基于大语言模型的加密流量分析方法研究综述

沈 蒙[1*],刘佩凯[1],艾俊宇[1],任琛琛[1],沈泓烨[1],李 琦[2],徐 恪[3],祝烈煌[1]
(1. 北京理工大学网络空间安全学院,北京 100081;2. 清华大学网络科学与网络空间研究院,北京 100084;
3. 清华大学计算机科学与技术系,北京 100084)

摘 要: 随着网络安全技术的快速演进,主流应用、社交平台及内容服务已广泛采用加密协议来保护用户隐私。这一趋势使得加密流量逐步成为现代网络通信的主导形态,但也导致传统依赖明文载荷的深度包分析(Deep Packet Inspection,DPI)技术全面失效。在此背景下,加密流量分析作为保障网络基础设施安全、优化网络资源调度与维持服务质量的关键技术,已成为当前网络空间安全领域的研究重点。面对复杂多变的加密流量,传统机器学习方法高度依赖专家先验知识进行繁琐的手工特征设计,难以适应当前加密流量动态伪装的特点;而基于深度学习的分析方法虽在特征提取方面表现优异,但其高计算成本与大规模标注数据需求限制了在资源受限场景下的部署。近年来,大语言模型(Large Language Model,LLM)凭借其卓越的自监督预训练范式和强大的长序列上下文建模能力,为突破上述瓶颈提供了全新的技术路径。LLM将非结构化的流量字节或统计特征视为一种特殊的网络语言,实现了从专有小模型向通用基础模型底座的范式跃迁。本文聚焦流量分类、恶意流量检测、流量预测和流量生成四大核心下游任务,对基于大语言模型的加密流量分析方法进行了系统性总结和深入分析。首先,本文梳理了从流量采集、数据处理、特征提取、模型构建与微调到推理评估的通用技术流程。其次,针对四大下游任务,本文深入剖析了不同研究的技术演进脉络与架构创新:在流量分类方面,探讨了从基于BERT的判别式微调到基于生成式大模型零样本推理的演进,展现了模型在跨协议与少样本场景下的强大泛化适配能力;在恶意流量检测方面,详细论述了LLM作为特征编码器、分类器甚至可解释性决策器的多重角色,突出了其在应对多阶段隐蔽攻击与未知威胁时展现出的深层语义捕获优势;在流量预测方面,对比了传统时序模型与大模型自回归生成在捕获网络高维时空动态依赖上的差异;在流量生成方面,总结了模型如何跨越单纯的统计分布拟合,迈向受控、高保真且严格遵循协议状态机的流量生成。最后,本文总结了当前领域面临的数据分布偏移、部署资源受限等关键挑战,并从构建多模态数据生态、端云协同轻量化部署等维度展望了未来研究方向。

关键词: 加密流量分析;大语言模型;流量分类;恶意流量检测;流量生成;流量预测
基金项目: 国家重点研发计划(No.2023YFB2703805);国家自然科学基金(No.U23A20304,No.U25A20428)
中图分类号: TP393.08 **文献标识码:** A **文章编号:** 0372-2112(2026)04-1425-21
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20251032

A Review of Encryption Traffic Analysis Methods Based on Large Language Models

SHEN Meng[1*], LIU Yukai[1], AI Junyu[1], REN Chenchen[1], SHEN Hongye[1], LI Qi[2], XU Ke[3], ZHU Liehuang[1]
(1. *School of Cyberspace Science and Technology*, *Beijing Institute of Technology*, *Beijing* 100081, *China*;
2. *Institute for Network Sciences and Cyberspace*, *Tsinghua University*, *Beijing* 100084, *China*;
3. *Department of Computer Science and Technology*, *Tsinghua University*, *Beijing* 100084, *China*)

Abstract: With the rapid evolution of cybersecurity technologies, mainstream applications, social platforms, and content services have widely adopted encryption protocols to protect user privacy. This trend has made encrypted traffic the dominant form of modern network communication, but it has also rendered traditional deep packet inspection (DPI) techniques—which rely on plaintext payloads—completely ineffective. Against this backdrop, encrypted traffic analysis has emerged as a critical technology for safeguarding network infrastructure security, optimizing network resource allocation, and maintaining service quality, becoming a key research focus in the field of cyberspace security. Confronted with complex and dynamic encrypted traffic, traditional machine learning approaches heavily rely on expert-driven, cumbersome manual feature design, struggling to adapt to the dynamic obfuscation characteristics of modern encrypted traffic. While deep learning-based analysis methods excel in feature extraction, their high computational costs and requirements for large-scale labeled data limit deployment in resource-constrained scenarios. In recent years, large language models (LLMs) have

收稿日期:2026-01-17;录用日期:2026-02-24;责任编委:胡宇翔
*通信作者:沈蒙

emerged as a novel technical pathway to overcome these limitations, leveraging their exceptional self-supervised pre-training paradigms and robust long-sequence context modeling capabilities. LLMs treat unstructured traffic bytes or statistical features as a specialized network language, enabling a paradigm shift from specialized small models to universal foundational models. This paper systematically summarizes and deeply analyzes LLM-based cryptographic traffic analysis methods across four core downstream tasks: traffic classification, malicious traffic detection, traffic prediction, and traffic generation. First, it outlines the general technical workflow from traffic collection, data processing, feature extraction, model construction and fine-tuning, to inference evaluation. Second, for each downstream task, this paper delves into the technical evolution and architectural innovations across different studies: for traffic classification, it explores the progression from discriminative fine-tuning based on BERT to zero-shot reasoning using generative large models, demonstrating robust generalization capabilities across protocols and sparse data scenarios. For malicious traffic detection, we detail the multifaceted roles of LLMs as feature encoders, classifiers, and even explainable decision-makers, highlighting their superiority in capturing deep semantic patterns when confronting multi-stage covert attacks and unknown threats; For traffic prediction, we contrast traditional time-series models with large-model autoregressive generation in capturing high-dimensional spatiotemporal dependencies within networks. Regarding traffic generation, we summarize how models transcend mere statistical distribution fitting to achieve controlled, high-fidelity traffic generation that strictly adheres to protocol state machines. Finally, this paper summarizes key challenges facing the field, such as data distribution skew and constrained deployment resources, and outlines future research directions from dimensions including building a multimodal data ecosystem and enabling lightweight deployment through edge-cloud collaboration.

Keywords: encrypted traffic analysis; large language models; traffic classification; malicious traffic detection; traffic generation; traffic forecasting

Foundation Item(s): National Key Research and Development Program of China (No.2023YFB2703805); National Natural Science Foundation of China (No.U23A20304, No.U25A20428)

0 引言

随着全球数字化进程的不断深入,互联网基础设施已全面融入社会经济各个关键领域,网络流量正呈现爆发式增长。根据第56次《中国互联网络发展状况统计报告》[1],截至2025年6月,我国网民数量已达11.23亿,互联网普及率达79.7%。庞大的用户基数和高频的在线活动推动主流应用、社交平台、内容服务及远程交互系统广泛采用端到端加密等安全机制,用户行为和数据交互普遍处于加密环境中。这一广泛而深入的技术演进,使得加密流量逐步成为网络通信的主导形态,因此,加密流量的分析和安全治理已成为网络安全领域不可避免的核心议题。

加密流量分析作为保障网络安全与服务质量的关键技术,已发展出涵盖流量分类、恶意流量检测、流量预测、流量生成等多维度的研究体系。具体而言,应用识别通过提取加密流量中的统计特征与时序模式,实现对终端应用的非侵入式判定[2];协议分类则有助于网络资源调度与合规性审计[3-4];而攻击检测与入侵检测技术致力于在不解密的前提下,通过挖掘流量的统计特征、时序模式和行为异常,识别隐蔽在加密流量中的恶意活动,已成为构建现代网络安全防御体系的核心支撑[5]。此外,基于历史流量建模的预测方法可为网络管理系统提供前瞻性决策支持[6];加密流量生成技术通过构建高保真、可控制的模拟流量,服务于测试验证与数据增强,推动分析模型的持续优化。这些方法共同构成了面向网络行为的理解框架,为在网络流量透明度受限条件下实现智能化的网络运维与服务优化提供了技术路径。

在深度学习与预训练模型发展的双重推动下,流量分析技术逐步实现了从人工经验驱动转向数据智能驱动。当前研究主要围绕三类技术路线展开:传统机器学习方法[7-8]、深度学习方法[9]以及预训练模型方法[10]。传统机器学习方法高度依赖专家经验设计特征,难以适应加密流量动态伪装的特点;深度学习方法虽在特征提取方面表现优异,但其高计算成本与大规模标注数据需求限制了在边缘设备、实时检测等资源受限场景下的部署;而以Transformer[11]、Mamba[12]等架构为核心的大语言模型(Large Language Model,LLM)方法,通过预训练-微调的范式革新与计算架构升级,显著提升了模型在不同任务间的迁移能力,为加密流量分析提供了新的解决路径。

近年来,已有众多文章对加密流量的分析方法展开综述性研究,如表1所示。这些工作从不同的角度对加密流量分类、恶意加密流量检测等技术进行分析与总结。与现有工作不同,本文首次从加密流量分类、恶意加密流量检测、加密流量预测和加密流量生成四个角度系统性研究基于大语言模型的加密流量分析方法。具体而言,本文的主要贡献有如下三点:

(1)系统梳理了基于大语言模型的加密流量分析方法的典型流程。本文从数据采集、特征提取、模型预训练与微调到性能评估全面梳理了加密流量分析的完整流程,明确了各环节在整体流程中的功能定位与协同作用。在此基础上,本文分析了现有评价指标及适用场景,为研究人员提供了清晰的方法论框架。

(2)讨论了四种下游任务的技术演进脉络。本文聚焦加密流量分析中流量分类、恶意流量检测、流量预测与流量生成四个下游任务,系统性分析了现有研究的技术范式演进,揭示了四个任务在流量表征、时序建模与跨任务迁移中的挑战与创新路径,为构建面向动态网络环境的加密流量智能分析框架提供了理论依据与技术路线参考。

(3)展望了加密流量分析的未来发展方向与待解的关键挑战。基于对大语言模型驱动的加密流量分析方法研究现状和发展趋势的全面分析,本文展望了流量分类、恶意流量检测、流量预测和流量生成的未来研究方向,并提出了若干仍待研究和克服的研究问题。

表1　现有综述对比分析

Table 1　Comparative analysis of existing reviews

论文名称	问题领域	论文主题
Machine Learning-Powered Encrypted Network Traffic Analysis: A Comprehensive Survey[13]	加密流量分析	从网络资产识别、网络表征、隐私泄露检测和异常检测四个方面对基于机器学习的加密流量分析方法进行了系统性综述
加密恶意流量检测及对抗综述[14]	加密恶意流量检测	从主动检测和被动检测两个方面对加密恶意流量检测方法进行了总结,并研究了对抗加密恶意流量检测的可实施性
基于机器学习的加密流量分类研究综述[15]	加密流量分类	总结了基于机器学习的加密流量分类方法,涵盖流量采集与处理、流量表征和模型分类及性能评估三个部分
基于深度学习的加密恶意流量检测方法研究综述[16]	加密恶意流量检测	总结了基于深度学习的加密恶意流量检测方法,从监督学习、半监督学习和无监督学习三个方面对比分析了现有方法
本文	加密流量分析	从流量分类、恶意流量检测、流量预测和流量生成四个方面对基于大语言模型的加密流量分析方法研究进行了系统性综述

尽管在自然语言处理领域,典型的LLM通常指代参数量达到百亿级且具备涌现能力的生成式模型,但本文的“大模型”主要指代基于Transformer或Mamba等深度序列建模架构,利用大规模无标注流量数据进行自监督预训练的通用基础模型。为了系统梳理技术演进脉络,本文将双向编码模型(Bidirectional Encoder Representations from Transformers,BERT)与解码器架构统一纳入综述范畴。前者属于理解型预训练模型,后者属于生成式大语言模型,两者均建立在“网络加密流量即语言”的同构假设之上,共享词元化与自注意力机制的技术内核,且都致力于通过预训练-微调范式突破传统深度学习严重依赖有监督标注数据的瓶颈,共同构成了加密流量分析从专用模型向通用表征底座演进的完整技术架构。

1　加密流量分析基本流程

本节基于加密流量分析各步骤之间的层层递进关系,将其一般流程划分为加密流量采集、数据预处理、流量特征提取、模型构建及应用四个关键步骤。首先通过工具捕获原始流量并构建数据集,随后进行数据清洗与类别平衡等预处理操作;继而提取统计、序列、协议或原始字节等多粒度特征;最终对模型进行训练与微调并应用于下游任务。现有基于大语言模型的加密流量分析基本流程如图1所示。下文将围绕这四个核心环节,系统阐述各阶段的技术要点与发展现状。

