基于日志语义分析与领域自适应预训练的网络入侵检测方法

杨宏宇[1,2]，崔恒宇[2]，赵子文[2]，刘坤[2]
（1.中国民航大学 计算机与人工智能学院，天津 300300；2.中国民航大学 安全科学与工程学院，天津 300300）

摘要：网络入侵检测旨在从高维、异构且动态变化的网络流量中识别正常行为与多类攻击，是保障网络系统安全的重要手段。现有方法主要存在流量字段关联建模不足、预训练模型与日志分布存在子领域分布差异及跨数据集泛化差等问题，针对上述不足提出一种基于日志语义分析与领域自适应预训练的网络入侵检测方法。首先依据网络连接行为逻辑，将结构化流量字段重构为由六类语义模块组成的模块化日志语义序列，并结合全局特征重要性确定模块内字段表达优先级；随后，设计面向日志语义结构的领域自适应预训练策略，通过行为字段优先掩码、完整字段片段掩码和语义组平衡加权掩码语言建模损失，增强模型对协议、端口、速率、时序和TCP状态等字段联合依赖的学习能力；最后，利用领域适配后的SecureBERT提取日志全局语义表示，完成正常流量与多类攻击流量的分类判别。在CIC-IDS2017和CSE-CIC-IDS2018数据集上，所提方法的F1分数分别提升0.35%～14.28%和0.08%～3.5%；跨数据集迁移实验中，准确率和F1分数分别提升4.94%～12.20%和5.26%～11.79%，检测性能与泛化能力较好。

关键词：网络入侵检测；日志语义重构；领域自适应预训练；SecureBERT；迁移学习；掩码语言建模
中图分类号：TP393.08
文献标志码：A
doi：10.11959/j.issn.1000

A Network Intrusion Detection Method Based on Log Semantic Analysis and Domain-Adaptive Pre-Training

Yang Hongyu[1,2], Cui Hengyu[2], Zhao Ziwen[2], Liu Kun[2]
1. School of Computer Science and Artificial Intelligence, Civil Aviation University of China, Tianjin 300300, China
2. School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China

Abstract: Network intrusion detection was aimed at identifying benign behavior and multiple attack categories from high-dimensional, heterogeneous, and dynamically evolving network traffic, and was an important means of safeguarding network systems. Existing methods often suffered from insufficient modeling of dependencies among traffic fields, subdomain distribution mismatches between pretraining corpora and network logs, and limited cross-dataset generalization. To address these issues, a network intrusion detection method based on log semantic analysis and Domain-Adaptive Pre-Training was proposed in this paper. First, structured traffic fields were reconstructed into modular log semantic sequences composed of six semantic modules according to the logic of network connection behavior, while global feature importance was used to determine the expression priority of fields within each module. Next, a log-structure-aware Domain-Adaptive Pre-Training strategy was designed by incorporating behavior-field-prioritized masking, complete field-span masking, and a semantic-group-balanced weighted masked language modeling loss, thereby enhancing the model's ability to learn joint dependencies among protocols, ports, transmission rates, temporal patterns, and TCP states.

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：杨宏宇，yang_hy2006@126.com
基金项目：国家自然科学基金民航联合研究基金重点项目(No U2433205)；国家重点研发计划项目(No 2025YFB3110500)
Foundation Items: Civil Aviation Joint Research Fund Project of the National Natural Science Foundation of China (No. U2433205); National Key Research and Development Program of China (No 2025YFB3110500)

Finally, the domain-adapted SecureBERT was used to extract global semantic representations from logs for the classification of benign traffic and multiple attack categories. Experiments on the CIC-IDS2017 and CSE-CIC-IDS2018 datasets showed that the F1-score of the proposed method was improved by 0.35%–14.28% and 0.08%–3.50%, respectively. In cross-dataset transfer experiments, the Accuracy and F1-score were improved by 4.94%–12.20% and 5.26%–11.79%, respectively. The proposed method demonstrates favorable detection performance and cross-dataset generalization capability.

Key words: Network intrusion detection, Log semantic reconstruction, Domain-adaptive pre-training, SecureBERT, Transfer learning, Masked language modeling

0 引言

随着互联网技术的迅猛发展，网络通信的体量持续增长。当前网络流量表现出维度高、动态性强且异构性突出的特点。复杂多变的流量环境大幅增加网络行为建模的难度，也为攻击者隐藏其恶意活动留下更多可乘之机。网络入侵检测[1]是识别异常流量和恶意行为的一项关键技术，它通过对网络流量或日志进行建模分析，能够实时检测并预警攻击，在保障网络安全方面具有重要价值。

针对网络入侵检测任务，已有方法大致可以归为基于机器学习的方法、基于深度学习的方法以及基于预训练语言模型的方法三大类[2]。基于机器学习的方法大多依赖人工设计的流量统计特征与基础字段特征，再结合支持向量机、随机森林或朴素贝叶斯等分类器实现网络入侵检测[3]。这类方法在一定程度上能够刻画网络行为模式。为减轻人工设计特征带来的制约，深度学习方法逐步被引入该任务。在一定程度上提升对复杂流量数据的建模能力[4]。近年来，随着预训练语言模型研究的进步，部分研究工作尝试把网络日志或流量数据转化为文本序列，借助预训练语言模型的上下文建模能力执行网络入侵检测[5]，然而，在复杂的网络环境与跨数据分布场景下，现有方法仍面临以下问题。

（1）网络流量字段之间的组合依赖关系导致行为建模更加复杂。以传输速率为例，较高的流字节速率既可能出现在正常的大文件传输中，也可能对应拒绝服务攻击。正常传输通常具有较完整的双向交互、较长的持续时间和以ACK为主的稳定会话状态。拒绝服务攻击则常伴随较短的流间隔、异常的前后向报文比例以及SYN、RST等状态字段组合。

（2）网络流量日志与通用预训练语料之间的领域分布差异导致日志语义建模难度增加。SecureBERT虽然已经在网络安全领域语料上进行了预训练，能够学习漏洞描述、威胁情报等安全文本中的语义知识，但其原始训练语料主要表现为自然语言或半结构化安全文本，句子结构相对符合自然语言表达规律。与之不同，本文使用的网络流量日志语料由协议类型、端口号、持续时间和TCP标志位等字段构成。此类语料中的“语义”来自多个流量字段在同一连接上下文中的组合关系。字段结构和字段共现模式的特殊性使预训练语言模型难以充分适配网络流量日志数据，从而影响模型的语义表征能力。

（3）日志语义表示与入侵检测判别目标之间的适配不足导致攻击类别区分变得更加困难，语义特征对类别差异信息的表达不充分，从而影响多类型攻击流量检测的准确率。在不同网络环境和数据采集条件下，上述适配不足还会进一步表现为模型跨数据集泛化能力下降，难以稳定应对目标域流量分布与攻击行为特征的变化。

针对上述挑战，本文提出一种基于日志语义分析与领域自适应预训练的网络入侵检测方法（Log Semantic Analysis and Domain-Adaptive Pre-Training，LSADA）。该方法首先构建一种类自然语言的日志语义重构机制，将原本结构化的网络流量字段转化为语义连贯的日志文本序列，在统一的上下文语境中，端口、协议、包数等字段的协同变化关系得以整体表达，从而强化对流量字段关联关系的表征。然后利用网络流量日志语料对SecureBERT实施领域自适应预训练（Domain-Adaptive Pre-Training，DAPT）[6-7]，引导模型逐步掌握日志的字段结构、共现规律和攻击行为语义模式，以此缓解通用预训练语料与网络流量日志之间的领域偏差。完成领域适配后，面向入侵检测任务对SecureBERT进行下游微调，进一步将日志语义表示优化为适用于正常流量与多类型攻击流量判别的任务特征，进而提升模型识别复杂攻击行为的能力与

检测稳定性。

考虑到真实网络环境中不同场景的流量分布差异显著，且目标域数据往往难以与源域保持同分布，本文设计迁移实验检验模型迁移能力和泛化表现。通过上述设计，模型在维持较高检测精度的同时，能够保持较好的迁移能力。

本文的主要工作及贡献如下：

（1）提出一种类自然语言的网络流量日志语义重构方法。该方法遵循网络连接行为的描述逻辑，对原始的结构化流量字段实施对齐与模板化表达，将端口、协议、包数、速率、时序间隔和标志位等关键字段转化为语义连贯的日志文本序列。通过将离散字段映射为序列，模型能够捕获字段间的上下文信息，从而增强流量数据的语义表达能力，克服传统数值特征难以建模字段间深层关联的局限。

（2）针对通用预训练语料与网络流量日志语料分布差异，将类自然语言重构的日志序列作为领域语料，对 SecureBERT 开展 DAPT，推动模型从通用网络安全文本领域向网络流量日志领域进行知识迁移与语义适配。该过程显著增强模型对入侵检测数据的领域适配能力，为后续区分正常流量与多类型攻击流量提供更具辨别力的语义表示。

（3）提出一种面向入侵检测任务的日志语义判别优化方法。在领域自适应预训练的基础上，提取日志文本序列的全局语义表示，将其输入任务分类层进行监督微调，引导模型捕捉正常流量与各类攻击之间的判别边界。通过将领域语义表示与入侵检测分类目标深度融合，预训练阶段积累的日志结构知识和字段依赖关系被进一步转化为攻击类别判别特征，有效缓解语义表示与检测任务目标之间适配不足的问题，进而提高模型对多类型攻击流量的检测精度与稳定性。

1　相关工作

1.1　基于机器学习的网络入侵检测方法

在基于机器学习的方法中，入侵检测通常被建模为特征工程与分类模型相结合的问题。早期研究主要通过构建高质量特征提升模型性能。Gu 等[8]引入朴素贝叶斯特征转换与嵌入机制，对原始流量特征进行重构，以此提高数据表示质量。后续研究进一步关注特征选择与优化问题，Hosseini 等[9]融合遗传算法与粒子群优化方法，对特征空间进行优化搜索，有效改善检测精度与训练效率。然而，这类方法的核心依然离不开人工特征工程，对数据质量与分布变化较为敏感，面对高维复杂数据及异构网络环境时泛化能力有限，同时难以刻画攻击行为中深层的结构信息。

1.2　基于深度学习的网络入侵检测方法

为了应对日益复杂的网络攻击，基于深度学习的网络入侵检测方法逐渐成为研究热点。Giuseppina 等[10]从数据分布建模的角度切入，将生成对抗网络与卷积神经网络结合，借助生成少数类样本缓解训练偏置，为后续研究在数据层面奠定改进基础。不过，该方法对生成模型分布拟合能力的依赖较强，复杂流量环境中可能引入噪声样本，从而影响检测稳定性。此后，研究的焦点转向特征表达能力的增强。Lin 等[11]构建多尺度卷积结构，对不同粒度特征进行联合建模，从特征层面提升对复杂流量模式的刻画能力。但其建模仍主要依靠局部卷积操作，捕捉长程依赖关系的能力有限。为弥补这一不足，Abdel-Basset 等[12]提出 HAE-HRL 方法，利用自编码器与 LSTM-CNN 残差结构联合建模流量特征，在统一框架内协同捕捉局部与时序特征。在特征优化方面，Jeyavim Sherin 等[13]提出 SSA-BOA-CNN 入侵检测方法。该方法首先利用樽海鞘群算法（Salp Swarm Algorithm，SSA）筛选流量特征，减少冗余与无关信息；随后采用蝴蝶优化算法（Butterfly Optimization Algorithm，BOA）优化深度 CNN 的关键超参数，并由 CNN 完成攻击特征提取与分类。该方法通过特征选择与参数优化改善检测性能，但其优化过程依赖多阶段启发式搜索，训练效率和稳定性容易受到种群规模、迭代次数及搜索参数设置的影响。针对类别不平衡问题，Cao 等[14]提出 A-NIDS，通过聚类与堆叠式 CTGAN 机制增强少数类样本的表达能力，使模型在复杂流量场景下具备更强的适应性。Bamber 等[15]设计了一种混合 CNN-LSTM 框架，结合特征筛选与时序建模来提升复杂攻击流量的检测能力。该方法依然以数值特征为主要输入，未能刻画流量语义信息及字段间的上下文关系。

随着网络环境日趋复杂，仅依靠固定数据分布训练的深度模型很难满足跨场景部署的需求，迁移学习因此成为该领域的重要延伸。Bierbrauer 等[16]从原始网络流量建模出发，提出基于字节序列的迁

移学习方法，在降低人工特征依赖的同时提升模型的迁移能力。但原始字节序列缺乏明确的语义结构，限制模型对复杂行为模式的理解。为增强跨场景表达能力，Xu等[17]提出多模态融合方法，通过整合不同模态特征来提升判别能力，然而其融合策略依赖人工设计，模型复杂度较高。进一步，Latif等[18]构建基于注意力机制的Siamese网络，利用度量学习强化样本间的相似性建模，从而提升模型在分布变化条件下的适应能力。但该方法依赖样本对构建策略，对数据组织形式较为敏感。

近年来，面向流级网络数据的迁移学习研究进一步向少数类知识迁移、跨域表征对齐和部署导向评估方向发展。Wang等[19]提出QuCAD-IDS，利用层次对比自编码学习源域可迁移表示，并结合队列式对比蒸馏、最大均值差异对齐和目标域微调提升异构网络环境下的跨域检测能力。Ha Thanh[20]提出面向NetFlow入侵检测的TAN-IDS框架，通过统一的流级特征接口组织同域检测、跨数据集迁移、混合域训练和少量目标域微调场景，进一步表明同域高性能并不能直接代表跨数据集鲁棒性。

上述研究从特征迁移、域间表示对齐和统一评价等角度改善了流级入侵检测的跨环境适应能力，但多数方法仍直接对数值流量特征或低维NetFlow特征进行建模，尚未充分利用协议、端口、速率、流间隔和TCP状态等字段在统一日志语义上下文中的组合关系。

1.3 基于预训练语言模型的网络入侵检测方法

近年来，预训练语言模型的发展催生基于语义建模的入侵检测方法，研究者开始将网络日志视为类自然语言文本进行建模，研究重心从“特征表示”逐步转向“语义理解”。Sun等[21]提出的LogSwift方法通过掩码语言建模学习日志序列中的语义模式，用于异常检测，在一定程度上提升对日志语义的刻画能力。然而，其建模过程主要依赖轻量化结构，对复杂上下文关系与长程依赖的表达能力依然有限。在此基础上，Li等[22]进一步引入大语言模型，设计LLM-IDS与RLFE-IDS框架，通过融合检索增强机制来提升对未知攻击的识别能力，但该方法依赖外部知识检索模块，系统复杂度较高，且对检索质量敏感。围绕语义理解能力的进一步提升，Qi等[23]提出LogGPT方法，利用大语言模型的上下文推理能力加强日志序列的语义建模，在异常识别上取得较好效果。但其性能较大程度依赖于预训练模型的规模和通用语料，缺乏针对网络安全领域的专门适配。针对数据受限场景，研究尝试将语义建模与决策优化机制结合，Ren等[25]提出一种基于预训练语言模型与深度强化学习的入侵检测方法，引入策略优化以提升复杂攻击识别能力。

SecureBERT虽然已经在网络安全领域文本上进行预训练，但其原始语料主要由漏洞描述、威胁情报及其他自然语言或半结构化安全文本组成，与本文处理的结构化流量日志仍存在明显的子领域差异。为定量分析这种差异，本文分别从SecureBERT安全文本语料和CSE-CIC-IDS2018训练划分构建的流量日志语料中等量抽取100,000条文本，统一使用SecureBERT分词器进行词元化，并将最大序列长度设置为256。

SecureBERT安全文本的平均词元数为182.35，结构化流量日志为137.07；两类语料的数字词元占比分别为6.18%和42.17%，非字母符号词元占比分别为4.32%和12.75%，高频前100个词元的覆盖率分别为34.02%和72.99%，词型—词次比分别为0.220%和0.125%。此外，两类语料一元词元分布的Jensen–Shannon散度为0.4437，表明二者的词元概率分布存在较为明显的差异。

上述结果表明，结构化流量日志包含更密集的端口、持续时间、速率、包数和TCP标志等字段取值，其数字词元占比约为SecureBERT安全文本的6.82倍。受固定日志模板和重复字段描述的影响，结构化流量日志的高频前100个词元覆盖率达到72.99%，较安全文本高38.97个百分点，而其词型—词次比相对较低，表现出更稳定的模板结构和字段共现模式。SecureBERT安全文本则具有更丰富的自然语言词汇和句法结构。两类语料在词元组成、数值密度、词汇多样性及高频词分布方面存在明显差异。因此，SecureBERT已有的安全文本知识不能直接等同于对流级网络日志字段关系的建模能力，需要进一步利用无标签流量日志开展DAPT，使模型适应流量日志的字段结构和行为语义分布。

