计算机科学
COMPUTER SCIENCE

基于图神经网络的SSL/TLS加密恶意流量检测算法研究

唐瑛, 王宝会

引用本文

唐瑛, 王宝会. 基于图神经网络的SSL/TLS加密恶意流量检测算法研究[J]. 计算机科学, 2024, 51(9): 365-370.

TANG Ying, WANG Baohui. Study on SSL/TLS Encrypted Malicious Traffic Detection Algorithm Based on Graph Neural Networks [J]. Computer Science, 2024, 51(9): 365-370.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230800079

基于图神经网络的 SSL/TLS 加密恶意流量检测算法研究

唐 瑛 王宝会
北京航空航天大学软件学院 北京 100191
(1543616175@qq.com)

摘 要 为实现 SSL/TLS 加密恶意流量的精准检测,针对传统机器学习方法过分依赖专家经验的问题,提出一种基于图神经网络的恶意加密流量检测模型。通过对 SSL/TLS 加密会话进行分析,利用图结构对流量会话交互信息进行表征,将恶意加密流量检测问题转化为图分类问题。生成的模型基于分层图池化架构,通过多层卷积池化的聚合,结合注意力机制,充分挖掘图中节点特征和图结构信息,实现了端到端的恶意加密流量检测方法。基于公开数据集 CICAndMal2017 进行验证,实验结果表明,所提模型在加密恶意流量二分类检测中,准确率高达 97.1%,相较于其他模型,准确率、召回率、精确率、F1 分数分别提升了 2.1%,3.2%,1.6%,2.1%,说明所提方法对于恶意加密流量的表征能力和检测能力优于其他方法。

关键词:SSL/TLS;恶意加密流量;图神经网络;图分类;分层池化
中图分类号 TP393.08

Study on SSL/TLS Encrypted Malicious Traffic Detection Algorithm Based on Graph Neural Networks

TANG Ying and WANG Baohui
School of Software,Beihang University,Beijing 100191,China

Abstract In order to achieve precise detection of SSL/TLS encrypted malicious traffic,a graph neural network-based model for malicious encrypted traffic detection is proposed,to address the issue of excessive reliance on expert experience in traditional machine learning methods. Through the analysis of SSL/TLS encrypted sessions,the interactive information within traffic sessions is characterized using a graph structure,transforming the problem of detecting malicious encrypted traffic into a graph classification task. The proposed model is based on a hierarchical graph pooling architecture,which aggregates through multiple layers of convolutional pooling,incorporating attention mechanisms to fully exploit node features and graph structure information,resulting in an end-to-end approach for malicious encrypted traffic detection. The proposed model is evaluated on public CICAndMal2017 dataset. Experimental results demonstrate tha it achieves an accuracy of 97.1% in binary classification of encrypted malicious traffic detection,outperforming other models with an accuracy improvement of 2.1%,recall improvement of 3.2%,precision improvement of 1.6%,F1 score improvement of 2.1%. These results indicate that the proposed method exhibits superior representational and detection capabilities for malicious encrypted traffic in comparison to other methods.

Keywords SSL/TLS,Malicious encrypted traffic,Graph neural network,Graph classification,Hierarchical pooling

1 概述

随着互联网的普及,以及数字化业务的蓬勃发展,大量涉及个人隐私和敏感数据的交易和传输需要加密保护,以防止信息泄露和欺诈行为。SSL(Secure Socket Layer)和 TLS(Transport Layer Security)协议是最广泛使用的加密通信协议[1],用于保护网络传输中的敏感数据。Google 的报告显示,从 2016 年至 2023 年,加密流量的应用比例逐年上升,超过 95%的流量数据经过 SSL/TLS 加密[2]。随着 SSL/TLS 加密协议的广泛应用,恶意加密流量也呈现爆炸性增长趋势。恶意流量通过各种手段达到加密传输[3],如伪造 SSL/TLS 证书或者滥用有效证书,从而隐藏身份,利用 SSL/TLS 加密通道来传输恶意数据,绕过网络安全检测和拦截机制[4]。在无法对加密流量数据进行解密的情况下,传统的基于内容和签名的恶意流量分析技术在处理加密流量方面受到限制,识别和分析加密流量的安全性变得愈加困难。攻击者不断创新和改进攻击手法,使得恶意流量的特征和行为模式多样且不断变化,这种多元化的变化趋势给传统的恶意流量识别带来了挑战。本文以 SSL/TLS 加密流量为研究对象,利用网络流量会话中数据包之间的时空关系,提出一种基于图神经网络与注意力机制的恶意加密流量检测方法;并利用数据集 CICAndMal2017[5]对模型的效果进行验证。