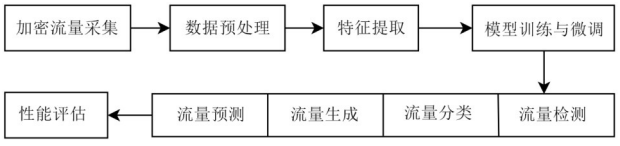

图1　基于大语言模型的加密流量分析基本流程

Figure 1　Basic workflow for encrypted traffic analysis using large language models

1.1　加密流量采集

高质量的数据集是支撑模型有效训练与可靠评估的底层基石,其构建始于系统化、可复现的流量采集过程。在加密流量分析研究中,原始加密流量通常通过部署于关键网络节点的嗅探设备或软件探针进行捕获,并经由预处理与脱敏等工序,最终形成标准化的基准数据集。当前阶段,研究人员普遍采用开源工具与商用探针相结合的方案,以应对多层协议解析与高吞吐量环境下的精准捕获需求。领域内广泛使用的典型流量采集与元数据提取工具包括Libp-

cap[17]、Wireshark[18]、NetFlow[19]等。

基于上述采集与构建手段,在实际研究中选用公开数据集还需兼顾隐私合规性、场景多样性与数据规模,以确保大模型在复杂真实网络中能够稳定收敛并具备强泛化能力。本文梳理了加密流量分析领域中现有的主流数据集,并依据其主要流量类型与研究目标,将数据集划分为三大类:攻击行为流量数据集、恶意软件流量数据集与互联网服务类型流量数据集,现有典型数据集的关键特征及适用任务如表2所示。

表2 相关数据集描述

Table 2 Description of relevant datasets

数据集名称	年份	数据格式	类别	适用任务	具体内容
USTC-TFC2016	2016	PCAP文件	恶意软件流量	恶意流量检测 流量生成	包含良性流量,Tinba、Neris、Nsis-ay、Rbot、Virut、Zeus、Geodo、Htbot、Miuref、Shifu共计10种恶意家族流量
ISCX VPN-nonVPN	2016	PCAP文件 CSV文件	服务类型流量	流量分类 流量生成	包含12类应用如Facebook、Gmail、Outlook、HTTP/HTTPS网页浏览等非VPN流量和通过OpenVPN隧道传输的12类VPN流量
CIRA-CIC-DoHBrw-2020	2020	PCAP文件	服务类型流量	流量分类 恶意流量检测	包含良性DoH、使用恶意软件或DoH隧道工具产生的恶意DoH和传统DNS三类流量
CIC-AndMal2017	2017	PCAP文件	恶意软件流量	恶意流量检测	包含良性流量,Adware、Ransomware、Scareware、SMS Malware 4种恶意软件流量,恶意软件流量又被细分为42种恶意软件家族流量
CICIoT2023	2023	PCAP文件	攻击行为流量	恶意流量检测	包含良性流量,DDoS、Brute Force、Spoofing、DoS、Recon、Web-based、Mirai共计7大类恶意流量
CICIDS-2017	2017	PCAP文件	攻击行为流量	恶意流量检测	包含良性流量,Brute Force、DoS、DDoS、Web Attack、Infiltration、Botnet、Heartbleed、PortScan共计8大类恶意流量
Kitsune	2018	PCAP文件	攻击行为流量	恶意流量检测	包含正常通信和Video Injection、OS Scan、SYN DoS、SSDP Flood、SSL ARP MitM、Active Wiretap、Fuzzing等9种不同类型的网络攻击流量
CSTNET-TLS 1.3	2021	PCAP文件	服务类型流量	流量预测 流量生成	包含了120类不同网站,涵盖了Alexa Top-500网站中部署有TLS 1.3协议的部分网站,共包含46 372条会话流和581 709个数据包

攻击行为流量数据集可支撑入侵检测、异常检测等研究任务,涵盖DDoS、端口扫描、暴力破解等多种网络攻击流量。典型数据集有:CICIoT2023[20]、CICIDS 2017[21]、Kitsune[22]等。此类数据集的优势在于攻击类型覆盖全面,能够支撑研究者在多种攻击场景中评估方法有效性。然而,其局限性在于数据多产自受控的沙箱或仿真环境,背景流量缺乏真实骨干网的高并发与长尾分布特性。

恶意软件流量数据集可用于支撑恶意软件检测、加密恶意通信识别等研究任务。典型数据集有:USTC-TFC2016[23]、CIC-AndMal2017[24]等。此类数据集通常在隔离的实验环境或真实环境中进行,部署采集设备来获取流量,包含由病毒、木马、勒索软件等恶意软件产生的网络流量,同时配有相应的良性软件或正常业务流量作为对照,以便模型在训练过程中实现有效地区分与判别,为研究者在加密流量识别、恶意软件家族分类以及加密通信异常检测等方向提供了重要基础。其缺点在于恶意软件变种迭代极快,数据集已难以覆盖当下的新型恶意软件。同时,USTC-TFC2016等数据集采集环境相对单一,流量背景较为纯净,缺乏复杂网络环境下的噪声干扰。

互联网服务类型数据集涵盖云存储、社交媒体等各类应用程序在正常使用过程中产生的通信流量。典型数据集有:ISCX VPN-nonVPN[25]、CSTNET-TLS 1.3[26]、CIRA-CIC-DoHBrw-2020[27]等。此类数据集的优势在于明确了应用与协议类型,通常被用于加密流量识别、网页与应用分类、协议识别等任务,为理解多样化互联网服务在HTTP、TLS、DoH等不同加密协议下的行为特征提供了基础。

综上所述,这些数据集在采集环境、流量类型和标注粒度等方面各具特点,为模型的训练与评估提供了多样化的基准支持。然而,现有公开数据集在反映

真实网络环境方面仍存在一定局限性:一方面,部分经典数据集的时效性滞后,缺乏 TLS 1.3、QUIC 等现代高隐私协议的流量样本,难以全面评估模型应对新型加密技术的有效性;另一方面,数据集多在受控的实验室环境中采集,背景噪声单一且流量分布相对均衡,而在真实骨干网环境中,流量往往呈现极度的长尾分布,并伴随着复杂的丢包、重传及多径路由现象,这种数据分布偏移是导致模型在实际部署中泛化能力下降的关键因素之一。

1.2　加密流量预处理

真实网络环境中的链路抖动、底层协议重传及分片机制,会导致采集的原始流量中存在畸形报文或残缺会话等问题。此类噪声数据不仅会扭曲流量原本的特征分布,更会显著削弱下游模型训练的收敛稳定性与泛化能力。因此,在提取流量特征前需对原始流量进行系统性预处理,主要包括数据清洗、长度统一和数据整理三个阶段。

数据清洗策略高度依赖下游任务目标:在加密应用识别场景中,通常过滤域名系统(Domain Name System, DNS)、因特网控制报文协议(Internet Control Message Protocol, ICMP)等数据包,并对以太网、互联网协议(Internet Protocol, IP)头及端口号等字段进行匿名化,以避免引入特定设备或会话的统计偏差[28];然而,在恶意流量检测场景中,DNS 协议与 ICMP 协议往往是核心威胁媒介,必须完整保留并提取其深层特征。此外,通用的数据清洗还包括剔除不包含有效载荷的空数据包、过滤因网络抖动产生的重传报文以及移除由抓包异常导致的畸形流,从而为后续的特征工程构建纯净的数据基础[29]。长度统一阶段是为满足深度学习模型对输入维度一致性的要求。常见策略包括:按固定字节数截取或补零、基于数据包数量的截断、基于时间窗口划分流量片段等。数据整理阶段主要根据需要进行归一化与标准化处理,以加速模型收敛,并提升数值稳定性[30]。

值得注意的是,经过上述基础处理后的加密流量数据集往往仍呈现显著的长尾分布特性:正常流量占据主导,而恶意或异常流量占比低。这种类不平衡现象极易削弱模型对关键威胁的感知与识别能力。因此,在完成清洗与规整后,通过合理的采样策略或数据增强手段缓解长尾分布带来的样本倾斜,是构建高质量训练数据的必要环节。总体而言,系统严谨的预处理流水线不仅有效过滤数据噪声,更对齐了模型输入的规范,为后续模型深入挖掘流量特征奠定了坚实的数据基础。

1.3　加密流量特征提取

在基于大语言模型的加密流量分析框架中,加密流量特征提取是连接原始网络流量与下游任务的关键环节。然而,加密协议的快速演进改变了网络流量特征的可获得性边界。随着 TLS1.3 和 QUIC 等新一代加密协议的广泛部署,传统分析方法所依赖的明文握手元数据逐渐失效。例如,TLS1.3 对握手阶段的 Server Hello 进行了加密,而 QUIC 协议则进一步屏蔽了传输层报头信息。这对模型的特征表征能力与抗干扰性提出了更高要求,推动研究者们聚焦加密流量的时序模式、流量突发行为等隐式侧信道特征。基于这一技术背景,根据所使用特征的粒度和来源,现有研究可将流量特征划分为以下四类:统计特征、流序列特征、协议特征及原始字节特征。

(1)统计特征旨在从宏观维度捕捉加密流量的全局规律与分布特性[31]。该类特征主要包括:数据包总长度,以及包大小的均值、标准差、中位数等;数据包到达时间间隔、流持续时间等。

(2)流序列特征是数据包或流处于交互状态下的属性特征,强调对序列结构的建模。这类特征包括包长度序列的变化趋势、数据包发送频率、上下行流量序列等[32]。

(3)协议特征聚焦于网络协议本身的结构性信息,主要关注五元组、传输控制协议(Transmission Control Protocol, TCP)标志位、用户数据报协议(User Datagram Protocol, UDP)头部字段、握手过程等[33-34],是实现语义层面流量理解的关键。

(4)原始字节特征是目前基于大语言模型的流量分析方案中最常采用的方式[35]。该方法直接以原始就加密流量中的二进制字节流作为输入,不需要进行任何特征提取。其优势在于能够保留完整的数据包结构,并且可以同时学习协议头、载荷模式与时间间隔方面的特征。

随着大语言模型被广泛应用于直接处理原始流量负载,研究者们关于"加密负载中究竟包含何种可利用信息"的探讨日益深入。从密码学理论角度看,利用加密协议产生的密文具有高熵特性,理论上难以通过统计方法与随机噪声区分。基于大语言模型的分类器之所以在加密流量上表现优异,并非是因为模型破解了加密内容,而是因为其敏锐地捕捉到了隐藏在负载中的非密文侧信道信息。此外,Wickramasinghe 等人[36]的研究进一步指出,模型性能的提升还可能源于对环境性偏差或特定伪特征的利用。例如,某些模型可能实际上是在拟合过时加密套件的特征或特定实验环境下的噪声,而非学习真正的应用行为模式。

综上所述,基于统计、序列、协议等特征提取方法各有侧重,分别从宏观分布、动态时序、语义结构等

维度刻画流量特性,为流量理解提供了多层次视角。然而,基于原始包二进制的特征提取方式因其通过神经网络自动提取特征,能够完整保留网络流量的细微结构与时空特性,避免信息损失和先验假设,成为大模型驱动下流量生成的主流选择,也能适应不断动态变化的网络环境。

1.4 模型构建与下游任务

在加密流量采集、数据预处理及特征提取后,核心步骤在于构建并优化适用于特定下游任务的大语音模型。这一过程通常遵循“预训练-微调”的范式,即首先利用海量数据在预训练阶段学习流量通用表征,随后针对特定下游任务进行微调。这种预训练机制使得模型能够形成具备强泛化能力的通用流量表征,有效缓解了特定领域标注数据稀缺的瓶颈。根据分析目标与决策边界的差异,加密流量分析的下游任务主要被归纳为以下四类。

(1)加密流量分类。该任务旨在构建从流量特征空间到预定义标签集合的映射关系。其核心目标是对加密流量所属的具体应用类型或网络协议进行精细化识别,例如流媒体、即时通信应用以及 TLS 与 QUIC 等传输协议,广泛应用于网络资源切片与服务质量保障。

(2)恶意流量检测。该任务侧重于安全性的二元或多元判别,即区分正常流量与包含僵尸网络、勒索软件等网络威胁的恶意流量。与流量分类不同,恶意流量检测更关注对偏离正常基线的异常分布或特定攻击签名的敏锐捕捉,需在样本极度不平衡环境下的精准检测网络威胁。

(3)加密流量预测。该任务属于典型的时序分析范式,目标是基于历史时刻的流量状态序列推演未来的演化趋势。其关键在于建模长周期的时序依赖与突发性模式,主要服务于网络运维中的主动式带宽调度与拥塞控制。

(4)加密流量生成。该任务利用生成式大模型的概率分布拟合能力,通过学习真实加密流量的隐式分布,生成符合协议规范与统计特性的流量样本。其核心价值在于:解决训练数据匮乏问题、支持对抗场景下的防御测试等。

