基于多表征融合的物联网恶意加密流量分类

王纶羽，顾益军

（中国人民公安大学信息网络安全学院，北京 100032）

摘　要：在恶意加密流量分类领域，模型通过增加流量特征维度来扩展判别表征的丰富性，但仍然存在选择模型与恶意加密流量数据特征不匹配以及特征选择不充分的问题，同时缺乏对加密流量数据特征的讨论研究。为此，针对物联网（IoT）恶意加密流量分类问题，提出一种基于多表征融合的分类模型。一方面，使用抽象表征学习模型学习流量会话的数据包级字节关联表征与会话统计表征；另一方面，使用明文表征学习模型学习未加密明文的会话连接表征。最后，根据抽象表征学习模型对分类结果的置信分数融合 2 个模型，获得最终的恶意流量分类结果。为验证模型的先进性，与 7 种基准模型进行实验比较，结果表明，所提模型的 F1 值达到 0.769 4，相较其他基准模型均有大幅提升。同时，为验证模型中各个模块与流量表征学习的适配性、选择特征所含判别表征之间的互补性，生成 10 种基于不同输入与不同模型架构的变体模型进行比较，结果表明，该模型具有更优的检测性能，证明了模型架构的适配性与表征之间的互补性。

关键词：恶意加密流量分类；深度学习；机器学习；表征融合；置信度

中图分类号：TP393　　**文献标志码：**A　　**DOI：**10.19678/j.issn.1000-3428.0252127

Malicious Encrypted Traffic Classification in IoT Based on Multi-Representation Fusion

WANG Guanyu, GU Yijun

(College of Information and Cyber Security, People's Public Security University of China, Beijing 100032, China)

【Abstract】 In malicious encrypted traffic classification, models expand the richness of discriminative representations by increasing traffic feature dimensions. However, issues such as the mismatch between the selected models and the data characteristics of malicious encrypted traffic, as well as insufficient feature selection still persist. Moreover, research and discussion on the data characteristics of encrypted traffic is lacking. To address these challenges, this study proposes a multi-representation fusion-based classification model for Internet of Things (IoT) malicious encrypted traffic classification. An abstract representation learning model is employed to learn packet-level byte-association representations and statistical representations of traffic sessions. Additionally, a plaintext representation learning model is used to learn session connection representations from unencrypted plaintext. Finally, the two models are fused based on the confidence scores of the abstract representation learning model to obtain the final malicious traffic classification results. The superiority of the proposed model is validated by comparing it with seven baseline models. The results show that the proposed model achieves an F1 value of 0.769 4, representing a significant improvement over all baseline models. Furthermore, to verify the compatibility of each module with traffic representation learning and the complementarity among discriminative representations contained in the selected features, ten variant models based on different inputs and model architectures were generated for comparison. The results demonstrate that the proposed model achieves superior detection performance, thereby proving the compatibility of the model architecture and complementarity among the representations.

【Key words】 malicious encrypted traffic classification; deep learning; machine learning; representation fusion; confidence level

0　引言

网络流量分类是实现网络安全的重要途径。网络流量分类有利于安全监管部门检测恶意流量，维护网络安全，确保网络的可用性。以深度包检测（DPI）为代表的基于有效负载的传统识别方法通过预先确定模式扫描数据包内容匹配来识别恶意流量，基于端口的识别方法通过由互联网号码分配局

基金项目：中央高校基本科研业务费专项资金（2023JKF01ZK14）。

作者简介：王纶羽，男，硕士研究生，主研方向为加密流量分类、网络安全；顾益军（通信作者），教授、博士。

收稿日期：2025-02-14　　**修回日期：**2025-04-11　　**E-mail：**guyijun@ppsuc.edu.cn

计算机工程 www.ecice06.com

(IANA)分配的传输控制协议/用户数据报协议(TCP/UDP)熟知端口号,以识别流量使用的协议类型[1]。随着加密流量的广泛应用,网络流量分类面临新的挑战[2]。流量加密技术的设计初衷是通过使用加密算法对载荷进行加密,以提高用户的隐私性与安全性。但是,也有不法分子利用加密协议传输恶意流量,攻击网络服务器,导致网络服务不可用[3]。传输载荷的加密、加密协议的复杂性以及可能存在的流量混淆[4-5],使得传统基于有效负载或端口的流量分类方法失效。此外,QUIC、HTTP/3、TLS1.3 等新兴加密协议压缩流量中有效明文的占比,这给恶意加密流量分类带来了严峻的挑战[6]。因此,在加密流量背景下,有效、准确地区分加密恶意流量类别,对实现网络的安全性与可用性具有重要意义。

当前恶意加密流量分类方法使用的主要特征包括原始载荷特征、侧信道特征与明文特征[7]。原始载荷特征是指被加密的流量原始载荷;侧信道特征是指不检查流量载荷而直接获得的特征,如统计信息等;明文特征是指流量中直接获取的有效明文信息,如证书、加密算法、密钥信息等。根据分类模型的输入流量特征的不同,现有恶意加密流量分类方法主要分为基于特征工程的恶意加密流量分类、基于表征学习的恶意加密流量分类和基于混合输入的恶意加密流量分类。基于特征工程的恶意加密流量分类方法根据应用场景进行特征工程,依赖专家经验,使用经典机器学习方法,导致在复杂场景下的实现效果有限。基于表征学习的恶意加密流量分类方法使用深度学习模型实现端到端的恶意加密流量分类,结果表明,即使是在加密的密文字节中仍然存在能够被捕捉的类别判别表征,但其忽略了深度学习架构与恶意加密流量数据结构的适配性,导致分类效果存在瓶颈。基于混合输入的恶意加密流量分类方法整合以上两种方法的优点,通过多维度表征融合进行分类决策,但未充分利用多维度特征之间的互补性来提高方法的分类效果,未讨论特征融合设计的合理性,如各模块选择的特征重合度,部分冗余无关特征的存在,在一定程度上影响了分类性能,增加了计算成本[8]。另一方面,多维度特征的选择缺乏合理的解释与明确的界限,如明文特征相较于原始加密载荷而言更加浅显和易学习,更适合轻便的经典机器学习,而原始加密载荷的字节关系、表征信息难以捕捉,适合表征学习能力更强的深度学习模型。不当的模型设计使得检测性能存在提升空间。

针对现有方法不能充分利用多维度特征来学习恶意加密流量表征、缺乏恶意加密流量表征互补性研究的问题,本文设计了一种基于多维度表征融合的物联网(IoT)恶意加密流量分类模型,该模型针对会话流量进行静态攻击类别分类。具体而言,在收集到完整的流量会话后,选择将流量的原始载荷特征、侧信道特征与明文特征作为 3 个不同维度的输入,分别在数据包粒度通过 BERT 学习流量数据包级字节关联表征,在会话粒度分别通过 BiLSTM 和 XGBoost 学习恶意加密流量的会话统计表征与会话连接表征,融合抽象表征学习模型学习的表征,使用抽象表征学习模型与明文表征学习模型作为 2 个分类器,最后定义标准化置信分数(Normalized Confidence Score),以置信分数为判别依据,融合不同分类器的分类结果,得到最终的恶意加密流量分类结果。本文的主要贡献如下:

1)提出基于多维度表征融合的 IoT 恶意加密流量分类模型,在原始载荷、侧信道、明文 3 个维度学习流量表征,对加密流量进行建模,结合表征选择学习模型,提高分类的准确性。

2)定义标准化置信分数,根据置信分数融合不同的分类结果,验证在加密流量分类领域深度学习模型的结果置信度与分类准确性之间的关联性。

3)使用公开数据集进行实验,对比多个基线模型,验证模型与表征学习的适配性与合理性、不同特征所含表征之间的互补性,同时证明模型的结构与复杂度是影响表征学习效果的重要因素。

1 相关研究

基于特征工程的恶意加密流量分类方法通过特征工程选择合适的流量特征,如数据包到达时间间隔、数据包载荷长度、协议的标志位信息等,然后应用经典机器学习模型实现对恶意加密流量的分类。文献[9]筛选流量的侧信道特征和统计特征,分别使用支持向量机(SVM)、随机森林(RF)、XGBoost(eXtreme Gradient Boosting)3 种经典机器学习方法进行恶意流量分类,最后分析比较不同特征之间的潜在关联,以及对恶意流量分类结果的贡献度。在此基础上,文献[10]分析恶意流量与正常流量在侧信道特征、明文特征维度上的区别,并增加 TLS 版本号、密钥信息等明文特征辅助分类恶意加密流量。针对前人依靠经验选择流量特征的问题,文献[11]使用连接特定特征来检测恶意加密 HTTPs 流量,采取 Fisher Score 算法、Select K Best 特征选择、RF 3 种不同的特征选择方法来筛选最具代表性的特征,并比较了 SVM、卷积神经网络(CNN)、长