2 相关研究

现有一些研究基于机器学习的方式,对加密流量的特征进行提取,利用传统机器学习模型进行分类。Chen 等[6]通过

到稿日期:2023-08-14 返修日期:2023-12-01
通信作者:王宝会(wangbh@buaa.edu.cn)

提取网络流特征和数据包特征，利用传统的随机森林、K 近邻、决策树等方法进行检测。Anderson 等[7]通过增加上下文流量中的有效负载来帮助检测，如 DNS 响应和 HTTP 头部信息等上下文信息，提升了模型的效果。Huo 等[8-9]针对单模型算法对多粒度特征适用性差的问题，集成多个分类器，分别利用 Stacking 策略和投票法原理，提升了模型的性能。

深度学习的方式主要基于 CNN(Convolutional Neural Network)和 RNN(Recurrent Neural Network)，利用模型的特征自学习，自动挖掘流量特征进行检测。Chen 等[10]将网络流量转换为网络指纹图像，然后将图像输入卷积神经网络进行预测，实现恶意流量的分类。Yang 等[11]将文本识别技术应用到加密流量识别领域，自定义 TextCNN 网络结构，通过多组一维卷积自动地从原始流量中提取上下文特征。Zhou 等[12]使用网络层的传输包序列和时间序列识别流量行为，基于长短期记忆网络(Long Short-Term Memory，LSTM)建立检测模型，分析流量传输模式的长距离依赖关系，从而进行恶意加密流量检测。Lopez-Martin 等[13]提出的混合模型使用 CNN 和 RNN 来捕获流的空间特征和时间特征。Jiang 等[14]结合长短时记忆网络和 Text CNN，整合加密流量的多尺度局部特征和双层全局特征，提升了模型的检测效果。

以上研究方法在恶意加密流量检测上取得了不错的效果，但是也存在一些局限。机器学习过分依赖专家经验对特征的选择，且对数据集的泛化能力不足。CNN 主要关注局部特征，难以捕捉全局依赖关系，无法考虑 SSL/TLS 流量的全局交互信息和时序关系。虽然 RNN 适合处理序列化的数据，但是它主要关注序列中的顺序信息，对于网络中存在的通信双方数据包交互关系的表征能力相对有限。

SSL/TLS 通信过程涵盖了网络中通信双方的复杂数据包交互，这种交互关系适合用图结构来表示。本文提出将会话转化为流量交互图，利用图神经网络(Graph Neural Network，GNN)处理流量交互图数据；基于会话的粒度，构建流量交互图，通过改进的图神经网络模型对 SSL/TLS 加密的恶意流量进行检测。本文的主要贡献如下：

(1)提出了一种新型流量表征方式，利用数据包的大小、时序和方向等信息构建流量拓扑图，充分体现通信会话中数据包之间的交互信息；

(2)采用分层图池化模型，并将不同池化层的输出串联，通过多次卷积池化操作，逐渐捕捉图数据的不同尺度和层次的特征，使模型能够同时考虑全局和局部信息；

(3)利用注意力机制学习节点的重要性得分，保留重要的节点和关键信息，增强图数据的表征能力。

3 基于图分层池化的恶意加密流量检测

利用图神经网络对 SSL/TLS 恶意加密流量进行检测，整体解决方案如图 1 所示。首先，根据五元组信息将原始流量切割为会话级别，并利用会话中数据包的大小、方向和时间序列信息，构建流量交互图；然后，采用图神经网络分层池化的方法，结合注意力机制，对流量图进行表征学习；最后，通过分类器对流量会话进行分类，从而实现对 SSL/TLS 加密恶意流量的有效检测。

