98.08	98.30
可微分连续邻域筛选机制	√	√	×	98.55	98.40	98.05	98.22

的性能表现有重要作用，三者协同能够更有效地刻画节点间关系并优化邻域结构，从而提升整体检测性能。

4.8 特征可视化

为分析图像特征与向量化特征的分布差异，本文采用t-SNE方法对高维特征进行二维可视化。考虑到样本数量过多会导致特征点分布过于密集，影响观察效果，因此分别从两类特征中随机选取2000个节点进行可视化，结果如图8所示。

由图8可知，图(a)向量化特征在低维空间中呈现出连续分布特征，部分样本形成带状结构。这是因为向量化特征基于流量统计属性构建，反映流量行为之间的整体变化关系。但部分恶意样本之间仍存在一定交叠现象，说明仅依赖统计特征时，对复杂恶意流量的区分能力仍存在一定局限。

图(b)中的图像特征呈现出聚类结构，不同类别之间的边界更加清晰。这是因为图像特征能够保留原始字节序列中的局部空间分布信息，从而增强模型对流量细粒度差异的表达能力。

综合来看，两类特征在低维空间中的分布形式存在明显差异，分别从统计行为与空间结构两个角度描述流量信息，具有较好的互补性。

4.9 模型泛化性与鲁棒性实验

4.9.1 泛化性实验

为进一步评估所提出方法在不同数据环境下的性能表现，本文在保持模型参数与训练策略不变的条件下，引入额外数据集进行跨数据集测试。本文选取CICIDS2017与USTC-TFC2016两个公开数据集进行泛化性实验，结果如表7所示。

表7 泛化性实验结果

数据集	Accuracy(%)	Precision(%)	Recall(%)	F1-score(%)
CICIDS2017	98.33	98.10	97.61	97.79
USTC-TFC2016	97.61	97.38	96.72	97.05
流量填充	99.18	99.14	98.32	98.73
时序混淆	98.74	98.69	97.85	98.26

由表7可以看出，模型在不同数据集上的性能存在一定差异，其中CICIDS2017数据集上的结果相对更优。这是因为不同数据集在流量构成及通信行为模式上存在一定差异，其中CICIDS2017与合

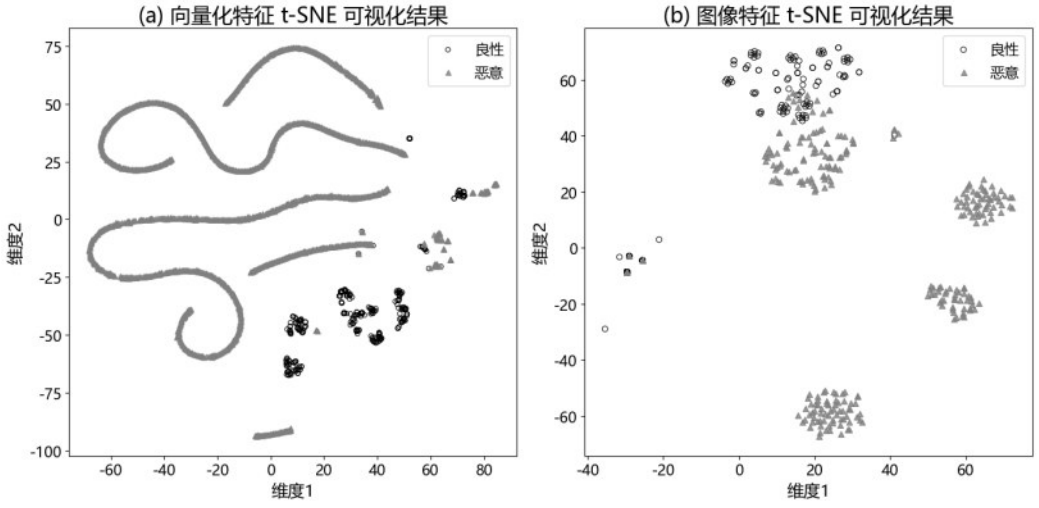

图8 特征可视化分析

并数据集的数据分布更为接近，因此能够取得更好的检测效果；USTC-TFC2016 与训练数据之间的差异相对更大，因此模型性能出现一定下降。

由上述分析可知，模型在不同数据分布条件下仍能够保持较好的检测性能。

4.9.2　鲁棒性实验

在上述跨数据集泛化实验的基础上，在合并数据集的测试集中增加扰动，并使用训练好的模型对扰动后的测试集进行检测，验证方法在真实网络环境中的抗干扰能力。实验采用了两种典型扰动方式：