短时记忆(LSTM)网络等不同分类模型的效果。文献[12]提取 TLS 握手到验证阶段的流量统计特征和证书特征,将流量特征向高维拓展,并引入层次聚类方法来压缩数据运算量。文献[13]注意到分类计算资源占用问题,提出基于网格筛选和自定义中心决策值的密度峰值聚类算法,来提取加密流量的 36 个特征,并依次依据链接、会话、证书类别特征对加密流量进行聚类,有效降低了计算资源占用。

基于特征工程的恶意加密流量分类方法根据应用场景的不同,筛选不同的流量特征,减轻了加密算法对流量分类的消极影响,取得了较好效果。但是,特征工程依赖于分类场景特点,同时也依赖于专家经验,并且经典机器学习算法更加适用于低维度特征、轻量级运算的场景[14],在高维度特征的应用场景中效果有限。

基于表征学习的恶意加密流量分类方法克服依赖专家经验的缺点,以原始载荷作为输入,使用深度学习模型实现端到端的恶意加密流量分类。文献[15]通过截断的方式截取会话字节流的前 784 Byte,以字节对应图像像素的方式,将字节流量转换成二维图像,并使用 1D-CNN 学习图像的表征。基于此,文献[16]为防止在提取空间特征时破坏数据包之间的时间关系,使用 BiLSTM 提取时间特征,然后使用 1D-CNN 提取空间特征,保证了特征的鲁棒性。考虑到加密流量特征的丰富性与全局性,文献[17]应用 BiLSTM 与 TextCNN 整合流量的局部与全局特征,引入注意力机制增强分类准确性。考虑到预训练模型的出色表现,文献[18]设计适合原始流量表示的模型和预训练任务,应用 BERT 模型[19]用于流量分类,方法泛化能力得到较大提升。文献[20]尝试基于网络服务关联与通信行为关联构建加密流量表示图,使用图注意力网络(GAT)识别恶意加密流量。

基于表征学习的恶意加密流量分类方法使用深度学习模型自动从原始载荷中学习流量表征,证明了即使是在加密的密文字节中仍然存在能够被捕捉的类别判别表征。虽然表征学习带来了较大的性能提升,但是该类方法只考虑单一的原始流量特征,同时忽略了深度学习架构与恶意加密流量数据结构的适配性,导致分类效果存在瓶颈。

基于混合输入的恶意加密流量分类方法结合以上 2 种方法的优点,接受多种维度的输入,包括原始载荷特征、侧信道特征与明文特征,通过融合特征工程与表征学习的输出得到分类结果。文献[21]将传输层字节载荷与数据包的行为序列分别输入 1D-CNN 和 BiGRU,然后拼接特征,对加密流量进行分类。在此基础上,文献[22]增加会话统计特征,使用 LSTM+CNN 模型融合多维度的特征,以检测恶意加密流量。文献[23]在字段粒度上基于词向量提取局部语义,在数据包粒度上基于时空提取全局语义,分别使用 BiLSTM 和 LSTM 模型作为语义学习模型,在 2 个粒度上提取语义,从而挖掘关键表征同时解决单一维度语义表征能力不足的问题。

基于混合输入的恶意加密流量分类方法结合特征工程针对性强以及深度学习鲁棒性好的优势,设计表征融合机制得到综合的分类结果,取得了性能上的突破。但是该类方法未充分利用多维度特征之间的互补性来提高分类效果,同时未讨论特征融合设计的合理性。

2 基于多特征融合的恶意加密流量分类模型

2.1 模型概述

本文方法针对静态流量进行分类,即收集到完整的会话后对会话进行分类。常见的分类粒度有数据包级、单向流级和会话(双向流)级。相对于数据包和单向流而言,会话具有更完整、更准确的流量类别信息,在会话层次对流量表征进行建模的过程,可以捕获更丰富的判别信息。

本文基于多特征融合的恶意加密流量分类模型架构如图 1 所示,包括数据预处理、多维度表征学习、分类信息融合 3 个阶段。在数据预处理阶段,原始流量数据包经过会话切分、流量清洗、数据分析、数据转换等步骤转换成 3 个表征学习子模块需要的数据形式。在多维度表征学习阶段,经过预处理的数据分别输入原始载荷表征学习模块、侧信道表征学习模块、明文表征学习模块,经过模块的推理预测后得到 2 个分类器的分类结果。在分类信息融合阶段,将分类器分类结果进行融合,得到最终的分类结果。

2.2 数据预处理

对于 pcap 包,以五元组(源 IP 地址,源端口,目的 IP 地址,目的端口,协议)作为会话的划分依据,将数据包划分为独立的会话,同一个会话中包含相同的五元组,五元组中的源 IP、目的 IP 与源端口、目的端口可以互换。

在训练数据集中,存在部分与类别无关的噪声会话,如 MDNS(Multicast Domain Name System)、STUN(Simple Traversal of UDP over NAT)等协议下的会话内容与攻击过程无关,属于物联网设备正常工作产生的会话,因此,这些会话将被舍弃。另

计算机工程 www.ecice06.com

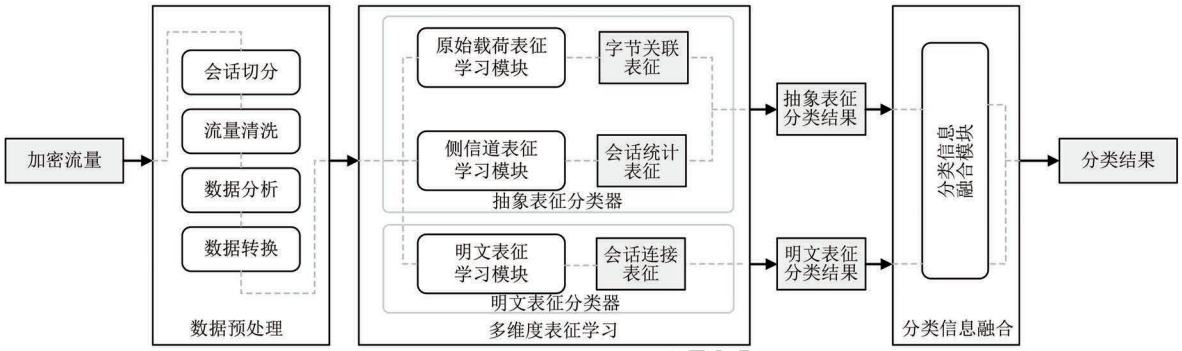

图 1 基于多表征融合的恶意加密流量分类模型架构

Fig. 1 Architecture of the malicious encrypted traffic classification model based on multi-representation fusion

外，在训练数据集的多个类别中同时存在的会话为背景正常流量，过短的会话（会话内数据包小于 4 个）因为不包含足量有效信息，所以也被当作噪声删除。

多维度表征学习的子模块输入数据差异较大，因此本节只对会话切分、流量清洗进行介绍，数据预处理中的数据分析与数据转换部分将在第 2.3～第 2.5 节中与各个子模块合并讨论。最后，若类别中因不包含传输层内容（如 ARP 攻击）而不存在会话，则只提取 IP 层的负载内容并仅依靠数据包级别的表征学习模块得到最终分类结果，不提取侧信道特征与明文特征。

在一个会话 s_{session} 中，第 k 个数据包内容用 p_k 表示。经过会话切分、流量清洗的会话表示如式(1)所示：

$$s_{\text{session}}=\{p_1,p_2,\cdots,p_k\} \tag{1}$$

2.3 原始载荷表征学习

原始载荷表征学习模块设计目的是学习加密载荷的字节关联表征，不同类型的加密恶意流量在载荷内容、加密方式的选择上有所区别，因此，可以通过捕捉字节之间的关联差异来学习流量判别信息。

模块输入的数据基本单元由会话中单个数据包的传输层首部和载荷组成。使用数据包粒度表征学习的原因在于：1）数据包级别的字节关联表征学习能捕捉更加细粒度的表征，与其他 2 个会话级表征学习子模块互为补充；2）数据集中的会话通常是较少的，而数据包是充分的，数据包粒度学习可以缓解普遍存在的数据集不平衡现象[24]。

加密流量有效载荷的长度变化需要通过截断统一输入长度来实现，这有助于提升模型的计算效率和训练稳定性。综合考虑协议数据包长度与模型性能，为保留传输层首部信息和部分有效载荷字节流，选择截断长度为 512 Byte。这一步骤减少了输入数据的维数，提高了模型训练和推理的效率。

