加密视频流量识别研究综述

赵航宇[1)] 吴桦[1),2),3)] 成果[1)] 银晓悦[1)] 刘嵩涛[1)] 程光[1),2),3),4)] 胡晓艳[1),2),3)]

[1)](东南大学 网络空间安全学院,南京 211189)

[2)](网络通信与安全紫金山实验室,南京 211111)

[3)](网络空间国际治理研究基地(东南大学),南京 211189)

[4)](江苏省泛在网络安全工程研究中心,南京 211189)

摘 要 随着互联网视频内容消费的持续增长，加密视频流量在全球网络中的比例显著提升，给网络空间治理带来了前所未有的挑战。视频平台采用端到端加密和各类隐私增强技术，极大增强了用户隐私保护能力，但同时削弱了基于明文内容解析的传统监管手段的有效性。如何在保障隐私前提下，实现对加密视频流量中公害视频的精准识别，已成为网络空间安全领域亟需突破的核心科学问题。本文系统梳理了加密视频流量识别领域的发展脉络，围绕特征构建、识别模式与应用场景，构建了基于特征工程的匹配识别法和基于机器学习的分类识别法两大技术体系。基于特征工程的匹配识别法主要包括基于密文指纹的匹配识别法与基于明文指纹的匹配识别法两类方法，前者侧重于从加密流量中直接提取传输特征，后者结合应用层信息与流量侧信道特征，均实现了在不同协议和复杂环境下的加密视频流量识别。基于机器学习的分类识别法则通过提取统计特征，采用分类模型对加密视频流量进行识别，提升了加密视频流量识别的适应性和泛化能力。近年来，研究者们围绕HTTP/1.1、HTTP/2、HTTP/3、加密代理、隧道、匿名网络等多种协议和场景，提出了具有针对性的特征设计与匹配算法。针对短视频、社交视频、无线网络等新兴场景，研究还提出了混合特征建模和协议自适应等创新方案，进一步拓展了技术适用边界。本文总结了当前研究在理论方法、实验设计与系统实现等方面的最新进展，归纳了加密视频流量识别面临的四大核心挑战：理论研究体系薄弱、研究假设理想化、实际部署难度大与隐私增强场景中识别研究不足。最后，展望了未来发展方向，提出需加强特征理论与模型匹配机制研究、推动识别假设的弱化与促进实际场景的应用、完善数据采集与模型增量更新机制和深化隐私增强场景中识别研究，促进加密视频流量识别技术的持续创新与工程落地。

关键词 加密流量识别；视频识别；隐私保护；网络空间安全；流量分析；内容监管

中图法分类号 TP309

A Review of Research on Encrypted Video Traffic Recognition

ZHAO Hang-Yu[1)] WU Hua[1),2),3)] CHENG Guo[1)] YIN Xiao-Yue[1)] LIU Song-Tao[1)]

CHENG Guang[1),2),3),4)] HU Xiao-Yan[1),2),3)]

[1)] (School of Cyber Science and Engineering, Southeast University, Nanjing, 211189)

[2)] (Purple Mountain Laboratories for Network and Communication Security, Nanjing, 211111)

[3)] (International Governance Research Base of Cyberspace (Southeast University), Nanjing, 211189)

[4)] (Jiangsu Province Engineering Research Center of Security for Ubiquitous Network, Nanjing, 211189)

Abstract With the continuous growth of the video content industry, encrypted video traffic has become an increasingly important component of global Internet traffic, posing significant challenges to cyberspace

---

本课题得到国家自然科学基金面上项目(No. 62572117)资助.赵航宇，博士研究生，计算机学会(CCF)学生会员，主要研究领域为加密流量分析、加密视频识别.吴桦(通信作者)，博士，副教授，计算机学会(CCF)会员, 主要研究领域为加密流量分析、网络空间安全、网络态势感知. 成果，硕士研究生，主要研究领域为加密流量分析、加密视频识别.银晓悦，硕士研究生，主要研究领域为加密流量分析、加密视频识别.刘嵩涛，博士研究生，主要研究领域为加密流量分析、细粒度网页识别.程光，博士，教授，计算机学会(CCF)会员,主要研究领域为网络空间安全监测与防护、网络流量大数据分析、僵尸网络和APT攻击检测.胡晓艳，博士，副教授，计算机学会(CCF)会员,主要研究领域为加密流量分析,网络空间安全、未来网络体系结构.

governance and network security supervision. The widespread adoption of end-to-end encryption and privacy-enhancing technologies by video platforms has substantially strengthened user privacy protection. However, it has also weakened conventional supervision methods that depend on plaintext inspection, deep packet inspection, or direct content analysis. Therefore, accurately identifying harmful videos from encrypted traffic without decrypting communication payloads or compromising user privacy has become a critical scientific problem in cyberspace security. This paper systematically reviews the development of encrypted video traffic identification and organizes existing studies from the perspectives of feature construction, identification paradigm, and application scenario. Two major technical systems are summarized: feature-engineering-based matching identification and machine-learning-based classification identification. Feature-engineering-based matching methods mainly include ciphertext-fingerprint-based and plaintext-fingerprint-based approaches. The former directly extracts observable transmission characteristics from encrypted traffic, such as packet sizes, packet directions, burst patterns, timing relationships, and traffic volume distributions, and compares them with predefined reference fingerprints. The latter constructs video fingerprints from application-layer information or plaintext video objects and then establishes mappings between plaintext fingerprints and side-channel features observable in encrypted traffic. Although they differ in fingerprint construction and matching mechanisms, both approaches enable encrypted video identification under different protocols and complex network conditions without directly accessing encrypted video content. Machine-learning-based classification methods extract statistical, sequential, or structural features from encrypted traffic and employ classification models to infer which ID the video belongs to. Compared with deterministic fingerprint matching, such methods can learn latent traffic representations from data and generally provide stronger adaptability and generalization under heterogeneous platforms, heterogeneous protocols, and privacy-enhancing measures. In recent years, researchers have proposed protocol-specific feature representations and matching algorithms for a wide range of protocols and communication environments, including HTTP/1.1, HTTP/2, HTTP/3, encrypted proxies, tunneling protocols, virtual private networks, and anonymous communication networks. These studies have investigated how protocol multiplexing, packet aggregation, retransmission, adaptive bitrate streaming, and additional encryption layers affect observable traffic patterns. For emerging scenarios such as short-video platforms, social-media video services, mobile and wireless networks, researchers have further introduced hybrid feature modeling, cross-layer information fusion, protocol-adaptive representation, and robust sequence matching. These techniques alleviate the adverse effects of unstable bandwidth, background traffic, content adaptation, and protocol heterogeneity, thereby extending the applicability of encrypted video traffic identification to more realistic and complex environments. This paper summarizes recent advances in theoretical methods, feature representation, experimental design, identification models, and system implementation. It identifies four fundamental challenges: the lack of a mature theoretical framework, the dependence on idealized research assumptions, the difficulty of practical deployment, and insufficient investigation of privacy-enhanced communication scenarios. Finally, this paper discusses future directions. Future research should strengthen the theoretical analysis of traffic features and model matching mechanisms, study the problem under weaker identification assumptions, promote deployment in realistic cross-network and cross-device environments, improve scalable data acquisition and incremental model updating, and deepen research on encrypted proxies, nested tunnels, anonymous networks, traffic obfuscation, and other privacy-enhanced scenarios. Progress in these directions will promote the continuous innovation, engineering implementation, and responsible application of encrypted video traffic identification technologies.