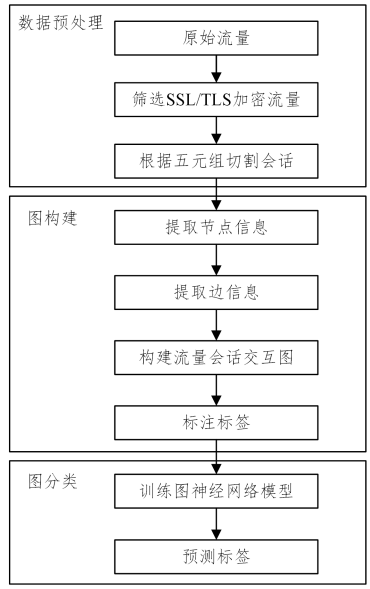


图 1 总体解决方案

Fig. 1 General solution

3.1 流量表示与预处理

基于会话粒度对 SSL/TLS 加密恶意流量进行检测，会话由源主机和目的主机之间的一系列数据包组成。相比单向流，会话由通信双方的双向流构成，包含更多的流量交互信息[15]。预处理阶段，将原始流量文件的 PCAP 格式转换为适合图神经网络处理的 CSV 文件，主要流程如下。

(1)协议筛选：利用 Wireshark 工具对流量协议进行筛选，仅保留 SSL/TLS 加密的流量。

(2)流量切分：利用 SplitCap 工具，根据五元组信息(源 IP 地址、源端口、目的 IP 地址、目的端口、网络协议)将网络流量切分为会话[16]。

(3)流量清洗：去除数据包中的冗余信息，删除空流和重复文件。

(4)填充 UDP 字段：由于 TCP 和 UDP 协议具有不同的连接方向和不同的标头大小(UDP=8，TCP=20)，因此用零填充 UDP 标头以匹配 TCP。

(5)数据标准化：数据包的最大传输单元(MTU)通常是 1 500 字节，为了涵盖数据包的全部信息，本文提取 1 500 字节。对于小于 1 500 字节的数据包，用 0 进行填充，保证每个数据包的长度统一。最后，将每个字节转换为十进制并除以 255，将每个字节标准化至[0，1]范围内。

3.2 流量交互图构建

SSL/TLS 协议主要包含 TLS 握手协议和 SSL/TLS 记录协议。其中，SSL/TLS 握手过程是用来在客户端和服务器端传输应用数据之前验证通信双方的身份，建立安全的通信机制，具体过程如图 2 所示[11]。

整个握手过程会进行一系列的数据包传输，如 Client Hello 数据包、Server Hello 数据包、Certificate 数据包、Client Key Exchange 数据包等，其中包含 TLS 版本、加密套件、密钥、证书等信息，这些信息在恶意和良性加密流量之间通常存在明显的差异，对于 SSL/TLS 恶意加密流量的检测有非常重要的意义[17]。

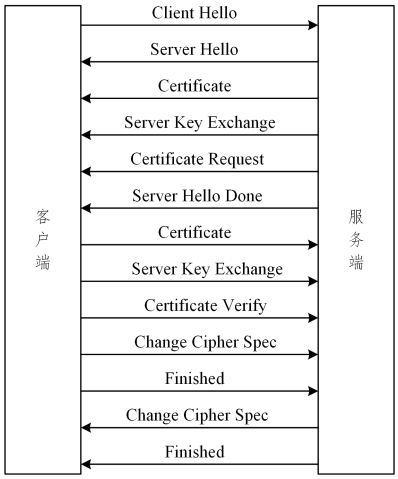


图 2　SSL/TLS 握手协议的消息交换过程

Fig. 2　Message exchange process of SSL/TLS handshake protocol

为了充分反映 TLS 握手过程以及后续加密应用程序的数据交换，提取会话的前 30 个数据包，并构建具有 30 个节点的流量拓扑图[18]。流量会话交互图如图 3 所示。

(1)节点：节点表示数据包，节点属性是数据包的负载长度。

(2)边：首先，按照数据包之间的时序关系构建边信息，由源节点指向目标节点，形成链式图结构；然后，根据数据包的 IP 信息，区分数据包的方向后，分别将发送端和接收端的数据包按时间顺序进行边连接。