BERT[19]是一种基于 Transformer 的双向编码表示模型，因自注意力机制和预训练步骤的设计而具有强大的上下文表征学习能力。此外，BERT 复杂的结构使之能够有效捕捉加密流量独立字节之间的关联模式[18]。因此，原始载荷表征学习模块选择使用 BERT 模型[19]进行预训练与微调。模型基本架构是 BERT-base，由 12 个 Transformer Encoder 模块[25]构成，每个 Encoder 的输入和输出向量一一对应且尺寸保持不变。Encoder 模块内部包括 2 个串联子层，第一个子层是一个多头自注意力机制，第二个子层是一个前馈神经网络，每个子层都使用残差连接，然后进行层归一化。

为了将有效载荷的字节流转换为类似自然语言的基本字符单元序列，本文采用字节对编码方法[26]，它通过将相邻的 2 个字节连接成一个基本字符单元（0000-ffff），进而将字节流转化为双元字符串。使用 Word Piece 算法[27]将双元字符串转换为令牌嵌入，并在 Word Piece 算法生成的字典中加入特殊令牌[CLS]、[SEP]、[PAD]、[MASK]，每个令牌序列第一个位置上的令牌都是[CLS]，该令牌对应的是用于分类任务最后输出的隐藏层状态，[SEP]标记子序列的结束，[PAD]是未满足最小长度时的序列填充标识，[MASK]在预训练任务中用于遮盖已有的令牌，供预训练任务使用。经过 Word Piece 算法转换后的令牌嵌入与段、位置嵌入相加得到输入序列。

2.3.1 预训练

预训练使用无标签流量数据进行训练，以学习恶意加密流量的通用表征，包括数据包加密字节之间以及与加密字节流之间的关联关系。根据预训练阶段的设计目的，模块的预训练任务参考通用预训练模型，分别设计 MLM（Masked Language Model）和改进 NSP（mNSP）2 个任务。

MLM 任务沿用原始训练方法设计与自然语言

处理任务的概率设置。具体而言，MLM 任务通过随机遮蔽部分输入令牌来学习令牌嵌入的上下文关系，被遮蔽令牌对应的输出隐藏层向量经过全连接层预测原令牌。

自然语言处理任务中的遮蔽率 15%是基于语言模型中词语的语义完整性考虑的，而加密流量的字节序列的特征分布有别于自然语言处理词语的关联关系，加密流量中的字节序列没有明显的语义结构，而是以二进制数据的形式存在，另外，流量的加密方式以块进行，块之间的加密相对独立，明文语义或存在关联，而加密关联性较小。虽然两者之间存在差异，但经过实验[18,26]验证，15%的遮蔽率在加密流量表征学习中仍能有效平衡模型的学习效率和上下文信息的完整性。遮蔽率过高会导致模型难以学习上下文关系，过低则会导致模型学习效率低下。因此，输入序列令牌以 15%的概率被遮蔽，被遮蔽令牌有 80%的概率替换为[MASK]，10%的概率替换为任意其他令牌，10%的概率保持原有令牌不变。使用负对数似然函数作为该任务的损失函数，如式(2)所示：

$$L_{\mathrm{MLM}}(\theta,\theta_{\mathrm{MLM}})=-\sum_{i=1}^{M}\ln P(t_{\mathrm{token},i}\mid\theta,\theta_{\mathrm{MLM}}) \tag{2}$$

式中：θ 是 Encoder 部分的参数；θ_{MLM} 是 MLM 任务在 Encoder 后接的全连接层参数；M 是随机遮蔽的令牌个数；$t_{\mathrm{token},i}$ 代表模型在序列 i 位置预测的令牌。

mNSP 任务通过判断 2 个子序列是否属于同一个数据包来学习载荷字节流的匹配关系。与自然语言不同，加密流量载荷是经过加密的连续字节流，没有明确的断句或者独立的意思表达，无法使用类似自然语言的标点符号划分语句，但不同明文加密后的密文载荷具有不同的密文特征分布，因此，仍可通过学习字节流的特征分布来判断两段字节序列是否属于同一个载荷。本文将流量载荷分割成 2 个长度相近的子字节序列，作为与自然语言类似的 2 个子句，在每个子序列末尾插入的[SEP]标志着子序列的结束，第二个子序列以 50%的概率被替换成其他数据包子序列，完整的输入字节序列由首部的[CLS]和 2 个子序列构成，[CLS]对应输出隐藏层向量，通过连接二分类器来计算 2 个子序列是否属于同一个数据包。使用负对数似然函数作为该任务的损失函数，如式(3)所示：

$$L_{\mathrm{mNSP}}(\theta,\theta_{\mathrm{mNSP}})=-\sum_{j=1}^{N}\ln P(y_j\mid\theta,\theta_{\mathrm{mNSP}}) \tag{3}$$

式中：θ_{mNSP} 是 mNSP 任务在 Encoder 后接的二分类器参数；N 是输入序列的数量；j 是二分类器的输出结果，0 代表未替换，1 代表已替换；y_j 代表序列实际是否替换。

将 2 个任务的损失函数之和作为模型的损失函数，计算梯度更新模型权重，如式(4)所示：

$$L(\theta,\theta_{\mathrm{MLM}},\theta_{\mathrm{mNSP}})=L_{\mathrm{MLM}}(\theta,\theta_{\mathrm{MLM}})+L_{\mathrm{mNSP}}(\theta,\theta_{\mathrm{mNSP}}) \tag{4}$$

2.3.2 微调

在微调阶段，模型重用预训练好的 Encoder 的参数，在给定任务场景的小规模有标签数据集上进行训练，通过对参数的微小调整来学习下游任务数据集的专有表征，适应下游微调任务。微调的优势在于经过预训练的模型参数，使微调模型能够快速适应下游任务，且效果优于随机初始参数直接训练的效果。

具体而言，使用预训练的 BERT 模型进行微调，并通过下游任务数据集进行训练，然后进行反向传播，更新模型参数。同时，仅使用序列第一个令牌[CLS]对应的输出向量，经过一个全连接层得到降维至 256 维的向量，将这个向量作为字节关联表征。

原始载荷表征学习模块的计算过程如式(5)所示：

$$\boldsymbol{R}_{\mathrm{Rep,B}}=\mathrm{model}_{\mathrm{BERT}}(\mathrm{load}(p_m)) \tag{5}$$

提取数据包 p_m 的传输层载荷，输入 $\mathrm{model}_{\mathrm{BERT}}$ 得到字节关联表征 $\boldsymbol{R}_{\mathrm{Rep,B}}$。原始载荷表征学习模块具体架构如图 2 所示。

2.4 侧信道表征学习

不同恶意流量在通信行为上存在差异，如洪泛攻击表现为瞬时的大规模无效数据包同时发送给被攻击者，基于 Web 的攻击则表现为与服务器的间歇交互数据包。因此，本文将通信行为差异定义为数据包到达时间与数据包长度在统计维度的不同，并使用侧信道表征学习模块学习恶意流量通信行为的差异。侧信道表征学习模块的设计目的是通过会话双方发送、接收数据包行为的统计信息，来学习恶意流量会话的侧信道表征。

以会话作为数据基本单位进行处理，学习流量的会话统计表征。为统一数据长度，选择会话中前 128 个数据包代表一个会话，并使用会话中数据包的到达时间间隔序列作为时间序列，将 IP 数据包长度序列与传输层载荷长度序列作为该模块的输入，如式(6)～式(8)所示：

$$\boldsymbol{s}_{\mathrm{seq,t}}=\{\mathrm{time}(p_1),\mathrm{time}(p_2),\cdots,\mathrm{time}(p_n)\} \tag{6}$$

$$\boldsymbol{s}_{\mathrm{seq,packet}}=\{\mathrm{len}(p_1),\mathrm{len}(p_2),\cdots,\mathrm{len}(p_n)\} \tag{7}$$