Key words encrypted traffic recognition; video recognition; privacy protection; cyberspace security; traffic analysis; content regulation

1 引言

在数字化社会快速发展的背景下，互联网环境日趋复杂，信息的传播速度和规模均呈指数级增长。在线视频作为重要的信息承载形式，不仅革新了信息传播的方式，也逐渐成为人们获取知识与表达情感的主要媒介。伴随视频内容消费的持续增长，其所产生的网络流量在整体互联网流量中的占比显著提升。根据相关统计数据显示，截至 2024 年底，视频流量已占全球互联网移动数据总流量的 74%[1]，且该比例仍呈持续上升趋势。

然而，随着视频流量的快速增长，网络空间中不可避免地滋生了公害视频（如暴力、色情、虚假信息等不良内容），并迅速渗透至社会生活的各个层面。这类公害视频不仅对社会秩序造成严重扰乱，还因其制作门槛低、传播速度快，进一步加剧了网络治理的复杂性和难度。在互联网发展早期，管理域内视频平台的协作机制相对容易达成，同时，管理域外的视频平台大多以明文形式传输视频流量，监管机构可直接对内容进行解析与判别[1]-[5]，从而对不良视频实现有效管控。对于管理域内的视频平台，在平台协作的前提下，服务端可引入计算机视觉（Computer Vision, CV）与人工审核等手段，对视频帧内容进行算法判别或人工复核，实现对公害视频内容的识别与标注。但对于管理域外的视频平台，监管机构可以使用深度报文检测（Deep Packet Inspection, DPI）技术，依据明文载荷中的关键特征定位视频流，并通过关键词匹配实现公害视频标定。总体而言，依托端侧的协同审查机制或者明文可观测性，监管体系可以实现对视频内容的持续监测与干预，从源头抑制了公害视频的传播与扩散。

随着用户隐私保护需求的不断提升，端到端加密技术日益普及。根据 W3Techs[2]的最新统计，截至 2025 年 3 月，全球互联网约 88.1%的网站已默认采用超文本传输安全协议（Hypertext Transfer Protocol Secure, HTTPS）等加密方式，视频流量的加密比例更是显著提升，目前几乎所有主流视频平台均已实现对视频流量的加密传输[6]，这种变化导致公害视频的传播渠道变得更加隐蔽。

尽管端到端加密技术在提升通信机密性与用户隐私保护方面具有积极意义，但其客观上削弱了监管侧对内容的可观测性，使内容治理面临新的技术约束。以 DPI 为代表的明文审查机制依赖载荷解析与关键词匹配，在加密传输成为主流后难以获取有效明文信息，因而在加密视频流量监管中适用性显著下降。同时，随着视频服务的全球化与跨域部署日益普遍，监管机构往往难以与跨管理域的视频平台建立稳定的协同审查机制，导致依赖平台侧配合获取媒体数据的 CV 与人工审核等方法在跨管理域场景下难以落地。上述变化共同造成了传统监管体系在加密与跨管理域双重约束下的能力缺口，亟需探索不依赖平台协作且适配流量加密传输特性的替代性技术手段。

与此同时，部分用户对网络隐私有着更高的诉求，促使研究者提出并推动了匿名网络、加密代理和隧道协议等多种隐私增强技术（Privacy Enhancing Technologies, PETs）[7]-[16]的应用与发展。以洋葱路由（The Onion Router, Tor）为代表的匿名网络[17]-[27]，通过构建在公共网络之上的覆盖网络，实现了不可关联、不可辨识与不可观测等特性。Tor 采用定长封装、流量加密、多重中继与动态路由等机制，极大增强了通信的匿名性。加密代理[28]-[36]则通过在用户与服务端之间引入中继节点，实现加密流量的转发与匿名通信，常见形式包括 SSH 代理、Shadowsocks 代理和 Vmess 代理等。隧道技术[36]-[44]则通过协议封装，支持在异构网络间安全通信，典型应用是虚拟专用网络（Virtual Private Network, VPN）。尽管上述三类 PETs 在部分技术实现上有所重叠，但各自聚焦的隐私保护方向有所差异。通过流量加密、单跳或多跳代理、载荷固定长度填充等手段，这些隐私增强技术极大提升了用户网络使用的安全性和匿名性，但也进一步提升了公害视频传播的隐蔽性，令网络空间治理面临新的严峻挑战。

加密视频流量中的公害视频监管已成为网络空间安全治理面临的核心挑战。需特别强调的是，现有的部分监管措施存在治理效能与民生负担之间的显著冲突。以 2025 年东南亚某国在边境区域

11 Ericsson Mobility Report,https://www.ericsson.com/en/reports-and-papers/mobility-report 2025,6

22 Usage Statistics of Default protocol https for Websites,https://w3techs.com/technologies/overview/site_element 2025,5,9

实施的阶段性基础设施管制措施为例，该措施涉及网络、电力和能源供应的全面管控。虽然在短期内有效遏制了特定非法活动，但也导致了该地区民生保障体系的严重紊乱。这一案例充分展示了高强度干预措施能够即时调控目标行为，但随之而来的系统性社会成本往往显著抵消了治理收益。因此，监管部门亟需发展更加精准的识别型监管技术，在确保治理效能的基础上，尽可能减少对社会运行的扰动，实现用户无感知的网络管理。具体到公害视频的治理场景，亟需依赖精细化的监管手段，实现治理效能与社会成本的平衡。加密视频流量识别技术以其细粒度的流量分析能力，为实现这一目标提供了全新的技术方案和实践基础。

基于加密流量分析的视频识别技术本质上属于一种基于流量特征分析的被动式网络行为推断方法。该方法通过分析视频传输过程中产生的特定流量模式，实现对用户观看内容的非侵入式识别。当前主流视频平台普遍采用基于 HTTP 的自适应流媒体（HTTP Adaptive Streaming, HAS）技术，该技术通过将视频内容编码为多分辨率等级（如 360P 至 1080P）并分割为时序片段，使客户端能够根据实时网络状况动态请求最优质量片段，这种技术机制导致加密视频流量在传输过程中呈现出可观测的特征模式。

从技术特征维度分析，用于加密视频识别的流量特征主要分为两类：1）一维统计特征，主要采用反映数据分组载荷的局部统计特性，包括峰值比特率（Bits Per Peak, BPP）、字节率序列（Byte Rate Sequence, BRS）、滑动窗口字节累计和以及视频片段传输长度等；2）多维统计特征，源于对数据分组传输参数的量化分析，包括基于时间戳、传输顺序、数据流向和分组大小等原始参数的直接统计量，以及通过数学变换获得的衍生特征。

在典型应用场景中，加密视频识别者通常位于近用户端网络节点（如基站或局域网网关），其技术实现需满足两个基本约束条件：1）网络拓扑约束，要求识别者具备捕获流量的能力；2）行为约束，限定识别操作必须保持被动性，即不得实施流量解密、内容篡改、数据分组丢弃等主动干预行为。这种被动识别模式在保障用户隐私数据加密安全性的同时，仍能通过流量特征分析实现加密视频流量识别。

加密视频流量识别研究作为网络空间安全治理的关键技术方向，已历近二十年的学术积累与技术演进，形成了具有重要理论价值和应用前景的研究体系。鉴于该技术在平衡用户隐私保护与网络内容监管方面的独特作用，对现有研究成果进行系统性归纳和技术脉络梳理具有显著的学术意义。然而，目前的加密流量分析综述大多聚焦于更为宏观的研究方向，如 Ahn 等人 **Error! Reference source not found.** 关注加密算法与协议漏洞，Zhou 等人[46]聚焦网站指纹识别及防御，Shen 等人[47]主要探讨机器学习和深度学习在加密流量分析中的应用，而 Sharma[48]、Feng**Error! Reference source not found.** 与 Papadogiannaki[50]等人的研究则集中在加密流量分类领域。尽管部分综述[47],**Error! Reference source not found.**,[50]涉及加密视频流量，但大多偏向于用户体验质量评估，尚未形成针对加密视频流量识别的系统性综述。基于此，本研究首次构建了基于识别模式的加密视频流量识别技术分类体系，其创新性体现在：1）提出基于识别模式的一级分类标准，将现有方法划分为基于特征工程的匹配识别法和基于机器学习（涵盖传统机器学习和深度学习）的分类识别法两大范式；2）建立基于特征维度与识别场景的二级分类维度，通过特征空间解构实现技术路线的可比性分析；3）提炼出当前研究面临的四大核心挑战，并基于技术发展规律预测四个潜在突破方向。本文采用结构化综述方法组织研究内容：第 2 节建立理论基础，涵盖研究意义论证、基本概念定义、识别假设条件和威胁模型构建；第 3 节实施技术解构，对国内外研究进展进行系统性分析与对比；第 4 节总结研究挑战与展望，着重分析当前加密视频流量识别研究面临的核心挑战，并给出加密视频流量识别未来研究方向的几点思考和前瞻。