(1)流量填充：在字节序列中随机选取约 10% 的位置，将其数值替换为取值范围(0，255)之间的随机字节。

(2)时序混淆：将字节序列按每 32 字节划分为固定长度的子块，并对这些子块的顺序进行随机重排，从而破坏原有的局部时序结构。

具体实验结果如表 7 所示。

由表 7 可以看出，在流量填充与时序混淆扰动下，模型性能分别下降约 0.4% 和 0.8%，其中时序混淆对模型影响更为明显。这是因为时序混淆会破坏原有字节序列的空间结构，从而对模型特征表示产生更直接的影响；相比之下，流量填充主要改变局部字节分布，对整体结构影响相对较小。

4.10　模型效率实验

4.10.1　ANR 与固定邻域效率对比

考虑到 ANR 模块在结构上引入了边关系估计及邻域自适应选择机制等，提升了模型的检测能力，但同时也带来了额外的计算开销。因此，本节将从推理时间、显存占用及参数规模三个方面，对模型效率进行评估，以验证方法在检测性能与计算开销之间的平衡性。为保证实验公平性，两种模型在除邻域建模方式外的网络结构均保持一致，实验结果如表 8 所示。

表 8　　效率实验对比结果

模型选择	推理时间	显存占用	参数量
ANR	23.91ms	1.35GB	4.28M
固定邻域选择	17.84ms	1.02GB	3.32M

从表 8 可以看出，引入 ANR 模块后，模型在推理时间、显存占用及参数规模上均有一定程度的增加。这主要来源于 ANR 中边相似性估计、邻域大小估计及连续邻域筛选机制所引入的额外计算过程，同时结合性能实验的结果，计算开销的增加相比于模型在准确率等指标上取得的提升，结果仍处于可接受的范围内。同时，如何在保持性能增益的前提下进一步降低计算复杂度，使模型更加轻量化，将是后续值得深入探索的研究方向。

4.10.2　可扩展性分析

为验证本文方法在大规模流量场景下的可扩展性，在原始数据集基础上，通过逐步扩充流量样本规模的方式，构建不同规模的流量交互图，并在相同实验环境下对模型性能及成本开销情况进行测试。

具体而言，在原始流量会话集合中随机抽取部分会话样本，并在保留其流量统计属性与字节分布特征的基础上，为新增样本重新分配节点标识，并重新参与特征提取及 TSG 图构建过程，从而模拟并发流量规模逐步增加的场景。为避免重复样本在不同数据划分中引入数据泄漏问题，在规模扩展后重新划分训练集、验证集与测试集，以避免重复样本同时出现在不同数据划分中所带来的数据泄漏问题。

实验分别构建 125%、150%、175% 及 200% 规模的数据集，并分析模型在不同图规模下的检测性能、推理时间及显存占用变化情况。实验结果如表 9 所示。

表 9　　可扩展性分析

数据规模	Accuracy(%)	Precision(%)	Recall (%)	F1-score (%)	推理时间	显存占用
125%	99.59	99.55	98.79	99.16	25.14	1.43
150%	99.59	99.56	98.77	99.16	26.82	1.52
175%	99.58	99.54	98.78	99.15	28.47	1.63
200%	99.59	99.56	98.77	99.16	30.26	1.75

由表 9 可知，随着流量规模的增加，模型推理时间与显存占用均有所增长。这是因为图节点数量增加后，邻域传播与边关系计算的数据量随之增加。但各项资源消耗增长较为平稳，未出现计算开销突增的现象。同时，在数据规模扩大的情况下，模型各项检测指标仅出现小幅波动，整体仍保持较好的检测性能。

5 总结

本文提出一种基于自适应邻域学习的加密恶意流量检测方法。通过构建TSG流量交互图融合图像与向量化特征，实现流量交互关系建模；引入ANR模块动态学习邻域关系及规模，增强模型对复杂流量交互模式的表达能力。在CTU-13与MCFP数据集上，Accuracy、Precision、Recall与F1-score分别达99.59%、99.56%、98.78%与99.16%。当前方法采用全局图训练策略，在大规模图数据下存在计算与存储开销问题，且ANR模块引入额外计算复杂度。未来将研究轻量化自适应邻域学习机制，提升模型在超大规模复杂网络环境中的计算效率与适用性。

参考文献：

[1] Shekhawat A S, Di Troia F, Stamp M. Feature analysis of encrypted malicious traffic [J]. Expert Systems with Applications, 1999, 28(6): 1998-2029.

[2] Taylor V F, Spolaor R, Conti M, et al. Robust smartphone app identification via encrypted network traffic analysis[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(1): 63-78.

[3] Shen M, Liu Y, Zhu L, et al. Fine-grained webpage fingerprinting using only packet length information of encrypted traffic [J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 2046-2059.

[4] Wang Z, Fok K W, Thing V L L. Machine learning for encrypted malicious traffic detection: Approaches, datasets and comparative study[J]. Computers & Security, 2022, 113: 102542.