(3)图属性：根据会话的标签，为流量图进行标签的映射。

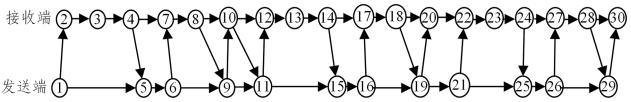


图 3　流量会话交互图

Fig. 3　Traffic session interaction graph

3.3　图分类模型的构建

基于分层图池化架构的图分类模型的整体结构如图 4 所示。该模型由学习器和分类器两部分组成，其中学习器由多个子学习器组成，每个子学习器包含图卷积、图池化和读出操作 3 个部分。

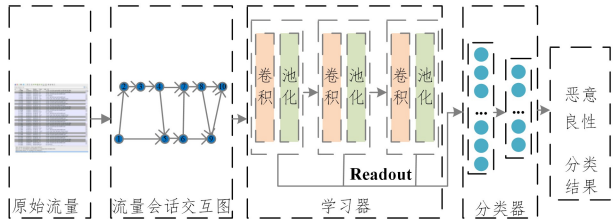


图 4　图分类模型架构

Fig. 4　Graph classification model architecture

首先，图卷积对节点的特征进行卷积操作，结合节点特征和图的拓扑结构，计算每个节点的重要性得分。随后，图池化根据节点重要性排名，保留重要性得分较高的节点，实现图的下采样。接着，读出操作针对采样后的图进行节点信息聚合，形成该层的图表示。学习器经过多层卷积池化的叠加，逐渐提取到更高阶的图表示。最后，将各层子学习器读出操作的输出进行聚合，得到整个图的综合表示，将其输入分类器中，输出图所属的类别标签，完成图分类任务。

(1)图卷积

使用图卷积层(Graph Convolutional Network Convolution，GCNConv)对流量交互图中的节点特征进行提取。GCNConv层是一种基于邻接矩阵的图卷积操作，通过聚合邻居节点的特征来更新每个节点的表示。图卷积过程如式(1)所示[19]：

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

其中，σ 为激活函数，本文使用 Relu 激活函数；$\boldsymbol{H}^{(l)}$ 是第 l 层的节点特征矩阵；$\tilde{\boldsymbol{A}}$ 是邻接矩阵；$\tilde{\boldsymbol{D}}$ 是 $\tilde{\boldsymbol{A}}$ 的度矩阵；$\boldsymbol{W}^{(l)}$ 为可训练的权重矩阵。

(2)图池化

图池化操作是为了进一步减小图的规模，提取图的全局特征。本文采取基于自注意力机制的图池化方法 SAGPool[20](Self-Attention Graph Pooling)。利用自注意力机制，通过学习每个节点的重要性得分并进行重要性排名，保留重要性得分较高的节点，从而生成具有更高重要性的子图，保留重要的节点和关键信息。图池化操作的结构示意图如图 5 所示。

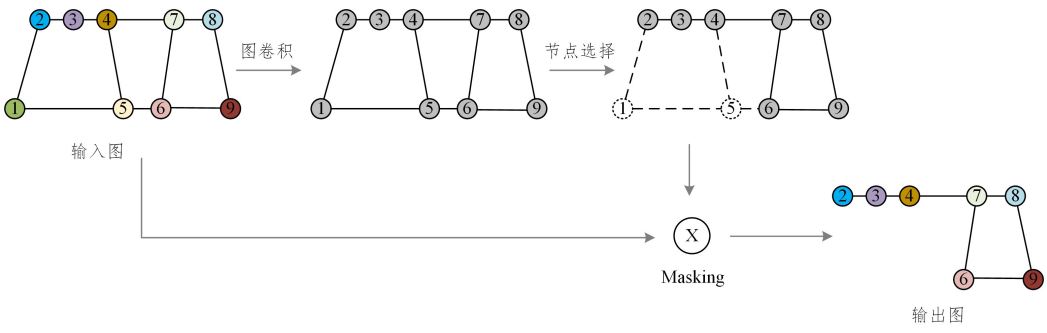


图 5　图池化示意图

Fig. 5　Graph pooling diagram

SAGPool 用图卷积方法得到 self-attention 分数，因此池化的结果是基于图的特征和拓扑结构的。计算公式如式(2)所示：

$$\boldsymbol{Z}=\boldsymbol{\sigma}(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{X}\boldsymbol{\Theta}_{\text{att}}) \tag{2}$$