2 研究方法

2.1 网络入侵检测问题形式化定义

为明确本文研究对象与优化目标，首先对网络

入侵检测问题进行形式化定义。设网络流量数据集为$\mathcal{D}=\left(\mathbf{x}_i,y_i\right)_{i=1}^{N}$。其中，$\mathbf{x}_i$表示第$i$条网络流量记录，$y_i\in\mathcal{Y}$表示其对应类别标签，$d_f$表示原始流量字段数量，$N$为样本数量，$\mathcal{Y}=\{1,2,\cdots,C\}$为类别集合，$C$表示类别总数。网络流量记录由多个结构化字段组成，可表示为:

$$\mathbf{x}_i=\left(x_{i,1},x_{i,2},\cdots,x_{i,d_f}\right)^{\mathrm{T}} \tag{1}$$

其中，$x_{i,j}$表示第i条流量记录的第j个原始字段。

网络入侵检测任务的目标是学习参数为$\boldsymbol{\theta}$的分类模型，使其能够根据输入流量记录预测对应类别，并最小化预测结果与真实标签之间的分类损失：

$$\boldsymbol{\theta}^{*}=\mathrm{argmin}_{\boldsymbol{\theta}}\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\ell_{\mathrm{cls}}\left(\boldsymbol{\psi}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{\mathrm{i}}\right),\mathrm{y_i}\right) \tag{2}$$

其中，$\boldsymbol{\theta}^{*}$表示优化后的模型参数，$\boldsymbol{\psi}_{\boldsymbol{\theta}}$表示参数为$\boldsymbol{\theta}$的分类函数，$\ell_{\mathrm{cls}}$表示分类损失函数。对于一条流量样本，其判别信息不仅来自各字段的单独取值，也在多个字段协同变化的模式之中。如果将流量日志仅仅看作相互独立的数值特征集合，字段间的行为关联就很容易被忽略，模型对复杂攻击行为的表征能力也会因此受限。

由此，本文所聚焦的问题可以进一步形式化为，当结构化网络流量字段之间存在错综复杂的关联时，如何学习一种有效的表示$\mathbf{z}_i=\phi_{\boldsymbol{\theta}_{\mathrm{enc}}}\left(\mathbf{x}_i\right)$，使其既能保留字段的取值信息，又能刻画字段间的协同关系，并最终服务于正常流量与多类型攻击流量的分类判别：$\hat{y}_i=\psi_{\boldsymbol{\theta}_{\mathrm{cls}}}\left(\mathbf{z}_i\right)$。其中，$\phi_{\boldsymbol{\theta}_{\mathrm{enc}}}$表示编码函数，$\boldsymbol{\theta}_{\mathrm{enc}}$表示编码器参数，$\boldsymbol{\theta}_{\mathrm{cls}}$表示分类层参数，$\mathbf{z}_i$表示第$i$条流量的高层语义表示，$\hat{y}_i$表示预测类别。

2.2　检测方法框架

本文方法的整体框架如图1所示，由日志数据预处理、日志语义转换、领域自适应预训练和模型微调四个环节构成。

（1）日志数据预处理。首先，对原始网络流量实施清洗，剔除缺失值、异常点及冗余信息，以保证数据质量。随后，对不同数据源的字段进行统一解析与标准化，同时完成攻击标签的映射，由此得到结构一致的数据表示。

（2）日志语义转换。依据网络连接行为的描述顺序组织字段，并将不同类型的流量特征划入对应的语义模块。然后，把每条样本的字段取值填入模板，生成结构规整、语义流畅的类自然语言日志序列。最后，每条样本被整理为包含text和label两个字段的记录，其中text将用于领域自适应预训练和语义编码，label则服务于下游入侵检测的监督微调。

（3）领域自适应预训练。首先使用SecureBERT分词器将日志文本切分为词元，并转换为Token ID序列。随后，按照设定的掩码比例和日志语义结构生成动态掩码，构建掩码语言建模(Masked Language Modeling, MLM)任务。模型根据未被遮蔽的上下文预测原始词元，并通过交叉熵损失更新参数。

（4）基于下游攻击任务的模型微调。日志文本样本先经分词器生成Token ID、注意力掩码和标签编号，输入编码器提取序列上下文表示。随后，抽取序列级全局表示作为该样本的语义向量，送入全连接分类层，输出各类别的预测概率。模型根据分类层输出概率的最大值确定样本类别，完成正常流量与多类攻击的判别。

2.3　日志语义分析

为使结构化流量字段在统一上下文中表达连接行为，本文首先构建固定模块化日志模板，再利用预训练语言模型学习日志序列中的跨字段依赖关系。前者解决字段语义组织问题，后者实现上下文语义表示学习。

2.3.1　类自然语言语义表达构建

为赋予网络流量数据更强的语义表达，本文将原始流量样本中的多维结构化字段按固定语义模板转化为文本序列，使原本彼此孤立的统计特征在统一的语境下协同描述网络连接行为，进而为后续预训练语言模型提供语义信息更密集的输入。

基于日志语义模板的类自然语言语义表达构建过程如图2所示，这里采用模板化映射方式生成文本样本，输入一侧为原始流量特征向量，输出一侧则是对应的类自然语言日志序列。

类自然语言语义映射函数记作：

$$S_i=\mathcal{T}_{\mathcal{P}}\left(\mathbf{x}_i\right) \tag{3}$$

其中$\mathcal{T}_{\mathcal{P}}:\mathbb{R}^{d_f}\rightarrow\mathcal{S}$表示由模板参数集合$\mathcal{P}$确定的日志语义映射函数。$\mathbf{x}_i$为第$i$条原始流量样本，$\mathcal{P}$包含字段名称、字段离散化规则、语义模块划分、模块内字段顺序及固定描述短语。$\mathcal{S}$表示日志文本空间，

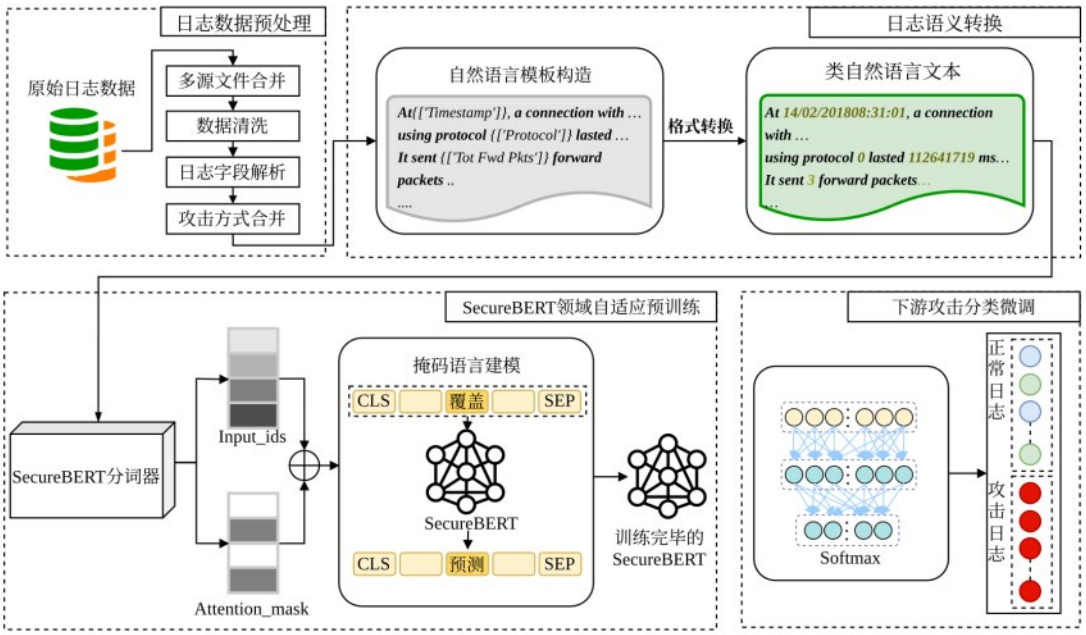


图1　基于日志语义分析的网络入侵检测方法框架

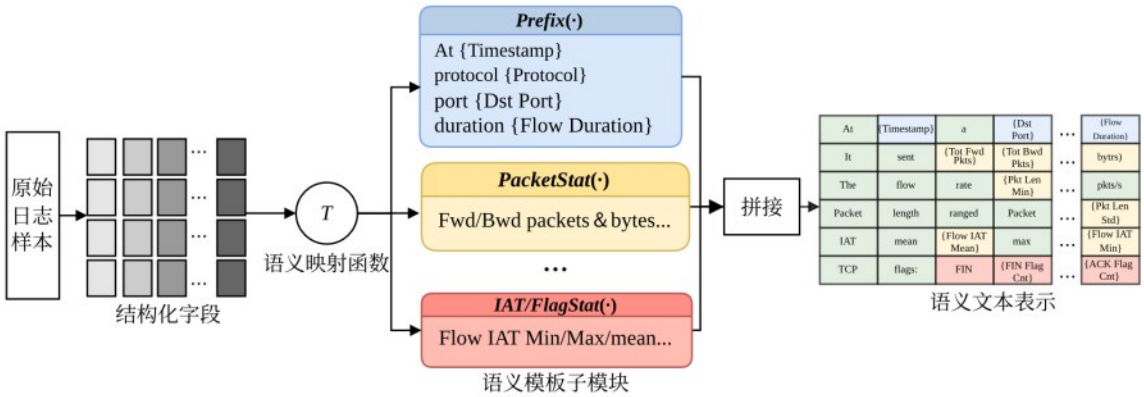


图2　日志语义模板结构化重构过程

S_i表示第i条流量重构后的原始日志文本序列。

此外，语义映射函数$\mathcal{T}_p$的具体实现依赖于模板的构造方式，而模板设计必须在信息表达能力和结构一致性之间寻求平衡。为此，在模板构造过程中，本文坚持每条语义槽位与原始字段一一对应，直接引用结构化特征并加以格式化表达，以保证文本表示能够追溯原始流量行为。模板构造的具体步骤如下。

（1）语义字段选取

在模板构造过程中，首先从原始网络流量字段中筛选出能够刻画网络行为的关键字段，主要包括时间信息、协议与端口信息、双向流量统计字段、速率字段、包长分布字段、流间隔时间字段以及TCP标志位字段。这些字段并非简单拼接，而是围绕攻击行为建模的需要加以选择，使其覆盖网络连接的上下文信息、方向特性、流量强度以及会话状态等多个维度。为更清晰地展示所选字段的语义功能，具体字段及其作用由表1给出。

（2）基于特征重要性加权的语义模板设计

在完成关键字段选取后，本文进一步构建特征重要性加权的固定语义模板，将结构化网络流量字段转换为语义边界明确、字段位置稳定的类自然语言日志。与按照原始数据表列顺序直接拼接字段的方式不同，本文以关键字段在攻击行为刻画中的判别贡献为依据确定模板排布，同时结合网络连接行为的固有逻辑保持字段之间的上下文关系。该机制

表1　日志语义模板中关键字段及其功能描述

字段类别	具体字段	功能描述
时间与上下文特征	Timestamp, Flow Duration	描述连接时间与持续性
协议与端口特征	Protocol, Dst Port	表征通信语境与服务类型
双向流量特征	Tot Fwd Pkts, TotLen Fwd Pkts, Tot Bwd Pkts, TotLen Bwd Pkts	描述流量交互结构
速率特征	Flow Byts/s, Flow Pkts/s	表征流量传输强度
包长分布特征	Pkt Len Min, Pkt Len Max, Pkt Len Mean, Pkt Len Std	描述包长分布特性
流间隔特征	Flow IAT Mean, Flow IAT Max, Flow IAT Min	表征流量时间间隔特性
TCP 标志位特征	FIN Flag Cnt, SYN Flag Cnt, RST Flag Cnt, ACK Flag Cnt	描述连接状态特征

不针对不同流量样本动态改变文本结构，而是在模板构建阶段离线确定字段的模块归属、相对位置和表达粒度，使同一字段在不同日志中具有一致的语义位置。

设$\mathcal{K}$表示参与日志构建的关键字段下标集合，κ_j表示第j个关键字段在前述字段选取阶段获得的综合重要性得分。该得分反映字段对攻击类别区分及网络行为描述的贡献程度。为消除不同字段得分尺度的影响，本文对其进行归一化，得到字段重要性权重：

$$\chi_j = \frac{\kappa_j}{\sum_{q \in \mathcal{K}} \kappa_q}, \qquad j \in \mathcal{K} \tag{4}$$

其中，χ_j表示第j个关键字段的归一化重要性权重，χ_j越大，说明对应字段在日志语义表达中具有更高的优先级。该权重仅用于模板的离线构建，不作为模型训练参数参与迭代更新，也不使用验证集、测试集或目标域标签信息，从而避免数据泄漏。

考虑到网络连接行为具有从连接属性到通信交互、传输特征、时间动态和会话状态的自然描述过程，本文并未将全部字段完全按照χ_j由高到低进行线性排列，而是将特征重要性与网络行为逻辑共同作为模板组织约束。具体而言，关键字段被划分为Prefix、PacketStat、RateStat、LengthStat、IATStat和FlagStat六个语义模块，其字段集合满足：

$$\begin{gathered} \mathcal{K} = \bigcup_{u=1}^{6} \mathcal{K}_u \\ \mathcal{K}_u \cap \mathcal{K}_v = \varnothing \\ u \neq v \end{gathered} \tag{5}$$

其中，$\mathcal{K}_u$表示第u个语义模块对应的字段下标集合，$u \in \{1,2,\cdots,6\}$依次对应Prefix、PacketStat、RateStat、LengthStat、IATStat和FlagStat。Prefix模块组织时间、端口、协议和持续时间等连接上下文；PacketStat模块描述前向与后向报文数量及字节规模；RateStat、LengthStat和IATStat模块依次刻画传输强度、报文长度分布和流量时间动态；FlagStat模块描述TCP标志及会话控制状态。模块间的固定顺序遵循网络连接行为的描述逻辑。

协议、目标端口和流持续时间等关键连接属性采用完整描述并优先表达；前向与后向报文统计保持方向配对；包长和流间隔字段保留最小值、最大值、均值及离散程度之间的统计关系。

基于上述设计，日志模板是以全局特征重要性为依据确定模块内字段的表达优先级，并通过固定模板保持跨样本的语义一致性。相较于随单条流量动态调整字段顺序或文本句式的表达方式，该方法能够为协议、端口等高判别性字段提供稳定的语义位置，同时减少文本表述差异对领域自适应预训练的干扰。表2给出了各语义模块的字段组成及固定日志表述。

2.3.2　语义表示机制

在类自然语言语义表达构建完成之后，原始网络流量数据被统一转换为文本形式的语义序列，并通过字段“text”进行存储与组织。需要对这些语义文本序列实施向量化编码，从中获取能够刻画网络行为语义特征的连续表示。

设第i条语义文本样本表示为$S_i = (\omega_{i,1},\omega_{i,1},\ldots\omega_{i,n_i})$,其中$\omega_{i,t}$表示第$i$条日志中位置$t$的词元，$n_i$表示该日志的有效词元数量。该序列通过词嵌入映射至连续向量空间，得到输入表示：$\mathbf{E}_i = [\mathbf{e}_{i,1},\mathbf{e}_{i,2},\cdots,\mathbf{e}_{i,ni}]^{\mathrm{T}}$。其中，$\mathbf{E}_i$表示第$i$条日志的词元嵌入矩阵，$\mathbf{e}_{i,t}$表示词元$\omega_{i,t}$对应的嵌入向量。

语义编码阶段利用自注意力机制捕捉序列中不

表2 日志语义模板结构化组成

语义模块	选取字段	日志语义表述
Prefix	Timestamp，Protocol Dst Port，Flow Duration	At {Timestamp}, a connection with destination port {Dst Port} using protocol {Protocol} lasted {Flow Duration} ms.
PacketStat	Tot Fwd/ Bwd Pkts，TotLen Fwd/ Bwd Pkts	It sent {Tot Fwd Pkts} forward packets ({TotLen Fwd Pkts} bytes) and {Tot Bwd Pkts} backward packets ({TotLen Bwd Pkts} bytes).
RateStat	Flow Byts/s，Flow Pkts/s	The flow rate was {Flow Byts/s} bytes/s and {Flow Pkts/s} pkts/s.
LengthStat	Pkt Len Min/ Max，Pkt Len Mean，Pkt Len Std	Packet length ranged from {Pkt Len Min} to {Pkt Len Max} with mean {Pkt Len Mean} and deviation {Pkt Len Std}.
IATStat	Flow IAT Mean/ Max，Flow IAT Min	IAT mean {Flow IAT Mean}, max {Flow IAT Max}, min {Flow IAT Min}.
FlagStat	FIN/SYN/RST/ACK Flag Cnt	TCP flags: FIN={FIN Flag Cnt}, SYN={SYN Flag Cnt}, RST={RST Flag Cnt}, ACK={ACK Flag Cnt}.