2 加密视频流量识别研究概述

加密视频流量识别研究的核心科学问题在于：如何通过加密网络流量中可观测的流量特征，实现对用户观看视频内容的准确推断。为建立系统的理论分析框架，本章首先明确研究意义并定义关键概念体系：2.1 节研究意义从网络空间治理维度，论证该技术在隐私保护与内容监管平衡中的独特价值，总结其研究意义；2.2 节理论基础阐述基于 HAS 协议的视频传输机制，分析加密视频流量中保留的特征维度及其可识别性基础。重点揭示视频片段长度、传输时序等特征与视频内容的映射关系；2.3 节识别假设构建研究的前提条件体系，包括场景假

设、模板假设和识别者能力假设等，通过假设条件来界定研究问题的边界约束；2.4 节威胁模型设定了加密视频流量识别的基本场景，定义识别者知识、识别者能力、识别者目标和识别者位置等信息。

加密视频流量识别研究建立在严谨的概念体系基础之上，其核心概念包括网络传输、流量分析和识别模型三个相互关联的组成部分。网络传输相关概念主要包含五元组流和数据流向：五元组（源 IP、源端口、目的 IP、目的端口和传输协议）作为网络流量的标识符，为流量分析提供了基础框架；数据流向涵盖上行分组（客户端到服务器的数据传输单元）、下行分组（服务器到客户端的数据传输单元）、上行流（特定连接中所有上行分组的集合）、下行流（特定连接中所有下行分组的集合）以及双向流（上行流与下行流的完整集合）。

流量分析相关概念包含突发（Burst）、迹（Trace）、实例（Instance）和元数据（Meta-data）。突发作为同方向连续数据分组构成的传输序列，能够反映视频传输的时序特征；迹则记录了单次视频会话产生的完整数据分组序列，是分析的基本单位；实例是特定视频类别对应的一条迹，是流量处理的基本单位；元数据即数据分组，是加密视频流量识别研究的最小研究对象。

识别模型相关的概念体系更为丰富，包含指纹（Fingerprint）、模板假设（Template Assumption）、应用层数据单元（Application Data Unit）和敌手（Adversary）。指纹是视频内容的鉴别性特征集合，分为明文指纹和密文指纹，其中明文指纹来自于视频文件本身，而密文指纹则直接或间接来自于实例；模板假设是加密视频流量识别研究的一个基本假设，它假设同一视频在不同传输过程中能够保持特征稳定性，该假设在一定的时期内是近似成立的；应用层数据单元作为视频传输的最小应用层单元，在自适应流媒体协议中通常表现为视频片段（Segment）；在加密视频流量识别研究中，敌手即是加密视频流量识别者。

加密视频流量识别研究的技术框架可划分为输入输出要素、系统识别场景和核心处理流程三个关键部分。在输入输出要素方面，系统以用户观看视频时产生的迹作为输入数据，经过特征提取和模型分析后，输出识别出的视频标题或内容标识。系统识别场景方面，如图 1 所示，典型的识别场景包含六个基本要素：1）用户（观看视频的普通网民）；2）识别者（识别用户是否观看目标视频的敌手）；3）加密流量数据（经过 TLS/SSL 加密的流量数据；）4）近用户端网络节点（识别者实施加密视频流量识别的节点）；5）网络；6）视频服务器（涵盖源服务器和 CDN 边缘节点）。

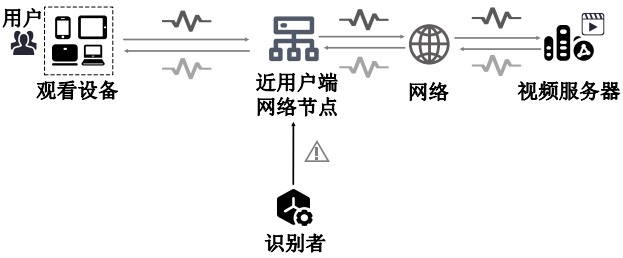

图 1 加密视频流量识别研究一般场景模型

识别过程的核心处理流程可以概括为三个阶段：第一阶段实施流量采集，识别者通过在近用户端网络节点部署监测模块，捕获用户与视频服务器之间的加密通信数据；第二阶段执行特征提取，从加密流量中分离出两类关键特征：一维统计特征和多维统计特征，这些特征的提取严格遵循非解密原则；第三阶段进行模型应用，基于预先构建的数学模型对提取的流量特征进行分析，当检测到与目标视频匹配的特征模式时触发告警机制。整个技术体系的本质特征是采用完全被动式的识别机制，严格遵守非解密原则。

加密视频流量识别研究的理论基础建立在 HAS 传输机制的特征稳定性之上。通过分析 HAS 协议下视频分段传输的规律性特征，研究者发现加密流量中仍保留着可识别的模式特征[51]。然而，实际网络环境中的识别效果受到三重关键因素制约：首先，用户观看行为的个体差异性导致流量模式呈现显著多样性；其次，复杂的背景流量会干扰目标视频流量的特征提取；最后，识别者对目标用户的先验知识获取不足，增加了准确建模的难度。为建立可量化分析的理论框架，该研究领域需要设置一些基本假设条件，即识别假设，为研究划定明确的边界条件。同时，识别者的位置是加密视频流量识别能否实现的关键之一。只有当识别者处于加密协议覆盖范围之内，能够实际截获加密视频流量时，加密视频流量识别研究才具有现实意义。因此，威胁模型的构建需要充分刻画识别者的能力、位置等相关信息，这些内容共同为加密视频流量识别方法的有效性与适用性提供理论支撑。

2.1 研究意义

加密视频流量识别技术的研究对当前网络空间治理和安全监管具有重要的理论价值和现实意义，具体体现在以下三个方面：

首先，加密视频流量识别技术为实现非合作视频平台上的公害视频治理提供了切实可行的解决方案。随着互联网的快速发展，大量视频平台出现，其中部分视频平台位于本网络管理域范围内，可以通过平台协作进行公害视频的治理。但是另一部分视频平台不在本网络管理域范围内，难以依赖平台方实现有效监管。加密视频流量识别方法能够在平台方不配合的情况下，实现对特定公害视频内容的及时识别，从而增强了治理的主动性与全面性。

其次，加密视频流量识别技术为网络空间安全治理提供了一种新型且高效的监管手段。当前，主流视频平台普遍采用端到端加密传输，导致传统依赖计算机视觉的监管手段因延迟高、资源消耗大以及对加密流量无能为力而难以满足实际监管需求。在此背景下，发展具备实时性和低资源开销的加密视频流量识别技术，成为破解加密场景下视频内容监管难题的关键路径。加密视频流量识别技术通过分析加密流量的外部特征实现对视频流量的有效识别，有助于提升监管的覆盖范围与响应能力，为复杂网络安全环境下的内容管理与治理提供了坚实的技术支撑。

最后，兼顾了监管需求与用户隐私保护的双重目标，符合当前监管技术发展的伦理要求。在现有网络治理实践中，监管效能与用户隐私保护常常存在矛盾。加密视频流量识别方法通过对加密流量统计特征的建模与分析，无需解密流量内容即可实现针对性识别，大幅降低了用户隐私泄露的风险。这种非侵入式的识别方式，仅针对特定目标视频进行分析，不影响用户对合法内容的正常使用与传播，在技术实现和合规性方面均具备高度可操作性，契合了全球互联网治理对数据隐私与合规性的基本要求。