www.ecice06.com 计算机工程

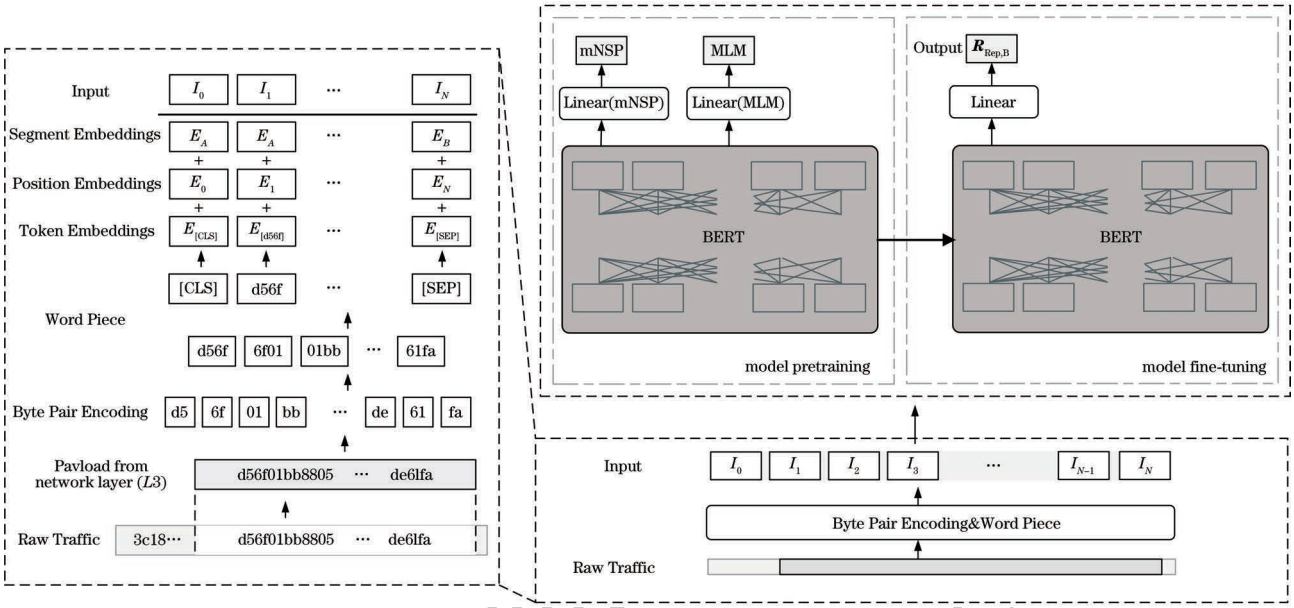

图 2 原始载荷表征学习模块架构

Fig. 2 Architecture of the original load representation learning module

$$\boldsymbol{s}_{\text{seq,load}}=\{\text{len}(\text{load}(p_1)),\cdots,\text{len}(\text{load}(p_n))\} \quad (8)$$

在会话的第 n 个包 p_n 中，$\text{time}(p_n)$ 表示与前一个包的到达时间间隔，$\text{len}(p_n)$ 表示数据包的长度，$\text{len}(\text{load}(p_n))$ 表示数据包传输层载荷的长度，其中，正数值表示数据包的方向是由发起连接 IP 发往响应 IP，负数值则反向。

本文使用 BiLSTM 学习会话统计表征。考虑到时间序列与长度序列在表达含义以及量纲上的差异，本文选择用 2 个 BiLSTM 分别对时间序列与长度序列表征进行学习，最后将 2 个 BiLSTM 的输出隐藏层向量进行拼接，然后接全连接层，融合向量得到一个 256 维的向量，作为会话统计表征，过程如图 3 所示。时间序列 BiLSTM 的输入尺寸是 1×128，长度序列 BiLSTM 的输入尺寸是 2×128。

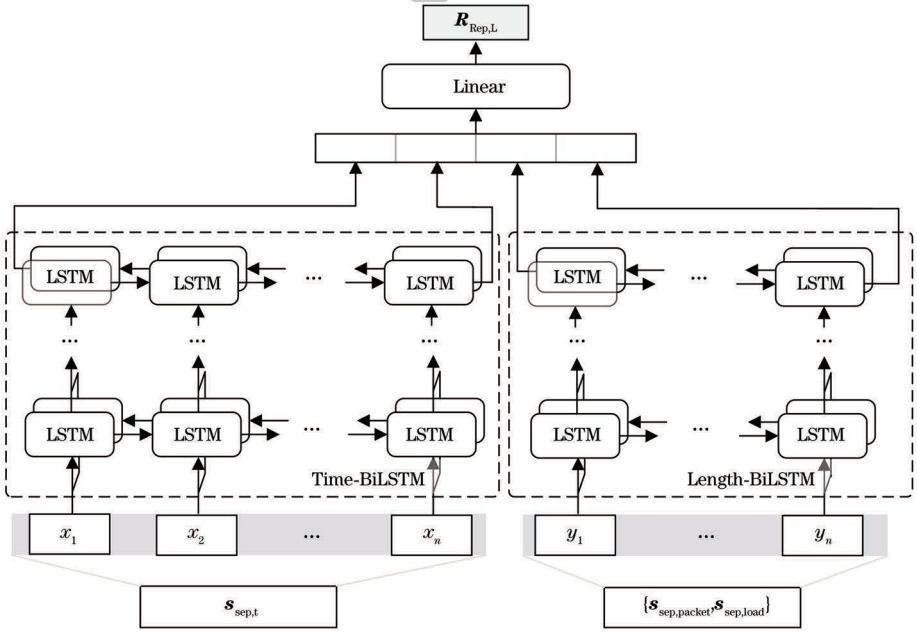

图 3 侧信道表征学习模块架构

Fig. 3 Architecture of the side-channel representation learning module

侧信道表征学习模块的计算过程如式(9)所示：

$$\boldsymbol{R}_{\text{Rep,L}}=\text{model}_{\text{BiLSTM}}(\boldsymbol{s}_{\text{seq,t}},\{\boldsymbol{s}_{\text{seq,packet}},\boldsymbol{s}_{\text{seq,load}}\}) \quad (9)$$

分别输入的时间序列与长度序列经过 $\text{model}_{\text{BiLSTM}}$ 得到会话统计表征 $\boldsymbol{R}_{\text{Rep,L}}$。

2.5 明文表征学习

在加密通信过程中，会话需使用未加密明文来协商加密协议中的部分具体信息，如会话协议、连接状态、证书版本、证书密钥算法等。区别于普通流

量，恶意流量使用不受信任的证书或利用存在漏洞的较低版本协议开展攻击行为，部分字段信息既与正常流量不同，又与其他类别攻击的字段存在差异。因此，本文设计的明文表征学习模块主要学习恶意加密流量明文有效信息包含的表征，进而获得相关恶意流量类别信息。把这些表征定义为会话连接表征，明文表征学习模块用于学习恶意流量会话的明文连接表征。

本文使用开源工具 Zeek[28] 对流量进行分析，得到日志文件（conn. log、ssl. log、x509. log）。conn. log 主要包括会话的连接信息，ssl. log 包含与 SSL/TLS 协议相关的会话信息，x509. log 包括 x. 509 证书的信息。3 个日志文件每条记录之间均存在唯一标识互相对应，未提取到则为空字段。接下来，根据会话信息整合日志文件字段，并保留关键字段用于训练与测试模型。保留字段及其含义如表 1 所示。

表 1　日志文件的保留字段

Table 1　Reserved fields in log files

字段	描述	来源	字段类型
proto	传输层使用协议	conn. log	文本
service	应用协议的标识	conn. log	文本
duration	连接持续时间	conn. log	浮点数
orig_bytes	连接发起地址发送的有效负载字节	conn. log	整数
resp_bytes	响应地址发送的有效负载字节	conn. log	整数
conn_state	会话连接状态	conn. log	文本
local_orig	连接是否源于本地	conn. log	布尔整数
local_resp	是否本地响应连接	conn. log	布尔整数
missed_bytes	会话内容间隙中丢失的字节	conn. log	整数
history	连接的状态历史记录	conn. log	文本
orig_pkts	连接发起地址发送的数据包	conn. log	整数
orig_ip_bytes	连接发起地址发送的 IP 层字节	conn. log	整数
resp_pkts	响应地址发送的数据包	conn. log	整数
resp_ip_bytes	响应地址发送的 IP 层字节	conn. log	整数
version	服务器 SSL/TLS 版本	conn. log	文本
cipher	服务器 SSL/TLS 密码套件	ssl. log	文本
curve	使用 ECDH/ECDHE 时服务器选择的椭圆曲线	ssl. log	文本
server_name	SSL/TLS 扩展的 SNI(Server Name Indicator)	ssl. log	文本
resumed	会话是否重新使用在早期连接中使用的密钥材料	ssl. log	整数
next_protocol	服务器选择的应用层下一个协议	ssl. log	文本
established	SSL 会话是否成功建立	ssl. log	整数
ssl_history	SSL 状态历史记录	ssl. log	文本
sni_matches_cert	SNI 主机名与证书是否匹配	ssl. log	布尔整数
validation_status	连接的证书验证结果	ssl. log	文本
certificate. version	证书版本	x509. log	整数，类文本
certificate. key_alg	证书使用的密钥算法	x509. log	文本
certificate. sig_alg	证书使用的签名算法	x509. log	文本
certificate. key_type	密钥类型	x509. log	文本
certificate. key_length	密钥长度	x509. log	整数
san. dns	SAN(Subject Alternative Name)中的 DNS 条目列表	x509. log	文本
basic_constraints. ca	证书的基本约束扩展	x509. log	布尔整数
host_cert	是否为终端主机证书	x509. log	布尔整数
client_cert	是否为客户端证书	x509. log	布尔整数