1.5 模型性能评估

在加密流量分析领域,量化评估是验证模型有效性的核心环节。然而,具体的评价指标并非一成不变,而是高度依赖于下游任务的类型与目标。例如,面向应用识别的分类任务关注类别判别的准确性,而异常检测任务则更强调对恶意行为的敏感性。鉴于不同任务对模型性能的定义存在本质差异,本部分将依据典型的下游任务场景分别系统介绍其常用的评价指标及其适用条件。

在流量分类和恶意流量检测任务中,通常通过混淆矩阵对模型性能进行评估。其中 TP 表示被正确预测为目标类别的样本数,FP 表示被错误预测为目标类别的其他类别样本数,FN 表示实际属于目标类别但未被正确识别的样本数,TN 表示被正确预测为非目标类别的样本数。常见的评价指标如下。

(1)准确率(Accuracy)。准确率用于反映模型整体分类的正确程度,即所有被正确识别样本占全部样本的比例。计算公式如式(1)所示:

$$\text{Acurracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{1}$$

(2)精确率(Precision)。精确率描述模型在预测为目标类别的样本中,真正属于该类别的比例。该指标侧重评估模型对正类样本的“纯度”,用于评价模型的可靠性,计算公式如式(2)所示:

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{2}$$

(3)召回率(Recall)。召回率衡量模型识别出目标类别样本的能力,即实际属于该类的样本中被成功检测出的比例,召回率尤为关键,高召回值意味着模型能够更全面地识别出各类样本,覆盖性更强,计算公式如式(3)所示:

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{3}$$

(4)误报率(FPR)。误报率反映模型将非目标类别样本误判为该类别的比例,过高的误报率会导致检测或分类系统的实际可用性下降,因此在模型设计中需兼顾降低误报风险,计算公式如式(4)所示:

$$\text{FPR}=\frac{\text{FP}}{\text{FP}+\text{TN}} \tag{4}$$

(5)漏报率(FNR)。漏报率描述模型未能识别出的目标类别样本比例,即目标类样本被错误判为其他类别的概率,较低的漏报率表示模型具备更强的安全检测敏感度,计算公式如式(5)所示:

$$\text{FNR}=\frac{\text{FN}}{\text{FN}+\text{TP}} \tag{5}$$

(6)F1 值。F1 值是精确率与召回率的调和平均,用于在两者之间取得平衡,该指标在样本分布不平衡的任务中尤为重要,能反映模型在整体性能上的协调性,计算公式如式(6)所示:

$$\text{F1}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{6}$$

(7)ROC 曲线与 AUC 值。ROC 曲线展示了不同判定阈值下模型真阳性率与假阳性率的变化关系,曲线越靠近左上角,代表模型的判别能力越强。AUC 值为 ROC 曲线下的面积,用于量化整体分类性能,AUC 值越接近 1,模型越接近理想检测器,AUC 计算

公式如式(7)所示:

$$\mathrm{AUC}=\int_0^1 \mathrm{TPR}(x)\mathrm{d}\left(\mathrm{FPR}(x)\right) \tag{7}$$

在加密流量预测任务中,当前研究通常利用时序预测误差、序列相似性度量等指标来评估模型的效果。常见的评价指标如下。

(1)平均绝对误差(MAE)。MAE用于反映预测值与真实值之间的误差。MAE的值越小,说明模型预测结果与真实值之间的偏离程度越低,预测性能越好。MAE计算公式如式(8)所示:

$$\mathrm{MAE}=\frac{1}{Q_N}\sum_{i=1}^{Q_N}|Y_{Pi}-\hat{Y}_i| \tag{8}$$

其中,Q_N为样本个数,Y_{Pi}为第i个样本的真实值,$\hat{Y}_i$为第i个样本的预测值。

(2)均方根误差(RMSE)。RMSE能够更全面地反映模型在预测过程中可能出现的较大误差。RMSE值越小,表明模型预测结果与真实值之间的偏差越小,整体预测性能越优。RMSE计算公式如式(9)所示:

$$\mathrm{RMSE}=\sqrt{\frac{1}{Q_N}\sum_{i=1}^{Q_N}(Y_{Pi}-\hat{Y}_i)^2} \tag{9}$$

其中,Q_N为样本个数,Y_{Pi}为第i个样本的真实值,$\hat{Y}_i$为第i个样本的预测值。

(3)Wasserstein距离。Wasserstein距离能够反映预测结果与真实值在分布上的整体差异,可评估模型对动态或噪声环境下数据分布的拟合能力。在一维情况下,Wasserstein距离可以简化为

$$W_1(p,q)=\int_0^1|F_p^{-1}(u)-F_q^{-1}(u)|\mathrm{d}u \tag{10}$$

其中,p和q分别表示预测分布与真实分布,F_p^{-1}和F_q^{-1}分别为分布p和q的分位数函数。

在流量生成任务中,当前研究普遍采用定量与定性相结合的评估策略,以确保生成流量在保真度、功能性与实用性方面满足实际应用需求。常见的评价指标如下。

(1)Jensen-Shannon散度(JSD)。JSD是一种基于KL散度的对称性概率分布相似性度量方法,取值范围为[0,1],值越小表示两个分布越接近,被广泛应用于量化生成数据与真实数据在统计分布层面的一致性。JSD计算公式如式(11)所示:

$$\mathrm{JSD}(P/\!/Q)=\frac{1}{2}D_{\mathrm{KL}}(P/\!/M)+\frac{1}{2}D_{\mathrm{KL}}(Q/\!/M) \tag{11}$$

其中,P和Q为两个概率分布,$M=\frac{1}{2}(P+Q)$是P和Q的逐点平均分布,$D_{\mathrm{KL}}(P/\!/M)$是从分布P到M的KL散度。

(2)总变差距离(TVD)。TVD用于衡量两个概率分布之间的差异,取值范围[0,1],值越小表示两个分布越相似,常用于评估生成流量在包长、流持续时间等关键统计特征上与真实流量的逼近程度。TVD计算公式如式(12)所示:

$$\mathrm{TVD}(P,Q)=\frac{1}{2}\sum_x\left|p(x)-q(x)\right| \tag{12}$$

其中,P和Q是两个定义在相同可测空间上的离散概率分布,其概率质量函数分别为$p(x)$和$q(x)$。

(3)经验模态分解(EMD)。EMD是一种衡量两个概率分布之间差异的度量方法,相较于JSD或TVD,EMD能更敏感地捕捉分布的几何结构和偏移,在流量生成评估中常用于衡量包到达间隔、流量强度等连续特征的分布相似性。EMD计算公式如式(13)所示:

$$\mathrm{EMD}(P,Q)=\int_{-\infty}^{+\infty}\left|F_P(x)-F_Q(x)\right|\mathrm{d}x \tag{13}$$

其中,P和Q是两个一维概率分布,其累积分布函数分别为$F_P(x)$和$F_Q(x)$。

在协议结构合规性方面,研究者通常利用自动化解析工具验证五元组完整性、TCP标志位合法性、TLS握手序列正确性。部分研究进一步利用特定规则剔除违反网络通信协议的无效流量。

2　加密流量分析

随着网络加密技术的广泛应用,传统基于明文内容的流量分析方法面临严峻挑战。在此背景下,研究者转向利用包长、时序、方向等流量的元信息及其统计或语义模式,在不依赖载荷解密的前提下实现对加密流量的深度理解与智能推断。当前,加密流量分析已从单一识别任务逐步拓展为分类、检测、预测与生成等多维技术体系。本章围绕四大核心下游任务展开系统阐述:流量分类旨在识别加密会话所属的应用类型或协议类别;恶意流量检测聚焦于从海量正常通信中精准发现异常或攻击行为;流量预测致力于建模流量的动态演化规律,以支持资源调度与网络规划;而流量生成则通过合成高保真加密流量,服务于数据增强、隐私保护与仿真测试等场景。这四类任务既相互独立,又在特征表示、模型架构与评估范式上存在紧密关联,共同构成了现代加密流量智能分析的技术全景。

具体而言,尽管上述四个下游任务的应用目标各异,但在底层技术架构上呈现出高度的通用性。从模型底座来看,各类任务均广泛复用Transformer、Mamba等架构作为核心特征提取器,利用其自注意力机制捕获加密流量中的长距离依赖关系;从输入处理

来看，基于原始字节或数据包头的词元化策略在不同任务间具有显著共享性。得益于独特的设计框架，一个经过大规模无监督预训练的流量基础模型，能够灵活适配多种下游任务场景。具体而言，该模型既能分类网络实现流量识别与恶意检测，也能通过结合解码器实现流量生成与预测，极大地提升了模型的跨任务迁移能力与实际部署效率。

2.1 加密流量分类

随着网络规模的不断扩大和应用服务的日益多样化，加密流量分类作为网络管理、安全监控和质量保障的基础任务，其重要性日益凸显。流量分类旨在识别网络中数据流所属的应用类别、服务类型或用户行为[37]，将复杂的网络活动映射到预定义的标签集合上。

传统的流量分类主要依赖端口映射与明文载荷匹配，但随着动态端口分配以及 TLS 1.3、QUIC 等强加密协议的全面普及，此类方法极易被规避或直接失效，这使得基于载荷的分类技术失效。这迫使研究者转向利用数据包长度、时间序列、流统计特征等流量侧信道信息实现加密流量分类。然而，早期基于深度学习和浅层 Transformer 的模型依然面临结构化信息捕捉不足、跨域泛化性差等瓶颈。这些局限性驱动着分类技术向大模型范式演进，推动研究人员转向更深层次的领域适配和更强大的通用表示学习。本节将从预训练模型的引入与 Transformer 上下文编码、BERT 编码器与混合架构、通用 LLM 模型的应用三个方面，系统综述大模型在流量分类任务中的技术脉络。基于大模型的流量分类基本框架如图 2 所示。

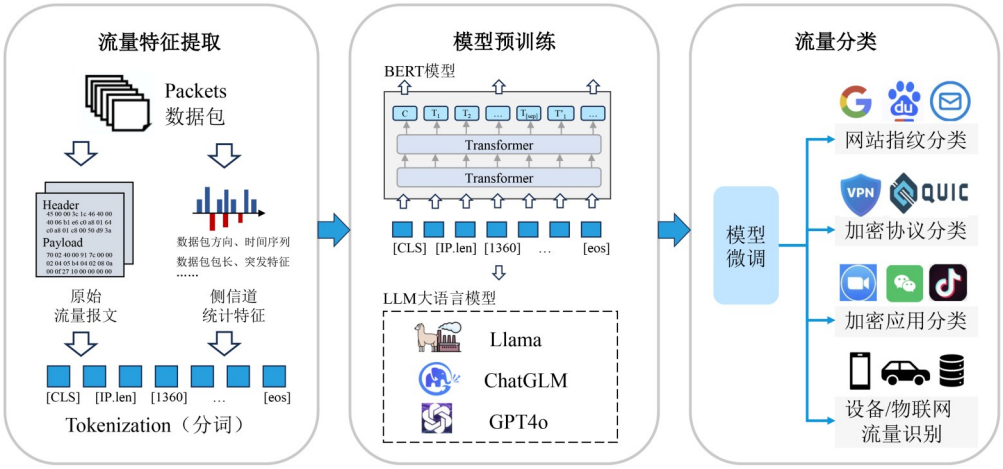

图 2 基于大模型的网络流量分类框架

Figure 2 A large-model-based network traffic classification framework

2.1.1 预训练模型的引入与 Transformer 上下文编码

在基于大模型的加密流量分类技术发展初期，研究人员确立了流量即序列的预训练思想。这一阶段的核心是借鉴自然语言处理领域中的 Transformer 架构，对流量进行序列化编码，捕捉包长、到达时间间隔等非载荷信息中的时序规律与上下文依赖关系。例如，Liu 等人[38]提出的 TransECA-Net 利用流量序列中的长距离依赖关系和上下文信息生成更具判别力的流量表示，从而提高分类的精度。TransECA-Net 的优势在于模型能够对流量序列进行全局建模，相比局部特征提取方法，其分类精度和鲁棒性通常更高，尤其适用于捕捉复杂的加密协议行为。

除了传统的应用类型，网站指纹分类是加密流量分类的子领域，旨在通过流量特征识别用户访问的匿名网站。有研究者将 Transformer 架构引入这一挑战性任务。例如，Zhou 等人[39]提出了 WF-Transformer 方法，该方法将加密流量视为时间序列，并利用 Transformer 网络强大的时间特征提取能力，以克服传统深度卷积神经网络在捕捉流量序列的长距离依赖性方面的不足。