其中，σ 为激活函数；$\tilde{\boldsymbol{A}}$ 是具有自连接的邻接矩阵；$\tilde{\boldsymbol{D}}$ 是 $\tilde{\boldsymbol{A}}$ 的度矩阵；$\boldsymbol{X}$ 是具有 N 个节点和 F 维特征的图的输入特征；$\boldsymbol{\Theta}_{\text{att}}$

是 SAGPool 层中唯一的参数，代表输入特征空间的卷积权重。

本文采用基于 Top-K 机制的节点采样，在计算完节点的重要性之后，根据节点重要性得分排名取 top $k(0<k<1)$ 的节点作为池化结果，得到保留的节点 idx，如式(3)所示：

$$\begin{cases} idx = top\text{-}rank(\boldsymbol{Z}, \lceil kN \rceil) \\ \boldsymbol{Z}_{\text{mask}} = \boldsymbol{Z}_{\text{idx}} \end{cases} \tag{3}$$

其中，池化比率 $k \in (0,1]$ 是一个超参数，决定了要保留的节点数量；$top \lceil kN \rceil$ 的节点是根据 $\boldsymbol{Z}$ 的值来选择的，$top\text{-}rank$ 是返回 $top \lceil kN \rceil$ 的节点的索引函数；idx 是一个索引操作；$\boldsymbol{Z}_{\text{mask}}$ 是特征注意力掩码。

输入图由图 5 中标记为 masking 的操作处理，如式(4)所示：

$$\begin{cases} \boldsymbol{X}' = \boldsymbol{X}_{idx,:}, \boldsymbol{X}\text{out} = \boldsymbol{X}' \odot \boldsymbol{Z}_{\text{mask}} \\ \boldsymbol{A}\ \text{out} = \boldsymbol{A}_{idx,idx} \end{cases} \tag{4}$$

其中，$\boldsymbol{X}_{idx,:}$ 是按行索引的特征矩阵，$\boldsymbol{X}_{\text{out}}$ 和 $\boldsymbol{A}_{\text{out}}$ 是新的特征矩阵和对应的邻接矩阵，$\boldsymbol{A}_{idx,idx}$ 是按行和按列索引的邻接矩阵。

(3)Readout 机制

Readout 读出机制是一种图数据表示方式，将池化后的图结构数据转化为特征向量[21]。每一层的池化输出都会进入 Readout 网络层。Readout 网络层操作如式(5)所示：

$$\boldsymbol{s} = \sum_{i=1}^{N} \underset{N}{mean}\, \boldsymbol{x}_i \parallel \sum_{i=1}^{N} \underset{N}{\max}\, \boldsymbol{x}_i \tag{5}$$

其中，N 代表图中节点的数量，$\boldsymbol{x}_i$ 代表第 i 个节点的特征向量，$\underset{N}{mean}\, \boldsymbol{x}_i$ 表示 $\boldsymbol{x}_i$ 的平均值，$\underset{N}{\max}\, \boldsymbol{x}_i$ 表示 x_i 的最大值。

最后，将每一层的 $\boldsymbol{s}$ 相加，生成多级图表示信息 $\boldsymbol{S}$，如式(6)所示：

$$\boldsymbol{S} = \sum_{i=1}^{n} \boldsymbol{s}_k \tag{6}$$

(4)分类器

分类器由多层感知机(Multi-Layer Perceptron，MLP)和 Sigmoid 激活函数构成。MLP 包含一个输入层、一个输出层和多个隐藏层。所有层之间是全连接的，每个连接都有一个对应的权重，每个神经元有一个偏置项。隐藏层的输出项如式(7)所示：

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{\sigma}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}) \tag{7}$$

其中，σ 表示 Sigmoid 激活函数，如(8)所示：

$$\sigma(\boldsymbol{x}) = \frac{1}{1+\mathrm{e}^{-x}} \tag{8}$$

本模型在包含 2 层隐藏层时取得了最好的效果。第一个隐藏层输入通道为 256，第二个隐藏层输入通道为 128。为了防止过拟合问题，在以上两层中引入了概率为 0.5 的 Dropout 层。