对于保留字段中的缺失字段，本文将其赋值为－1，对于数字类型字段，则直接保留。另外，对于文本字段或类文本字段，使用独热编码方式进行处理，并删除出现频率较低的字段信息，以减少

计算机工程 www.ecice06.com

数据特征的维度，最后，将这些特征拼接成一个会话的明文字段特征。考虑到字段信息之间关联度较弱，以及明文特征的高维非正交特点，本文使用 XGBoost 方法作为表征学习模块，模块架构如图 4 所示。

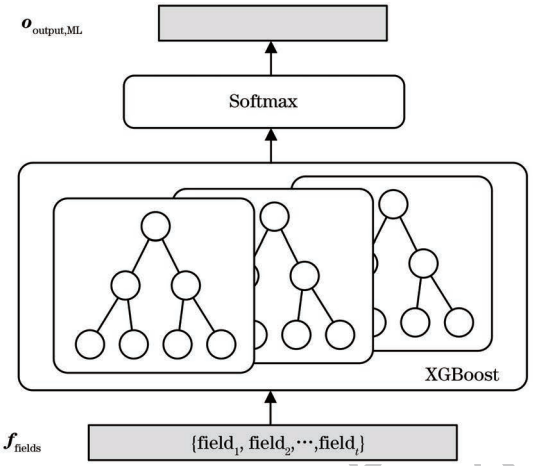

图 4 明文表征学习模块架构

Fig. 4 Architecture of the plaintext representation learning module

明文表征学习模块单独作为明文表征分类器，具体计算过程如式(10)、式(11)所示：

$$\boldsymbol{f}_{\text{fields}}=\{\text{field}_1(s_{\text{session}}),\cdots,\text{field}_t(s_{\text{session}})\} \tag{10}$$

$$\boldsymbol{o}_{\text{output,ML}}=\text{Softmax}(\text{model}_{\text{XGBoost}}(\boldsymbol{f}_{\text{fields}})) \tag{11}$$

通过 field$_t$ 提取会话的明文字段信息，组成样本的明文字段向量 $\boldsymbol{f}_{\text{fields}}$，经过 model$_{\text{XGBoost}}$ 与 Softmax 计算得到明文表征分类结果 $\boldsymbol{o}_{\text{output,ML}}$，分类结果是一维向量，元素数量与类别数量一致，每一个元素数值对应模型预测属于该类别的概率。

2.6 分类信息融合

分类信息融合模块分为 2 个部分：1)原始载荷表征学习模块与侧信道表征学习模块的表征合并；2)抽象表征分类器结果的校准。

原始载荷表征学习模块与侧信道表征学习模块共同组成抽象表征分类器，通过合并 2 个模块的输出表征得到抽象表征分类结果，并在训练时让分类器 2 个模块的模型参数同时接受梯度更新，实现抽象表征学习模型表征的有效融合。具体的合并操作是将字节关联表征与会话统计表征进行拼接，然后接一个全连接层，将维度降至与分类类别数量一致，得到关联统计结果，如式(12)所示：

$$\boldsymbol{o}_{\text{output,DL}}=\text{Softmax}(\text{Linear}(\text{Concat}\{\boldsymbol{R}_{\text{Rep,B}},\boldsymbol{R}_{\text{Rep,L}}\})) \tag{12}$$

本文结合熵和抽象表征学习模型置信度[29]定义分类结果的标准化置信分数，如式(13)所示：

$$c_{\text{conf_score}}=1-\frac{\text{entropy}(\boldsymbol{o}_{\text{output,DL}})}{\ln(n_{\text{num_classes}})} \tag{13}$$

式中：entropy 表示计算分类结果的熵函数；$n_{\text{num_classes}}$ 表示分类结果包含的类别数量。标准化置信分数数值越接近 1，表示结果的置信程度越大；越接近 0，表示结果的置信程度越小。

熵可以衡量分类结果的混乱程度，分类器的预测类别概率越集中，则熵越小；类别概率越分散，则熵越大。为方便后续计算，将分类结果的熵归一化到[0,1]的区间内，以此定义标准化置信分数，并量化模型对预测结果的置信程度。

在实验中将设定具体的置信分数阈值，当分数超过阈值时，认定抽象表征分类器的预测结果是准确的，直接使用抽象表征分类结果作为最终分类结果。若未超过阈值，则结合明文表征分类结果进行最终结果校准，具体计算如式(14)所示：

$$\boldsymbol{o}_{\text{output}}=\text{argmax}(\boldsymbol{o}_{\text{output,ML}}+k\times c_{\text{conf_score}}\times\boldsymbol{o}_{\text{output,DL}}) \tag{14}$$

式中：$\boldsymbol{o}_{\text{output}}$ 是最终分类结果；argmax 表示按照概率最大索引计算预测结果函数；k 是抽象表征学习模型的缩放系数，用于控制校准结果中抽象表征分类结果的比重。

3 实验结果与分析

3.1 数据集

CIC IoT Dataset 2023[30]是一个物联网攻击流量数据集，在其构建的由 105 个物联网设备构成的拓扑结构上收集正常物联网流量和 33 种由恶意物联网设备发起的网络攻击流量。33 种攻击分为 7 个类别，即 DDoS、Brute Force、Spoofing、DoS、Recon、Web Based、Mirai，包含 QUIC、TLS1.3 等主流加密协议。数据集具体类别情况如表 2 所示。其中，DDoS-ICMP_Flood、DDoS-ICMP_Fragmentation 类别只有网络层 ICMP 数据包，因此，仅提取以上类别的原始载荷特征学习数据包级流量表征。

Browser 2020[31]是一个浏览器访问流量数据集，包含使用谷歌、火狐、Samsung Internet 和 UC 等 4 种浏览器访问 Alexa 排名前 1 000 的网站产生的流量，其中，每个网站访问时间持续 15 s，访问时使用脚本模拟随机的点击和浏览行为。

CIC-IDS 2017[32]是一个网络攻击流量数据集，包含基于 HTTP、HTTPs、FTP、SSH 等协议的良性网络流量和网络攻击流量，这些流量以时间段为划分依据，不同的时间段产生不同类型的流量。

计算机工程 www.ecice06.com

表 2 CIC IoT Dataset 2023 数据集类别概述

Table 2 Overview of categories in the CIC IoT Dataset 2023

攻击类别	攻击
DDoS	ACK Fragmentation, HTTP Flood, ICMP_Flood, ICMP_Fragmentation, PSHACK Flood, RSTFIN Flood, Slow Loris, Synonymous IP Flood, SYN Flood, TCP Flood, UDP Flood, UDP Fragmentation
Brute Force	Dictionary Brute Force
Spoofing	ARP Spoofing, DNS Spoofing
DoS	HTTP Flood, SYN Flood, TCP Flood, UDP Flood
Recon	Host Discovery, OS Scan, Ping Sweep, Port Scan, Vulnerability Scan
Web Based	Backdoor Malware, Browser Hijacking, Command Injection, SQL Injection, Uploading Attack, XSS
Mirai	Greeth Flood, GREIP Flood, UDPPlain

Browser 2020 和 CIC-IDS 2017 用于原始载荷表征学习模块的预训练，CIC IoT Dataset 2023 用于模拟物联网攻击场景，具体包括模型的训练与测试。预训练数据集包含 6.15×10^5 条无标签流量数据，共 6.1 GB。在正式训练阶段，从数据集的每个类别选取 1 500 个会话或 5 000 个数据包作为样本，训练集、验证集和测试集的比例为 8∶1∶1。

3.2 实验环境与参数设置

实验 GPU 型号是 NVIDIA Tesla V100，Python 版本为 3.10.12，CUDA 版本为 11.7，PyTorch 版本为 1.13.0。

原始载荷表征学习模块使用由 12 个 Transformer Encoder 组成的 BERT-base 模型进行实验，Encoder 中每个多头自注意力子层包含 12 个注意力头，嵌入向量的维度设置为 768，输入向量序列的最大长度设置为 512。如图 5 所示，对所有数据集的数据包进行长度统计，发现大部分数据包长度集中分布在 512 Byte 以内，少部分在 1 500 Byte 左右，如果选择过大的输入向量序列长度，则大部分数据需要用大量[PAD]填充，如果选择过小的输入向量序列长度，则会损害模型的表征学习能力。考虑到数据集的整体分布与预训练模型的通用性设计目的，选择 512 Byte 作为输入向量序列的最大长度设置。TCP 数据包具有 20 Byte 的固定封装报头，UDP 具有 8 Byte 的报头，这种选择能够保留完整的运输层报头信息与大部分应用层加密载荷，同时减少过多[PAD]对模型分类结果的影响。模型预训练阶段的批量大小为 32，总步长为 3×10^5，学习率为 2×10^{-5}。