[5] Wang W, Zhu M, Wang J, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]//2017 IEEE international conference on intelligence and security informatics (ISI). IEEE, 2017: 43-48.

[6] Shapira T, Shavitt Y. Flowpic: Encrypted internet traffic classification is as easy as image recognition[C]//IEEE INFOCOM 2019-IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, 2019: 680-687.

[7] Lotfollahi M, Jafari Siavoshani M, Shirali Hossein Zade R, et al. Deep packet: A novel approach for encrypted traffic classification using deep learning[J]. Soft Computing, 2020, 24(3): 1999-2012.

[8] Lin K, Xu X, Gao H. TSCRNN: A novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of IIoT[J]. Computer Networks,2021,190:107974.

[9] Yao H, Liu C, Zhang P, et al. Identification of encrypted traffic through attention mechanism based long short term memory[J]. IEEE Transactions on Big Data, 2019, 8(1): 241-252.

[10] Lin X, Xiong G, Gou G, et al. Et-bert: A contextualized datagram representation with pre-training transformers for encrypted traffic classification[C]//Proceedings of the ACM Web Conference 2022. 2022: 633-642.

[11] Barut O, Luo Y, Li P, et al. R1dit: Privacy-preserving malware traffic classification with attention-based neural networks[J]. IEEE Transactions on Network and Service Management, 2022, 20(2): 2071-2085.

[12] Wang W, Shang Y, He Y, et al. BotMark: Automated botnet detection with hybrid analysis of flow-based and graph-based traffic behaviors [J]. Information Sciences, 2020, 511: 284-296.

[13] Shen M, Zhang J, Zhu L, 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.

[14] Li Y, Li R, Zhou Z, et al. Graphddos: Effective ddos attack detection using graph neural networks[C]//2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2022: 1275-1280.

[15] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]// International Conference on Learning Representations. Toulon, France, 2017.

[16] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs[C]// Advances in Neural Information Processing Systems 30 (NIPS 2017. LongBeach, CA, USA, 2017: 1024-1034.

[17] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks [C]// International Conference on Learning Representations, 2018.

[18] Yang J, Jiang X, Lei Y, et al. MTSecurity: Privacy-preserving malicious traffic classification using graph neural network and transformer [J]. IEEE Transactions on Network and Service Management, 2024, 21 (3): 3583-3597.

[19] Li Z, Zhao H, Zhao J, et al. SAT-Net: A staggered attention network using graph neural networks for encrypted traffic classification[J]. Journal of Network and Computer Applications, 2025, 233: 104069.

[20] Zhao W, Wu Q, Yang C, et al. Graphglow: Universal and generalizable structure learning for graph neural networks[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023: 3525-3536.

[21] Zou D, Peng H, Huang X, et al. Se-gsl: A general and effective graph structure learning framework through structural entropy optimization [C]//Proceedings of the ACM web conference, 2023. 2023: 499-510.

[22] Zou X, Li K, Chen C, et al. DGSLN: Differentiable graph structure learning neural network for robust graph representations[J]. Information Sciences, 2023, 626: 94-113.

[23] Xie X, Chen W, Kang Z. Robust graph structure learning under heterophily[J]. Neural Networks, 2025, 185: 107206.

[24] Ang E, Gu S, Poole B. Categorical reparameterization with gumbel-softmax[C]// International Conference on Learning Representations. Toulon, France, 2017.

[25] Kingma D P, Welling M. Auto-encoding variational bayes[C]// International Conference on Learning Representations. Banff, AB, Canada, 2014.

[26] Prillo S, Eisenschlos J. Softsort: A continuous relaxation for the argsort operator[C]//International Conference on Machine Learning. PMLR, 2020: 7793-7802.

[27] Xie Y, Dai H, Chen M, et al. Differentiable top-k with optimal transport [C]// Advances in Neural Information Processing Systems 33 (NeurIPS 2020. Virtual, 2020: 20520-20531.

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

[29] Wang R, Zhao J, Zhang H, et al. Network Traffic Analysis Based on

Graph Neural Networks: A Scoping Review[J]. Big Data and Cognitive Computing, 2025, 9(11): 270.

[30] Zhang H, Xiao X, Yu L, et al. One train for two tasks: An encrypted traffic classification framework using supervised contrastive learning [EB/OL]. arXiv:2402.07501, 2024.

[作者简介]

张丽娜（1981—），女，西安科技大学副教授，主要研究方向为信息安全和密码学。

杨阳（2001—），男，西安科技大学硕士研究生，主要研究方向为信息安全和密码学。

鲁亦群（2002—），男，西安科技大学硕士研究生，主要研究方向为图像安全和深度学习。

贾弘瑜(2001–)，女，西安科技大学硕士研究生，主要研究方向为图像安全和深度学习。

段爱卓伦（2002–），女，西安科技大学硕士研究生，主要研究方向为图像安全和深度学习