2.2 理论基础

加密视频流量识别研究的理论基础源于 HAS 技术所呈现的规律性传输特征。当前主流的 HAS 技术标准包括由 MPEG 组织牵头开发的基于 HTTP 的动态自适应流（Dynamic Adaptive Streaming over HTTP, DASH）[3]技术和苹果公司的 HTTP 实时流媒体（HTTP Live Streaming， HLS）[4]技术。HAS 技术的核心特征体现在其分层传输架构上，如图 2 所示，视频内容首先经过可变比特率（Variable Bitrate, VBR)编码处理，生成多个质量等级的视频文件(如 360p、720p、1080p 等)，每个视频文件被进一步分割为等长的视频片段，这些片段通过 HTTP 协议按需传输。客户端根据实时网络状况，通过自适应比特率(Adaptive Bitrate, ABR)算法动态选择最优质量片段进行请求和播放。

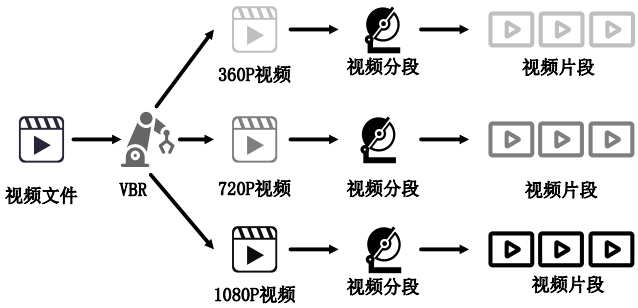

图 2 HAS 技术解析图

视频传输过程呈现出周期性的 ON/OFF 状态转换特征。如图 3 所示，在初始阶段，客户端会连续下载多个视频片段以填充播放缓冲区。进入稳定阶段后，视频平台为了确保视频播放的连续性与稳定性，客户端会在缓冲区降低至某一阈值以下时请求视频片段，传输模式转变为周期性的 ON/OFF 状态：在 ON 阶段，客户端会请求视频片段；在 OFF 阶段则保持低传输活动状态，且持续时长远长于 ON 阶段。这种传输模式在不同视频平台间存在一定差异，大部分视频平台（如 Twitter、Netflix 等）都遵循一个 ON 阶段只传输一个视频片段的原则，而小部分平台（如 YouTube）会采用视频片段组合传输策略来增加识别难度，在 ON 阶段传输由多个视频片段组合成的视频块（Chunk）。

33 Information technology — Dynamic adaptive streaming over HTTP (DASH),https://www.iso.org/standard/83314.html 2022

44 RFC 8216: HTTP Live Streaming,https://datatracker.ietf.org/doc/html/rfc8216 2017,8

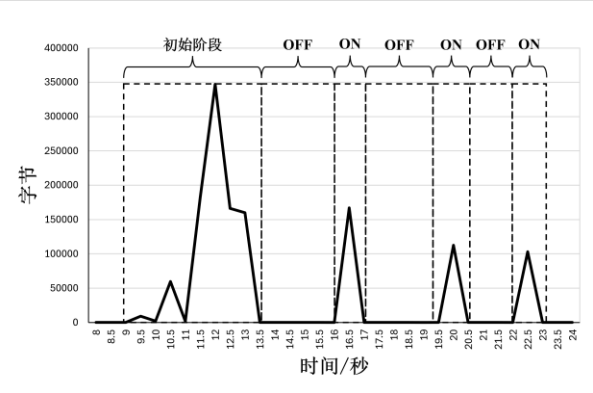

图 3 ON/OFF 模式示意图

这种 ON/OFF 模式使得视频在应用层形成了一种显著的数据特征，由于 ON 阶段传输的数据量与视频内容相关，使得该特征与视频内容强相关，当视频编码方式和分发逻辑不变时，这一特征具有稳定性和唯一性[52]。加密视频流量识别技术的关键在于从加密视频流量中提取与这种应用层数据特征直接或间接关联的流量特征。主要提取的特征包括峰值比特率、字节率、滑动窗口字节累计、视频片段传输长度、到达时间、传输顺序、数据流向和分组大小等。这些特征经过一定的数学模型转化，可以获得能够表征 ON/OFF 模式下应用层数据特征的特征向量，通过分类模型或匹配模型处理后，可以实现对加密视频流量的有效识别。

2.3 识别假设

加密视频流量识别研究通常需要在预设的理论框架下进行，这种研究范式主要基于两个重要的学术考量：首先，通过设定标准化的实验条件，可以确保不同研究方法之间的可比性，使实验结果具有可重复性和可验证性；其次，这种方法有助于研究者集中精力于流量特征的本质属性分析，从而更有效地识别和提取具有区分度的关键特征。

2.3.1 场景假设

加密视频流量识别研究采用两种标准化的场景假设作为评估基础，以系统验证识别方法的有效性。封闭世界假设将研究对象限定在预先定义的监控视频集合内，这种设置能够有效控制实验变量，便于进行方法间的性能比较。开放世界假设则更贴近实际网络环境，其包含监控集和非监控集两类视频内容，且非监控集的规模通常远大于监控集。

两种假设条件具有明确的差异。封闭世界假设主要用于验证方法的理论可行性，其核心价值在于提供可重复的实验基准。开放世界假设则侧重于评估方法的实用性能，要求识别系统具备区分目标视频和非目标视频的能力。两种假设条件都需要严格定义视频集合的组成和规模。监控集通常选取具有代表性的视频样本，而非监控集则需要覆盖足够多样的视频类型。这种差异化的实验设置能够全面评估识别方法的各项性能指标，包括准确率、精确率、F1 分数和误报率等。

值得注意的是，两种假设条件的选用与研究阶段密切相关。初期研究多采用封闭世界假设进行方法验证，而成熟的方法评估则需要通过开放世界测试。这种递进式的评估策略已成为该领域的标准研究范式，在近年来的相关研究中得到广泛应用。

2.3.2 模板假设

加密视频流量识别研究默认假设同一视频在不同时间点的应用层数据特征保持稳定，这种稳定性是特征提取和模型建立的必要前提条件。然而，实际网络环境中存在多种因素会影响这一假设的成立。

首先，视频平台的技术演进会直接改变应用层数据特征。例如视频片段分割时长的调整，当平台将分割窗口从 2 秒改为 5 秒时，应用层数据特征会发生显著变化。其次，平台安全机制的升级和协议规范的更新也会影响应用层数据特征的表现形式。这些变化导致同一视频在不同时期的应用层数据特征可能存在明显差异。

尽管如此，在相对较短的时间范围内，模板假设具有合理性。不同视频平台的应用层数据特征稳定性存在差异，这与各平台的技术架构和更新策略密切相关。研究者根据目标平台的特点合理设置时间窗口能够有效平衡假设的合理性和实验的可操作性。

2.3.3 识别者能力假设

识别者能力假设主要包含两个维度的前提条件：在知识维度方面，研究假设识别者能够获取目标用户的充分先验信息，包括网络配置参数、设备特征和使用习惯等。这种假设使得研究者能够建立与用户环境高度一致的实验条件，从而排除无关变量的干扰。需要说明的是，这些先验知识的获取途径通常被视为研究的外部条件，不属于加密视频流量识别方法本身的组成部分。在技术能力维度，研究假设识别者具备精确的流量分割能力。具体表现为：1）能够有效区分视频流量与其他应用产生的背景噪声；2）可以从混合流量中准确提取特定视频会话的完整数据流；3）能够获取足够长度的流

量数据以支持特征提取。这种技术假设为特征分析和模型构造提供了数据质量保障。

加密视频流量识别假设体现了科学研究中"简化-深化"的认知规律。通过合理简化实际问题中的复杂因素，研究者能够聚焦于核心特征的发现和利用。随着研究的深入，近年来出现了在弱化假设条件下的探索性工作，这些研究尝试在更接近真实场景的条件下验证方法有效性，如：Wu 等人[6]探索了自适应分辨率情况下的加密视频流量识别。表 1 系统总结了加密视频流量识别研究中的主要假设条件及其应用情况。从发展趋势来看，假设条件的逐步弱化反映了该领域研究正从理论验证向实际应用过渡。这种演进路径既符合科学研究的普遍规律，也体现了加密视频流量识别技术的成熟过程。