同位置间的依赖关系。这一机制的核心在于通过注意力权重自适应地为各个词元分配重要性，使模型得以从全局范围挖掘网络流量字段之间潜在的关联。具体而言，任意位置t的上下文表示由所有位置的加权组合给出：

$$\mathbf{h}_{i,t}=\sum_{r=1}^{n_i}\alpha_{i,t,r}\mathbf{v}_{i,r} \tag{6}$$

其中，$\mathbf{h}_{i,t}$表示第i条日志中位置t的上下文表示向量，$\mathbf{v}_{i,r}$表示位置r的值向量，$\alpha_{i,t,r}$表示位置t对位置r分配的注意力权重，且满足:

$$\sum_{r=1}^{n_i}\alpha_{i,t,r}=1 \tag{7}$$

经过多层语义编码后，第i条日志的上下文表示矩阵为：$\mathbf{H}_i=(\mathbf{h}_{i,1},\mathbf{h}_{i,2},\cdots,\mathbf{h}_{i,ni})^{\mathrm{T}}$,其中，$\mathbf{H}_i$表示日志序列的上下文表示矩阵。本文使用首位特殊词元[CLS]的上下文向量作为整条日志的全局语义表示：$\mathbf{z}_i=\mathrm{Pool}(\mathbf{H}_i)=\mathbf{h}_{i,[\mathrm{CLS}]}$。这里的Pool(·)表示序列级聚合算子，$\mathbf{z}_i$表示第$i$条日志的全局语义表示向量。这一全局表示综合序列中不同语义片段的信息，从而完成对一次网络连接行为的整体建模。

结合实际数据构建方式可以看出，该语义表示机制直接作用于无标签文本数据，仅依赖“text”字段进行表示学习。这种设计使得模型能够在不借助类别标签的条件下，自主探索网络流量中的语义结构和字段共现模式，从而获得更具通用性的表示能力。

综上，本文所设计的语义表示机制实现对网络流量语义信息的深层表示学习。该机制有效整合多维字段信息，并通过上下文依赖建模增强不同攻击行为之间的可分性，从而为后续入侵检测任务提供更具判别力的语义表示。

2.4 领域自适应预训练

本文在传统DAPT框架的基础上，结合日志数据自身的特点进行适配性改进，以有效建模网络日志的语义结构，具体设计流程见图3。

本文通过日志语义模板构造领域预训练语料，使输入序列具有清晰的语义模块结构。记语义重构后的日志文本序列为$S_i=(\omega_{i,1},\omega_{i,2},\cdots,\omega_{i,n_i})$。DAPT的目标是在不引入监督信号的条件下，通过日志语料对模型参数进行再优化，使其从初始参数$\boldsymbol{\theta}_0$更新为领域参数$\boldsymbol{\theta}_{\mathrm{d}}$：

$$\boldsymbol{\theta}_{\mathrm{d}}=\mathrm{argmin}_{\boldsymbol{\theta}}\frac{1}{\left|\mathcal{D}_{\log}\right|}\sum_{\mathrm{S_i}\in\mathcal{D}_{\log}}\mathcal{L}_{\mathrm{MLM}}\left(\mathrm{S_i};\boldsymbol{\theta}\right) \tag{8}$$

其中，$\mathcal{D}_{\log}$表示日志域语料分布，$\boldsymbol{\theta}_0$表示SecureBERT的初始参数，$\boldsymbol{\theta}$表示预训练过程中待优化的模型参数。

$\boldsymbol{\theta}_{\mathrm{d}}$表示完成DAPT后的领域适配参数，$\mathcal{L}_{\mathrm{MLM}}$表示掩码语言建模损失函数。该过程在保持原有语言建模能力的同时，引导模型学习日志语料中的字段共现规律和结构依赖模式，从而实现语义空间的领域对齐。

在训练任务阶段，本文以MLM作为核心目标。训练过程中，从输入序列里结构化动态掩码构成掩码集合M，并对相应词元进行扰动，由此得到掩码序列。模型依据未被掩码的上下文信息预测被遮蔽的词元，其优化目标定义为：

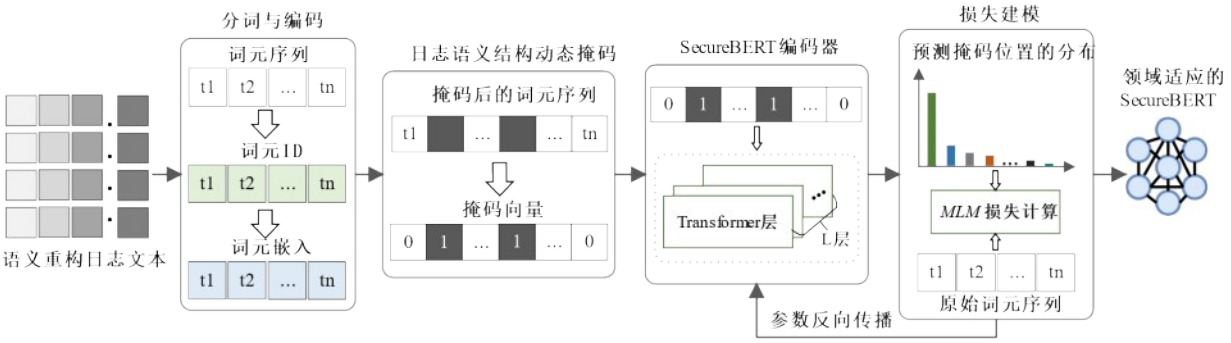


图3　面向日志语义结构的改进领域自适应预训练流程

$$\mathcal{L}_{\mathrm{MLM}}\left(S_i;\boldsymbol{\theta}\right)=-\frac{1}{\left|\mathcal{M}_i\right|}\sum_{t\in\mathcal{M}_i}\log p_{\boldsymbol{\theta}}\left(\omega_{i,t}\middle|\tilde{S}_i\right) \tag{9}$$

其中，$\mathcal{M}_i$表示第i条日志的掩码位置集合，$\tilde{S}_i$表示掩码后的日志序列，$\omega_{i,t}$表示位置t处的原始词元，$p_{\theta}(\cdot)$表示参数为$\boldsymbol{\theta}$的模型预测概率。对于单个被掩码位置t，其预测过程可表示为：

$$p_{\boldsymbol{\theta}}\left(\omega_{i,t}\middle|\tilde{S}_i\right)=\mathrm{softmax}\left(\mathbf{W}_{\mathrm{MLM}}\mathbf{h}_{i,t}+\mathbf{b}_{\mathrm{MLM}}\right)_{\omega_{i,t}} \tag{10}$$

其中，$\mathbf{W}_{\mathrm{MLM}}$表示MLM预测头的权重矩阵，$\mathbf{b}_{\mathrm{MLM}}$表示MLM预测头的偏置向量，$\mathbf{h}_{i,t}$表示位置$t$的上下文表示向量。考虑到网络日志具有“长序列、高冗余、强结构”的特点，本文对预训练过程进行两方面的适配性改进。第一，在输入层面，直接采用语义模板重构后的日志文本作为预训练语料。第二，在训练层面，对动态随机掩码策略[26]进行适配修改，设计面向日志语义结构的动态掩码策略，使模型在不同迭代轮次中接触差异化的字段组合，从而避免对固定模板模式的机械记忆，并增强对字段共现关系的建模能力。

2.5　面向日志语义结构的动态掩码策略

通用DAPT采用均匀随机掩码，易使高频模板词占据较多预测位置，模型可能依赖固定字段顺序完成词元恢复，难以充分学习协议、端口等关键行为字段之间的关联。为此，LSADA将日志语义结构引入DAPT过程：将主要掩码预算优先分配给行为字段，以完整字段片段替代孤立词元作为掩码单位，并依据语义组频率对预训练损失进行平衡加权。该策略能够抑制模板结构带来的预测捷径，强化模型对网络流量行为关联的建模，使预训练目标由通用词元恢复转化为结构化行为字段恢复。

设第i条网络流量经日志语义重构后形成词元序列S_i。依据日志模板的功能差异，LSADA将协议、端口等内容划分为Prefix等六类行为语义组；字段名称、连接词和分隔符等固定内容归入模板词组Tmp。该划分使掩码策略能够识别不同词元所对应的行为语义，而非仅根据其在序列中的位置进行随机选择。

设Ω_i表示第i条日志中可参与掩码的有效词元位置集合，不包含特殊词元和填充词元。整体掩码比例记为ρ，本文取ρ=0.15。LSADA首先计算总掩码预算，再将其划分为行为字段掩码预算与模板词掩码预算：

$$B_i=\max\left(1,\left\lfloor\rho\left|\Omega_i\right|\right\rfloor\right) \tag{11}$$

$$B_i^{\mathrm{beh}}=\left\lfloor\eta B_i\right\rfloor \tag{12}$$

$$B_i^{\mathrm{tmp}}=B_i-B_i^{\mathrm{beh}}$$

其中，B_i表示第i条日志的总掩码词元数，$|\Omega_i|$表示可掩码词元数量，η表示行为字段掩码比例，本文设置为η=0.90。因此，约90%的掩码预算用于协议、端口等行为字段，仅保留约10%的预算用于固定模板词。该分配方式的理论依据在于，行为字段包含更多与攻击模式相关的不确定信息，而模板词主要承担字段组织作用。将有限的预测监督优先分配给行为字段，能够提高模型对网络行为依赖关系的学习强度，同时保留少量模板词掩码以维持日志表达结构的完整性。

不同语义组的字段数量和词元频率存在差异。若在行为字段内部继续进行均匀采样，高频字段组仍会获得更多掩码机会，导致模型优化偏向频繁出现的常规行为模式。为此，LSADA根据各语义组

在DAPT训练语料中的词元频率构造动态采样概率：

$$\pi_u = \frac{\left(f_u + \varepsilon\right)^{-\beta}}{\sum_{v \in \mathcal{U}_{\text{beh}}} \left(f_v + \varepsilon\right)^{-\beta}}, \qquad u \in \mathcal{U}_{\text{beh}} \tag{13}$$

其中，π_u表示行为语义组u被选中的概率，$\mathcal{U}_{\text{beh}}$表示六类行为语义组构成的集合，f_u表示语义组u在DAPT训练语料中的词元频率，ε为防止分母为零的平滑常数，β为频率平衡系数。逆频率平方根形式能够适度降低高频字段组的采样优势，使端口等相对低频但具有较强行为指示作用的字段获得更多恢复机会；同时避免极低频字段被过度放大，从而保证掩码分布与训练过程的稳定性。

在确定语义组后，LSADA不以单个词元作为最小掩码单位，而是以完整字段片段进行掩码。LSADA将同一字段对应的子词整体选中或整体保留。设$\mathcal{M}_i^{\text{beh}}$和$\mathcal{M}_i^{\text{tmp}}$分别表示行为字段与模板词的掩码位置集合，则最终掩码集合定义为$\mathcal{M}_i = \mathcal{M}_i^{\text{beh}} \cup \mathcal{M}_i^{\text{tmp}}$。

其中，$\mathcal{M}_i$表示第i条日志的最终掩码位置集合，$|\mathcal{M}_i^{\text{beh}}| \approx B_i^{\text{beh}}$和$|\mathcal{M}_i^{\text{tmp}}| = B_i^{\text{tmp}}$分别表示两类掩码词元数量。由于字段片段长度并不一致，行为字段掩码数量允许在单条日志上与预算值存在轻微偏差，但在批量训练层面保持预设掩码比例。该机制消除了子词残留形成的预测捷径，迫使模型结合协议、端口等上下文信息恢复被遮蔽字段。

此外，LSADA在每个训练迭代中重新生成掩码集合，使同一条日志在不同训练轮次中对应不同的字段遮蔽组合。动态重采样相当于从不同局部视角观察同一网络连接，可降低模型对固定字段位置的记忆，并增强其对跨字段语义关系的鲁棒建模能力。

仅改变掩码采样仍无法完全消除高频字段组在损失计算中的梯度优势。为此，LSADA进一步构造语义组平衡权重，并将其引入加权MLM损失：

$$\lambda_u = \min\left\{2.0, \max\left\{0.5, \left(\frac{\bar{f} + \varepsilon}{f_u + \varepsilon}\right)^{\gamma}\right\}\right\} \tag{14}$$

$$\mathcal{L}_{\text{LSADA - MLM}} = -\frac{\sum_{t \in \mathcal{M}_i} \lambda_{\zeta(i,t)} \log p_{\boldsymbol{\theta}}\left(\omega_{i,t} \middle| \tilde{S}_i\right)}{\sum_{t \in \mathcal{M}_i} \lambda_{\zeta(i,t)}} \tag{15}$$

其中，$\mathcal{L}_{\text{LSADA - MLM}}$表示语义组平衡加权的MLM损失；$\zeta(i,t)$表示第$i$条日志中位置$t$的词元所属语义组编号；$\lambda_{\zeta(i,t)}$表示该词元对应的语义组平衡权重。权重下限设置为0.5、上限设置为2.0，用于避免高频字段权重过低或低频字段权重过高而引起梯度失衡；分母采用权重归一化，使不同日志长度、字段分布和掩码组合下的损失尺度保持可比。

综上，LSADA通过行为字段优先采样提升关键网络行为的预测比例，通过完整字段片段掩码避免子词残留造成的信息泄漏，并通过语义组平衡损失抑制高频模板词和高频字段组对参数更新的主导作用。该策略将日志结构先验转化为可优化的预训练约束，使模型更加关注协议、端口、速率、时序和TCP状态之间的联合依赖，为后续入侵检测提供具有更强判别能力的领域语义表示。

2.6 基于网络入侵检测的下游任务微调

为执行具体的入侵检测任务，本文引入下游任务微调策略，对预训练模型进行任务适配，使其具备网络行为的分类判别能力。

不同于预训练阶段对语义建模的侧重，微调(fine-tuning)阶段的目标在于利用已有的语义表示能力，将输入样本导向任务相关的映射，实现从通用表示到任务判别的转换。

本文的微调阶段面向入侵检测任务，把语义重构后的日志文本作为输入，基于经过领域自适应预训练的模型进行参数优化。具体而言，输入日志序列S_i经模型编码后得到隐藏表示$\mathbf{H}_i$，各个位置的向量均已融合日志上下文的语义信息。本文选用序列级语义表示作为分类的依据，具体而言，抽取首位的全局向量$\mathbf{h}_{i,[\text{CLS}]}$作为该样本的整体语义表示$\mathbf{z}_i$。将语义表示$\mathbf{z}_i$输入分类层，得到各攻击类别的预测概率：

$$\hat{\mathbf{p}}_i = \text{softmax}\left(\mathbf{W}_{\text{cls}}\mathbf{z}_i + \mathbf{b}_{\text{cls}}\right) \tag{16}$$

$$\hat{y}_i = argmax_{r \in y}\hat{p}_{i,r} \tag{17}$$

其中，$\hat{\mathbf{p}}_i \in \mathbb{R}^C$表示第$i$条日志属于各类别的预测概率向量，$\hat{p}_{i,r}$表示该日志被预测为类别$r$的概率，

$\mathbf{W}_{cls} \in \mathbb{R}^{C \times d_h}$表示分类层的权重矩阵，$\mathbf{b}_{cls} \in \mathbb{R}^{C}$表示分类层的偏置向量，$\hat{y}_i$表示最终预测类别。

微调期间，预训练模型参数和分类层参数共同参与优化，使得模型在保持语义表示能力的同时，进一步捕捉入侵检测任务中的判别边界。由于预训练阶段已借助DAPT学到日志中的字段关联关系与语义结构信息，微调阶段无须由原始特征重新建模，而是在已有的语义空间上进行任务适配，从而有效增强模型对复杂网络行为的识别能力。