然而不可忽视的是，Transformer 庞大的参数规模与高昂的算力开销会导致显著的推理延迟，使得上述方法在计算资源受限或实时性要求高的网络环境中面临严峻的部署瓶颈。为了解决资源受限问题，Meng 等人[40]提出了轻量级、浅层的 Transformer 架构 LETformer，并引入锐度感知优化机制来平衡高性能和高效率，克服了传统大型预训练模型参数量巨大，不适合在低计算资源环境下实时分类的问题。研究者通过浅层架构将模型参数降低到 1 760 万，使得 LETformer 能够应对移动设备、实时分类等场景。

2.1.2 BERT 编码器与混合架构

随着 Transformer 架构的成熟，研究的核心转向如

何将BERT等强大的编码器架构适配到流量分类领域,并通过设计领域特定的预训练任务,学习更通用、上下文信息更丰富的流量表示。这一阶段的工作极大地提升了模型对加密流量特征的提取能力。Lin等人[26]提出的ET-BERT是基于BERT编码器架构的代表性工作,该方法从大规模未标注数据中学习数据报级别的上下文特征表示,以缓解现有方法在特征泛化性方面的局限及依赖大规模标注数据的问题。Pan等人[41]提出的FlowBERT将加密流量分解为数据包长度序列和数据包载荷两个维度进行特征学习,进而捕获流量的内在时序关系,学习比传统统计特征或单一序列信息更为鲁棒和丰富的流量信息。然而,这两种方法主要侧重于数据报级别的表示,在一定程度上弱化了流或会话级别的长距离结构化依赖特征的学习。

随着TLS 1.3等加密技术的发展,载荷内容的随机性会越来越高,仅依赖载荷的编码模型鲁棒性受限。因此,研究人员开始寻求增强模型对流量模式的捕捉粒度和抗干扰能力。Chen等人[42]假设流量的关键模式源于流内数据包之间的交互关系,进而通过多实例学习范式,捕捉流内数据包之间的交互特征和流级别模式,成功地将特征学习的粒度从令牌提升到数据包-流级别,更有效地捕捉了加密流量的关键模式。虽然该方法在分类准确率方面取得了不错的效果,但模型对流长度变化的鲁棒性以及对短流的特征捕捉能力依旧不足。

为进一步克服单一编码器架构的局限性,研究人员探索了BERT编码器与其他深度学习架构的优势互补。Shi等人[43]提出的BFCN结合了BERT上下文提取的优势与卷积神经网络(Convolutional Neural Network,CNN)局部模式识别的优势,构建了性能更优的混合分类模型。该模型首先利用BERT对流量序列或数据包的上下文信息进行深度编码,然后将BERT的输出作为特征图输入给CNN层,通过CNN的卷积操作来提取局部、精细的模式特征,最终实现对加密流量的准确分类。

上述基于BERT等编码器的流量分类技术总结如表3所示,通过领域适配和自监督预训练,显著提升了模型特征提取深度和泛化能力。然而,这些模型在特定任务数据集上训练,难以跨越多种任务和未见数据集。为了实现更高级别的通用性、多任务支持和零样本学习能力,研究前沿迈向了基于更大、更通用的LLM基础模型的范式。

表3　基于BERT编码器的流量分类工作总结

Table 3　Summary of traffic classification work based on the BERT encoder

方法	年份	流量表征	模型架构	评估数据集	评价指标
FlowBERT[41]	2023	加密负载序列与包长序列	BERT	ISCX-VPN-nonVPN CIRA-CIC-DoHBrw-2020	准确率、精确率、召回率、F1值
ET-BERT[26]	2022	原始流量报文	BERT	ISCX VPN-nonVPN CSTNET-TLS 1.3 Dataset USTC-TFC2016	准确率、精确率、召回率、F1值
MIETT[42]	2025	原始流量报文	BERT	ISCX VPN-nonVPN ICIOT2023	准确率、F1值
BFCN[43]	2023	包级全局特征 字节级局部特征	BERT+CNN	ISCX VPN-nonVPN	准确率、精确率、召回率、F1值

2.1.3　通用LLM模型的应用

当前的加密流量分类研究已迈向基于LLM的流量基础模型范式,旨在通过更大规模的预训练、统一的架构以及指令微调,实现跨任务、跨数据集的通用性和零样本学习能力。例如,Cui等人[33]提出的TrafficLLM是这一领域的代表,该方法引入双阶段微调框架,首先利用流量领域特定的词元化机制对异构原始流量数据进行处理,然后在两个阶段进行调优,使其能够从异构数据中学习通用的流量表示,并通过指令微调赋予模型任务理解能力,使其能够执行流量分类等多种下游任务。解决了传统方法难以跨越不同任务和未见数据的问题,在泛化性上表现出突破性进展。这种基础模型范式的优势不仅在于其通用性,更在于其统一架构对多样化任务的支持。

鉴于LLM基础模型的巨大计算开销和单一序列建模的局限性,研究者们着眼于LLM架构与其他模型的融合与优化,以提高分类的实时性和对复杂网络结构信息的捕捉能力。Farrukh等人[44]提出了XG-NID框架,将流级和包级数据融合进异构图,利用异构图神经网络来捕获流量数据中复杂的结构关系和时序依赖关系,同时引入大型语言模型对数据包载荷信息进行处理和特征提取,全面捕捉网络流量的结构关系和深度语义信息,实现了双模态融合的流量分类与实时推理。

LLM强大的语义理解能力也被用于解决标注数据稀缺、新型威胁难检测等问题。Luo等人[45]研究了T5-LLM的生成式迁移学习潜力,该模型通过生成式

学习的机制，将网络流量的统计信息或序列特征编码并融合到T5模型的输入空间中，通过少量数据的微调实现高效的加密流量分类，同样将分类问题转化为文本到标签的映射问题。上述方法对加密流量的分类性能高度依赖于预训练LLM所学习到的嵌入特征的泛化能力，在TLS 1.3强加密的背景下，LLM从加密载荷中提取有效语义特征的能力下降，需进一步研究如何有效地将数值化、时序化的流量特征转化为LLM能够理解的文本或序列。

商用大语言模型的快速演进，不仅加速了流量分类技术从理论研究向现网部署的跨越，更使得模型在真实网络环境中的应用安全性评估成为亟待解决的核心议题。Antari等人[46]利用从网络设备中提取的19个流特征，对包括Web、IPSec、Backup和Email在内的五种流量进行分类，并通过零样本和少量样本学习的方式评估了GPT-4o、Gemini等商用LLM的性能，验证了商用LLM在加密流量分类中的可行性，也表明LLM的领域适应性和特征理解能力仍有局限性。

表4总结了基于通用LLM模型的流量分类典型方法。尽管基于通用LLM的流量分类方法在语义理解上表现优异，但在实际应用中仍面临显著挑战。首先是推理延迟与资源开销的矛盾，相比随机森林、XGBoost等轻量级的传统机器学习模型[47]，LLM庞大的参数量导致推理速度较慢，难以满足骨干网实时流量分类的需求。其次是非语义特征的处理瓶颈，当面对IPSec等缺乏直观文本语义的强加密流量或私有协议时，LLM的分类准确率往往不如针对特定统计特征优化的专用模型。

表4 基于通用LLM模型的流量分类工作总结

Table 4 Summary of traffic classification work based on general-purpose LLM models

方法	年份	流量表征	模型架构	评估数据集	评价指标
TrafficLLM[33]	2025	原始流量报文(包头+负载)	Llama2-7B	ISCX VPN-nonVPN CSTNET 2023	精确率、召回率、F1值
XG-NID[44]	2025	流持续时间、方向、包数、传输字节数等流级与负载信息等包级特征	GNN + LLM	CICIDS2017	精确率、召回率、F1值、时间开销
T5-LLM[45]	2025	原始流量报文(包头+负载)	T5	ISCX VPN-nonVPN	F1值
Antari, et al[46]	2025	时间、包长等19个数据流特征	GPT-4o、Gemini	自建数据集	准确率、F1值

2.2 恶意加密流量检测

恶意加密流量检测是加密流量分析的重要分支，旨在从高维、非结构化的流量数据中识别潜在的异常或恶意行为，涵盖恶意软件传播、入侵攻击、异常通信等多种威胁，对维护国家网络安全具有重要意义。在互联网发展的早期，传统的恶意流量检测方法主要依赖基于特征签名与基于规则匹配的技术路线。前者通过比对数据包或流中的特定字节序列、正则表达式或已知攻击模式来识别威胁；后者依赖人工定义的检测规则或专家经验，将协议行为或特征阈值映射为异常判定条件。这些方法解释性强、检测速度快，但严重依赖已知特征库，对未知或变体攻击几乎无能为力。针对此问题，研究者开始转向基于流量统计与机器学习的检测方法[48]，即利用加密流量中包长度分布、到达间隔、流持续时间及上下行数据模式等侧信道特征构建特征向量，并利用机器学习模型实现流量分类或异常检测。然而，人工构造的统计特征难以完整表达复杂协议或多阶段攻击的语义模式，并且模型泛化性受限，尤其在面对加密流量场景、样本稀缺或攻击多样化时，传统机器学习模型往往表现不稳定，无法适应快速演化的攻击环境。

近年来，随着大语言模型的快速发展，其强大的序列建模与上下文理解能力为恶意流量检测任务提供了新的技术路径。LLM能够从原始字节序列、流量级特征等信息中自动提取语义特征，实现从统计驱动向语义驱动的检测范式转变。根据检测目标的不同，本文将恶意流量检测分为三类典型任务：恶意软件检测、异常检测与入侵检测。以下将分别对这三类任务的代表性研究进行综述与分析，恶意流量检测的典型框架如图3所示。

2.2.1 恶意软件检测

在网络恶意流量检测领域，恶意软件检测任务面临特征提取困难、样本稀缺与攻击多样化等挑战。LLM凭借其强大的序列建模与上下文理解能力，被逐步引入恶意软件检测中。一方面，作为编码器，LLM能够将系统调用、数据包字节序列等非文本安全数据转化为语义化嵌入表示，从而有效捕捉恶意软件流量的长程依赖关系与上下文语义特征；另一方面，作为分类器，LLM通过迁移学习或微调机制在少量标注样本上实现恶意流量的精准识别与家族划分。基于LLM的恶意软件检测研究已在系统调用分析、少样本识别等多种任务场景中展现出显著优势。

为解决加密恶意流量检测中单一特征的局限性，Liang等人[49]提出了EM-BERT，该方法将网络流量视为字节序列，通过TLS会话重建、清除协议头等预处理后，使用基于BERT架构的EM-BERT作为核心编码

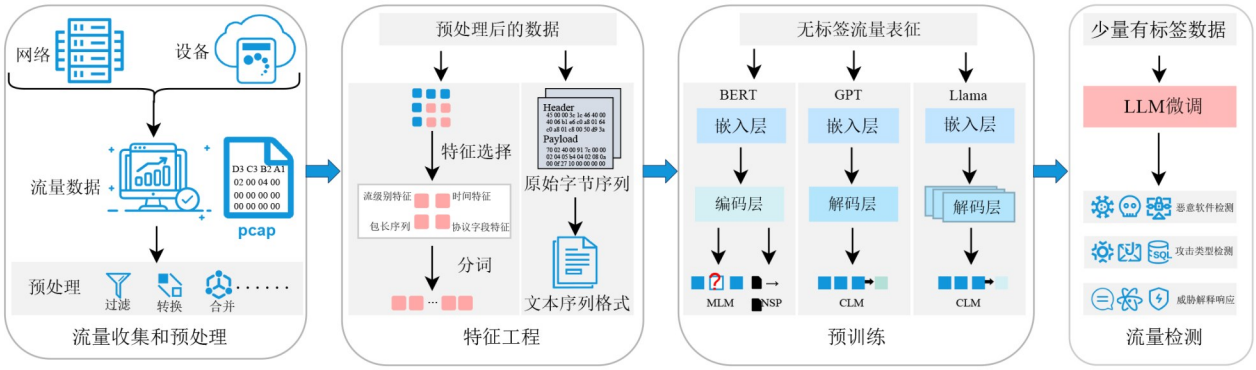

图3　恶意流量检测典型框架

Figure 3　Typical framework for malicious traffic detection

器,从原始字节中同时提取空间内容与时间序列特征,生成富含上下文语义的特征表示,最终通过分类器实现高效检测。在少样本检测方向,为解决恶意样本标注稀缺问题,Stein 等人[50]提出了一种基于 LLM 与少样本学习的恶意数据包识别方法,首次将 LLM 的序列建模能力成功迁移至网络流量分析,系统地引入了小样本学习范式,为样本稀缺场景下的流量检测提供了新思路。