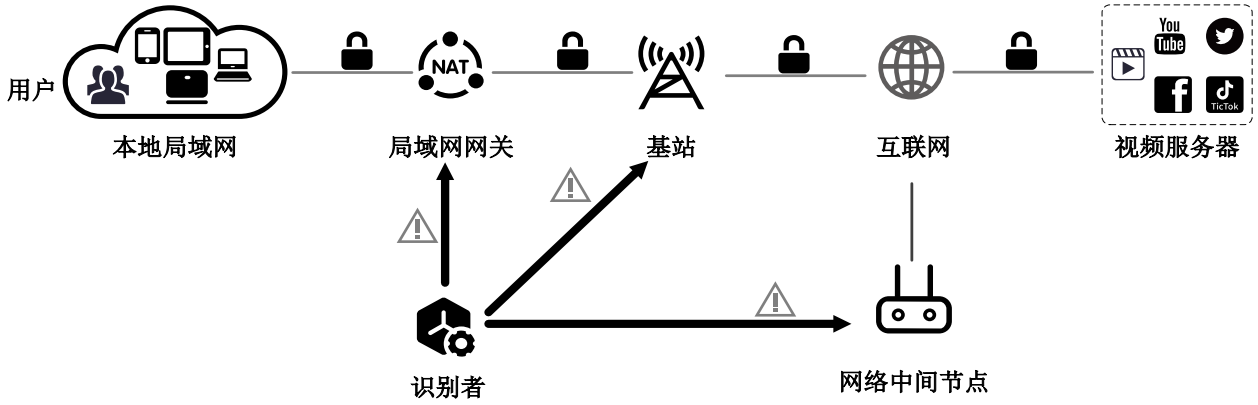

图 4 识别者位于 LAN 网关的加密视频流量识别威胁模型

表 1 加密视频流量识别文献使用的识别假设汇总

识别假设	具体假设	加密视频流量识别文献
场景假设	封闭世界	[6],[51]-[97]
场景假设	开放世界	[55],[59],[66],[71],[73],[81],**Error! Reference source not found.**,[85],[87]
模板假设	应用层数据特征稳定	[6],[51]-[97]
识别者假设	先验知识	[6],[51]-[53],[57],[59]-[97]
识别者假设	流量分割	[6],[51]-[56],[59]-[80],[83]-[97]

2.4 威胁模型

威胁模型描述了加密视频流量识别问题的核心场景。具体而言，威胁模型不仅阐明了识别者对目标用户所掌握的先验知识（即识别者知识），也明确了识别者实施有效识别所必备的能力（即识别者能力）以及所试图识别的视频对象及其规模（即识别者目标）。此外，威胁模型还定义了识别者可能存在的网络位置及来源。

加密视频流量识别方法的有效性，通常依赖于特定的威胁模型。不同的威胁模型会给识别工作带来不同的挑战。依据典型加密视频流量识别定义，当前研究多集中在本地网络中的被动识别者场景，即识别者位于近用户端网络节点，在该位置，识别者能够捕获用户产生的加密视频流量，从而为识别提供基础条件。因此，加密视频流量识别的威胁模型通常对识别者的位置提出了明确的限制要求。具体而言，在符合威胁模型定义的位置中，识别者的来源可能包括本地网络服务提供商（Internet Service Provider, ISP）的基站节点、网络中间节点以及本地局域网（Local Area Network, LAN）的网关设备等。

ISP 基站、网络中间节点和 LAN 网关这三种典型的识别者位置，在实际应用场景中主要区别在于其涉及的用户规模以及流量的数据量大小，其通用威胁模型如图 4 所示。目前，大多数研究倾向于采用用户数量最少、流量规模较小的 LAN 网关作为识别者的部署位置。以识别者位于 LAN 网关的典型场景为例，用户通常利用手机或计算机等终端设备观看特定的视频内容，这些设备通过网络连接与视频服务器建立通信。识别者则部署在本地局域网的出口网关节点，截获用户终端与远程视频服务器之间的双向网络流量，并利用所截获的数据进行视频流量识别。

3 加密视频流量识别研究进展

本节首先对加密视频流量识别领域的研究现状、主流识别方法、总体架构、流量特征、常用评价指标及数据集进行了简要回顾。在此基础上，本

文进一步依据不同的识别模式对现有研究进行了分类，并对各类方法进行了系统的对比与综合分析。

3.1 概述

加密视频流量识别的相关研究最早可以追溯到 2007 年。Saponas 等人[53]首次发现，经过 VBR 编码后，同一视频在传输过程中表现出的流量特征高度一致。基于这一观察，他们提出视频指纹的概念，将加密视频流量抽象为可以标识该视频的特征向量，即视频指纹，并构建视频指纹库。将收集到的视频流量通过数学方法转化为特征向量，并与指纹库中的指纹进行匹配，以相似度最高的指纹作为最终的识别结果。该类方法被称为基于特征工程的匹配识别法，因其计算开销较低，具备较好的实际应用与部署前景，近年来持续受到研究关注。

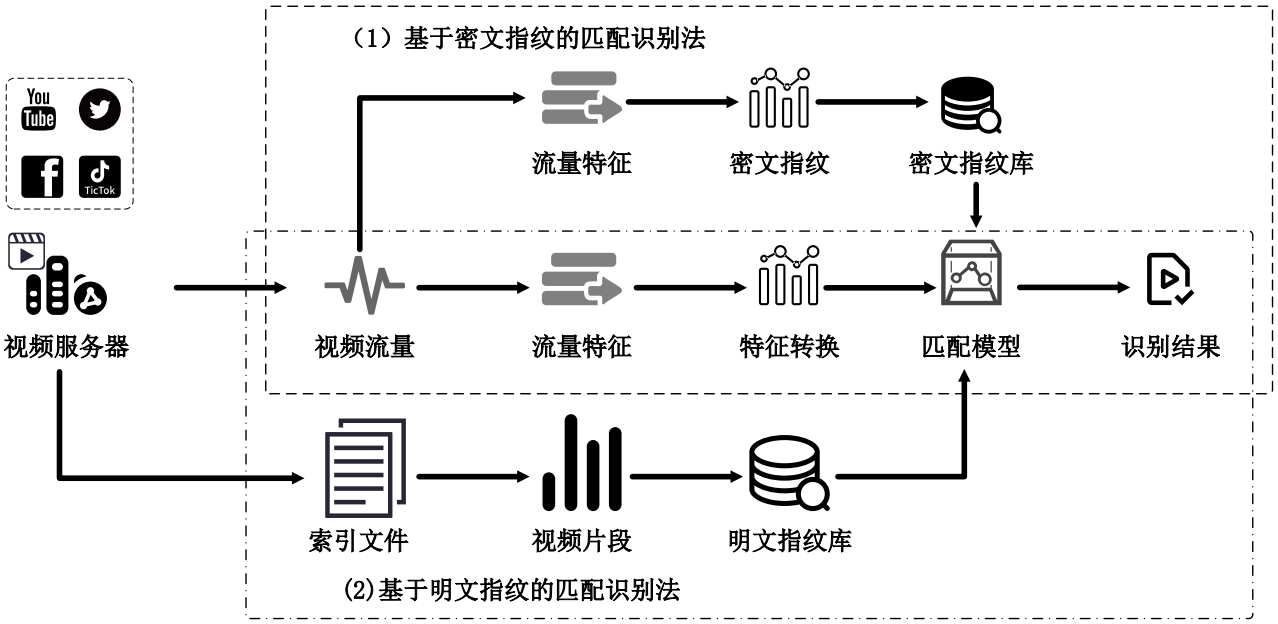

图 5 基于特征工程的匹配识别法架构图

2017 年，Dubin[76]和 Schuster[52]等学者开辟了另一条重要的技术路径—基于机器学习的分类识别法。他们在研究中发现视频流量传输过程存在显著的 ON/OFF 模式，利用这一模式划分流量特征，可以提取能够表征应用层特征的流量特征。随后，研究者分别提取了基于 ON/OFF 模式的统计特征，进一步构建特征向量，并首次将机器学习（涵盖传统机器学习和深度学习）技术引入到加密视频流量识别领域，采用机器学习方法训练分类模型，实现对加密视频流量的自动识别，为后续研究提供了新的方向。