最后输出层的输出通道数量为 2，与 SSL/TLS 加密恶意流量检测二分类任务相对应。

4 实验设置

为了验证本文恶意加密流量检测的效果，利用公开数据集进行模型训练。

4.1 实验环境及设置

本文在 Python 3.9.13 的 PyTorch 框架下实现分层图池化模型，损失函数采用交叉熵损失函数，并使用 Adam 优化器对模型进行优化。模型的超参数如下：batch size 为 1 000，常规训练的 epoch 数为 200，交叉验证时的 epoch 数为 100，池化比率为 0.7，学习率为 0.000 3。为确保实验的可比性和一致性，以下实验采用相同的训练规范。

4.2 实验数据集

本文选择 CICAndMal2017 数据集作为实验数据集。该数据集由加拿大网络安全研究所(Canadian Institute for Cybersecurity)团队创建，数据集一共包含 5 类：1 类良性软件样本，4 类恶意软件样本。数据集为 pcap 格式，总大小为 36.7GB，具体样本分布如表 1 所列。

表 1 数据集分布

Table 1 Dataset distribution

流量类别	类型名称	大小/GB	SSL/TLS 加密会话数	论文选取会话数
恶意流量	Adware	7.6	60373	10000
	Ransomware	2.8	42698	10000
	Scareware	4.4	64425	10000
	SMS	1.9	38252	10000
良性流量	Benign	16	215704	40000

本实验将每个类型的数据集按 7∶2∶1 的比例随机分成 3 部分，分别作为训练集、验证集、测试集。接着，采用十折交叉验证方式，重复进行 10 次数据集随机拆分过程，并且每次都采用不同的随机种子。对于每次实验，计算出分类器的准确度，并将 10 次实验的准确度取平均值。

4.3 实验评价方法

本文针对 SSl/TLS 恶意加密流量的检测，为了方便对检测效果进行评价，采用精确率(Precision)、F1 值(F1-Score)、召回率(Recall)和准确率(Accuracy)这 4 个评价指标对模型的分类效果和性能进行判断。计算公式如下所示：

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F1 = \frac{2 * Recall * Precision}{Recall + Precision} \tag{11}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

其中，TP 表示被正确识别为恶意会话的恶意样本数；FN 表示被错误识别为良性会话的恶意样本数；FP 表示被错误识别为恶意会话的良性样本数；TN 表示被正确识别为良性会话的良性样本数。

4.4 实验结果与分析

4.4.1 图卷积池化层数确定与结果分析

为了确定论文模型学习器中卷积池化层的最佳层数，对模型子学习器的数量进行了调整，并对比分析它们在测试集上的性能，以选择最合适的层数。论文设置 1～5 层子学习器，每层采用相同的超参数和图卷积操作。实验结果如图 6 所示。

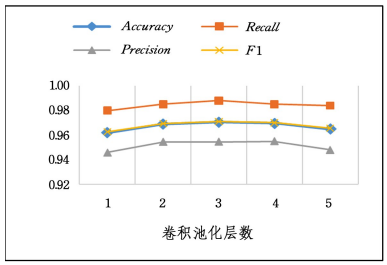


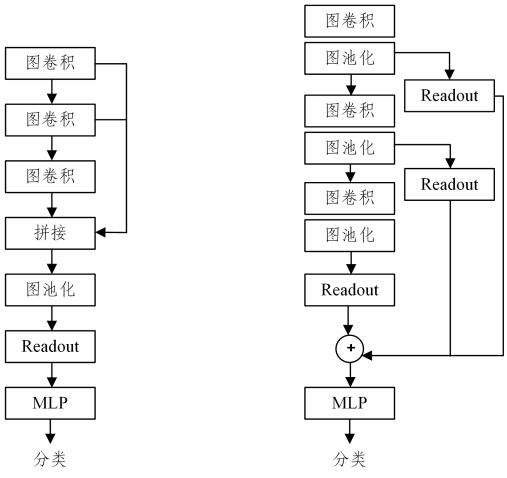
图 6　图卷积池化层数效果对比图

Fig. 6　Comparison of graph convolution pooling layer effects

由图 6 可以看出，随着卷积池化层数的增加，模型性能在一定程度上得到了提升，但当层数达到 4 层时，出现了性能饱和，效果呈现下降趋势。当子学习器数量为 3 时，模型的表现最佳，此时的精确率、F1 值、召回率和准确率 4 个评价指标均优于其他层数模型。实验结果说明，当图卷积池化层数为 3 时，所提模型能够在任务中捕捉到有效的特征，且不会过于复杂，较少的层数可能无法充分学习到图的高阶全局结构，而较多层数可能导致过拟合，使得模型性能下降。