为了解决恶意流量检测中数据不平衡问题,Nasser 等人[51]提出了一个结合生成式与判别式 LLM 的恶意流量检测框架 Syndetect。该框架旨在通过合成流量样本来缓解数据不平衡问题,并实现端到端的生成与分类闭环。在该框架中,LLM 分别承担生成器与分类器的角色:在生成阶段,利用 GPT-2 作为生成式 LLM,对恶意流量进行合成与扩增,生成高质量的攻击样本以平衡训练数据分布;在检测阶段,采用 BERT 作为判别式 LLM,提取合成与真实流量的语义特征,并在恶意软件家族检测任务中展现出良好性能,体现了生成式与判别式语言模型在安全检测任务中的协同潜力。

2.2.2　异常检测

网络流量异常检测的核心目标是识别偏离正常行为的流量模式。传统异常检测方法通常依赖手工设计特征或浅层学习算法,难以捕捉复杂通信行为中的深层语义与时序依赖。近年来,LLM 凭借强大的序列建模和上下文理解能力,能够从原始流量数据中自动提取语义特征,实现从统计驱动向语义驱动的检测范式转变。现有研究表明,LLM 在异常检测中主要承担两类角色:编码器与分类器。异常检测方法的总结如表 5 所示。

表 5　异常检测方法总结

Table 5　Summary of anomaly detection methods

方法	年份	流量表示	模型架构	数据集	评价指标
BERTAD[52]	2024	MQTT 协议流量	bert-base-uncased	MQTT-IoT	准确率、精确率
Yang, et al[53]	2025	带宽利用率、延迟、丢包率等多维特征时间序列	Transformer	CICIDS-2017	准确率、F1 值、训练时间、推理速度
Zhao, et al[54]	2024	设备指纹向量	Llama3-8B	自建数据集	准确率、F1 值
FedLLMGuard[55]	2025	数据包大小、熵值、持续时间等数值特征	Tiny-BERT	TII-SSRC-23 CICDDoS2019	准确率、精确率、F1 值、召回率、误报率、检测延迟

由于 LLM 具有强大的特征抽象与语义建模能力,能够将网络数据包或二进制字节流等非结构化数据映射到高维语义空间中,研究者们展开了将 LLM 作为编码器的研究,以便更好地捕获加密流量潜在的行为模式与上下文依赖关系。Chaves-Tibaduiza 等人[52]提出的 BERTAD 方法将 MQTT 协议流量文本化,直接利用 BERT 编码器的多层自注意力机制提取高维时序上下文特征,并通过附加全连接层实现端到端的异常分类。为应对复杂时序流量和未知攻击的检测挑战,Yang 等人[53]提出了一种基于 LLM 的云平台网络流量异常检测系统。该方法将网络流量视作时序序列,利用 Transformer 架构的自注意力机制捕捉长程依赖关系,进而提取加密流量深层特征。系统采用结合重构误差与注意力正则化的混合损失函数,并引入迁移学习增强模型对新环境的适应能力。

另一类研究侧重于将 LLM 直接作为分类器,依托少样本微调或提示工程,在高维语义空间中实现异常流量的精准判别。针对传统检测模型依赖特定协议、泛化性差、特征提取不全等问题,Zhao 等人[54]提出了一种基于 LLM 的工业控制系统异常检测方法。

该方法结合通信指纹分析与Llama3模型微调，构建了一个能够在主动与被动环境中识别设备异常的智能检测框架。系统首先通过ICS协议通信模型提取涵盖软件与硬件特征的设备指纹向量，随后将指纹数据转化为问答式训练样本，利用Unsloth框架对Llama3-8B模型进行高效微调，从而赋予模型跨厂商、跨协议的异常识别能力。

在此基础上，研究者们进一步将编码与判别功能融为一体，实现语义嵌入与分类任务的联合优化，从而兼顾特征泛化与检测精度。针对5G网络入侵检测中存在的高延迟、隐私风险和对抗攻击脆弱性等问题，Rezaei等人[55]提出了FedLLMGuard的框架，将联邦学习与LLM相结合。该框架采用轻量级TinyBERT模型，兼具编码器与分类器的双重功能：一方面通过语义分析将网络流量日志转换为上下文信息丰富的嵌入表示；另一方面基于语义特征实现精准的异常流量分类。通过软提示调优技术，仅需更新少量参数即可完成模型优化，显著降低了通信与计算开销。

2.2.3 入侵检测

入侵检测旨在识别外部攻击者或内部异常用户对网络系统的非法访问，是保障信息基础设施安全的关键环节。由于LLM不仅具备强大的语义建模和上下文推理能力，还能够在结构化特征与自然语言描述之间建立桥梁，从而提升检测方法的泛化性与可解释性。因此在该任务中，LLM不仅能作为编码器或分类器实现高精度的检测，还能作为解释器和决策器参与检测流程。

LLM作为特征编码器的核心在于把原始流量或多维特征映射为高维语义表示，实现攻击模式的自动挖掘。针对DDoS攻击中低速率、多向量导致流量难以检测的问题，Li等人[56]提出了一个基于编码器的检测框架DoLLM。该框架旨在将非序列化的网络流数据转化为具有上下文语义的序列输入，充分发挥LLM在语义理解与特征提取方面的优势，从而识别复杂的分布式攻击模式。在复杂加密与未知攻击场景中，Lin等人[57]提出了基于LLM指令调优的加密流量检测方法ETool。该方法将流量交互模式转化为图结构，通过构建包含突发与邻接关系的流量关系图，捕捉了网络攻击的固有协作模式。框架采用两阶段指令调优，首先通过自监督的BURST图匹配任务使LLM理解流量交互语义，再利用少量标注数据进行任务适配。

在强调可解释性与推理能力融合的方向上，Wu等人[58]将决策树与LLM相融合，首先训练多个决策树，从流持续时间、协议标志数等加密流量数值特征中学习攻击模式，并提取从根节点到叶节点的推理路径及其置信度。在此框架中，LLM被定位为核心的决策器与自然语言解释器，它接收多条序列化路径，通过综合推理完成最终攻击类型的判断并生成详细解释。随着可解释性与推理能力的进一步融合，Xu等人[59]提出了APTSniffer，将检索增强生成架构应用于APT加密流量检测。APTSniffer通过多级检索机制获取相似样本，并创新性地采用轻量级辅助权重机制替代全参数微调，通过深度神经网络学习各检索模块的置信权重。作为决策器，LLM综合检索信息完成APT流量二分类；作为解释器，LLM为分类结果提供可理解的推理依据，展现出优异的决策鲁棒性与语义可解释性。

现有典型入侵检测方法的总结如表6所示，主要分为有监督和无监督两种类别。有监督方法侧重于利用注意力机制捕捉长距离的攻击载荷签名，在已知威胁的精准归因上表现出色；而无监督方法通过建模正常流量分布并计算重构误差，能够有效发现偏离基线的未知攻击。在实际部署中，前者适合用于威胁确认与取证，后者则常作为第一道防线进行异常筛选。然而，LLM虽然提升了对复杂攻击的识别能力，但也引入了新的安全与隐私风险。其一是数据隐私泄露问题[60]，将敏感的加密流量上传至云大模型可能违反数据合规要求，而私有化部署又面临高昂的算力成本。其二是对抗样本与规避攻击的脆弱性[61]，攻击者可能通过微小的流量扰动或注入特定的文本模式，诱导LLM产生误判从而绕过检测。此外，现有研究多基于特定数据集进行微调，当面对现实网络中不断

表6 入侵检测方法总结

Table 6 Summary of intrusion detection methods

方法	年份	流量表示	模型架构	数据集	评价指标
DoLLM[56]	2024	类别特征、流量统计特征	Llama2-7B	CIC-DDoS2019	准确率、精确率、召回率、F1值
ETool[57]	2025	原始字节序列、包长序列	Vicuna-7B-v1.5 (LLaMA)	ISCX-Botnet 自建数据集	准确率、精确率、召回率、F1值、训练时间、推理延迟
Wu, et al[58]	2025	结构化特征	GPT系列	CICIoT2023	准确率、精确率、召回率、F1值
APTSniffer[59]	2025	包长序列、流量关联特征	GPT-4o	DAPT2020 EarlyCrow-APT	准确率、精确率、召回率、F1值

演进的零日攻击或分布外流量时,大模型容易出现过度自信的误判,其跨域泛化能力仍需进一步验证。

2.3　加密流量预测

在复杂的网络环境中,网络流量呈现出高度的周期性、突发性和时空相关性。流量预测旨在基于历史观测数据,预测网络未来某一时间段的流量值或模式,是网络管理和优化的关键环节。准确的流量预测是实现网络资源动态分配、拥塞控制、故障预警和质量保障优化的前提。

传统的流量预测方法主要基于统计学和基于深度学习模型。基于统计学的模型擅长处理具有明显周期性和趋势性的流量,但难以捕捉网络流量中的非线性、突发性以及复杂的长距离时序依赖关系。循环神经网络(Recurrent Neural Network,RNN)和长短期记忆网络(Long Short-Term Memory,LSTM)等深度学习模型虽在一定程度上解决了非线性建模问题,但正如Kong等人[62]所指出的,这些模型在处理长序列时易出现梯度消失、信息丢失等问题,难以有效捕捉网络流量时间序列中的时序依赖特征。此外,网络流量的演变不仅受时间因素影响,还与网络拓扑结构紧密相关,传统方法难以有效融合这些复杂的空时依赖。为解决上述挑战,研究者们开始将自然语言处理领域的Transformer等大模型架构,引入加密流量预测领域。

流量预测的基本框架如图4所示。本节将从大模型架构在流量预测中的早期应用与基础、基于基础模型的流量预测、基于集成大型语言模型的精准预测三个方面对基于大模型的流量预测任务进行综述。

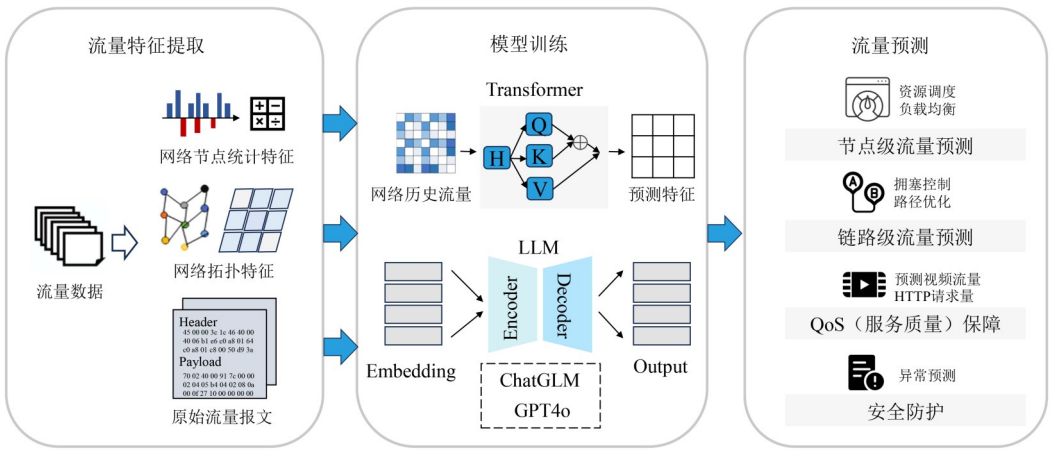

图4　流量预测基本框架

Figure 4　Basic framework for traffic forecasting

2.3.1　大模型架构在流量预测中的早期应用与基础

作为大模型的基石,Transformer架构凭借其核心的自注意力机制,擅长捕捉序列中的长距离依赖,这成为其应用于网络流量时间序列预测的基础。Kong等人[62]利用Transformer架构中自注意力机制的优势,解决了RNN和LSTM在处理长序列时易出现的梯度消失和信息丢失问题,提升了网络链接的质量和效率。然而,该方法虽然解决了长距离依赖的捕捉难题,但其标准架构在处理超长连续数值序列时,计算开销随序列长度呈二次方增长,且对局部细粒度波动特征的提取能力相对较弱。

除了单一节点的预测精度,跨区域的模型迁移也是边缘网络流量预测中的一大挑战。针对该难题,Xu等人[63]提出了TransMUSE框架,该框架通过服务聚类和节点分组,在每个群组内部署基于Transformer的多服务流量预测模型,该模型可以在同一群组的节点间直接迁移而无需微调。然而,TransMUSE依赖初始的服务聚类和节点分组,如果网络流量模式随时间动态变化剧烈,导致预先划分的群组不再准确时,模型的迁移性能可能会受到影响。

上述基于Transformer及其混合架构的研究工作,为加密流量预测提供了强大的序列建模基础。然而,这些研究工作仍依赖于足够的标注数据或特定的下游任务微调。为进一步解决泛化性和数据依赖性问题,研究者转向了通用基础模型的预训练范式。

2.3.2　基于基础模型的流量预测

通用基础模型的理念,是将大规模未标注数据作为知识源,通过预训练获得通用表征,以支持多种下游任务。这一范式为克服加密流量分析中数据稀疏和标注成本高昂的挑战提供了新的解决路径。Lin等人[26]率先提出了ET-BERT模型。该模型利用Transformer架构,在海量未标注数据上进行预训练,学习