自此，加密视频流量识别领域的研究主要围绕基于特征工程的匹配识别法和基于机器学习的分类识别法两大技术路线展开，推动了该领域理论和实践的不断发展，并取得了丰富的研究成果。

3.1.1 加密视频流量识别总体架构

加密视频流量识别领域的两大技术路线在相互促进与借鉴的过程中，逐步发展出了基于特征工程的匹配识别法总体架构和基于机器学习的分类识别法总体架构。

基于特征工程的匹配识别方法的总体架构如图 5 所示，其基本流程包括多个步骤：首先，提取视频指纹并构建视频指纹库；然后，从加密视频流量中提取初步流量特征；接着，通过特征转换方法获得高辨识度的流量特征；最后，利用匹配模型对高辨识度流量特征与指纹库中的指纹进行相似度匹配，从而实现加密视频流量的识别。

根据指纹类型的不同，基于特征工程的匹配识别方法可进一步分为基于密文指纹的匹配识别法和基于明文指纹的匹配识别法，如图 5(1)和图 5(2)所示。密文指纹是通过从特定的加密视频流量中提取初步流量特征，经过特征转换形成密文指纹，即具有标签的高辨识度流量特征。明文指纹则是基于 HAS 技术原理，从视频服务器中获取视频索引文件，基于索引文件中记录的视频片段长度构建视频片段长度序列。初步流量特征是通过 ON/OFF 模式

对视频流量进行局部字节聚合得到的，典型的初步流量特征为视频片段传输长度序列，该序列通过累加每个 ON 阶段中的数据分组长度来生成，依次得到每个视频片段的传输长度。特征转换通过对 HAS 技术及传输协议的深入理解，结合专家知识，进一步提升流量特征对应用层数据特征的表征能力。长度修正是典型的特征转换过程，研究者通过建立数学模型，逐层消除传输协议和加密协议引起的视频片段传输长度偏差，从而获得与视频片段长度几乎相同的修正长度。最后，匹配过程通过匹配模型对高辨识度流量特征与指纹库中的指纹进行相似度计算，找到最相似的指纹，指纹对应的标签即为视频的识别结果。

基于机器学习的分类识别方法总体架构如图 6 所示，其主要流程包括多维统计特征获取、多维特征融合、分类模型训练和模型识别四个核心环节。首先，在多维统计特征获取阶段，系统会提取流量中的局部与全局统计特征，涵盖时间和空间两个维度，既包含原始统计量（如包长度、到达间隔等），也包含由这些原始量衍生出的高阶统计特征。随

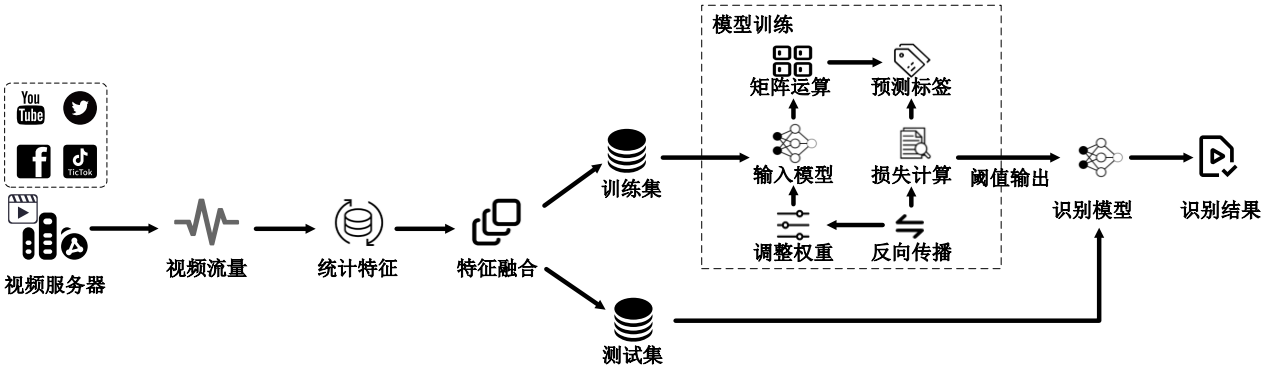

图 6 基于机器学习的分类识别法架构图

后，多维特征融合环节利用数学建模方法，将各类统计特征有效整合并映射为统一的特征向量，为后续建模提供基础。在分类模型训练阶段，通过收集多轮特征向量组成的训练集，采用传统机器学习方法或深度学习方法对模型进行训练，得到适用于加密视频流量识别任务的分类器。最后，分类模型识别阶段将待识别的未知视频流量对应的特征向量输入训练好的模型，模型依据已学习到的特征判别规则输出识别结果，实现对加密视频流量的自动化分类与判别。

3.1.2 流量特征与视频指纹

流量特征是指从加密视频流量中提取，用于描述加密视频流行为的各类特征信息；而视频指纹则是依据 HAS 技术原理获取或直接从加密视频流中提取，能够唯一表征特定视频内容的特征序列。在基于特征工程的匹配识别方法中，视频指纹具备区分不同视频的唯一性，通过对比加密视频流量中提取的流量特征与已知视频指纹之间的相似度，实现对加密视频流量的准确识别。

在加密视频流量识别领域的发展过程中，研究者提出了多种类型的流量特征，每种特征均展现出独特的属性与应用价值。综合现有研究成果，可以将这些流量特征归纳为两大类别：一维统计特征与多维统计特征。

一维特征通常是用以反映视频流 ON/OFF 模式的数据序列，这些特征一般以等长时间窗口或 ON 阶段作为分割单元，对每一分割区间内的数据分组载荷进行累加，顺序排列后形成局部字节和序列。为了提升特征表征能力，不同研究进一步对局部字节和序列进行了深度处理，发展出包括峰值比特率序列[76],[77],[52]、字节率序列[81],**Error! Reference source not found.**、视频片段传输长度序列[57],[58],[65],[66]、视频片段修正长度序列[6],[51],[61],[62],[67]-[75]、差分长度序列[54]-[56]等多种特征形式，从而更有效地刻画视频流的时序与结构特征。

多维统计特征则关注于流量的全局属性，涵盖如数据分组的方向（上行分组为正，下行分组为负）、时间分布、每秒数据分组数、数据分组长度的均值、极值与方差等信息。多数研究[84]-[89],[91]-[93]会提取全流量的统计特征，但也有部分工作[90]选择仅对视频流初始阶段的数据分组进行统计分析，以减少维度和计算复杂度。总体而言，多维统计特征的维度往往较高，能细致描述流量分布与波动。

自视频指纹概念提出以来，相关技术持续演

进，逐步形成了具有不同特点的视频指纹类型。经过梳理现有研究，视频指纹主要可划分为密文指纹和明文指纹两大类，各自具备独特的生成方式和应用场景。

密文指纹是指直接从加密视频流量中提取，能够唯一标识特定视频内容的局部字节和特征序列。获取密文指纹的典型流程如下：首先将加密视频流量按照等长时间窗口或 ON 阶段进行分段，对每个分段内的全部载荷进行聚合，从而生成滑动窗口字节和序列或视频片段传输长度序列。部分研究[53],[58]将上述滑动窗口字节和序列或视频片段传输长度序列直接作为密文指纹，另一些研究则对初步特征进一步处理，例如利用相邻数据单元长度的差分构建差分长度序列[54],[56]，基于传输长度序列修正生成的视频片段修正长度序列[59]，或通过马尔科夫链方法将传输长度序列转化为张量[57]等。