侧信道表征学习模块 BiLSTM 与明文表征学习模块 XGBoost 的参数设置如表 3 所示。微调阶段批量大小为 32，学习率为 8×10^{-4}，微调训练 10 个 Epoch。分类信息融合阶段的标准化置信分数阈值设置为 0.92，缩放系数设置为 0.006。

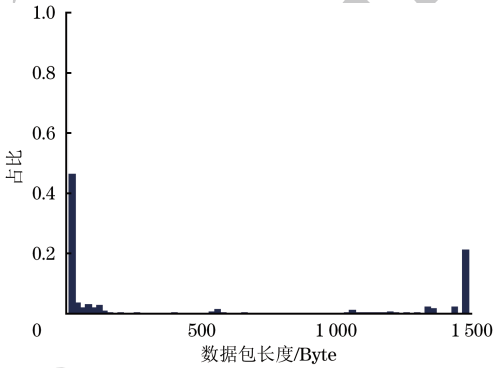

图 5 数据包长度分布统计

Fig.5 Packet length distribution statistics

表 3 部分模型相关超参数设置

Table 3 Setting of some model-related hyperparameters

模块	输入	参数	设置
侧信道表征学习模块 BiLSTM	时间序列	隐藏层	5
		隐藏层尺寸	128
	长度序列	隐藏层	4
		隐藏层尺寸	256
明文表征学习模块 XGBoost	明文特征	树最大深度	25
		树数量	300
		随机样本比例	0.8
		随机特征比例	0.8
		L1 正则化权重	0.2
		L2 正则化权重	0.5

3.3 评价指标

使用经典指标准确率(Accuracy，A)、精确率(Precision，P)、召回率(Recall，R)和 F1 值(F1，F_1)来评估与比较模型性能。在二分类问题中，上述指标的计算公式如下：

$$A=\frac{N_{TP}+N_{TN}}{N_{TP}+N_{FP}+N_{FN}+N_{TN}} \tag{15}$$

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}} \tag{16}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}} \tag{17}$$

$$F_1=2\times\frac{P\times R}{P+R} \tag{18}$$

式中：N_{TP} 表示预测正确的正例；N_{FP} 表示预测错误的正例；N_{TN} 表示预测正确的负例；N_{FN} 表示预测错误的负例。

在多分类问题中，选择使用 Macro Average[33] 方法计算精确率、召回率和 F1 值，即按照每个类别计算对应指标后取平均值，准确率指标不受多分类问题影响。

3.4 对比实验

为验证本文方法在恶意加密流量分类任务中的有效性与先进性，在物联网攻击分类场景中比较本文方法与基线模型的表现。实验选择的基线模型如表 4 所示，选择的基线模型包括常见的深度学习与机器学习模型，以及现有先进表征学习模型。

表 4 基线模型概述

Table 4 Overview of baseline models

基线模型	类别	描述
RF	多特征/机器学习	集成学习方法，通过独立构建多个决策树完成分类任务，输入特征经过特征提取、特征筛选进行处理
1D-CNN	单一特征/深度学习	一维卷积神经网络，适合处理一维序列数据，输入特征为原始流量载荷
BiLSTM	单一特征/深度学习	循环神经网络，广泛用于处理序列数据，输入特征为侧信道特征
DNN[34]	多特征/深度学习	基于侧信道特征和明文特征的全连接神经网络
BERT[19]	单一特征/深度学习	基于 Transformer 架构的深度学习方法，输入特征为原始流量载荷
Distiller[21]	多特征/深度学习	基于原始流量载荷特征和侧信道特征多维度输入的融合模型(1D-CNN、BiGRU)
MalDIST[22]	多特征/深度学习	基于原始流量载荷特征和侧信道特征多维度输入的融合模型(1D-CNN、BiGRU、2D-CNN、BiLSTM)

对比实验的结果如表 5 所示，最优结果加粗标注。实验结果表明，本文提出的方法在所有评价指标下都明显优于其他方法。对比单一特征输入与多维度特征方法，基于单一特征输入的方法表现明显低于多维度特征的方法，表明单一输入维度包含的流量类别判别表征有限，而多维度信息能够有效增加判别表征。对比同样使用原始流量载荷作为输入的方法，发现不同的模型即使使用相同的输入，准确率上仍然存在一定的差异，证明模型的架构、加密流量数据的匹配程度、模型自身的复杂度，都是影响模型性能的重要因素。比较基于多维度特征融合的深度学习方法，本文方法要优于 Distiller[21] 和 MalDIST[22]。MalDIST[22] 相比 Distiller[21] 的主要差别在于加入了关于会话不同阶段的特征信息，因此带来了性能上的小幅提升，对比之下，本文方法的显著性能提升来源于更加多元的特征输入以及更加适合加密流量表征学习的模型选择。Distiller[21] 和 MalDIST[22] 都选择使用 1D-CNN 学习原始流量载荷的表征，但 BERT[19] 的字节嵌入转换以及复杂网络结构能更加有效地学习到原始流量载荷的表征。本文方法通过分别选择基于独立特征的表征学习架构专注于不同维度的有效信息挖掘，有效提高了模型的分类性能。

表 5 不同方法的对比实验结果

Table 5 Comparative experimental results of different methods

方法	Accuracy	Precision	Recall	F1
RF	0.577 2	0.467 7	0.394 3	0.394 3
1D-CNN	0.432 7	0.443 5	0.470 5	0.427 1
BiLSTM	0.499 6	0.476 0	0.489 6	0.466 1
DNN[34]	0.705 3	0.700 0	—	0.630 0
BERT[19]	0.707 0	0.727 5	0.704 2	0.707 2
Distiller[21]	0.574 8	0.592 9	0.569 4	0.573 2
MalDIST[22]	0.588 1	0.612 0	0.580 0	0.586 6
本文方法	**0.769 3**	**0.791 2**	**0.768 1**	**0.769 4**

3.5 超参数实验

为讨论超参数对实验的影响，本文分别改变侧信道表征学习模块与明文表征学习模块的超参数，以进行实验。另外，本文还比较了标准化置信分数阈值与缩放系数设定对分类结果的影响。本文的模型设计总体策略是使用贪婪策略获得模块的最优超参数选择，最后组合所有模块得到最终的模型，这种策略影响了模型超参数的选择。

具体而言，本文分别调整侧信道表征学习模块的 2 个 BiLSTM 子模块(隐藏层数、隐藏层尺寸)、明文表征学习模块(树最大深度、树数量)、分类信息融合模块(阈值、缩放系数)的参数，计算分类效果的 F1 值。本文基于过去的研究经验[21-22]，选择了具有代表意义的超参数范围，在更广泛的参数范围内进行实验，选择最具代表性的实验结果进行展示，如图 6 所示。

从图 6 可以看出，BiLSTM 的超参数设置与模型效果之间整体上随着隐藏层尺寸的增加而增加，但存在与输入序列包含信息量相匹配的最佳模型参数组合，过于复杂的模型会出现性能下降的表现。时间序列和长度序列分别从时间与空间 2 个统计维度刻画了流量会话的表征，其中时间序列包含的类