本文建立的技术路线遵循“语义重构—领域自适应预训练—下游任务微调”的递进逻辑。借助语义重构，结构化的流量字段被编排成连贯的日志文本，使多字段协同变化得以在统一语境下呈现，而不是退化为彼此孤立的离散特征。随后，领域自适应预训练环节把通用语言模型的知识向网络日志域迁移，在不依赖标注的条件下逐步拟合日志独有的字段结构、共现规律和攻击行为语义模式，完成语义空间的领域对齐。下游任务微调将此语义表示引向攻击类别判别，联合优化编码器与分类层，促使模型在语义空间中自主捕捉正常流量与各类攻击之间的决策边界。使模型既具备从语义层面解析不同攻击模式差异的能力，又能在复杂网络环境中保持更稳定的检测表现与泛化水平。

3　实验与结果分析

3.1　实验数据集

为验证本文所提出方法在网络入侵检测任务中的有效性与跨数据集泛化能力，本文选取CIC-IDS2017[27]、CSE-CIC-IDS2018[28]以及UNSW-NB15[29]三个公开数据集开展实验。其中，CIC-IDS2017与CSE-CIC-IDS2018均用于全域多分类实验和跨数据集迁移实验；UNSW-NB15仅用于跨数据集迁移实验，用于评估模型在复杂网络环境下的迁移适应能力。

CIC-IDS2017数据集由加拿大网络安全研究所发布，是入侵检测领域常用的基准数据集，包含DDoS攻击、拒绝服务攻击（Denial of Service，DoS）、暴力破解（Brute Force），僵尸网络（Botnet）等多种攻击类型，具有较好的流量多样性与现实性。

CSE-CIC-IDS2018数据集由加拿大通信安全机构（Communications Security Establishment，CSE）与加拿大网络安全研究所联合发布，规模较大且攻击类型丰富，覆盖DDoS攻击、拒绝服务攻击（DoS）、Brute Force、Botnet和渗透攻击（Infiltration）等典型攻击行为，同时包含大量正常流量（Benign）。

UNSW-NB15数据集由澳大利亚新南威尔士大学网络安全实验室构建，包含模糊测试攻击（Fuzzing）、分析攻击（Analysis）等攻击类型，能够反映复杂网络环境下的攻击行为特征。考虑到该数据集与CIC系列数据集在流量分布、攻击类型和特征空间上存在差异，本文仅将其用于跨数据集迁移实验。本文将CIC-IDS2017与CSE-CIC-IDS2018样本集按8:1:1的比例划分为训练集、验证集和测试集，具体样本分布如表3和表4所示。

表3　CIC-IDS2017数据集样本分布

攻击类型	总样本数	训练集	验证集	测试集
Benign	2273097	1818478	227310	227309
DoS	252661	202129	25266	25266
Infiltration	159650	127720	15965	15965
DDoS	128027	102422	12803	12802
BruteForce	15342	12274	1534	1534
Botnet	1966	1573	197	196
总计	2830743	2264596	283075	283072

表4　CSE-CIC-IDS2018数据集样本分布

攻击类型	总样本数	训练集	验证集	测试集
Benign	6112137	4889710	611214	611213
DDoS	1263933	1011146	126393	126394
DoS	654300	523440	65430	65430
Botnet	286191	228953	28619	28619
BruteForce	381877	305501	38188	38188
Infiltration	161934	129547	16193	16194
总计	8860372	7088297	886037	886038

为保证不同数据集之间标签体系的一致性，并降低细粒度攻击子类差异对模型学习的干扰，本文对CIC-IDS2017和CSE-CIC-IDS2018中的原始攻击标签进行统一归并，最终形成Benign、DDoS、DoS、Botnet、BruteForce和Infiltration六类检测任务。

其原因在于，本文采用的SecureBERT属于预

训练语言模型，其优势在于通过上下文建模学习网络行为的语义模式和字段间关联关系。若直接使用过于细粒度的原始攻击标签，不同数据集中相同攻击家族可能因命名方式、采集环境或攻击变体不同而被划分为多个子类，容易造成标签语义不一致、类别边界模糊和样本分布稀疏等问题，从而影响模型对攻击行为共性特征的学习。将攻击类型归并到攻击家族层面，有助于增强标签语义的一致性，使模型更加关注不同攻击类别在端口、协议、包数、速率、IAT特征和标志位等字段组合上的整体行为差异，减少对数据集特定标签形式的依赖。同时，六分类设置能够缓解部分细粒度攻击类别样本不足带来的类别不平衡问题，提高预训练语言模型在下游微调阶段的类别判别稳定性，并为后续跨数据集迁移实验提供统一的标签空间。

3.2 实验设置与评估指标

实验使用的CPU为Intel Core i9-14900K处理器，64GB内存，NVIDIA GeForce RTX 4090GPU。采用Python 3.11作为编程语言，基于深度学习框架PyTorch 2.1实现本文方法，并结合Transformers工具库完成模型训练与推理。模型具体训练设置如表5所示。

表5 微调训练设置

参数	设置
优化器	AdamW
学习率	2×10^{-5}
批次大小	16
训练轮次	5
损失函数	交叉熵损失

为系统验证LSADA的检测性能、关键机制有效性、跨数据集泛化能力和工程部署可行性，本文按照实验目的将后续实验划分为以下五组。

（1）基础检测性能实验。在CIC-IDS2017和CSE-CIC-IDS2018数据集上开展六分类检测实验，并设置极端类别不平衡实验，分别用于验证LSADA在常规多分类场景和低比例恶意流量场景下的检测性能与鲁棒性，对应第3.4节和第3.5节。

（2）跨数据集迁移与泛化实验。在CIC-IDS2017、CSE-CIC-IDS2018和UNSW-NB15之间设置少样本迁移和目标域无标注样本迁移任务，用于分析模型在数据分布变化和目标域标注数据受限条件下的迁移适应能力，对应第3.6节。

（3）关键机制与参数有效性实验。通过不同掩码策略对比、超参数敏感性分析、不同日志语义建模策略对比以及模块消融实验，分别验证面向日志语义结构的动态掩码机制、掩码比例与学习率设置、模块化日志语义表达和各组成模块的实际作用，对应第3.7节至第3.10.2节。

（4）语义表征与可解释性实验。通过类别表征空间可视化、字段组自注意力分析和跨数据集联合表征可视化，观察日志语义重构与领域自适应预训练对类别可分性、关键字段关联学习和域间表征偏移的影响，对应第3.10.3节和第3.10.4节。

（5）模型效率与工程适用性实验。统计模型参数量、领域自适应预训练开销、单流实时推理时延和离线批量检测吞吐量，用于评估LSADA在实际网络入侵检测场景中的资源成本和部署可行性，对应第3.11节。

上述实验由“检测性能验证—跨域泛化分析—关键机制验证—语义表征解释—工程效率评估”逐层展开，从性能、机制、泛化、解释性和部署开销等方面对本文方法进行综合评价。

为评估模型的检测性能，本文采用Accuracy、Precision、Recall、F1分数、FPR和FNR作为评价指标。对于二分类任务,TP表示被正确识别为攻击流量的样本数，TN表示被正确识别为正常流量的样本数，FP表示被错误判定为攻击流量的正常样本数，FN表示被错误判定为正常流量的攻击样本数：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{18}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{19}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{20}$$

$$F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{21}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{22}$$

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}} \tag{23}$$

3.3 数据预处理

为增强网络流量数据的语义表达能力，本文设

计一套面向日志语义建模的数据预处理流程。该流程接收多天网络流量数据，将原始结构化特征转化为统一的类自然语言文本序列，从而为后续的领域自适应预训练和下游任务微调提供规范的输入。

（1）数据整合。将多天的流量数据合并，统一处理字段名称、大小写、空格以及命名不规范之处，建立一致的字段集合。并为每条样本生成唯一的哈希标识，以备后续去重与一致性校验使用。

（2）数据清洗。重复记录被移除，缺失值、异常值和异常字段经滤除或标记处理。面对类别分布不均衡的状况，进一步结合随机欠采样与过采样策略调整样本分布，以此改善数据集的完整性、一致性与类别均衡性。

（3）标签统一。针对原始攻击标签粒度参差不齐的问题，对细粒度攻击标签进行归并映射，构建一套统一的攻击类别体系，从而削减类别冗余，统一多分类任务的标签空间。

（4）语义重构。在完成字段规范化和标签统一后，选取时间戳、端口、协议、流量统计特征以及TCP标志位等能刻画网络行为的关键字段，按照固定的日志语义模板将其转换为类自然语言文本序列。

（5）格式转换。把生成的语义文本及对应标签统一存储为JSONL格式。每条样本由语义文本字段text和标签字段label组成，可直接用于领域自适应预训练、下游微调和模型评估，并按照实验设定划分出训练集、验证集和测试集，为后续实验提供统一的数据输入。

3.4　网络入侵检测多分类实验

为验证本文方法在多分类入侵检测任务中的有效性，本文分别在CIC-IDS2017和CSE-CIC-IDS2018数据集上开展六分类实验，并选取HAE-HRL[12]、SSA-BOA-CNN[13]、A-NIDS[14]、BERT[24]、BERT-RL[25]和SecureBERT作为对比方法。SecureBERT保留本文日志语义模板和下游分类配置，但不执行日志领域DAPT，仅利用带标签训练样本进行监督微调。该对照用于分析SecureBERT已有网络安全文本知识在流级网络日志场景下的适用性，并验证DAPT对缩小网络安全文本语料与结构化网络流量日志之间领域分布差异的作用。为保证实验公平性，各方法使用相同的训练集与测试集进行训练和评估，评价指标的统计口径保持一致。对比方法保留其原始模型结构及关键训练配置，LSADA在相同数据划分条件下完成日志语义重构、领域自适应预训练和下游微调；DAPT语料仅由训练划分中的无标签日志构成，不使用验证集和测试集信息。各方法在两个数据集上的精确率和F1分数结果如表6至表9所示，其中表6和表7对应CIC-IDS2017，表8和表9对应CSE-CIC-IDS2018。

从表6和表7可以看出，LSADA在CIC-IDS2017数据集上取得最优性能。精确率方面，该模型在Benign、Botnet、BruteForce、DDoS、DoS和Infiltration类别上均保持较高水平，宏平均精确率和加权平均精确率分别达到0.9993和0.9988，展现出较低的误报率和稳定的分类能力。F1分数方面，LSADA的Macro-F1和Weighted-F1分别达到0.9989和0.9982，明显优于所有对比方法。表明日志语义表示与领域适配机制能够有效减轻类别不平衡所带来的性能退化。

表6　CIC-IDS2017数据集的精确率实验结果

攻击类型	HAE-HRL	SSA-BOA-CNN	A-NIDS	BERT	BERT-RL	SecureBERT	**LSADA**
Benign	0.9513	0.9846	0.9898	0.9677	0.9812	0.9948	**0.9986**
Botnet	0.7468	0.8037	0.8651	0.7884	0.8279	0.9710	**0.9997**
BruteForce	0.8276	0.9224	0.9361	0.8847	0.9126	0.9810	**0.9993**
DDoS	0.9901	0.9934	0.9967	0.9728	0.9876	0.9988	**0.9999**
DoS	0.9307	0.9752	0.9874	0.9336	0.9701	0.9965	**0.9999**
Infiltration	0.9094	0.9547	0.9733	0.9236	0.9491	0.9859	**0.9981**
Macro Avg	0.8927	0.9390	0.9581	0.9118	0.9381	0.9880	**0.9993**
Weighted Avg	0.9482	0.9818	0.9888	0.9617	0.9780	0.9945	**0.9988**

表7 CIC-IDS2017数据集的F1分数实验结果

攻击类型	HAE-HRL	SSA-BOA-CNN	A-NIDS	BERT	BERT-RL	SecureBERT	**LSADA**
Benign	0.9490	0.9838	0.9891	0.9668	0.9806	0.9916	**0.9979**
Botnet	0.7315	0.7924	0.8568	0.7746	0.8217	0.9820	**0.9996**
BruteForce	0.8126	0.9157	0.9304	0.8742	0.9073	0.9864	**0.9991**
DDoS	0.9888	0.9927	0.9964	0.9698	0.9855	0.9990	**0.9999**
DoS	0.9274	0.9739	0.9862	0.9307	0.9681	0.9975	**0.9999**
Infiltration	0.9035	0.9516	0.9725	0.9189	0.9467	0.9800	**0.9978**
Macro Avg	0.8855	0.9350	0.9552	0.9058	0.9350	0.9894	**0.9989**
Weighted Avg	0.9454	0.9808	0.9878	0.9604	0.9773	0.9918	**0.9982**

由表8和表9可知，LSADA在CSE-CIC-IDS2018数据集上同样表现最优。在精确率方面，LSADA在Benign、Botnet、BruteForce、DDoS、DoS和Infiltration类别上分别达到0.9993、0.9999、0.9999、0.9999、0.9999和0.9711，整体误报控制能力优于对比方法。与其他模型相比，LSADA在多数攻击类别上均保持接近1.0000的精确率，说明该方法能够较准确地区分正常流量与不同类型攻击流量，减少不同类别之间的误判。

其中，Infiltration类别由于攻击行为隐蔽通常更难识别，但LSADA仍取得0.9718的F1分数。该结果表明，LSADA不仅在高频攻击类别上具有较强检测能力，也能够在复杂、边界模糊的攻击类型识别中保持较好的鲁棒性。

综合表6至表9的结果，LSADA在两个公开数据集上均取得稳定优势。其原因在于，LSADA通过日志语义重构将端口、协议等字段组织到统一上下文中，增强多字段关联关系的表达能力；同时，领域自适应预训练进一步缩小SecureBERT与网络流量日志分布之间的差异，使模型能够更充分地学习日志字段共现关系和攻击行为语义模式。

3.5 极端不平衡场景下的检测鲁棒性分析

为模拟真实网络中恶意流量占比较低的检测场景，本文将Botnet、BruteForce、DDoS、DoS和Infiltration合并为恶意样本类别，并与正常样本构成二分类任务。训练集总规模固定为100000条，恶意样本占比分别设置为5%、1%、0.5%和0.1%，其余样本均为正常样本。恶意样本按照原始攻击类别进行分层抽样，以保证各类攻击行为均被保留。各模型采用相同的不平衡训练集、训练参数和固定测试集进行评估，不使用过采样、类别权重或阈值重标定等额外平衡策略。F1分数用于衡量模型对恶意流量的识别能力，FPR表示正常样本被错误判定为恶意样本的比例。表10给出了不同恶意样本占比下各模型的检测结果。

由表10可知，随着恶意样本占比持续降低，

表8 CSE-CIC-IDS2018数据集的精确率实验结果

攻击类型	HAE-HRL	SSA-BOA-CNN	A-NIDS	BERT	BERT-RL	SecureBERT	LSADA
Benign	0.9425	0.9532	0.9887	0.9740	0.8928	0.9987	**0.9993**
Bot / Botnet	0.9932	0.7724	0.9723	0.9996	**0.9999**	0.9956	**0.9999**
BruteForce	0.7287	0.9845	0.9189	0.8393	0.9016	0.9938	**0.9999**
DDoS	0.9976	0.9637	0.9962	0.7243	0.8867	0.9979	**0.9999**
DoS	0.9654	0.9925	0.9985	0.7815	0.6798	0.9961	**0.9999**
Infiltration	0.8076	0.4928	0.7568	0.4693	**0.9945**	0.9309	0.9711
Macro Avg	0.9058	0.8599	0.9386	0.7980	0.8926	0.9855	**0.9950**
Weighted Avg	0.9437	0.9420	0.9836	0.9229	0.8874	0.9968	**0.9989**

表 9　　CSE-CIC-IDS2018 数据集的 F1 分数实验结果

攻击类型	HAE-HRL	SSA-BOA-CNN	A-NIDS	BERT	BERT-RL	SecureBERT	LSADA
Benign	0.9426	0.9710	0.9906	0.9829	0.9392	0.9988	**0.9993**
Botnet	0.9894	0.8489	0.9546	0.9991	**0.9999**	0.9971	**0.9999**
BruteForce	0.7607	0.9786	0.9019	0.8861	0.9218	0.9953	**0.9999**
DDoS	0.9893	0.9683	0.9979	0.8094	0.9207	0.9983	**0.9999**
DoS	0.9207	0.9818	0.9991	0.6286	0.7330	0.9967	**0.9999**
Infiltration	0.5797	0.6250	0.7897	0.1675	0.9363	0.9574	**0.9718**
Macro Avg	0.8637	0.8956	0.9390	0.7456	0.9085	0.9906	**0.9951**
Weighted Avg	0.9363	0.9607	0.9850	0.9207	0.9326	0.9976	**0.9989**