与之相对，明文指纹是从视频文件本身结构信息中直接提取的特征。2016 年，Reed 等人[63]提出，将同一质量等级下的视频片段长度序列按顺序组合，可稳定地标识该视频内容，解释了早期学者关于流量特征稳定性的发现。视频片段长度序列因此成为一种具有高度稳定性和辨识能力的应用层数据特征，被称为视频明文指纹。2017 年，Schuster 等人[52]从理论与实验层面进一步证明，足够长的视频片段长度序列具有全局唯一性，即视频明文指纹具备唯一标识视频的能力。明文指纹的提取过程如下：主流视频平台通常采用 HAS 技术传输视频，HAS 会将原始视频通过 VBR 编码为不同质量等级，并进一步切分为等时长的视频片段，这些信息被记录在索引文件中，如 DASH 的媒体呈现描述（Media Presentation Description, MPD）文件和 HLS 的媒体播放列表（m3u8）。通过分析这些索引文件，可获得各质量等级下的视频片段长度序列，用作视频的明文指纹。

3.1.3 评价指标

在加密视频流量识别领域，研究初期通常采用准确率（Accuracy）作为最主要的性能评估标准。准确率也常与召回率（Recall）或真正例率（True Positive Rate, TPR）等价，具体定义如式(1)所示，其中，TP 代表被检测为阳性并且实际也为阳性的样本数量，FN 为检测为阴性但实际为阳性的样本数量。

$$Accuracy = Recall = TPR = \frac{TP}{TP + FN} \quad (1)$$

随着研究的深入，学者们逐步引入了精确率（Precision）、F1 分数（F1-Score）、假阳率（False Positive Rate, FPR）以及面向开放世界场景的贝叶斯检测率（Bayes Detection Rate, BDR）等多元化评价指标。各项指标的计算方式分别如式(2)-(5)所示。精确率衡量的是被判定为阳性的样本中实际为阳性的样本比例，其中 FP 为被检测为阳性但实际为阴性的样本数。F1 分数作为精确率与召回率的加权调和平均，是衡量模型整体表现的重要指标，广泛应用于信息检索和分类模型评估。假阳率则反映了误报风险，是评估模型精准管控能力的关键参考指标。贝叶斯检测率在开放世界识别任务中尤为重要，$p(mon)$代表监控集样本数占总样本数的比例，$p(unmon)$为非监控集样本数占总样本数的比例。各项指标的引入和综合应用，有效提升了加密视频流量识别模型评估的科学性和全面性。

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

$$FPR = \frac{FP}{TN + FP} \quad (4)$$

$$BDR = \frac{TPR \times p(mon)}{TP \times p(mon) + FN \times p(unmon)} \quad (5)$$

在加密视频流量识别研究中，除了常见的评价指标外，识别时间也是一个重要的衡量标准。然而，不同研究对识别时间的定义存在差异。本文对涉及识别时间的相关研究进行了总结，并将识别时间划分为三类：输出时间、处理时间和嗅探时间。各类识别时间的具体定义如表 2 所示。由于识别时间没有标准定义，导致数值差别显著，因此，现有文献中该指标并不是主要指标。

表 2 三类识别时间定义

概念	定义	文献	范围
输出时间	基于流量特征输出识别结果的时间	[6],[51],[58],[61],[62],[67],[69],[71],[72],[78]	3.33 微秒-24.2 分钟
处理时间	从流量特征提取到输出识别结果的时间	[63]-[65]	5 分钟-10 分钟
嗅探时间	单个流量样本采集时间	[73],[75],[79],[80],[89],[91],[92]	10 秒-180 秒

3.1.4 数据集

在加密视频流量识别领域，为提升研究方法的可复现性，一些研究团队选择在互联网公开发布实验用数据集。以 Dubin Dataset[76]为代表的公开数据集，已被众多后续学者广泛引用[57],[58],[78],[81],**Error! Reference source not found.**。然而，由于加密视频流量的传输模式会随主流平台不断演进和变化，为适应最新的技术环境，多数学者更倾向于自采数据集以支撑其实验。现有论文中的加密视频流量识别数据集主要分为三类：一类是为了应对早期数据采集不便而模拟的数据集；另一类是适用于基于特征工程的匹配识别法的视频数量较多、但每个视频采集轮次较少的数据集；最后一类则主要应用于基于机器学习的分类识别法，特点是视频数量较少、但单个视频采集轮数较多。为便于后续阐述，本文对具有代表性的数据集进行了整理，详见表 3。

第一类数据集的典型代表是 Wu-H1 Dataset[70]，其通过代理采集 277 条 Facebook 视频流，并基于数据模拟扩展至 55,677 条。虽然该数据集在理论上验证了合理性，并证明模拟指纹会增加相似指纹比例，从而提高验证难度，但其与真实网络环境差距较大，限制了后续研究的实际应用价值。第二类数据集包括 Zhao Dataset[74]、Wu-H2 Dataset[6]和 Tang Dataset[67]，其中，Zhao Dataset[74]和 Wu-H2 Dataset[6]涵盖平台数量多、视频规模大，可用于平台泛化能力验证；Tang Dataset[67]则聚焦单一平台，虽缺乏跨

表 3 加密视频流量识别数据集概览

数据集名称[1]	平台	隐私增强（PETs）	指纹数	封闭世界数据集[2]				开放世界数据集[2]		
				视频流数	视频数	PETs 视频流数	PETs 视频数	视频流数	视频数	未知视频数
Wu-H1 Dataset[70]	YouTube	未考虑	55,677	55,677	277+55,400	无	无	55,677	55,677	150,000
Zhao Dataset[74]	Twitter	未考虑	302,358	17,781	17,781	无	无	无	无	无
	Facebook		304,198	3,100	3,100					
	Instagram		90,439	2,900	2,900					
Wu-H2 Dataset[6]	Facebook	未考虑	310,350	4,440	4,440	无	无	无	无	无
	Instagram		98,540	3,316	3,316					
	Triller		5,000	3,945	3,945					
	Twitter		5,000	3,500	3,500					
	芒果 TV		5,000	1650	1650					
Tang Dataset[67]	YouTube	VPN	62,232	48,391	5186	11,541	5186	无	无	无
Björklund Dataset[96]	Amazon	VPN	1,285,713	900	116,070	840	840	无	无	无
	Max		1,544,778		99,478					
	SVT Play		230,582		26,816					
Walsh Dataset[87]	YouTube	Tor	无	59,580	60	53,160	60	600	60	64,000
	Vimeo			59,580	60	53,160	60			
	Rumble			59,580	60	53,160	60			
	Facebook			59,580	60	53,160	60			
Xu-Proxy Dataset[83]	Twitter	Proxy	无	2,515	48	无	无	无	无	无
	Daylymotion			2472	49					
Dubin Dataset[76]	YouTube	未考虑	无	10000	100	无	无	5000	2000	无

注：[1]数据集名称根据论文作者和内容自拟

平台适用性，但其采集规模大，适合封闭世界与开放世界实验，并特别包含 VPN 场景数据，为隐私增强环境下的研究提供了支撑。然而，该类数据集多为一轮或少量轮次采集，难以满足基于机器学习

的分类识别方法对大规模训练数据的需求。第三类数据集以 Xu-Proxy Dataset[83]、Dubin Dataset[76]和 Walsh Dataset[87]为代表，分别采集了 50 轮、100 轮和 1000 轮流量，适合机器学习与深度学习模型的训练。其中，Xu-Proxy Dataset[83]和 Walsh Dataset[87]分别聚焦代理与 Tor 场景，验证了加密视频流量识别在隐私增强环境中的适用性；Dubin Dataset[83]作为目前应用最广泛的基准数据集，具备良好的研究延续性，便于与既有成果进行对比分析。

3.2 基于特征工程的匹配识别法

基于特征工程的匹配识别法将领域专家知识与先进的匹配算法相结合，核心思想是通过比对从加密视频流中提取的特征序列与预先构建的视频指纹库中的指纹，从而实现对加密视频流量的识别。根据指纹库中所存储指纹的类型，该方法可以进一步细分为基于密文指纹的匹配识别法和基于明文指纹的匹配识别法两种类型。

3.2.1 基于密文指纹的匹配识别法

基于密文指纹的匹配识别法，是指从加密视频传输流中提取侧信道特征，将这些特征直接作为密文指纹，或经过数学模型处理后形成密文指纹，进而构建密文指纹库，然后通过将待检测加密视频流的传输特征与指纹库中的密文指纹进行相似性比对，根据最高相似度原则判断密文指纹匹配结果，从而实现加密视频流量的识别。本文对基于密文指纹的匹配识别法进行了系统梳理，并将相关研究成果归纳于表 4 中。