4.4.2　与常规 GCN 的对比实验结果与分析

为评估本文模型的分层池化是否有效，将其与全局池化模型进行对比实验，卷积层统一采用 GCNConv。模型结构如图 6 所示，其中图 6(a)为全局池化架构，图 6(b)为本文采用的分层池化架构。



(a)全局池化架构　　(b)分层池化架构

图 7　全局池化和分层池化架构示意图

Fig. 7　Schematic of global pooling and hierarchical pooling architecture

图 8 给出了 GCN 和本文模型的准确率对比。

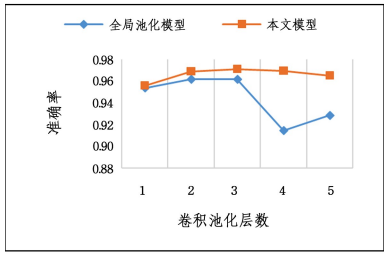


图 8　变体模型的准确率对比

Fig. 8　Accuracy comparison of variant models

可以看出，对于不同的卷积层数，论文模型的准确率始终高于常规 GCN 模型，且两个模型均在层数为 3 时达到了最佳效果。本文模型的准确率提升了 1%，验证了分层池化在图数据处理中的有效性，说明本文的分层池化模型在图结构数据的分析和处理中具有明显的优势，更加适用于恶意加密流量的检测。

4.4.3　与主流模型对比实验

为了验证论文所构建的模型在 TLS 加密恶意流量检测上的优势，将其与现在主流的检测方法进行比较，结果如表 2 所列。

表 2　模型测试结果的对比

Table 2　Comparison of model test results

(%)

方法	分类模型	精确率	准确率	召回率	F1
文献[3]	随机森林	—	95.0	95.0	95.0
文献[3]	决策树	—	93.0	92.0	92.0
文献[3]	KNN	—	85.0	86.0	84.0
文献[11]	TextCNN	90.0	—	87.9	88.9
文献[11]	BiLSTM	78.9	—	73.9	75.9
文献[11]	Text-BiLSTM	94.5	—	92.5	93.5
—	本文模型	96.1	97.1	98.2	97.1

基于 CICAndMal2017 数据集，本文模型准确率 97.1%，精确率 96.1%，召回率 98.2%，F1 值达到 97.1%，相比深度学习中表现最好的 Text-BiLSTM 结合的方法，F1 分数提升了 3.6%，精确率提升了 1.6%，召回率提升了 5.7%；相比传统机器学习表现最好的随机森林模型，准确率、F1 分数均提升了 2.1%，召回率提升了 3.2%，说明本文模型在所有模型中的分类效果最好。

结束语　本文针对 SSL/TLS 加密流量，提出了一种基于图神经网络分层池化的恶意流量检测方法。该方法利用流量交互图对 SSL/TLS 加密会话进行表征，将恶意流量检测问题转化为图分类问题。通过对图神经网络模型进行改进，采用图分层池化架构，并引入注意力机制以捕捉图数据的不同尺度和层次的特征，从而提升了模型对流量交互图中节点特征和结构特征的表征能力。实验结果表明，在针对 SSL/TLS 加密恶意流量的检测任务中，本文方法的准确率、精确率、召回率和 F1 分数等评价指标均显著优于其他方法。这表明分层池化图神经网络的结构在提高模型性能和图数据建模方面具有明显的优势。

然而，目前仅考虑了良性流量和恶意流量的二分类检测，且在平衡数据集上进行检验。在后续的工作中，将考虑对恶意流量种类的精细化识别以及在不平衡数据集上的处理，对于不平衡数据集的处理，将采用合适的采样策略或引入类别权重等方法来平衡数据分布，以获得更全面和鲁棒的模型效果。

参 考 文 献

[1]　ZHAO J J, LI Q, LIU S L. Towards traffic supervision in 6G: a graph neural network-based encrypted malicious traffic detection method[J]. Chinese Science: Information Science, 2022, 52(2): 270-286.