各模型的 F1 分数均出现下降，表明极少恶意样本监督会削弱模型对攻击行为的识别能力。LSADA 在各比例下均取得最高 F1 分数，在 0.1% 极端不平衡条件下仍达到 0.9586，较 A-NIDS[14]、BERT-RL[25]和 SSA-BOA-CNN[13]分别提高 0.1423、0.1805 和 0.2094。

在误报控制方面，LSADA 的 FPR 始终不高于 0.08%，明显低于其他对比方法。随着恶意样本比例降低，部分模型的 FPR 虽然有所下降，但其 F1 分数同步大幅降低，说明模型更倾向于将样本预测为 Benign，从而产生较多漏报。相比之下，LSADA 在维持较低误报率的同时仍保持较高 F1 分数，表明日志语义重构与领域自适应预训练能够提升极端类别不平衡条件下的恶意流量检测稳定性。

3.6　跨数据集迁移实验

3.6.1　少样本跨数据集迁移实验

为进一步验证本文方法在跨数据分布条件下的泛化能力，本文设计跨数据集迁移实验，对模型在目标域标注数据受限条件下的适应能力进行评估。与常规同域训练测试不同，本文在目标域中仅选取少量已标注的正常流量与攻击流量样本对模型进行快速适配，以模拟实际网络安全场景中目标域标注数据获取成本高、人工标注困难且数据更新频繁的应用条件。具体实验设置为：首先，使用源域中正常流量和攻击流量各 1000 条完成任务微调。其次，从目标域中随机选取正常流量和攻击流量各 5 条进行适配。最后，使用目标域中正常流量和攻击流量各 5000 条构建固定测试集，对模型进行评估，并计算 Accuracy、Recall、Precision 与 F1 分数等指标。

在 CIC-IDS2017，CSE-CIC-IDS2018 与 UNSW-NB15 三个数据集上共设计以下四个实验：

（1）以 CSE-CIC-IDS2018 作为源域进行模型训练，在 CIC-IDS2017 上进行迁移测试；

（2）以 CIC-IDS2017 作为源域进行模型训练，在 CSE-CIC-IDS2018 上进行迁移测试

表 10　　极端类别不平衡场景下不同模型的二分类检测性能

恶意样本占比	指标	HAE-HRL	SSA-BOA-CNN	A-NIDS	BERT	BERT-RL	**LSADA**
5%	F1 分数	0.9138	0.9374	0.9628	0.8856	0.9481	**0.9962**
5%	FPR	0.0541	0.0116	0.0701	0.0086	0.0101	**0.0008**
1%	F1 分数	0.8564	0.8912	0.9326	0.8073	0.9051	**0.9914**
1%	FPR	0.0472	0.0098	0.0059	0.0072	0.0840	**0.0006**
0.5%	F1 分数	0.8042	0.8563	0.8977	0.7458	0.8642	**0.9847**
0.5%	FPR	0.0415	0.0081	0.0048	0.0060	0.0640	**0.0005**
0.1%	F1 分数	0.6685	0.7492	0.8163	0.5947	0.7781	**0.9586**
0.1%	FPR	0.0321	0.0055	0.0029	0.0039	0.0380	**0.0003**

（3）以 CSE-CIC-IDS2018 作为源域进行模型训练，在 UNSW-NB15 上进行迁移测试；

（4）以 UNSW-NB15 作为源域进行模型训练，在 CSE-CIC-IDS2018 上进行迁移测试。

在对比方法选择上，本文选取 HCNN-LSTM[15]、1dCNN-RF[16]、MMF-FSL[17] 和 Att-Siamese[18] 作为基线方法。为保证对比结果的公平性与可比性，所有方法采用相同的源域训练样本、目标域五样本适配集和目标域测试集，并保持评价指标与实验环境一致。各基线保留其原始模型结构和关键训练配置。实验结果分别如表 11 至表 14 所示。

表 11　基于 CIC-IDS2017 训练的 CSE-CIC-IDS2018 跨数据集迁移结果

方法	Accuracy	Precision	F1 分数
HCNN-LSTM	0.7348	0.7431	0.7302
1dCNN-RF	0.6682	0.6215	0.7215
MMF-FSL	0.7458	0.7325	0.7391
Att-Siamese	0.8422	0.9107	0.8279
LSADA	**0.9141**	**0.9966**	**0.9135**

由表 11 可知，在以 CIC-IDS2017 数据集进行训练、CSE-CIC-IDS2018 数据集进行迁移测试的实验中，本文方法取得最优性能，其 Accuracy 为 0.9141，Recall 为 0.8310，Precision 为 0.9966，F1 分数为 0.9135。与性能较优的 Att-Siamese 相比，Accuracy 提升 0.0719，F1 分数提升 0.0856。该结果表明，本文方法在数据分布发生变化时，仍能够保持较高检测性能。

表 12　基于 CSE-CIC-IDS2018 训练的 CIC-IDS2017 跨数据集迁移结果

方法	Accuracy	Precision	F1 分数
HCNN-LSTM	0.5884	0.5614	0.6624
1dCNN-RF	0.8440	0.8707	0.8382
MMF-FSL	0.7458	0.7325	0.7391
Att-Siamese	0.8919	0.9215	0.8885
LSADA	**0.9413**	**0.9937**	**0.9411**

由表 12 可知，在以 CSE-CIC-IDS2018 数据集进行训练、CIC-IDS2017 数据集进行迁移测试的实验中，本文方法同样取得最优结果，与 Att-Siamese 相比，Accuracy 提升 0.0494，F1 分数提升 0.0526。该结果说明，在源域与目标域互换后，本文方法仍具有较好的稳定性与泛化能力。

由表 13 可知，在以 CSE-CIC-IDS2018 数据集进行训练、UNSW-NB15 数据集进行迁移测试的实验中，本文方法取得最优性能，其 Accuracy 为 0.9835，Recall 为 0.9718，Precision 为 0.9951，F1 分数为 0.9833，显著优于各对比方法。与性能较优的 Att-Siamese 相比，Accuracy 提升 0.1220，F1 分数提升 0.1179。说明本文方法在面对不同采集环境与特征体系的数据集时，仍能够保持优异的迁移检测能力。

表 13　基于 CSE-CIC-IDS2018 训练的 UNSW-NB15 跨数据集迁移结果

方法	Accuracy	Precision	F1 分数
HCNN-LSTM	0.6124	0.5483	0.6445
1dCNN-RF	0.7018	0.6485	0.7261
MMF-FSL	0.7843	0.7765	0.7872
Att-Siamese	0.8615	0.8894	0.8654
LSADA	**0.9835**	**0.9951**	**0.9833**

表 14　基于 UNSW-NB15 训练的 CSE-CIC-IDS2018 跨数据集迁移结果

方法	Accuracy	Precision	F1 分数
HCNN-LSTM	0.5746	0.5239	0.6192
1dCNN-RF	0.6835	0.6327	0.7071
MMF-FSL	0.7918	0.7763	0.7997
Att-Siamese	0.8427	0.8725	0.8490
LSADA	**0.9146**	**0.9518**	**0.9124**

由表 14 可知，在以 UNSW-NB15 数据集进行训练、CSE-CIC-IDS2018 数据集进行迁移测试的实验中，本文方法仍取得最佳结果。与 Att-Siamese 相比，Accuracy 提升 0.0719，F1 分数提升 0.0634。该结果进一步验证本文方法在异构数据源之间具有稳定的知识迁移能力。

综合四组实验结果可以看出，本文方法在所有跨数据场景下均取得最优性能，总体而言，本文方法在跨数据迁移条件下兼具较高分类准确性与攻击检测覆盖率。

产生上述结果的主要原因如下：（1）通过日志语义重构，将不同数据集中的结构化流量特征映射

为统一的类自然语言表示，缓解异构特征空间带来的分布差异；（2）通过领域自适应预训练，使模型能够学习日志中的字段共现关系、结构模式与行为语义规律，从而增强对未知数据分布的适应能力；（3）预训练语言模型具备较强的上下文建模能力，使模型能够在有限监督信息条件下快速完成目标域适配。综合来看，本文提出的LSADA方法在复杂跨数据环境下具有较好的泛化能力与实际应用价值。

3.6.2　目标域无标注样本跨数据集迁移实验

为评估LSADA在无法获得目标域标注样本、但能够获取目标域无标签流量日志条件下的跨数据集适配能力，本节以CIC-IDS2017为源域、UNSW-NB15为目标域开展二分类迁移实验。该实验属于目标域无标注样本条件下的领域适配，不属于完全不接触目标域数据的纯零样本迁移，也不涉及未知攻击类别拒识的开放集识别。从CIC-IDS2017训练划分中随机抽取正常样本与攻击样本各1000条，共2000条源域标注样本，用于各模型的监督训练。其中，攻击流量由源域中的各类攻击日志流量合并构成。目标域UNSW-NB15不提供任何标注样本参与模型训练、分类器微调、参数选择或阈值调整。

在实验流程上，首先使用源域训练集完成模型预训练与下游任务学习，获得具有初始判别能力的检测模型；随后在目标域中构建统一迁移任务。对于HCNN-LSTM[15]、1dCNN-RF[16]、MMF-FSL[17]和Att-Siamese[18]，均仅使用相同的2000条CIC-IDS2017源域标注样本完成训练后，直接迁移至UNSW-NB15进行检测，不接触目标域训练划分中的任何数据。

在目标域测试阶段，从UNSW-NB15测试划分中随机抽取正常与攻击样本各5000条，共10000条样本构建固定测试集。目标域测试标签仅用于最终的Accuracy、Precision、F1分数、FPR和FNR统计。为保证实验公平性，所有方法均采用相同的源域训练样本、目标域测试样本、评价指标和实验环境，实验结果如表15所示。

由表15可知，在目标域不提供标注样本、仅使用无标签日志完成领域适配的条件下，LSADA取得0.8750的Accuracy、0.9044的Precision和0.8703的F1分数，三项指标均优于对比方法。相较于综合性能最优的基线Att-Siamese，LSADA的Accuracy、Precision和F1分数分别提高0.1238、0.1263和0.1318，表明LSADA能够利用目标域无标签日志中的字段共现关系和连接行为模式缓解域间分布偏移，并在不使用目标域标注样本的条件下保持较稳定的攻击判别能力。

进一步分析误报与漏报指标可知，LSADA的FPR仅为0.0886，较Att-Siamese降低0.1118；其FNR为0.1614，较FNR最低的基线MMF-FSL进一步降低0.0640。LSADA在FPR和FNR上同时取得最优结果，说明其并非通过偏向预测某一类别获得较高检测性能，而是在攻击识别、正常流量判别和跨域稳定性之间实现了更优平衡。

上述结果表明，LSADA能够在不使用目标域标注样本的条件下，利用目标域无标签日志学习字段共现关系、流量统计分布和连接行为模式，从而减弱源域与目标域之间的表征偏移，并在攻击识别与误报控制之间取得较好平衡。

需要说明的是，本实验使用了目标域无标签日志开展领域适配，因此不属于完全不接触目标域数据的纯零样本迁移。同时，源域与目标域均被统一为正常与攻击二分类任务，模型不包含未知类别拒识机制，因此该实验也不属于严格意义上的开放集识别。后续将结合不确定性估计、异常分数建模和开放集识别方法，进一步研究全新攻击类别及攻击变体的发现能力。

表15　目标域无标注样本条件下的CIC-IDS2017→UNSW-NB15迁移结果

方法	Accuracy	Precision	F1分数	FPR	FNR
HCNN-LSTM	0.6232	0.6283	0.6155	0.3568	0.3968
1dCNN-RF	0.7174	0.7435	0.7014	0.2290	0.3362
MMF-FSL	0.6898	0.6623	0.7140	0.3950	0.2254
Att-Siamese	0.7512	0.7781	0.7385	0.2004	0.2972
LSADA	**0.8750**	**0.9044**	**0.8703**	**0.0886**	**0.1614**

3.7　面向结构化日志的掩码机制对比与结果分析

为验证不同掩码策略对日志领域自适应预训练的影响，本文构建 LSADA-SM、LSADA-DM 和 LSADA 三组对照模型。LSADA-SM 采用静态掩码策略；LSADA-DM 采用动态随机掩码策略；LSADA 在动态掩码基础上引入日志语义结构约束，将掩码预算优先分配给协议、端口等关键行为字段，并以完整字段片段作为掩码单位，避免子词残留形成预测捷径；同时，依据不同字段组的词元频率对 MLM 损失进行平衡加权，以增强低频关键行为字段对预训练优化的贡献。

三组模型均以 SecureBERT 为初始模型，使用相同的日志语义模板、领域预训练语料和训练参数，仅改变掩码策略及损失函数。为比较预训练收敛过程，使用固定验证掩码集记录不同模型的 MLM 损失；随后，将各预训练模型用于 CSE-CIC-IDS2018 六分类任务微调，并在相同测试集上评估下游检测性能。

图 4 给出了三组模型在固定验证掩码集上的 MLM 损失收敛曲线。LSADA-SM 在训练初期下降较快，但后期收敛趋于平缓；LSADA-DM 在后期取得更低的验证损失；LSADA 最终达到最低验证 MLM 损失 1.205，较 LSADA-DM 降低 0.075，较 LSADA-SM 降低 0.114。

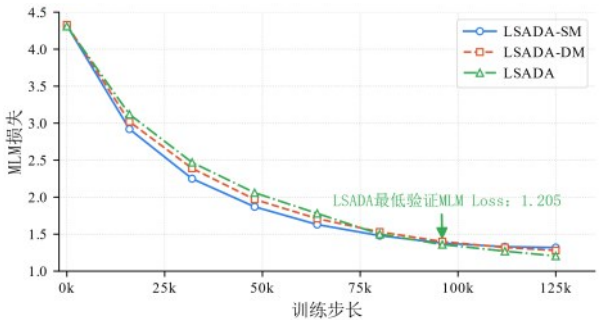


图 4　不同掩码策略在固定验证掩码集上的 MLM 损失收敛曲线

表 16 给出了三组模型的下游六分类检测性能。由表 16 可知，LSADA-DM 相较 LSADA-SM 在 Macro-F1 上提高 0.0037。LSADA 进一步取得最优性能，Macro-F1 达到 0.9951，较 LSADA-DM 提高 0.0027，较 LSADA-SM 提高 0.0064。结果表明，优先预测行为字段，并平衡不同字段组的损失贡献，有助于增强模型对复杂网络行为和低频攻击类别的语义表征能力。

表 16　不同掩码策略预训练模型六分类检测性能对比

预训练方法	Accuracy	Precision	Macro-F1
LSADA-SM	0.9974	0.9939	0.9887
LSADA-DM	0.9983	0.9946	0.9924
LSADA	**0.9989**	**0.9950**	**0.9951**

3.8　超参数实验与结果分析

为验证 LSADA 超参数设置的合理性，本文采用控制变量法考察 DAPT 掩码比例与下游微调学习率对模型检测性能的影响。本节超参数敏感性实验均在 CSE-CIC-IDS2018 数据集的六分类任务上开展，并以 Macro-F1 作为评价指标。实验中除待考察参数外，其余配置均与主实验保持一致。

如图 5 所示，随着 DAPT 掩码比例由 5% 增至 15%，模型 Macro-F1 由 98.68% 提升至 99.51%；当掩码比例继续提高时，整体性能逐步下降。其中，25% 至 30% 区间出现轻微波动，但不改变总体下降趋势。该结果说明，较低掩码比例下模型难以充分学习端口、传输速率、流间隔和 TCP 标志等字段之间的条件依赖；而较高掩码比例会削弱日志上下文的完整性，增加掩码恢复难度。综合来看，10%～20% 为较优掩码比例区间，15% 时取得最优 Macro-F1，与主实验配置一致。