深度且上下文感知的数据报级通用表征，虽然其核心是为分类任务服务，但其学习到的鲁棒特征为预测任务提供了高质量的输入基础。紧随其后，Van Langendonck等人[64]通过将加密流量建模为动态时空图，并利用自监督链路预测作为预训练任务，学习通用网络图框架中的时空动态和表示。这种预训练后的基础模型能够通过较小的微调适应多种网络任务或环境。通过预训练，流量模型获得了更强大的泛化性和更低的标注数据依赖。

2.3.3 基于集成大型语言模型的精准预测

大型语言模型拥有强大的序列生成能力和世界知识，将其应用于网络流量预测，代表了从专用模型到通用智能体的范式转变。Diaf等人[65]提出了BARTPredict框架，该框架利用LLM的强大能力对物联网网络中的网络威胁进行预测，并采用了双LLM协同的机制，其中双向自回归模型作为一种序列到序列的架构，能够有效地捕捉和学习时间序列数据中的复杂模式，而另一个经过微调的模型则被用于对预测出的流量进行评估。该框架最大的优点在于其利用双向自回归模型在序列生成方面的优势，使预测结果保持高准确度，为后续的安全评估提供了高质量的输入。该团队的另一项工作，Diaf等人[66]通过主动的方式来预测和缓解恶意活动，从而克服传统入侵检测系统的被动性。该方法率先将LLM强大的序列建模和文本理解能力应用于结构化的网络流量预测任务，实现了网络安全从被动检测向主动预测的范式转变。上述团队的两个方案最大的局限性在于其高昂的计算成本和部署复杂性，在一个反馈回路中需要同时训练和运行三个复杂的深度学习模型，在资源受限的物联网环境下将面临严峻的挑战。

上述方法揭示了流量预测从“确定性数值回归”向“概率性序列生成”的演进。相比适于短周期拟合的传统回归模型，生成式路线通过将流量状态词元化，在捕捉长尾分布与突发性方面展现出显著优势。然而，将大语言模型应用于连续数值预测仍面临三大固有挑战：一是模态适配鸿沟，数值词元化易破坏流量的连续性特征，导致预测精度在某些高频波动场景下不如LSTM等传统的时序模型；二是计算复杂度约束，处理长周期流量依赖时仍存在计算瓶颈；三是推理延迟高昂，LLM的高延迟推理特性可能导致控制回路滞后，难以满足现网资源调度的时效性需求。

2.4 加密流量生成

传统的流量生成方法多依赖于手工设计统计特征或规则模板[67]，难以捕捉复杂动态行为。而大模型通过从原始数据中自动提取高阶语义信息，显著提升了生成流量的真实性与泛化能力。这类生成技术不仅能够提升数据覆盖度与代表性，还能在多种场景中发挥作用，例如：构建高度还原的网络环境用于协议与系统性能评估、在网络安全测试中模拟攻击与异常行为验证防护策略、为机器学习任务提供丰富的训练数据实现数据增强，以及在对抗样本构造与流量仿真中探索潜在的脆弱点。通过大语言模型的通用表示与跨任务适配能力，流量生成可兼顾真实性与可控性，为网络基础设施研发、运维与安全保障提供有力支持。

流量生成的基本框架如图5所示。本节将系统梳理该领域的技术演进：从大语言模型在流量生成中的初步尝试，到协议感知与多任务协同的基础模型构建，最终迈向面向安全对抗、多协议集成与隐私约束等复杂场景的高阶生成方法，全面分析基于大语言模型的加密流量生成技术在网络空间建模中的前沿进展。

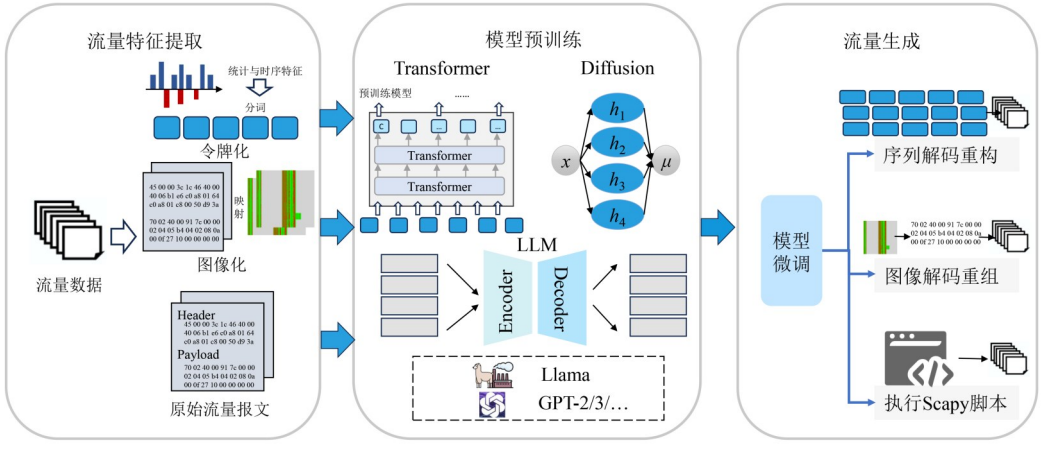

图5 流量生成基本框架

Figure 5 Basic framework for traffic generation

2.4.1　大模型架构在流量生成中的基础框架

随着生成式人工智能在自然语言处理领域的成功,研究者开始尝试将大模型架构迁移至网络流量领域,以突破传统生成方法的表达瓶颈。这一阶段的核心目标是验证大模型能否有效建模网络流量的时序依赖与统计规律,并建立端到端的流量生成范式。早期工作聚焦于如何将原始字节流或统计特征转化为适合语言模型处理的序列形式,并利用自回归机制实现逐令牌生成。

Meng 等人[68]基于 GPT-2 模型提出了 NetGPT,这是首个面向网络流量理解与生成任务的生成式预训练模型。该方法采用统一文本编码策略,将多模式的网络流量转化为连续令牌序列,并利用 GPT 系列模型的单向生成特性,通过最后一个令牌预测后续流量,实现了文本到文本的流量生成。此外,NetGPT 在预训练和微调阶段引入了报文分段、报头字段重排以及任务提示词设计,从而支持多种类型的流量生成。然而,该方法依赖单向上下文建模,在需要双向复杂流量信息任务中仍存在一定局限性。

在 NetGPT 工作的基础上,后续研究围绕 GPT-2 模型在序列长度与建模效率方面进行了进一步的优化。Qu 等人[69]提出的 TrafficGPT 延续了将网络流量转化为令牌序列的技术路线,但在架构上引入线性注意力与局部注意力机制,取代传统 Transformer 的二次复杂度自注意力,将可处理的令牌长度上限提升至 12 032 个。TrafficGPT 通过生成预训练与自监督学习,能够以输入一段令牌序列为起点,生成结构特征与分布均接近真实网络流量的数据。该方案在提升长序列建模能力方面表现突出,尤其适用于需要完整上下文的流量生成任务,但模型规模与计算成本较高,在资源受限场景下的落地仍面临挑战。

上述工作标志着网络流量生成正式迈入大模型时代,初步验证了 Transformer 架构在流量序列建模中的可行性。然而,这些早期方法多将流量视为黑盒字节序列,缺乏对协议语义结构的显示建模,也未充分挖掘生成模型在多任务协同与跨场景泛化方面的潜力。因此,后续研究逐步转向对流量内在结构的理解与统一表征学习,推动生成模型从简单的流量生成向懂协议、会推理、可复用的方向演进。

2.4.2　基于生成式基础模型的多任务流量生成

在 NetGPT 和 TrafficGPT 开创性工作的基础上,研究重心开始从单纯模仿流量分布转向对网络协议语义结构的深度建模。这一阶段的核心是如何使生成的流量不仅看起来像真实流量,更能从行为上符合协议规范。为此,研究者提出了协议感知的令牌化策略、统一特征嵌入空间以及新型基础模型架构,显著提升了生成流量的结构性保真度与任务适配能力。

为强化生成流量对协议语法与语义的遵循能力,Zhao 等人[70]提出了一种生成式预训练基础模型 GBC。该方法设计了协议感知的令牌化策略,将不同协议及其字段进行语义分割编码,以保留完整协议结构;在大规模无标签流量数据上,通过自回归方式进行生成式预训练,学习真实流量的时序与结构模式;在下游生成任务中,GBC 能够按 token 逐步生成符合协议规范的完整流量包,并可通过添加协议字段提示实现特定类型或任务定制化生成,从而有效用于少数数据增强。

面对异构流量日志的统一表征与多任务协同需求,Zhou 等人[71]提出了一种面向加密流量动态生成的预训练基础模型方法——NetFlowGen 框架。该框架针对 IP 地址、端口、协议、包数、字节数及时间戳等异构 NetFlow 数据,设计了统一的特征嵌入流程,将不同类型流量特征映射至同一向量空间;随后采用类 GPT 的 Transformer 解码器架构,在大规模无标签数据下进行生成预训练,并通过微调适配流量分类、拥塞预测、攻击检测等下游任务。该框架虽验证了多任务的有效性,但在跨域流量分布迁移以及对噪声数据的稳健性等方面仍存在优化空间。

为突破传统注意力机制在长序列建模中的效率瓶颈,并更好地捕捉流量会话的状态演化规律,Chu 等人[72]提出 NetSSM,一种基于状态空间模型的多流网络流量生成框架。该框架通过在大规模真实流量上预训练,并针对握手、数据传输、会话结束等阶段微调,实现状态感知的包级流量生成。实验表明,NetSSM 在多项分布一致性指标上显著优于现有方法,并在长序列多流场景保持生成模式稳定性,体现了大模型驱动高保真网络流量生成的优势。

2.4.3　面向安全对抗与复杂场景的流量生成方法

随着生成式基础模型能力的日趋成熟,其应用边界迅速正从实验室仿真向真实的复杂网络场景延伸。这一阶段的研究不再局限于生成逼真流量,而是聚焦于如何在安全对抗、多协议协同、隐私保护等高阶需求下,生成可控、可靠、可解释的加密流量。为此,研究者们引入混合专家系统、跨模态生成、条件引导等技术,构建了面向特定任务的集成化生成框架。这不仅显著提升了大模型在实际系统中的实用价值,更推动其在实战中落地应用。复杂场景的典型加密流量生成方法总结如表 7 所示。

为分析深度学习检测系统的脆弱性,Sun 等人[73]提出了一种对抗性流量生成方法 AdvTG。该框架将大型语言模型与强化学习相结合,首先构建多种基于文本和图像特征的深度学习检测器,作为攻击目标及

表7 面向安全对抗与复杂场景的流量生成工作总结

Table 7 Summary of traffic generation research for security countermeasures and complex scenarios

方法	年份	流量表示	模型架构	数据集	评估
AdvTG[73]	2025	原始流量包前784字节	Meta-Llama-3-8B	自收集数据集、CiCIDS、CICIoT	检测模型欺骗成功率
PAC-GPT[74]	2023	协议字段解析结果	GPT-3	DARPA、KDD'99、TON	生成数据包的有效性
MoE-based[75]	2025	会话协议字段	GPT-3.5 Turbo	TON、自生成数据集	生成数据包的有效性
NetDiffusion[76]	2024	原始流量包前1 024字节	Stable-Diffusion	自收集本地真实环境数据	Jensen-Shannon散度

奖励评估器;之后利用LoRA方法在超文本传输协议(HyperText Transfer Protocol, HTTP)等协议数据上微调语言模型,使其生成的流量在保留功能与遵循协议规范的同时具备语义合理性;最后引入强化学习,根据检测器的反馈不断调整生成策略,迭代优化生成结果以提升攻击成功率。该框架揭示了现有基于深度学习的流量检测方法的脆弱性,但该框架在跨协议泛化与生成效率方面仍存在提升空间。

为实现多协议流量的端到端自动化合成,Kholgh等人[74]提出了PAC-GPT方法。该方法将加密流量转换为结构化文本描述,并利用GPT-3系列模型进行序列生成,通过特定模块将生成结果准确恢复为可传输的数据格式。该方法通过链式调用多个大型语言模型,支持在同一流程中按需生成ICMP与DNS等异构协议流量,通过命令行接口实现端到端的自动化数据生成。然而,在面临复杂协议时,该方法生成的结构和字段一致性仍存在不足,限制了其在高复杂度、高保真流量场景中的可靠性与泛化能力。