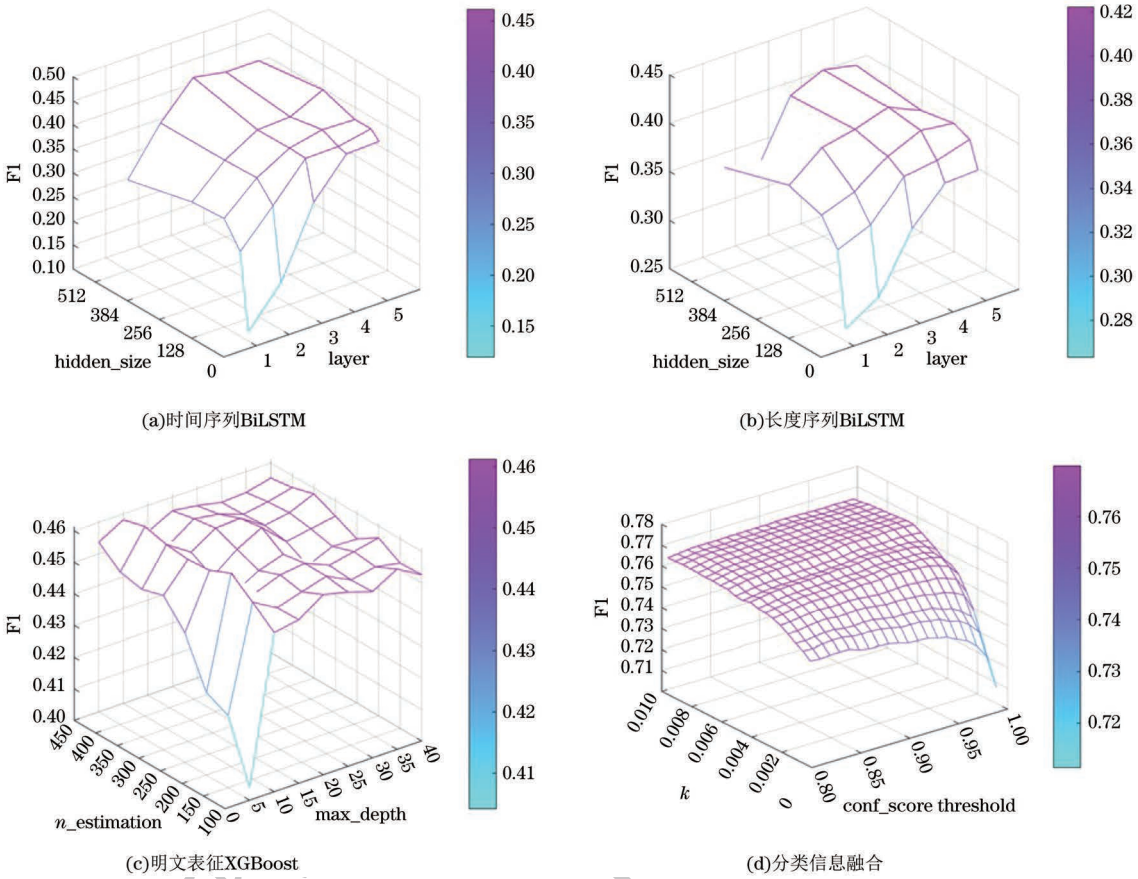

图 6　不同超参数的实验结果

Fig. 6　Experimental results with different hyperparameters

别判别表征更加丰富。XGBoost 与 BiLSTM 不同，在树深度超过 10 时，F1 值基本稳定在 0.45 以上，之后波动幅度较小，其中最佳效果出现在(25,300)。同时注意到，侧信道表征学习模块和明文表征学习模块捕捉的表征有限，均不能独立完成分类任务，因此，在整体模型中作为辅助手段来提升主分类模块的效果。

就分类信息融合模块而言，标准化置信分数阈值代表对抽象表征学习模型的抽象表征分类结果正确率的完全信任度，当置信分数高于阈值时，则将抽象表征分类结果作为最终结果，若低于阈值，则引入明文表征分类结果。缩放系数代表当分类信息融合模块对抽象表征学习模型的结果不信任时 2 个分类器结果占最终分类结果的比重，缩放系数越大，则抽象表征分类结果占比越高，反之则越低。发现置信度阈值在[0.82, 1]、缩放系数在[0, 0.005]时模型性能的变化幅度较大。其中，当缩放系数固定为 0(代表不信任抽象表征分类结果，仅使用明文表征学习结果作为分类结果)，随着置信度在一定范围([0.82, 1])内的降低，引入明文表征分类结果，性能有一定的提高，当固定为其他数值时，仍有类似的提高效果，说明在对抽象表征分类结果不信任的样本中，抽象表征学习模型缺乏足以判断样本类别的判别信息，而明文表征学习模型则有更加准确的类别判断，也意味着不同表征学习模型间存在学习表征的差异，对于不同的样本会出现不同的分类准确度，正确地融合表征可以提高模型分类效果。当固定标准化置信分数阈值，一定范围([0, 0.005])内的缩放系数的增加也会提高分类准确度，这意味着虽然抽象表征学习模型对结果信心不足并且确实分类错误，但仍能通过结合其他辅助信息来进行纠正，进而得到正确的结果。

3.6　消融实验

为验证模型各部分的有效性与先进性，本文选择不同输入组合以及不同模型组合进行效果测试，实验效果如表 6 所示。

消融实验在原始设计的基础上分别进行以下设置：

1)将载荷表征学习模块的模型(1D-CNN)进行替换。

2)将侧信道表征学习模块的输入替换为时间序列或长度序列。

3)将明文表征学习模块的机器学习模型(RF)进行替换。

4)删除上述任一模块。

表 6 消融实验结果

Table 6 Ablation experiment results

方法	Accuracy	Precision	Recall	F1
1D-CNN(原始载荷)	0.432 7	0.443 5	0.470 5	0.427 1
BERT(原始载荷)	0.707 0	0.727 5	0.704 2	0.707 2
BiLSTM(时间序列)	0.462 4	0.468 6	0.462 3	0.458 8
BiLSTM(长度序列)	0.434 2	0.469 6	0.434 2	0.428 6
BiLSTM(时间+长度)	0.499 6	0.476 0	0.489 6	0.466 1
RF(明文)	0.577 2	0.467 7	0.394 3	0.394 3
XGBoost(明文)	0.485 7	0.472 4	0.476 6	0.459 6
BERT+BiLSTM	0.739 1	0.764 0	0.738 4	0.739 1
BERT+XGBoost	0.736 6	0.753 9	0.732 7	0.735 5
BiLSTM+XGBoost	0.543 6	0.530 6	0.535 3	0.514 7
本文方法	**0.769 3**	**0.791 2**	**0.768 1**	**0.769 4**

对以上方法的分类性能进行综合比较，发现不论是更换模型、更改输入或者是删除模块，都会造成不同程度的性能指标下降，这说明对于相同输入，不同模型的表征学习能力有所不同，模型分类效果也存在差异性表现。原始加密载荷作为经过加密的字节序列，仍隐含协议特定的语法结构和字节分布模式，BERT 的自注意力机制能够全面捕获字节序列间的长程依赖关系，1D-CNN 的局部感受野难以建模这种跨越多跳的语法关联。对于侧信道特征，BiLSTM 在捕捉主机之间的心跳包周期性以及数据包突发传输的时序模式方面更具优势，其双向门控机制可以同时捕捉前向和后向的流量交互特征。恶意加密流量中的明文特征具有高维度、稀疏性和强判别性的特点，XGBoost 的梯度优化和正则化策略能够更精准地筛选关键特征，并减少弱特征的干扰，实现更高的分类精度。

对比单一模块，任意 2 个模块的融合都可以带来性能上的提升，而将 3 个模块进行融合得到的表现最为出色，这再次验证了每个表征学习模块都能在不同程度上学习恶意加密流量的表征，每个模块捕捉的流量表征存在一定的重合度，融合 3 个模块则可以获得流量类别判别表征的最大集合，获得最佳的检测性能表现。

4 结束语

本文提出一种基于多维度表征融合的恶意加密流量分类方法。该方法基于原始流量载荷特征，使用 Word Piece 算法将字节流转换成字节令牌，使用 BERT 捕捉恶意加密流量数据包的字节关联表征；基于侧信道特征，使用 BiLSTM 学习会话的统计表征。将以上 2 个深度学习模块进行融合，得到抽象表征分类结果，计算标准化置信分数，根据模型对预测结果的信心程度融合 XGBoost，通过学习会话连接表征得到明文表征分类结果。该方法考虑到流量本身的固有特征以及恶意流量之间存在的行为差异，结合模型与流量数据特征的适配性，使用深度学习与机器学习方法从多个维度抽取恶意流量的表征信息，然后设计合适的融合方式，将多种表征信息进行整合，有效增强了物联网环境下的恶意加密流量分类效果。同时，实验结果验证了多维度表征信息是单一维度表征信息的有效补充，模型结构与流量特征的适配度也是影响分类表现的重要因素。

同时，本文方法还存在以下局限：1)在模型设计方面，方法的特征处理与模型架构较复杂，特别是原始流量载荷表征学习模块训练占用较多计算资源，另外对于此模块输入的截断操作，可能丢失部分长数据包的信息；2)在设计原理方面，仅在宏观层面广泛讨论了模块的有效性，缺少微观层面不同类型恶意流量针对性适配差异的对比讨论；3)在模型效果方面，方法在其他非物联网攻击场景下的先进性仍有待验证。

在未来工作中，将进一步研究方法计算资源占用与实时性要求的平衡关系，如使用参数高效微调等更加轻便的训练方法、层次化建模等更灵活的长度处理策略，通过对比各种攻击与主机异常行为的关系，深层次探索恶意加密流量的表征分布与攻击类别的关联，针对性地设计更加轻便有效的模型架构，实现稳定、高效、平衡的恶意加密流量分类模型。

参考文献

[1] WANG Z H, 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.