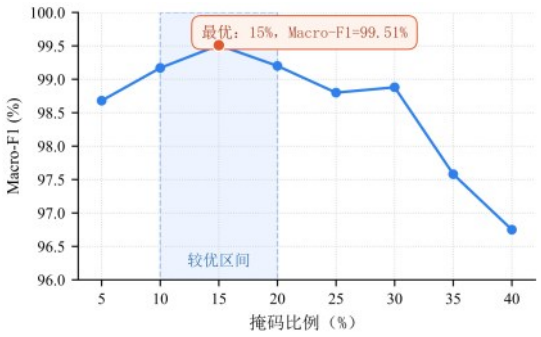


图 5　不同 DAPT 掩码比例下六分类任务的 Macro-F1

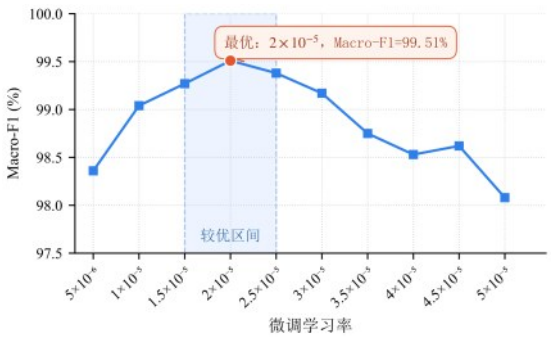


图 6　不同微调学习率下六分类任务的 Macro-F1

如图 6 所示，微调学习率由 5×10^{-6} 增加至 2×10^{-5} 时，Macro-F1 由 98.36% 提升至 99.51%；当学习率继续增大时，检测性能整体下降，5×10^{-5} 时降至 98.08%。在 4×10^{-5} 至 4.5×10^{-5} 之间虽存在局部波动，但模型性能已明显低于最优配置。较低学习率会限制预训练语义表示向下游判别任务的有效迁移；过高学习率则易造成参数更新过大，扰动 DAPT 阶段学习到的字段关联表示，降低优化稳定性。因此，1.5×10^{-5}～2.5×10^{-5} 为较优学习率范围，本文最终选取 2×10^{-5} 作为微调学习率。

总体而言，掩码比例 15% 与微调学习率 2×10^{-5} 能够在日志语义信息保留、掩码预测监督强度和下游优化稳定性之间取得较优平衡，其对应的 Macro-F1 为 99.51%。

3.9　日志语义建模策略的性能对比与迁移验证

为验证不同日志语义建模方式对网络流量行为表征的影响，本文选取 LogSwift 式语义表达、LogGPT 式语义表达和 LSADA 模块化语义表达进行对比。需要说明的是，LogSwift[21] 和 LogGPT[23] 原本主要面向系统运行日志，其原始数据对象与本文使用的网络流量记录存在差异。因此，本文并未直接将二者的完整系统日志分析流程应用于 CSE-CIC-IDS2018 数据集，而是借鉴其日志文本组织思想，构建相应的流量日志语义建模对照方案。

本文从 CSE-CIC-IDS2018 数据集中选取时间戳、协议、目标端口等能够反映网络连接行为的字段，并将其转换为流级日志语义文本。由此，原始流量数据不再以相互独立的数值特征形式输入模型，而是被组织为描述一次网络连接行为的日志语义序列。

在统一字段集合的基础上，LogSwift 式语义表达采用紧凑的“字段名称—字段取值”形式依次组织流量字段；LogGPT 式语义表达采用提示化自然语言描述，将不同流量字段转换为连续文本。

三种方案使用完全相同的原始流量样本、字段集合、字段取值、数据划分、SecureBERT 初始化模型、领域自适应预训练参数和下游分类器，仅改变结构化流量字段的日志语义组织方式。因此，本节实验用于比较不同日志语义建模策略对流量行为表示、攻击类别判别及跨数据集迁移能力的影响，而不是比较 LogSwift 和 LogGPT 在其原始系统日志任务中的检测性能。

首先，在 CSE-CIC-IDS2018 数据集上开展六分类检测实验，结果如表 17 所示。

表 17　不同日志转换方案的同域六分类检测性能对比

转换方案	Accuracy	Precision	Macro-F1
LogSwift	0.9967	0.9861	0.9851
LogGPT	0.9975	0.9896	0.9887
LSADA	**0.9989**	**0.9950**	**0.9951**

由表 17 可知，LSADA 模块化语义表达在 Accuracy、Precision 和 Macro-F1 三项指标上均取得最优结果。需要指出的是，三种方案使用的是相同的结构化流量字段，因此性能差异主要来源于日志语义组织方式。LogSwift 式语义表达能够保留字段名称和取值，但字段关系主要依赖固定排列位置，难以显式表示前后向报文之间的方向关系以及速率、流间隔和 TCP 状态之间的关联。LogGPT 式语义表达通过自然语言描述增强了字段上下文，但其文本组织方式仍以通用提示为主，缺少对网络连接行为结构的针对性建模。相比之下，LSADA 的组织方式有助于 SecureBERT 学习不同流量字段之间的联合变化关系，因此在同域六分类任务中表现出更好的攻击类别判别能力。

本文在 CSE-CIC-IDS2018→UNSW-NB15 迁移任务中比较三种转换方案的跨域泛化性能。目标域无标注适配设置不使用目标域标注样本，仅利用目标域无标签流量日志开展领域适配。五样本实验设置中，UNSW-NB15 目标域的 Benign 和 Attack 类别各随机选取 5 条标注样本进行适配，共使用 10 条目标域标注日志。结果如表 18 所示。

表 18　不同日志转换方案在迁移任务中检测性能对比

转换方案	目标域无标注 Accuracy	目标域无标注 F1 分数	五样本 Accuracy	五样本 F1 分数
LogSwift	0.7584	0.7263	0.9426	0.9401
LogGPT	0.8017	0.7846	0.9618	0.9597
LSADA	**0.8690**	**0.8571**	**0.9835**	**0.9833**

由表 18 可知，LSADA 在目标域无标注适配和五样本适配条件下均取得最优结果。在目标域无标注适配条件下，LSADA 的 F1 分数为 0.8571，较 LogSwift 式语义表达和 LogGPT 式语义表达分别提高 0.1308 和 0.0725；在五样本适配条件下，LSADA 的 F1 分数达到 0.9833，分别提高 0.0432 和 0.0236。

LSADA 模板的优势来源于对多维字段的联合语义建模。TCP 标志反映连接建立、确认、终止和异常复位等会话状态；流间隔刻画通信节奏、突发性和持续性；速率与包数、字节数共同描述单位时间内的传输强度。孤立字段的取值容易受到业务类型、网络负载和服务场景变化影响，而将 TCP 状态、流间隔、速率、双向流量和包长分布置于统一语义上下文中，能够使模型从连接状态、时间行为和传输强度的协同变化中识别攻击模式。因此，LSADA 在同域多分类检测和跨数据集迁移任务中均表现出更好的稳定性和泛化能力。

3.10 消融实验

本文从日志语义模板设计与整体方法结构两个层面开展消融实验。

3.10.1 日志模板设计消融实验

为进一步验证本文所提出日志语义模板的有效性，并系统分析不同语义子模块对网络入侵检测性能的贡献，本文开展日志模板设计消融实验。在保持模型结构、训练参数、数据划分方式及训练流程一致的条件下，仅调整输入模板中所包含的语义模块，通过比较不同模板变体下模型性能变化，评估各类特征信息在语义表示学习过程中的作用。

本文设置 6 种日志模板变体模型（Log Template Variants, LTV）进行对比实验，具体定义如下：

（1）LTV-Full：表示完整日志模板，包含 Prefix、PacketStat、RateStat、LengthStat、IATStat 与 FlagStat 六个语义模块，用于作为整体性能的基准参考。

（2）LTV\T 表示在完整模板基础上去除时间相关模块，用于分析时序信息对模型性能的贡献。

（3）LTV\P：表示去除双向统计模块，用于评估流量交互关系信息的重要性。

（4）LTV\R：表示去除速率统计模块，用于验证传输速率信息在攻击检测中的作用。

（5）LTV\L：表示去除包长分布模块，用于分析数据包长度统计信息的影响。

（6）LTV\F：表示去除协议状态模块，用于评估协议层状态信息对模型性能的贡献。

图 7 和图 8 中的实验均在 CSE-CIC-IDS2018 六分类任务上开展，使用相同的训练集、验证集和测试集划分。各模板变体均以 SecureBERT 为基础模型，采用一致的日志预训练语料、动态掩码策略、下游微调参数及评价流程；实验中仅移除对应语义模块及其日志表达，不改变其余字段、模型结构和训练配置，从而保证不同模板变体之间的比较公平。实验结果如图 7 和图 8 所示。图 8 横轴表示六类流量类别，纵轴表示不同日志模板变体，单元格数值为类别级 F1 分数；颜色由浅至深表示 F1 分数由低至高。

总体来看，采用完整日志模板时模型性能达到最优，准确率达到 0.9989，Macro-F1 为 0.9951，明显高于其他模板变体。图 7 进一步显示，每移除一个语义模块，准确率与 Macro-F1 都会出现不同程度的下滑，这说明通过统一的语义模板对时序信息、统计特征、交互关系以及协议状态等多维信息进行联合建模，能够有效强化模型对复杂网络行为模式的表达。

就各语义子模块的贡献而言，速率统计信息带来的影响最为突出。去掉速率模块后，准确率下降至 0.9784，Macro-F1 降至 0.9187，在所有消融设置中表现最差；与此对应，DDoS 和 DoS 类别的 F1 分数分别滑落到 0.8614 和 0.8927，表明单位时间内的数据传输强度能够有力刻画洪泛类攻击的特征，是识别拒绝服务攻击的依据。

双向统计信息同样起着重要作用。去除前向与后向流量统计特征后，Macro-F1 降到 0.9324，Botnet、DDoS 和 DoS 类别的 F1 分数分别降至 0.9017、0.9185 和 0.9441，反映出通信规模及双向交互关系能够有效描述异常连接行为，对僵尸网络通信和流量型攻击的识别意义显著。

时间相关信息主要作用于隐蔽型攻击。移除时间模块后，Macro-F1 降低到 0.9486，其中 Infiltration 类别的 F1 分数从 0.9718 骤降至 0.8412，这揭示出连接持续时间、请求间隔和交互节奏等时序特征能够反映攻击行为的潜伏性与渐进性，对隐蔽攻击检测具有不可忽视的价值。

包长分布与协议状态信息提供辅助判别信息。去除包长模块后，BruteForce 和 Infiltration 类别的 F1 分数分别降至 0.8814 和 0.9026；去掉 TCPFlags 模块后，准确率和 Macro-F1 分别降至 0.9838 和 0.9395，其中 DDoS 类别的 F1 分数降至 0.9016。以上结果表明数据包长度模式与连接状态变化可以为复杂攻击行为的识别补充有效线索。

3.10.2　组件消融实验

本文方法包含多个关键组成部分，为此通过消融实验来检验各模块对入侵检测性能的贡献。针对本文方法的设计，设置四种消融模型，并与完整模型 LSADA 进行对比分析，具体如下。

（1）LSADA\LD：作为基线，既不使用日志模板重构，也不引入领域自适应预训练，直接将原始网络流量数据输入 SecureBERT 进行训练，用于考

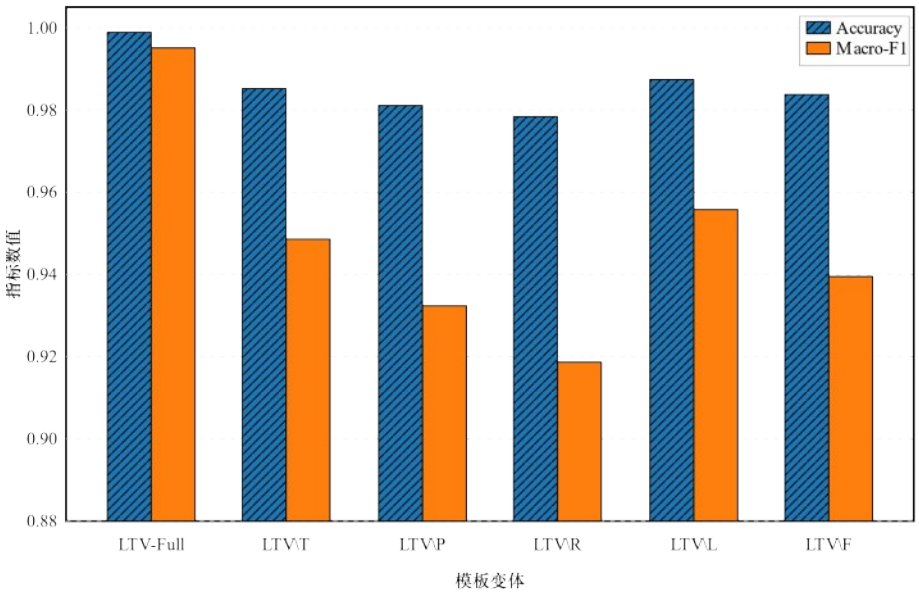


图 7　CSE-CIC-IDS2018 六分类任务中不同日志模板变体的总体检测性能对比

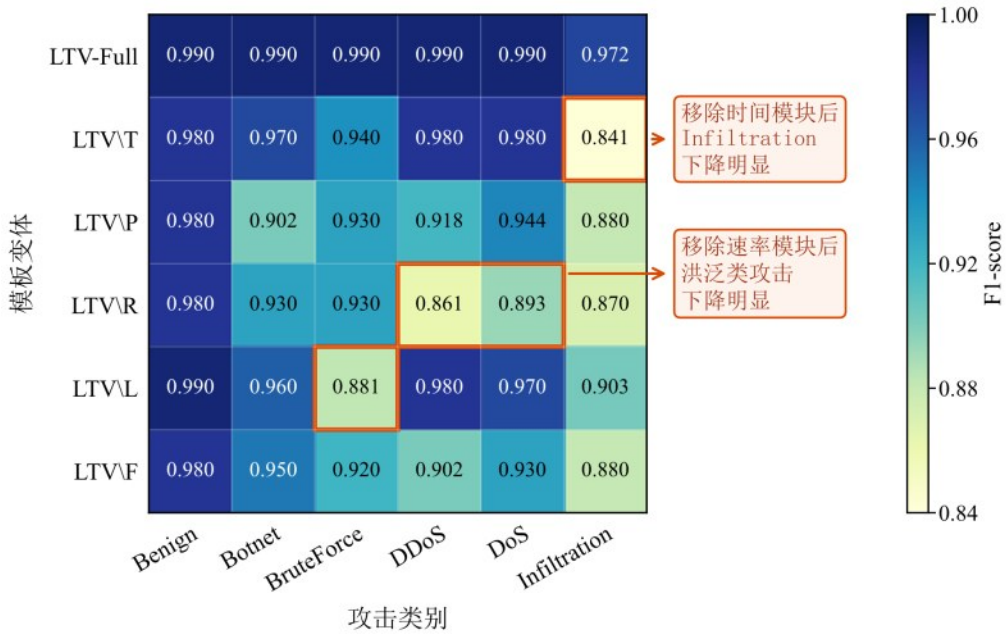


图 8　CSE-CIC-IDS2018 测试集中不同日志模板变体的类别级 F1 分数热力图

察无语义建模且无领域适配条件下模型的基础检测能力。

（2）LSADA\D：为验证 SecureBERT 在流级网络日志场景下进一步开展领域自适应预训练的必要性，本文设置 LSADA\D 作为无 DAPT 对照模型。该模型保留模块化日志语义模板，并以 SecureBERT 作为初始编码器，仅利用带标签训练样本进行下游微调，不接触无标签流量日志。相较于完整 LSADA，LSADA\D 仅移除了日志领域 DAPT 模块，二者其余输入形式、模型结构、数据划分和训练参数均保持一致。因此，LSADA\D 与完整 LSADA 的性能差异能够反映 SecureBERT 已有网络安全文本知识与结构化流级日志语义之间的适配差距，以及 DAPT 对学习字段位置、字段共现和行为组合关系的实际作用。

（3）LSADA\L：在 LSADA\LD 的基础上只引入领域自适应预训练，不依赖日志模板，利用原始数据对 SecureBERT 进行 DAPT 后再完成下游微调，用于检验领域自适应预训练对语义表示能力的提升作用。

（4）LSADA-ATG：在完整 LSADA 的基础上，以自适应文本生成策略替换固定模块化日志模板。该模型根据字段取值范围、预设字段重要性或样本局部特征动态选择字段描述短语和文本表达顺序，其余设置均与 LSADA 保持一致，包括输入字段集合、SecureBERT 初始化参数、DAPT 语料、动态掩码策略、字段组平衡 MLM 损失及下游分类训练过程。该对照模型用于验证固定语义模板相较于自适应文本表达的合理性。

完整模型 LSADA 则同时结合日志模板重构与领域自适应预训练，先在模板化日志语料上开展 DAPT，再面向入侵检测任务进行微调，从而验证两种策略协同作用下的整体性能提升。