为提升多协议协同生成的准确性与结构一致性,Delgado-Soto等人[75]提出了一种基于GPT-3.5 Turbo与混合专家的多协议流量生成方法。该方法通过少样本提示与GPT-3.5 Turbo微调,分别训练出面向ICMP、DNS、TCP、HTTP等协议的专家模型,将流量摘要精准映射为Scapy代码,并利用路由机制根据协议类别调用专家模型生成流量。尽管该MoE架构在多协议场景下显著优于单一模型,但其对提示工程的过度依赖与各专家独立训练的高昂成本,导致系统在扩展新协议时面临较重的定制化负担。

在基于Transformer的预训练模型为网络流量生成提供强大序列建模能力的同时,扩散模型凭借其在图像生成领域的突出表现,为流量数据的跨模态合成开辟了全新路径。为突破传统序列生成范式并实现跨模态条件控制下的高保真生成,Jiang等人[76]提出了NetDiffusion方法。该方法首先将原始数据包编码为二进制比特流,再映射为图像表示,从而使流量生成问题转化为图像生成任务;随后基于带标签的流量图像数据集微调生成模型,以提升特定协议下的内容生成质量;在生成阶段,该方法不仅对协议类型与包头字段进行精确控制,还通过嵌入领域专家规则对生成结果进行过滤与修正,以确保流量在结构和功能上的有效性。该方法依赖图像化编码过程,在跨协议泛化与高维流量特征保留方面仍存在不足。

为破解数据匮乏与隐私壁垒,现有加密流量生成技术主要衍生出属性条件生成与协议结构化合成两大范式。前者利用语义提示增强样本多样性以提升下游分类器鲁棒性;后者则通过拟合协议状态机合成规范报文。然而,尽管LLM在生成高保真流量方面取得了突破,但其在协议合规性与可控性上仍存在局限。由于LLM本质上是基于概率的生成模型,其生成的流量虽然在统计特征上逼近真实样本,却依然存在协议违规、生成虚假字段等问题,导致生成的流量无法在真实网络环境中传输。此外,生成模型可能会记忆并泄露训练数据中的真实IP等敏感隐私信息,这在用于数据增强时构成了潜在的安全隐患。

3 分析与讨论

根据前文的分析,基于大语言模型的加密流量分析方法已经在流量分类、恶意流量检测、流量预测和流量生成等方面取得了显著进展。在流量分类方面,LLM通过针对特定任务的预训练显著提升了跨协议泛化能力;在恶意流量检测方面,其强大的上下文建模能力增强了对检测隐蔽攻击的鲁棒性;在流量预测方面,Transformer架构能够有效捕捉流量中复杂的时序依赖信息,进而提高了预测精度;而在流量生成方面,生成式大模型逐步实现了协议合规的流量生成。然而,现有研究仍普遍存在数据表征不适配、计算资源开销高、可解释性缺失等共性瓶颈,制约了其在资源受限的真实网络环境中的部署可行性。综上所述,面对加密流量分析领域面临的挑战,未来的研究不应仅局限于单一模型的参数优化,而需从数据生态构建、可信安全防御、协议适应性及模型轻量化协同等维度构建系统的技术突破路径。

(1)构建高保真与多模态的基准数据集:当前研究普遍面临训练环境与实际部署环境存在数据分布偏移的挑战。未来的数据集构建需突破单一环境限制,构建包含真实网络噪声、长尾流量分布及多协议混合的数据集。同时,推动数据形态从单一的流量载荷向流量及日志相结合的多模态方向演进,利用多源

异构信息弥补单一模态在语义表征上的不足,为模型提供更丰富的上下文理解能力。

(2)迈向可解释与可验证的可信分析框架:针对大模型缺乏透明度带来的信任挑战,未来的研究需致力于提升模型的可解释性,使网络管理员能够直观理解模型判定背后的侧信道依据。同时,建立模型行为的可验证机制也至关重要,通过引入形式化验证或逻辑约束手段,精准识别新型网络威胁并修正模型在对抗环境下的非预期推理行为,确保其在关键安全决策中的可靠性。

(3)挖掘适应协议演进的新型鲁棒特征空间:面对网络协议全面加密化趋势,传统依赖显式元数据的特征提取方法逐渐失效。未来应挖掘独立于负载内容的深层隐式特征空间,利用大模型强大的时序建模能力,捕捉细粒度的流突发行为、包到达间隔序列及加密指纹的自适应表征,从而构建不受加密协议影响的鲁棒特征体系。

(4)平衡隐私保护与对抗安全的长效机制:随着大模型在攻防两端的同步应用,构建具备对抗鲁棒性的分析方法迫在眉睫。需研究如何防御流量扰动和提示注入等新型攻击手段。此外,探索联邦学习与大模型的深度结合,在严格保障数据主权的前提下实现跨域知识融合,将是打破数据孤岛限制并满足隐私合规要求的核心技术路径。

(5)面向异构场景的泛化部署与能效优化:通用大模型普遍存在参数量庞大与推理资源消耗高的问题,这与网络流量分析场景的异构性形成了显著矛盾。骨干网环境对实时吞吐量要求极高,而边缘网关与物联网设备则严格受限于计算功耗与存储空间,直接部署全量大模型往往缺乏实际可行性。未来的研究必须将部署可行性与泛化性纳入核心考量维度,致力于解决高资源投入与低延迟需求之间的适配难题。重点突破方向包括模型知识蒸馏、低比特量化及结构剪枝等轻量化技术,以及探索云端大模型负责复杂离线分析、边缘侧小模型负责实时在线推理的端云协同架构,从而在保障模型精度的同时显著降低算力与能耗开销,实现从理论模型到现网部署的有效落地。

4　结束语

本文对基于大语言模型的加密流量分析方法进行了系统性综述,从流量分类、恶意流量检测、流量预测和流量生成四个主要方向进行了全面梳理。研究表明,大语言模型方法通过预训练-微调的范式革新,显著提升了模型在不同任务间的迁移能力,为加密流量分析提供了新的解决路径。在流量分类方面,大语言模型能够有效提取加密流量中的统计特征与时序模式,实现对终端应用的非侵入式判定;在恶意流量检测方面,大语言模型可以挖掘流量的统计特征、时序模式和行为异常,识别隐蔽在加密流量中的恶意活动;在流量预测方面,基于大语言模型的预测方法可为网络管理系统提供前瞻性决策支持;在流量生成方面,大语言模型技术能够构建高保真、可控制的模拟流量,服务于测试验证与数据增强。

本文对当前基于大语言模型的加密流量分析方法面临的挑战与未来发展方向进行了分析。一是高质量标注数据的缺乏,导致模型训练困难;二是模型在边缘设备上的部署效率问题,限制了在实时检测等资源受限场景的应用;三是对加密流量动态伪装的适应性不足,需要更灵活的模型架构。未来研究应着重于开发更高效的预训练方法、构建高质量的加密流量数据集、优化模型在资源受限环境下的部署,以及探索大语言模型与加密流量分析的深度融合,以进一步提升加密流量分析的准确性和实用性,为构建更加安全、智能的网络环境提供技术支撑。

参考文献

[1] 中国互联网络信息中心. 第56次《中国互联网络发展状况统计报告》[R/OL]. (2025-07-21) [2025-10-30]. https://www.cnnic.cn/n4/2025/0721/c88-11328.html.
China Internet Network Information Center. The 56th statistical report on the development of the internet in China[R/OL]. (2025-07-21) [2025-10-30]. https://www.cnnic.cn/n4/2025/0721/c88-11328.html. (in Chinese)

[2] Shen Meng, Zhang Jinpeng, Zhu Liehuang, et al. Accurate decentralized application identification via encrypted traffic analysis using graph neural networks[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2367-2380.

[3] Zhou Guangmeng, Guo Xiongwen, Liu Zhuotao, et al. TrafficFormer: An efficient pre-trained model for traffic data[C]//2025 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2025: 1844-1860.

[4] Shen Meng, Wu Jinhe, Ye Ke, et al. Robust detection of malicious encrypted traffic via contrastive learning[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 4228-4242.

[5] Fu Chuanpu, Li Qi, Bertino E, et al. Training with only 1.0‰ samples: Malicious traffic detection via cross-modality feature fusion[C]//Proceedings of 2025 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2025: 3930-3944.

[6] Zhang Junbo, Zheng Yu, Qi Dekang, et al. DNN-based prediction model for spatio-temporal data[C]//Proceedings of

the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2016: 92.

[7] Li Haibin, Zhao Yi, Yao Wenbing, et al. Towards real-time ML-based DDoS detection via cost-efficient window-based feature extraction[J]. Science China Information Sciences, 2023, 66(5): 152105.

[8] Fu Chuanpu, Li Qi, Xu Ke, et al. Point cloud analysis for ML-based malicious traffic detection: Reducing majorities of false positive alarms[C]//Proceedings of 2023 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2023: 1005-1019.

[9] Shen Meng, Ji Kexin, Gao Zhenbo, et al. Subverting website fingerprinting defenses with robust traffic representation[C]//32nd USENIX Security Symposium. USENIX Association, 2023: 607-624.

[10] 康海燕, 张义钒, 王楠敏. 基于联邦大模型的网络攻击检测方法研究[J]. 电子学报, 2025, 53(6): 1792-1804.
Kang Haiyan, Zhang Yifan, Wang Nanmin. Research on network attack detection method based on federated large model[J]. Acta Electronica Sinica, 2025, 53(6): 1792-1804. (in Chinese)

[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2017: 6000-6010.

[12] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[PP/OL]. V2. arXiv (2024-05-31) [2025-10-30]. https://arxiv.org/abs/2312.00752.

[13] Shen Meng, Ye Ke, Liu Xingtong, et al. Machine learning-powered encrypted network traffic analysis: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 791-824.

[14] 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. DOI:10.13328/j.cnki.jos.006891.
Hou Jian, Lu Hui, Liu Fangai, et al. Detection and countermeasure of encrypted malicious traffic: A survey[J]. Journal of Software, 2024, 35(1): 333-355. (in Chinese)

[15] 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 Yu, Liu Taotao, Wang Kun, et al. Survey of research on encrypted traffic classification based on machine learning[J]. Journal on Communications, 2025, 46(1): 167-191. (in Chinese)

[16] 王钢, 高雲鹏, 杨松儒, 等. 基于深度学习的加密恶意流量检测方法研究综述[J]. 信息网络安全, 2025, 25(8): 1276-1301.
Wang Gang, Gao Yunpeng, Yang Songru, et al. A survey on deep learning-based encrypted malicious traffic detection methods[J]. Netinfo Security, 2025, 25(8): 1276-1301. (in Chinese)

[17] Bonelli N, Giordano S, Procissi G. Enabling packet fan-out in the libpcap library for parallel traffic processing[C]//Proceedings of 2017 Network Traffic Measurement and Analysis Conference. Piscataway: IEEE, 2017: 1-9.

[18] Chappell L. Wireshark network analysis[M]. 2nd ed. Laura: Laura Chappell University, 2012.

[19] Hofstede R, Čeleda P, Trammell B, et al. Flow monitoring explained: From packet capture to data analysis with Netflow and IPFIX[J]. IEEE Communications Surveys & Tutorials, 2014, 16(4): 2037-2064.

[20] 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.

[21] Sharafaldin I, Lashkari A H, Ghorbani A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//Proceedings of the 4th International Conference on Information Systems Security and Privacy. [S.l.]: SciTePress, 2018: 108-116.

[22] Mirsky Y, Doitshman T, Elovici Y, et al. Kitsune: An ensemble of autoencoders for online network intrusion detection[C]//25th Annual Network and Distributed System Security Symposium. [S.l.]: The Internet Society, 2018.

[23] Wang Wei, Zhu Ming, Zeng Xuewen, et al. Malware traffic classification using convolutional neural network for representation learning[C]//2017 International Conference on Information Networking. Piscataway: IEEE, 2017: 712-717.

[24] 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. Piscataway: IEEE, 2018: 1-7.

[25] Draper-Gil G, Lashkari A H, Mamun M S I, et al. Characterization of encrypted and VPN traffic using time-related features[C]//Proceedings of the 2nd International Conference on Information Systems Security and Privacy. SciTePress, 2016: 407-414.

[26] Lin Xinjie, Xiong Gang, Gou Gaopeng, et al. ET-BERT: A contextualized datagram representation with pre-training transformers for encrypted traffic classification[C]//Proceedings of ACM Web Conference 2022. New York: ACM, 2022: 633-642.

[27] MontazeriShatoori M, Davidson L, Kaur G, et al. Detection of doh tunnels using time-series classification of en-

crypted traffic[C]//2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress. Piscataway: IEEE, 2020: 63-70.

[28] Shen Meng, Wei Mingwei, Zhu Liehuang, et al. Classification of encrypted traffic with second-order Markov chains and application attribute bigrams[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(8): 1830-1843.

[29] Gao Li, Fu Chuanpu, Deng Xinhao, et al. Wedjat: Detecting sophisticated evasion attacks via real-time causal analysis[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1. New York: ACM, 2025: 342-353.