[2] 李志远, 吴安昊, 谭林, 等. 基于流序列特征融合与注意力机制的加密流量分类方法[J]. 小型微型计算机系统, 2025, 46(7): 1718-1726.
LI Z Y, WU A H, TAN L, et al. Stream sequence feature fusion and attention mechanism based method for encrypted traffic classification [J]. Journal of Chinese Computer Systems, 2025, 46(7): 1718-1726. (in Chinese)

[3] LIN K D, XU X L, GAO H 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.

[4] PAPADOGIANNAKI E, IOANNIDIS S. A survey on encrypted network traffic analysis applications, techniques, and countermeasures[J]. ACM Computing Surveys, 2022, 54(6): 1-35.

[5] 陈子涵, 程光, 徐子恒, 等. 互联网加密流量检测、分类与识别研究综述[J]. 计算机学报, 2023, 46(5): 1060-1085.
CHEN Z H, CHENG G, XU Z H, et al. A survey on Internet encrypted traffic detection, classification and identification[J]. Chinese Journal of Computers, 2023, 46(5): 1060-1085. (in Chinese)

计算机工程 www.ecice06.com

[6] 康鹏，杨文忠，马红桥. TLS 协议恶意加密流量识别研究综述[J]. 计算机工程与应用，2022，58(12)：1-11.
KANG P，YANG W Z，MA H Q. TLS malicious encrypted traffic identification research[J]. Computer Engineering and Applications，2022，58(12)：1-11.（in Chinese）

[7] 侯剑，鲁辉，刘方爱，等. 加密恶意流量检测及对抗综述[J]. 软件学报，2024，35(1)：333-355.
HOU J，LU H，LIU F A，et al. Detection and countermeasure of encrypted malicious traffic：a survey[J]. Journal of Software，2024，35(1)：333-355.（in Chinese）

[8] 付钰，刘涛涛，王坤，等. 基于机器学习的加密流量分类研究综述[J]. 通信学报，2025，46(1)：167-191.
FU Y，LIU T T，WANG K，et al. Survey of research on encrypted traffic classification based on machine learning[J]. Journal on Communications，2025，46(1)：167-191.（in Chinese）

[9] SHEKHAWAT A S，DI TROIA F，STAMP M. Feature analysis of encrypted malicious traffic[J]. Expert Systems with Applications，2019，125：130-141.

[10] ANDERSON B，MCGREW D. Identifying encrypted malware traffic with contextual flow data[C]//Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. New York，USA：ACM Press，2016：35-46.

[11] WENG Z Q，CHEN T M，ZHU T T，et al. TLSmell：direct identification on malicious HTTPs encryption traffic with simple connection-specific indicators[J]. Computer Systems Science and Engineering，2021，37(1)：105-119.

[12] YU T D，ZOU F T，LI L S，et al. An encrypted malicious traffic detection system based on neural network[C]// Proceedings of the International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery（CyberC）. Washington D.C.，USA：IEEE Press，2020：62-70.

[13] CHEN L C，GAO S，LIU B X，et al. THS-IDPC：a three-stage hierarchical sampling method based on improved density peaks clustering algorithm for encrypted malicious traffic detection[J]. The Journal of Supercomputing，2020，76(9)：7489-7518.

[14] LIU J Y，ZENG Y Z，SHI J Y，et al. MalDetect：a structure of encrypted malware traffic detection[J]. Computers，Materials & Continua，2019，60(2)：721-739.

[15] WANG W，ZHU M，ZENG X W，et al. Malware traffic classification using convolutional neural network for representation learning[C]//Proceedings of the International Conference on Information Networking（ICOIN）. Washington D.C.，USA：IEEE Press，2017：712-717.

[16] ZHU S Z，XU X L，GAO H H，et al. CMTSNN：a deep learning model for multiclassification of abnormal and encrypted traffic of Internet of Things[J]. IEEE Internet of Things Journal，2023，10(13)：11773-11791.

[17] 蒋彤彤，尹魏昕，蔡冰，等. 基于层次时空特征与多头注意力的恶意加密流量识别[J]. 计算机工程，2021，47(7)：101-108.
JIANG T T，YIN W X，CAI B，et al. Encrypted malicious traffic identification based on hierarchical spatiotemporal feature and multi-head attention[J]. Computer Engineering，2021，47(7)：101-108.（in Chinese）

[18] LIN X J，XIONG G，GOU G P，et al. ET-BERT：a contextualized datagram representation with pre-training transformers for encrypted traffic classification[C]// Proceedings of the ACM Web Conference 2022. New York，USA：ACM Press，2022：633-642.

[19] DEVLIN J，CHANG M W，LEE K，et al. BERT：pre-training of deep bidirectional transformers for language understanding[EB/OL].[2024-12-05]. https://arxiv.org/abs/1810.04805.

[20] 赵键锦，李祺，刘胜利，等. 面向 6G 流量监控：基于图神经网络的加密恶意流量检测方法[J]. 中国科学(信息科学)，2022，52(2)：270-286.
ZHAO J J，LI Q，LIU S L，et al. Towards traffic supervision in 6G：a graph neural network-based encrypted malicious traffic detection method[J]. Science in China（Information Sciences），2022，52(2)：270-286.（in Chinese）

[21] ACETO G，CIUONZO D，MONTIERI A，et al. DISTILLER：encrypted traffic classification via multimodal multitask deep learning[J]. Journal of Network and Computer Applications，2021，183：102985.

[22] BADER O，LICHY A，HAJAJ C，et al. MalDIST：from encrypted traffic classification to malware traffic detection and classification[C]// Proceedings of the IEEE 19th Annual Consumer Communications & Networking Conference（CCNC）. Washington D.C.，USA：IEEE Press，2022：527-533.

[23] 谷勇浩，徐昊，张晓青. 基于多粒度表征学习的加密恶意流量检测[J]. 计算机学报，2023，46(9)：1888-1899.
GU Y H，XU H，ZHANG X Q. Multi-granularity representation learning for encrypted malicious traffic detection[J]. Chinese Journal of Computers，2023，46(9)：1888-1899.（in Chinese）

[24] 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.

[25] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[EB/OL].[2024-12-05]. https://arxiv.org/abs/1706.03762.

[26] HE H Y，YANG Z G，CHEN X N. PERT：payload encoding representation from transformer for encrypted traffic classification[C]// Proceedings of the ITU Kaleidoscope：Industry-Driven Digital Transformation（ITU K）. Washington D.C.，USA：IEEE Press，2020：1-8.

[27] WU Y H，SCHUSTER M，CHEN Z F，et al. Google's neural machine translation system：bridging the gap between human and machine translation[EB/OL].[2024-12-05]. https://arxiv.org/abs/1609.08144.

[28] SHARMA A，KREIBICH C，WALA F B，et al. Zeek[EB/OL].[2024-12-05]. https://download.zeek.org/zeek-6.0.4.tar.gz.

[29] KADAVATH S，CONERLY T，ASKELL A，et al. Language models（mostly）know what they know[EB/OL].[2024-12-05]. https://arxiv.org/abs/2207.05221.

[30] NETO E C P，DADKHAH S，FERREIRA R，et al. CICIoT2023：a real-time dataset and benchmark for large-scale attacks in IoT environment[J]. Sensors，2023，23(13)：5941.

[31] VAN EDE T，BORTOLAMEOTTI R，CONTINELLA A，et al. FlowPrint：semi-supervised mobile-App fingerprinting on encrypted network traffic[EB/OL].[2024-12-05]. https://www.ndss-symposium.org/ndss-paper/flowprint-semi-supervised-mobile-app-fingerprinting-on-encrypted-network-traffic/.

[32] SHARAFALDIN I，HABIBI L A，GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[EB/OL].[2024-12-05]. http://cs.unb.ca/research-expo/expos/2018/submissions/20180403-14-56-isharafa-at-unb.ca-toward_generating_a_new_intrusion_detection_dataset_and_intrusion_traffic_characterization.pdf.

[33] LIU C，WANG W Y，WANG M，et al. An efficient instance selection algorithm to reconstruct training set for support vector machine[J]. Knowledge-Based Systems，2017，116：58-73.

[34] NKORO E C，NWAKANMA C I，LEE J M，et al. Detecting cyberthreats in Metaverse learning platforms using an explainable DNN[J]. Internet of Things，2024，25：101046.

文字编辑 吴云芳
栏目编辑 赖玉玲

计算机工程 www.ecice06.com