基于 CIC-IDS2017 和 CSE-CIC-IDS2018 数据集的消融实验结果分别如图 9 和图 10 所示。从中可以归纳出以下结论。

（1）日志模板重构的影响：通过对比 LSADA 与 LSADA\L 的检测结果可以看出。在两个数据集上，LSADA 的准确率、召回率、精确率和 F1 分数均优于 LSADA\L，说明仅仅依靠领域自适应预训练而缺少统一的语义输入形式时，模型对流量字段间上下文关系的利用依然不够充分。日志模板重构能够将离散异构的结构化特征组织成具有统一语义结构的文本序列，进而增强模型对网络行为模式的表征能力，并带来检测性能的显著提升。

（2）领域自适应预训练的影响：从 LSADA 与 LSADA\D 的实验结果对比中可以发现。LSADA 在两个数据集上的精确率和 F1 分数均高于 LSADA\D，表明在完成日志模板重构之后，进一步引入领域自适应预训练能够有效增强模型对网络日志领域词元分布、字段共现模式及结构依赖关系的建模能力，从而改善对异常流量的判别表现。这反映出 DAPT 在本文框架中充当一种有效的领域适配机制，对提升模型的鲁棒性与稳定性起到积极作用。

（3）日志语义建模与领域适配协同作用的影响：通过对比 LSADA\LD、LSADA\D、LSADA\L 与完整模型 LSADA 的整体检测结果可以看到，LSADA 在两个数据集上的各项指标均达到了最优水平。

这一结果表明，日志模板重构与领域自适应预训练在统一的语义输入空间下形成了明显的协同增益。一方面，日志模板重构为预训练语言模型提供一种更契合语言建模机制的输入形态；另一方面，DAPT 则推动模型进一步拟合网络日志领域的语义分布特征，两者合力提升对复杂攻击行为的识别能力。

（4）特征重要性加权固定语义表达的影响：由图 11 可知，LSADA 在 CIC-IDS2017 和 CSE-CIC-IDS2018 两个数据集上的各项检测指标均优于 LSADA-ATG。该结果表明，基于全局特征重要性确定模块内字段优先级，并保持字段模块、相对位置和描述边界稳定的表达方式，能够使协议、端口、速率、流间隔和 TCP 状态等关键行为字段形成更清晰的语义锚点。由此，模型可在不同流量样本间学习稳定的字段关联模式，减少文本表述变化对领域自适应预训练的干扰，从而提升日志语义表示的判别能力与下游检测性能。

综合上述结果，日志模板重构与领域自适应预训练对入侵检测性能的提升均发挥着积极作用。其中，日志模板强化模型对多字段关联关系和网络行为语义模式的刻画，领域自适应预训练则进一步改善模型对网络流量日志分布的适配水平。实验结果既验证各组成模块设计的合理性，也说明将日志语

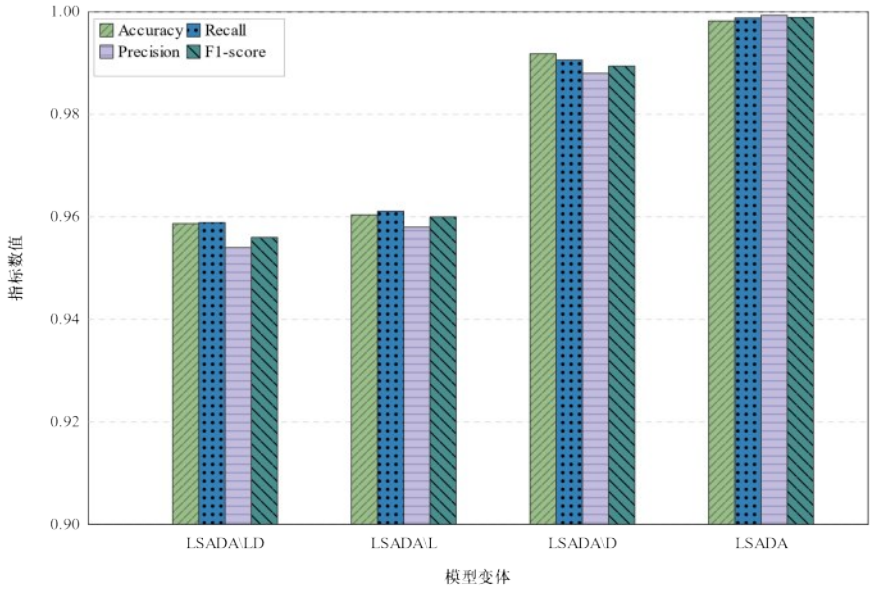


图9　基于CIC-IDS2017数据集的消融实验结果

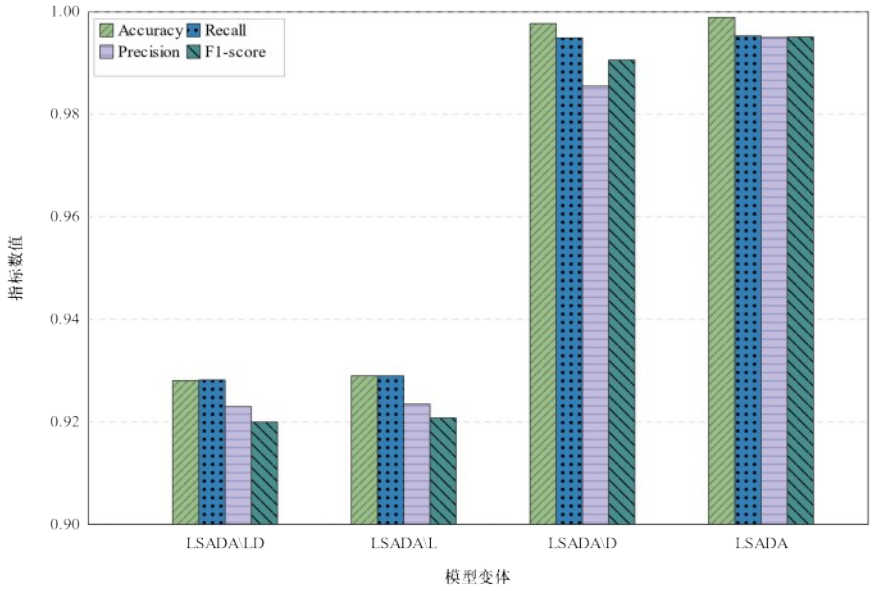


图10　基于CSE-CIC-IDS2018数据集的消融实验结果

义分析与领域自适应预训练进行联合建模，能够有效增强网络入侵检测的性能。

3.10.3　日志语义表征的可视化与注意力关联分析

为直观分析日志语义重构模块L和领域自适应预训练模块D对表征学习的影响，本文分别对LSADA\LD、LSADA\D、LSADA\L和完整模型LSADA的样本向量进行可视化。表征可视化实验使用CSE-CIC-IDS2018测试集，包含Benign、Bot-net、BruteForce、DDoS、DoS和Infiltration六类流量。为保证不同方案的可比性，从每类测试样本中随机抽取300条，共1800条样本，对LSADA\LD，直接使用标准化后的原始数值特征；对其余三种方案，提取下游分类模型最后一层的[CLS]向量，并采用t-SNE将高维特征映射至二维空间。图12展示了四种方案的样本向量分布。

由图12可知，LSADA\LD的类别样本存在较

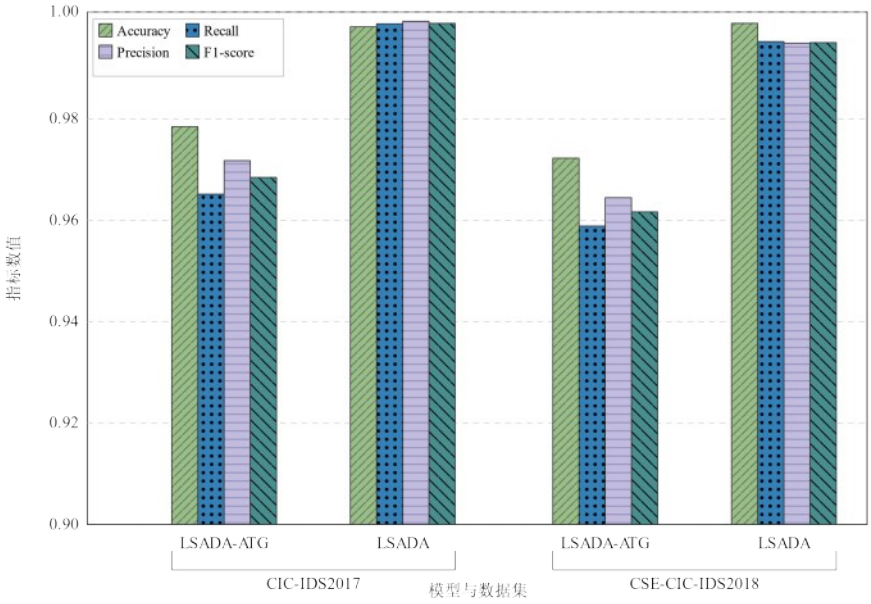


图 11　固定模块化日志模板与自适应文本生成策略的检测性能对比

明显重叠。引入日志语义重构模块后，LSADA\D 的类内聚集程度有所提升；仅采用领域自适应预训练的 LSADA\L 也能够改善部分攻击类别的表征分布。完整模型 LSADA 的同类样本分布更加紧凑，不同类别之间的边界更加清晰，说明日志语义重构与领域自适应预训练具有互补作用，能够共同提升网络流量表征的类别可分性。

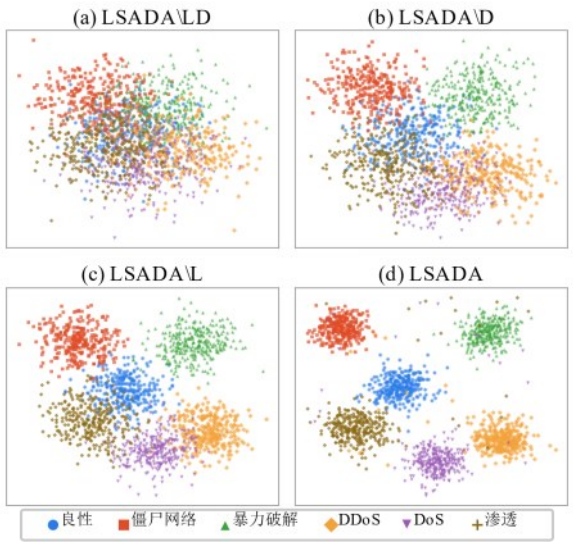


图 12　不同模块配置下网络流量语义表征空间的 t-SNE 可视化

为分析模型对关键攻击字段的关注方式，本文在 LSADA 模型上开展字段组自注意力分析。从 DDoS、BruteForce 和 Infiltration 三类测试样本中随机选取 100 条，提取模型最后 4 层全部注意力头的平均权重，并按照六个语义模块进行聚合。图 13 中，纵轴表示 Query 字段组，横轴表示 Key 字段组，颜色越深表示字段组之间的平均注意力权重越高。

由图 13 可知，不同攻击类型呈现出差异化的字段关联模式。DDoS 样本的高权重主要集中于 RateStat 与 IATStat、FlagStat 之间，体现传输速率、通信间隔和 TCP 状态的联合变化；BruteForce 样本中 Prefix 与 PacketStat、FlagStat 关联较强，反映目标端口、重复连接和会话状态的

组合特征；Infiltration 样本则主要关注 IATStat 与 LengthStat、PacketStat、FlagStat 之间的关系，体现隐蔽通信中的时间节奏、包长变化和连接状态。结果表明，LSADA 能够根据不同攻击行为关注相应的字段组合关系。

3.10.4　跨数据集表征分布可视化分析

为进一步验证领域自适应预训练对跨数据集表征分布对齐的作用，本文基于 CSE-CIC-IDS2018→UNSW-NB15 少样本迁移任务，对 LSADA\D 与完整 LSADA 的源域—目标域联合表征进行可视化分析。LSADA\D 保留日志语义重构模块，但移除领域自适应预训练；完整 LSADA 在相同日志语义模板、SecureBERT 初始化参数、源域训练样本和下

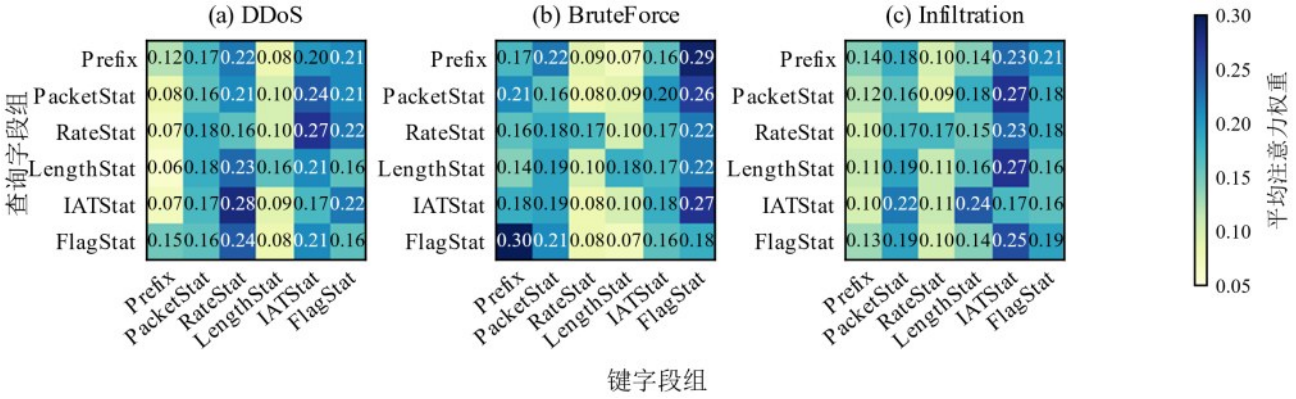


图 13　LSADA 对典型攻击流量的字段组间自注意力权重矩阵

游微调配置基础上，引入面向日志语义结构的 DAPT。两种模型均使用 CSE-CIC-IDS2018 中的正常和攻击样本各 1000 条进行源域监督训练，并使用 UNSW-NB15 中的正常和攻击样本各 5 条进行目标域少样本适配。

为保证可视化比较的一致性，从 CSE-CIC-IDS2018 源域测试集和 UNSW-NB15 目标域测试集中分别随机抽取正常和攻击样本各 500 条，共计 2000 条样本。对于 LSADA\D 和 LSADA，均提取分类层之前最后一层的[CLS]向量作为流量日志的全局语义表征，并将源域与目标域样本的表征联合输入 t-SNE 映射至二维空间。图中颜色表示真实类别，圆点和三角形分别表示源域和目标域样本。目标域测试标签仅用于事后可视化及性能评价，不参与模型训练、参数选择或阈值调整。

由图 14(a)可知，LSADA\D 中源域与目标域同类样本仍存在明显的表征偏移。尽管源域 Benign 和 Attack 样本能够形成相对集中的局部区域，但目标域样本整体分布更为离散，部分目标域攻击样本位于正常和攻击的边界区域，说明仅依赖日志语义重构和少量目标域标注样本，难以充分缓解不同数据集在流量采集环境、字段取值范围和攻击行为模式方面带来的分布差异。

由图 14(b)可知，完整 LSADA 在引入 DAPT 后，源域与目标域同类样本的主分布区域更加接近，目标域正常样本与源域正常样本之间的间隔有所减小，目标域攻击样本也更加集中于攻击语义区域。同时，正常和攻击样本之间的整体边界更加清