[30] Xu Songsong, Fu Chuanpu, Li Qi, et al. "One Model Fits All Nodes": Neuron activation pattern analysis-based attack traffic detection framework for P2P networks[J]. IEEE Transactions on Networking, 2025, 33(4): 1630-1645.

[31] Shen Meng, Liu Yiting, Zhu Liehuang, et al. Fine-grained webpage fingerprinting using only packet length information of encrypted traffic[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2046-2059.

[32] Fu Chuanpu, Li Qi, Shen Meng, et al. Detecting tunneled flooding traffic via deep semantic analysis of packet length patterns[C]//Proceedings of 2024 on ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2024: 3659-3673.

[33] Cui Tianyu, Lin Xinjie, Li Sijia, et al. TrafficLLM: Enhancing large language models for network traffic analysis with generic traffic representation[PP/OL]. V2. arXiv (2025-04-15) [2025-10-30]. https://arxiv.org/abs/2504.04222.

[34] Shen Meng, Zhang Jinpeng, Zhu Liehuang, et al. Encrypted traffic classification of decentralized applications on ethereum using feature fusion[C]//Proceedings of International Symposium on Quality of Service. New York: ACM, 2019: 18.

[35] Wu Songyun, Dong Enhuan, He Hong, et al. ALM: A two-stage traffic anomaly detection and analysis system via the large language model[C]//2025 IEEE Network Operations and Management Symposium. Piscataway: IEEE, 2025: 1-7.

[36] Wickramasinghe N, Shaghaghi A, Tsudik G, et al. SoK: Decoding the enigma of encrypted network traffic classifiers[C]//2025 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2025: 1825-1843.

[37] Shen Meng, Zhang Jinpeng, Xu Ke, et al. DeepQoE: Real-time measurement of video QoE from encrypted traffic with deep learning[C]//2020 IEEE/ACM 28th International Symposium on Quality of Service. Piscataway: IEEE, 2020: 1-10.

[38] Liu Ziao, Xie Yuanyuan, Luo Yanyan, et al. TransECA-Net: A transformer-based model for encrypted traffic classification[J]. Applied Sciences, 2025, 15(6): 2977.

[39] Zhou Qiang, Wang Liangmin, Zhu Huijuan, et al. WF-transformer: Learning temporal features for accurate anonymous traffic identification by using transformer networks[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 30-43.

[40] Meng Zhiyan, Liu Dan, Meng Jintao. LETFORMER: Lightweight transformer pre-training with sharpness-aware optimization for efficient encrypted traffic analysis[J]. International Journal of Innovative Computing, Information and Control, 2025, 21(2): 359-371.

[41] Pan Quanbo, Yu Yang, Yan Hanbing, et al. FlowBERT: An encrypted traffic classification model based on transformers using flow sequence[C]//2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications. Piscataway: IEEE, 2023: 133-140.

[42] Chen Xuyang, Han Lu, Zhan Dechuan, et al. MIETT: Multi-instance encrypted traffic transformer for encrypted traffic classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(15): 15922-15929.

[43] Shi Zhaolei, Luktarhan N, Song Yangyang, et al. BFCN: A novel classification method of encrypted traffic based on BERT and CNN[J]. Electronics, 2023, 12(3): 516.

[44] Farrukh Y A, Wali S, Khan I, et al. XG-NID: Dual-modality network intrusion detection using a heterogeneous graph neural network and large language model[J]. Expert Systems with Applications, 2025, 287: 128089.

[45] Luo Jian, Chen Zechao, Chen Wenxiong, et al. A study on the application of the T5 large language model in encrypted traffic classification[J]. Peer-to-Peer Networking and Applications, 2025, 18(1): 15.

[46] Antari A, Abo-Aisheh Y, Shamasneh J, et al. Network traffic classification using machine learning, transformer, and large language models[C]//2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI). Piscataway: IEEE, 2025: 1-5.

[47] Fu Chuanpu, Li Qi, Xu Ke. Flow interaction graph analysis: Unknown encrypted malicious traffic detection[J]. IEEE/ACM Transactions on Networking, 2024, 32(4): 2972-2987.

[48] Fu Chuanpu, Li Qi, Shen Meng, et al. Realtime robust malicious traffic detection via frequency domain analysis[C]//Proceedings of 2021 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2021: 3431-3446.

[49] Liang Yu, Li Pengchao, Lai Qiunan, et al. EM-BERT: A language model based method to detect encrypted malicious network traffic[M]//Proceedings of International Conference on Image, Vision and Intelligent Systems 2023 (ICIVIS 2023). Singapore: Springer, 2024: 580-589.

[50] Stein K, Mahyari A A, Francia G, et al. Towards novel malicious packet recognition: A few-shot learning approach[C]//2024 IEEE Military Communications Conference. Piscataway: IEEE, 2024: 847-852.

[51] Naseer M, Ullah F, Ijaz S, et al. Obfuscated malware detection and classification in network traffic leveraging hybrid large language models and synthetic data[J]. Sensors, 2025, 25(1): 202.

[52] Chaves-Tibaduiza J F, Becerra-Muñoz A I, Robledo-Giron A, et al. On the feasibility of using an encoder-only model for anomaly detection: The BERTAD approach[C]//2024 IEEE Colombian Conference on Communications and Computing. Piscataway: IEEE, 2024: 1-6.

[53] Yang Z, Jin Y, Liu J, et al. Research on cloud platform network traffic monitoring and anomaly detection system based on large language models[C]//2025 IEEE 7th International Conference on Communications, Information System and Computer Engineering (CISCE). Piscataway: IEEE, 2025: 1029-1032.

[54] Zhao Jianming, Jin Ziwen, Zeng Peng, et al. An anomaly detection method for oilfield industrial control systems fine-tuned using the llama3 model[J]. Applied Sciences, 2024, 14(20): 9169.

[55] Rezaei H, Taheri R, Shojafar M. FedLLMGuard: A federated large language model for anomaly detection in 5G networks[J]. Computer Networks, 2025, 269: 111473.

[56] Li Qingyang, Zhang Yihang, Jia Zhidong, et al. DoLLM: How large language models understanding network flow data to detect carpet bombing DDoS[PP/OL]. V1. arXiv (2024-05-13) [2025-10-30]. https://arxiv.org/abs/2405.07638.

[57] Lin Xinjie, Xiong Gang, Gou Gaopeng, et al. Respond to change with constancy: Instruction-tuning with LLM for Non-I.I.D. network traffic classification[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 5758-5773.

[58] Wu Dongming, Peng Zhiyuan, Chen Mingzhe, et al. Transforming network intrusion detection using large language models[C]//2025 IEEE 22nd Consumer Communications & Networking Conference. Piscataway: IEEE, 2025: 1-9.

[59] Xu Hongbo, Si Chengxiang, Zhouzhou, et al. APTSniffer: Detecting APT attack traffic using retrieval-augmented large language models[C]//2025 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2025: 1-5.

[60] Shen Meng, Ji Kexin, Wu Jinhe, et al. Real-time website fingerprinting defense via traffic cluster anonymization[C]//2024 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2024: 3238-3256.

[61] Liu Zixuan, Zhao Yi, Liu Zhuotao, et al. A hard-label black-box evasion attack against ML-based malicious traffic detection systems[PP/OL]. V1. arXiv (2025-10-16) [2025-10-30]. https://arxiv.org/abs/2510.14906.

[62] Kong Qian, Zhang Xu, Zhang Chongfu, et al. Network traffic prediction: Apply the transformer to time series forecasting[J]. Mathematical Problems in Engineering, 2022, 2022: 8424398.

[63] Xu Luyang, Liu Haoyu, Song Junping, et al. TransMUSE: Transferable traffic prediction in multi-service edge networks[J]. Computer Networks, 2023, 221: 109518.

[64] Van Langendonck L, Castell-Uroz I, Barlet-Ros P. Towards a graph-based foundation model for network traffic analysis[C]//Proceedings of the 3rd GNNet Workshop on Graph Neural Networking Workshop. New York: ACM, 2024: 41-45.

[65] Diaf A, Korba A A, Karabadji N E, et al. BARTPredict: Empowering IoT security with LLM-driven cyber threat prediction[C]//2024 IEEE Global Communications Conference. Piscataway: IEEE, 2024: 1239-1244.

[66] Diaf A, Korba A A, Karabadji N E, et al. Beyond detection: Leveraging large language models for cyber attack prediction in IoT networks[C]//2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things. Piscataway: IEEE, 2024: 117-123.

[67] Sommers J, Kim H, Barford P. Harpoon: A flow-level traffic generator for router and network tests[J]. ACM SIGMETRICS Performance Evaluation Review, 2004, 32(1): 392.

[68] Meng Xuying, Lin Chungang, Wang Yequan, et al. NetGPT: Generative pretrained transformer for network traffic[PP/OL]. V3. arXiv (2025-08-28) [2025-10-30]. https://arxiv.org/abs/2304.09513.

[69] Qu J, Ma X, Li J. TrafficGPT: Breaking the token barrier for efficient long traffic analysis and generation[PP/OL]. V2. arXiv (2024-03-18) [2025-10-30]. https://arxiv. org/abs/

2403.05822.

[70] Zhao Di, Jiang Bo, Liu Song, et al. Language of network: A generative pre-trained model for encrypted traffic comprehension[PP/OL]. V1. arXiv (2025-05-26) [2026-05-07]. https://arxiv.org/abs/2505.19482.

[71] Zhou Jiawei, Kim W, Xu Zhiying, et al. Netflowgen: Leveraging generative pre-training for network traffic dynamics[PP/OL]. V1. arXiv (2024-12-30) [2025-10-30]. https://arxiv.org/abs/2412.20635.

[72] Chu A, Jiang X, Liu S, et al. Netssm: Multi-flow and state-aware network trace generation using state-space models[J]. Proceedings of the ACM on Networking, 2026, 4(CoNEXT1): 1-24.

[73] Sun Peishuai, Yun Xiaochun, Li Shuhao, et al. AdvTG: An Adversarial traffic generation framework to deceive dl-based malicious traffic detection models[C]//Proceedings of the ACM on Web Conference 2025. New York: ACM, 2025: 3147-3159.

[74] Kholgh D K, Kostakos P. PAC-GPT: A novel approach to generating synthetic network traffic with GPT-3[J]. IEEE Access, 2023, 11: 114936-114951.

[75] Delgado-Soto J A, López de Vergara J E, González I, et al. GPT on the wire: Towards realistic network traffic conversations generated with large language models[J]. Computer Networks, 2025, 265: 111308.

[76] Jiang Xi, Liu Shinan, Gember-Jacobson A, et al. NetDiffusion: Network data augmentation through protocol-constrained traffic generation[J]. Proceedings of the ACM on Measurement and Analysis of Computing Systems, 2024, 8(1): 11.

作者简介

沈　蒙　男,1988年1月出生于山东省德州市。现为北京理工大学网络空间安全学院教授、博士生导师。主要研究方向为网络加密流量分析、数据隐私安全、区块链应用。中国电子学会会员编号:E190019899M。

E-mail: shenmeng@bit.edu.cn

刘俣凯　男,1998年6月出生于河北省石家庄市。北京理工大学网络空间安全学院博士研究生。主要研究方向为网络加密流量分析、恶意流量检测。中国电子学会会员编号:E190199178A。

E-mail: liuyukai@bit.edu.cn

艾俊宇　男,2002年7月出生于河北省唐山市。北京理工大学网络空间安全学院硕士研究生。主要研究方向为网络加密流量分析、恶意流量检测。

E-mail: zhaobufan@bit.edu.cn

任琛琛　女,2003年4月出生于河北省保定市。北京理工大学网络空间安全学院硕士研究生。主要研究方向为网络加密流量分析、网站指纹攻击。

E-mail: chenchenren@bit.edu.cn

沈泓烨　女,2004年10月出生于江苏省苏州市。北京理工大学网络空间安全学院本科生。主要研究方向为网络加密流量分析、恶意流量检测。

E-mail: 1120231080@bit.edu.cn

李　琦　男,1979年1月出生于山东省东营市。现为清华大学网络研究院副教授、博士生导师。主要研究方向为互联网安全、人工智能安全。中国电子学会会员编号:E190026656S。

E-mail: qli01@tsinghua.edu.cn

徐　恪　男,1974年12月出生于江苏省淮安市。现为清华大学计算机系教授、博士生导师。主要研究方向为新一代互联网体系结构、网络空间安全。中国电子学会会员编号:E190008630F。

E-mail: xuke@tsinghua.edu.cn

祝烈煌　男,1976年9月出生于浙江省衢州市。现为北京理工大学网络空间安全学院教授、博士生导师。主要研究方向为密码学、网络和信息安全。中国电子学会会员编号:E190010255M。

E-mail: liehuangz@bit.edu.cn