[2]　HTTPS encryption on the web(2023)[R/OL]. Google Transparency Report. https://transparencyreport. google. com/ht-

tps/overview?hl=en.
[3] KANG P, YANG W Z, MA H Q. TLS Malicious Encrypted Traffic Identification Research [J]. Computer Engineering and Applications, 2022, 58(12): 1-11.
[4] HU B. Research on the Detection of Malicious SSL/TLS Encrypted Traffic[D]. Shanghai: Shanghai Jiao Tong University, 2022.
[5] LASHKARI A H, KADIR A F A, TAHERI L, et al. Toward developing a systematic approach to generate benchmark android malware datasets and classification[C]// 2018 International Carnahan Conference on Security Technology (ICCST). IEEE, 2018: 1-7.
[6] CHEN R, LI Y, FANG W. Android malware identification based on traffic analysis[C]// International Conference on Artificial Intelligence and Security. Cham: Springer International Publishing, 2019: 293-303.
[7] ANDERSON B, MCGREW D. Identifying encrypted malware traffic with contextual flow data[C]// Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. 2016: 35-46.
[8] HUO Y H, ZHAO F Q. Encrypted Malicious Traffic Detection Based on Stacking and Multi-Feature Fusion[J]. Computer Engineering, 2023, 49(5): 165-172, 180.
[9] HUO Y H, ZHAO F Q, WU W H. Multi-feature fusion based encrypted malicious traffic detection method for coal mine network [J]. Journal of Mine Automation, 2022, 48(7): 142-148.
[10] CHEN J, HUANG J, LU X. Convolutional neural network-based identification of malicious traffic for TLS encryption [C]// 2022 7th International Conference on Intelligent Computing and Signal Processing(ICSP). IEEE, 2022: 1544-1549.
[11] YANG Z C, ZHU C W, CHOU J. Encrypted malicious traffic detection method based on TextCNN [J]. Journal of Guangzhou University(Natural Science Edition), 2022, 21(1): 1-9.
[12] ZHOU Y, ZHANG J, JIANG B. Detection of Malicious Encrypted Traffic Based on LSTM Recurrent Neural Network [J]. Computer Applications and Software, 2020, 37(2): 308- 312.
[13] LOPEZ-MARTIN M, CARRO B, SANCHEZ-ESGUEVILLAS A, et al. Network traffic classifier with convolutional and recurrent neural networks for Internet of Things[J]. IEEE Access, 2017, 5: 18042-18050.
[14] JIANG T T, YIN W X, CAI B . Encrypted Malicious Traffic Identification Based on Hierarchical Spatiotemporal Feature and Multi-Head Attention [J]. Computer Engineering, 2021, 47(7): 101-108.
[15] DAINOTTI A, PESCAPE A, CLAFFY K C. Issues and future directions in traffic classification [J]. IEEE Network, 2012, 26(1): 35-40.
[16] CHEN M H, ZHU Y F, LU B. Classification of Application Type of Encrypted Traffic Based on Attention CNN [J]. Computer Science, 2021, 48(4): 325-332.
[17] ZHANG X L, CHENG Q F, MA J F. Advance in TLS 1.3 Protocol Studies [J]. Journal of Wuhan University(Natural Science Edition), 2018, 64(6): 471-484.
[18] WANG Q F, ZHAI J T, CHEN W. An encrypted traffic classification method based on graph convolutional neural networks [J]. Electronic Measurement Technology, 2022, 45(14): 109-115.
[19] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv: 1609. 02907, 2016.
[20] LEE J, LEE I, KANG J. Self-attention graph pooling [C]// International Conference on Machine Learning. PMLR, 2019: 3734-3743.
[21] DENG H C. Research on Fake News Detection Based on Interaction Graph Hierarchical Pooling [D]. Wuhan: Huazhong University of Science and Technology, 2022.

TANG Ying, born in 1996, postgraduate. Her main research interests include network security and graph neural networks, etc.

WANG Baohui, born in 1973, senior engineer, master supervisor. His main research interests include network security, big data, artificial intelligence, etc.

(责任编辑:柯颖)