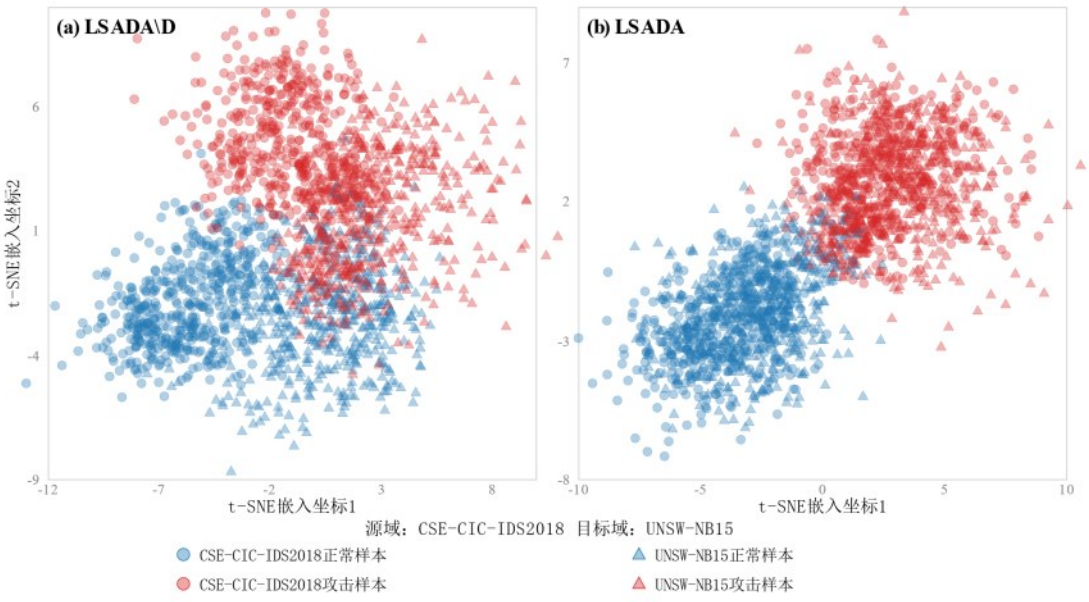


图 14　LSADA\D 与 LSADA 的 CSE-CIC-IDS2018→UNSW-NB15 联合表征分布 t-SNE 可视化

晰，但仍存在少量位于类别边界附近的目标域样本，表明DAPT能够有效减弱跨数据集表征偏移，但不能完全消除不同网络环境和攻击模式带来的域间差异。

综合图14可知，在相同源域监督训练和目标域五样本适配条件下，DAPT能够进一步利用日志中的协议、端口、速率、流间隔和TCP状态等字段联合语义，减弱源域与目标域同类流量之间的表征偏移，并保持正常流量与攻击流量之间的判别结构。该结果从表征空间层面说明了领域自适应预训练对LSADA跨数据集迁移能力的促进作用。

3.11　模型复杂度、训练开销与推理效率分析

为评估LSADA在实际部署中的资源开销，本文从在线推理复杂度、领域自适应预训练开销以及不同检测场景下的推理效率三个方面开展分析。首先，在统一硬件环境下对HAE-HRL[12]、SSA-BOA-CNN[13]、A-NIDS[14]、BERT[24]、BERT-RL[25]和LSADA的总参数量、单样本平均推理时延及峰值GPU显存进行统计，结果如表19所示。

由表19可知，数值特征深度模型的参数量、推理时延和显存占用整体较低，适用于资源受限的检测场景。HAE-HRL、SSA-BOA-CNN和A-NIDS直接以结构化数值流量特征作为输入；BERT、BERT-RL和LSADA则将流量记录转换为文本序列，并通过预训练语言模型进行语义编码。BERT、BERT-RL和LSADA均采用预训练语言模型编码器，模型参数量处于亿级，推理开销相对较高。LSADA虽然增加了日志语义重构与DAPT过程，但DAPT属于离线训练环节，未在部署阶段增加额外编码器层，其在线推理时延和显存占用仍与同类预训练语言模型处于相近水平。

表20给出了LSADA的模型复杂度与DAPT训练开销。LSADA总参数量为124.70M，未引入额外编码器层，DAPT不会改变部署阶段的模型结构和参数规模。采用冻结前6层编码器参数的训练策略后，DAPT阶段参与更新的参数占比为65.9%。在200万条日志语料上完成125000步预训练的训练时长为6.1h，峰值显存占用为9.8GB。可见，DAPT带来的额外成本主要集中于一次性的离线训练阶段。

表21进一步给出了两类检测场景下的推理效率。在离线批量检测场景中，LSADA的吞吐量达到1 851.85 flows·s^{-1}，适用于历史日志审计和流量回溯。在单流实时检测场景中，其平均时延为6.82ms。批量检测通过GPU并行计算降低单条样本的平均开销，单流检测时延则主要来自模板构建、分词和单样本前向计算。

针对资源受限部署场景，后续可采用知识蒸馏将LSADA的语义表征能力迁移至轻量学生模型，并结合FP16推理或INT8量化进一步降低显存占用和推理时延。需要说明的是，层冻结用于减少DAPT阶段的反向传播开销，并不直接降低在线推理时延。

3.12　方法局限性与工程优化思路

上述实验验证了LSADA在公开离线数据集上的检测性能与跨数据集迁移能力，但真实生产网络在流量规模、字段可观测性和协议演化等方面与离线实验环境存在差异。为客观分析方法的工程适用边界，本文进一步讨论LSADA在实际部署中可能面临的局限及相应优化方向。

（1）大规模流量场景下，日志文本转换与DAPT需要较多计算和存储资源。面对持续增长的生产流量，若反复使用全量历史数据进行预训练，将增加训练成本。后续可采用增量DAPT，仅利用新增或分布变化显著的日志进行适配，并结合代表性历史样本回放；同时通过层冻结、参数高效微

表19　各模型单流实时检测复杂度与推理效率对比

模型类别	方法	总参数量/M	单样本平均推理时延/ms·flow^{-1}	峰值GPU显存/GB
数值特征深度模型	HAE-HRL	1.76	1.24	0.54
数值特征深度模型	SSA-BOA-CNN	**0.62**	0.71	0.38
数值特征深度模型	A-NIDS	0.0122	**0.24**	**0.18**
预训练语言模型	BERT	109.49	6.08	1.34
预训练语言模型	BERT-RL	110.18	7.16	1.58
预训练语言模型	**LSADA**	124.70	6.82	1.48

表20　LSADA的模型复杂度与DAPT训练开销

指标	数值
总参数量/M	124.70
DAPT可训练参数占比/%	65.9
最大序列长度	256
DAPT语料规模/M	2
训练步数	125000
DAPT训练时长/h	6.1
峰值显存/GB	9.8

表21　LSADA在两类检测场景下的推理效率

指标	离线批量检测	单流实时检测
批量大小	256	1
测试样本数	10000	10000
平均时延/ms·flow^{-1}	0.54	6.82
P95时延/ms·flow^{-1}	—	8.93
吞吐量/flows·s^{-1}	1851.85	146.63
峰值显存/GB	6.36	1.48

调、知识蒸馏及离线预训练与在线检测分离部署等方式降低资源消耗。

（2）本文模板依赖协议、端口、持续时间、报文统计、速率、流间隔和TCP标志等流级字段。当部分字段缺失或流量加密导致可观测信息受限时，日志语义完整性和检测性能可能下降。对此，可在模板中引入统一的缺失字段标识与缺失状态描述，并针对加密流量利用包长、方向、时间间隔、会话持续时间、TLS握手信息等可获取元数据构建扩展语义表示。

（3）固定模块化模板能够保证跨样本字段表达的一致性，但面对新型协议、私有端口服务或设备特有字段时，现有模板可能无法充分覆盖新增业务语义。后续可构建模式约束下的动态模板扩展机制，在保持核心模块稳定的基础上，为新字段设置扩展槽位，并通过字段类型识别和语义映射将其接入对应模块。

总体而言，LSADA更适用于流级字段较完整、网络行为相对稳定且支持离线领域适配的场景。后续将围绕增量DAPT、加密与缺失流量语义建模、模板扩展及轻量化部署等方向进一步提升方法的工程适用性。

4　结束语

针对现有网络入侵检测方法在流量字段关联表达不足、预训练模型与网络流量日志领域分布不匹配以及检测任务适配不充分等问题，本文提出一种基于日志语义分析与领域自适应预训练的网络入侵检测方法。该方法首先借助类自然语言的日志语义重构，将结构化的流量特征转化为日志文本序列，使端口、协议、包数、速率、时序间隔和标志位等字段的上下文信息得以更充分地呈现。随后，在大规模网络流量日志语料上引入DAPT，引导SecureBERT进一步学习日志的字段结构、字段共现关系以及攻击行为的语义分布，从而增强模型对网络流量日志领域的适配。最后，面向入侵检测任务进行下游微调，完成正常流量与多类攻击流量的区分。在CIC-IDS2017和CSE-CIC-IDS2018数据集上的实验结果显示，本文方法在准确率、精确率、召回率和F1分数等指标上均优于对比方法，能够有效提升多分类入侵检测的性能。消融实验证实日志语义重构和领域自适应预训练的有效性，跨数据集迁移实验则进一步表明该方法具备良好的稳定性和泛化表现。从工程应用角度看，LSADA适用于具有较完整流级字段、支持离线领域适配训练且需要兼顾历史流量审计与在线检测的网络安全场景。

后续将从四个方向进一步开展研究：(1)通过知识蒸馏、参数高效微调、模型剪枝和量化推理降低DAPT与在线检测开销，提升模型在边缘端和资源受限设备上的部署能力；(2)结合开放集识别、不确定性估计和异常分数建模，提高对未知攻击和攻击变体的发现能力；(3)融合流量日志、主机日志、DNS日志、TLS握手信息和安全告警等多源异构数据，构建更完整的攻击行为语义表示；(4)探索增量DAPT与在线漂移检测机制，使模型能够适应长期运行过程中网络业务、协议类型和攻击模式的持续变化。

参考文献：

[1] 邓淼磊, 阚雨培, 孙川川,等.基于深度学习的网络入侵检测系统综述[J].计算机应用,2025,45(2):453-466.
Deng M L, Kan Y P, Sun C C, et al. Survey on deep learning-based network intrusion detection systems[J]. Journal of Computer Applications, 2025, 45(2): 453-466.

[2] ZHANG Y, MUNIYANDI R C, QAMAR F. A review of deep learning applications in intrusion detection systems: overcoming challenges in

spatiotemporal feature extraction and data imbalance[J]. Applied Sciences, 2025, 15(3): 1552.

[3] 刘静, 慕泽林, 赖英旭. 面向异构环境的物联网入侵检测方法[J]. 通信学报, 2024,45(4):114-127.
LIU J, MU Z, LAI Y. Intrusion detection method for IoT in heterogeneous environment[J]. Journal on Communications, 2024, 45(4): 114-127.

[4] HOZOURI A, MIRZAEI A, EFFATPARVAR M. A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges[J]. Discover Artificial Intelligence, 2025, 5(1): 314.

[5] JI Y, LIU Y, YAO F, et al. Adapting large language models to log analysis with interpretable domain knowledge[C]//Proceedings of the 34th ACM International Conference on Information and Knowledge Management. Seoul, Republic of Korea, 2025-11-10—2025-11-14. New York: Association for Computing Machinery, 2025: 1135-1144.

[6] AGHAEI E, NIU X, SHADID W, et al. SecureBERT: a domain-specific language model for cybersecurity[C]//LI F, LIANG K, LIN Z, et al, eds. Security and Privacy in Communication Networks: 18th EAI International Conference, SecureComm 2022, Virtual Event, 2022-10-25—2022-10-27, Proceedings, Part I. Cham: Springer Nature Switzerland, 2023: 39-56.

[7] GURURANGAN S, MARASOVIĆ A, SWAYAMDIPTA S, et al. Don't stop pretraining: adapt language models to domains and tasks[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020-07-05—2020-07-10. Stroudsburg: Association for Computational Linguistics, 2020: 8342-8360.

[8] GU J, LU S. An effective intrusion detection approach using SVM with naïve Bayes feature embedding[J]. Computers & Security, 2021, 103: 102158.

[9] HOSSEINI S, ZADE B M H. New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN[J]. Computer Networks, 2020, 173: 107168.

[10] ACETO G, CIUONZO D, DE DONATO W, et al. GAN augmentation to deal with imbalance in imaging-based intrusion detection[J]. Future Generation Computer Systems, 2021, 123: 108-127.

[11] 林伟, 洪容容. 基于多尺度一维卷积神经网络的入侵检测模型[J]. 中国电子科学研究院学报, 2023, 18(7): 656-662，670.
Lin W, Hong R R. Intrusion detection model based on multi-scale one-dimensional convolutional neural network[J]. Journal of China Academy of Electronics and Information Technology, 2023，18(7): 656-662, 670.

[12] ABDEL-BASSET M, HAWASH H, CHAKRABORTTY R K, et al. HAE-HRL: a network intrusion detection system utilizing a novel auto-encoder and a hybrid enhanced LSTM-CNN-based residual network [J]. Computers & Security, 2025, 151: 104328.

[13] SHERIN R C J, PARKAVI K. Enhancing network resilience: a hybrid meta-heuristic intrusion detection system leveraging SSA-BOA-CNN [J]. Journal of Network and Systems Management, 2026, 34: 17.

[14] ZHA C, WANG Z, FAN Y, et al. A-NIDS: adaptive network intrusion detection system based on clustering and stacked CTGAN[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 3204-3219.

[15] BAMBER S S, KATKURI A V R, SHARMA S, et al. A hybrid CNN-LSTM approach for intelligent cyber intrusion detection system[J]. Computers & Security, 2025, 148: 104146.

[16] BIERBRAUER D A, DE LUCIA M J, REDDY K, et al. Transfer learning for raw network traffic detection[J]. Expert Systems with Applications, 2023, 211: 118641.

[17] XU C, ZHAN Y, WANG Z, et al. Multimodal fusion based few-shot network intrusion detection system[J]. Scientific Reports, 2025, 15(1): 21986.

[18] LATIF S, AHMAD J, BOULILA W, et al. Few-shot learning for IoT intrusion detection: an attention-based Siamese network approach[J]. Procedia Computer Science, 2025, 270: 4553-4562.

[19] WANG M, YANG Y, YUAN J. QuCAD-IDS: cross-domain network intrusion detection via hierarchical contrastive autoencoding and queue-based adaptive distillation[J]. Journal of King Saud University-Computer and Information Sciences, 2026, 38: 266.

[20] HA THANH D. A transfer-aware, deployment-oriented evaluation framework for NetFlow-based intrusion detection systems: TAN-IDS [J]. PLOS ONE, 2026, 21(4): e0346801.

[21] SUN M, CHANG Y, WANG X, et al. LogSwift: an efficient log anomaly detection method with lightweight semantic modeling and hierarchical attention[C]//2025 Thirteenth International Conference on Advanced Cloud and Big Data. Tokyo, Japan, 2025-11-29—2025-12-01. Piscataway: IEEE, 2025: 163-168. DOI: 10.1109/CBD69312.2025.00037.

[22] LI X, ZHENG Z, ZHAO M, et al. RLFE-IDS: a framework of intrusion detection system based on retrieval-augmented generation and large language model[J]. Computer Networks, 2025, 268: 111341.

[23] QI J, HUANG S H, LUAN Z Z, et al. LogGPT: exploring ChatGPT for log-based anomaly detection[C]//2023 IEEE 25th International Conference on High Performance Computing and Communications; 9th International Conference on Data Science and Systems; 21st International Conference on Smart City; 9th International Conference on Dependability in Sensor, Cloud and Big Data Systems and Application. Melbourne, Australia, 2023-12-18—2023-12-20. Piscataway: IEEE, 2023: 273-280.

[24] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1. Minneapolis, MN, USA, 2019-06-02—2019-06-07. Stroudsburg: Association for Computational Linguistics, 2019: 4171-4186.

[25] REN K, LIU L, BAI H, et al. Intrusion detection system based on pre-trained language models and deep reinforcement learning[C]//2025 IEEE International Conference on Unmanned Systems. Changzhou, China, 2025-09-18—2025-09-19. Piscataway: IEEE, 2025: 606-612. DOI: 10.1109/ICUS66297.2025.11294779.

[26] SINHA K, JIA R, HUPKES D, et al. Masked language modeling and the distributional hypothesis: order word matters pre-training for little [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online, 2021-11-07—2021-11-11. Stroudsburg: Association for Computational Linguistics, 2021: 2888-2913.

[27] KHAN Z I, AFZAL M M, SHAMSI K N. A comprehensive study on CIC-IDS2017 dataset for intrusion detection systems[J]. International Research Journal on Advanced Engineering Hub, 2024, 2(2): 254-260.

[28] LEEVY J L, KHOSHGOFTAAR T M. A survey and analysis of intru-

sion detection models based on CSE-CIC-IDS2018 Big Data[J]. Journal of Big Data, 2020, 7(1): 104.

[29] THALJAOUI A. Intelligent network intrusion detection system using optimized deep CNN-LSTM with UNSW-NB15[J/OL]. International Journal of Information Technology, 2025[2026-07-01]. DOI: 10.1007/s41870-025-02416-0.



赵子文(2000-.)，男，天津人，中国民航大学博士研究生，主要研究方向为网络系统安全、软件安全。

杨　宏　宇(1969−)，男，吉林长春人，博士，中国民航大



刘坤(1998-.)，男，山东泰安人，中国民航大学博士研究生，主要研究方向为网络与系统安全。





崔恒宇(2002−)，男，山东济南人，中国民航大学硕士生，主要研究方向为网络与信息安全。