表 4 基于密文指纹的匹配识别法研究汇总

方法[1]	模型	特征	复杂环境	平台	视频数	封闭世界性能	开放世界性能
Saponas-DTF (2007) [53]	近似最近邻搜索	每 100 毫秒的吞吐量	未考虑	Slingbox Pro	26	准确率:62%~77%	未考虑
Gu-PDTW (2018) [54]	P-DTW 算法	每秒比特数的差分	不同质量编码	-	200	准确率:90%	未考虑
Afandi-SDF (2022) [56]	CNN 算法	每秒字节数、周期字节数	VPN	YouTube	86	准确率:90%	未考虑
Yang-markov (2020) [57]	最大似然估计	视频传输长度序列的马尔科夫张量	未考虑	YouTube	100	准确率:97.5%	未考虑
Yang-levenshtein (2021) [58]	谱聚类	视频传输长度序列	未考虑	YouTube	100	准确率:95.47%	未考虑
Huang-Mix (2025) [59]	逐轮匹配算法	数据分组载荷总修正长度	未考虑	微信 WhatsApp	3,500 3,000	准确率:98.05%~100% 精确率:97.54%~100% F1 分数:97.69%~100%	未考虑

注：[1]方法名根据作者和方法特点自拟

Saponas 等人[53]的工作标志着加密视频流量识别领域的开端，他们首次提出了无需解密即可对加密视频流量进行识别的方法。该方法基于 VBR 编码的特性，通过分析加密视频流量的吞吐量变化推断视频内容。研究者通过监测网络连接的数据传输速率，提取吞吐量序列，并采用离散傅里叶变换（Discrete Fourier Transform, DFT）将其转换为流量特征。为减小原始特征中的噪声影响，将多个流量特征合并后作为密文指纹，用于构建密文指纹库。在识别阶段，通过近似最近邻搜索算法，将待识别流量特征与指纹库中的密文指纹进行比对，实现加密视频内容的识别。在实验验证中，研究团队对 Slingbox Pro 设备播放的 26 部不同电影的加密视频流量进行了分析。实验结果显示，当使用 10 分钟的流量数据时，该方法平均识别准确率可达 62%；监测时间延长至 40 分钟时，准确率进一步提升至 77%。在部分特定电影场景下，识别准确率超过 98%。这些结果表明，该方法能够从加密视频流量中有效提取用于内容识别的特征信息。然而，该方法也存在局限性：其性能高度依赖 VBR 编码，针对固定比特率编码的视频识别效果有限；方法需提前采集大量加密视频流量以构建指纹库，存储和计算资源消耗较大；对网络噪声的敏感性较高，易受复杂环境影响识别准确率，并且在大规模数据集下的扩展性尚未得到充分验证。尽管如此，Saponas 等人[53]的研究开创了基于密文指纹的匹配识别新范式，验证了即便在流量加密条件下，通过流量特

征分析依然具备识别视频内容的可行性。

Gu 等人[54]提出了一种面向 YouTube DASH 加密视频流量的侧信道攻击方法。该方法基于 DASH 视频流的 VBR 编码特性，通过分析比特率随时间的变化趋势实现加密视频流量识别。研究者将捕获到的加密视频流量按照等长时间窗口进行分段，然后对各个分段内的载荷加总后进行差分处理作为视频的密文指纹。在识别阶段，该方法采用部分动态时间规整（Partial Dynamic Time Wrapping, P-DTW）算法对流量特征和密文指纹之间的相似性进行度量。该算法支持对流量特征与指纹的部分匹配，在捕获的流量数据不完整的情况下，也能实现加密视频流量识别。实验结果显示，在常规网络环境不同视频内容、分段长度和视频质量水平下，该方法使用三分钟的窃听时间，识别准确率达到 90%。与 Saponas-DTF[53]方法相比，该方法在计算效率和准确率上均有显著提升。尽管如此，该方法的识别率仍受限于可利用视频数据时长，对数据采集条件和网络环境的稳定性要求较高，易受复杂网络环境下的噪声和背景流量的影响，在复杂环境和大规模数据集下的适用性尚需进一步提升。值得强调的是，该方法实现了在 VPN 加密场景下的基于密文指纹的识别，展现了良好的协议兼容性和场景适应能力。

Afandi 等人[56]在 Gu-PDTW 方法[54]的基础上对密文指纹识别策略进行了进一步优化，尤其关注 VBR 编码带来的特征稳定性问题和零除异常。Afandi 等人[56]从 YouTube 加密流量中提取了每秒字节数（BPS）特征，并将其聚合为峰值比特率（BPP）特征，以缓解流量模式的不一致性，从而提升指纹在固定周期内的稳定性。针对零除异常，采用数值替换策略，将 BPP 序列中的“0”替换为“1”，从而避免零除异常。在指纹生成方面，研究团队采用了三种差分指纹生成方法：简单差分指纹（Simple Difference Fingerprint, SDF）、绝对差分指纹（Absolute Difference Fingerprint, ADF）以及差分指纹（Difference Fingerprint, DF）。其中，通过计算相邻峰值比特率的差值生成的 SDF 指纹识别效果最好。基于 SDF 指纹，研究团队构建并训练了优化的序列卷积神经网络（CNN）模型，实现了对 YouTube 加密视频流的精准识别，并具备一定的协议场景适应性，包括对 VPN 和非 VPN 环境的适配。实验数据表明，该方法在 HTTPS 环境下的识别准确率可达 90%，表现出良好的特征表达能力；但在 VPN 环境下，识别准确率下降至 59%，显示出模型在泛化性和抗干扰能力方面仍有待提升。总体来看，该方法在复杂网络场景下的效果尚有不足。

不同于 Gu-PDTW[54]和 Afandi-SDF[56]等差分指纹方法，Yang 等人[57]提出了一种基于马尔科夫概率指纹的加密视频流量识别方法。该方法通过对 DASH 协议下的视频片段传输长度序列进行分析，首先将连续的片段长度离散化为不同区间，并将传输长度序列转化为状态转移序列。随后，利用马尔科夫链计算得到状态转移张量，将其作为视频的指纹特征。在识别阶段，采用最大似然估计（Maximum Likelihood Estimation, MLE）计算待识别流量的状态转移张量与指纹库中指纹的相似度，从而判定加密视频流量的归属。实验结果显示，该方法在 Dubin Dataset[76]上的识别准确率达到 97.5%。然而，该方法主要基于公开数据集，未对复杂网络环境下的抗干扰能力进行充分验证，且模型在真实应用场景中的泛化性尚不明确。

为拓展加密视频流量识别在无标签数据场景下的应用，Yang 等人[58]引入了无监督学习方法，将聚类分析技术应用于加密视频流量识别。该方法首先提取加密视频流中的视频片段传输长度序列作为视频指纹，随后采用基于 Levenshtein 距离的谱聚类（Spectral Clustering, SC）算法对流量进行聚类，不依赖领域专家知识，实现了加密视频流量的无监督识别。实验结果显示，在 Dubin Dataset[76]数据集上，该聚类方法的准确率达到 95.47%，表明其在无标注场景下具有较好的实际效果。尽管相较于有监督的 Yang-markov[57]方法准确率略低，但显著优势在于降低了对数据标注的依赖，为实际应用提供了更高的灵活性。需要指出的是，该方案在面对网络延迟和丢包造成的序列乱序、数据丢失等复杂因素时，抗干扰能力仍有不足，且计算成本较高，对小样本类别的聚类表现也不够稳定。

Huang 等人[59]针对即时通讯（Instant Messaging, IM）软件在视频传输过程中使用多种复杂协议（如 GQUIC、HTTPS 及私有协议）的问题，首次提出了面向 IM 场景的基于密文指纹的匹配识别方法。与传统仅适用于 HAS 流媒体平台的方案不同，该方法针对 IM 场景中协议不统一和分段特征难以直接利用的挑战，针对不同协议提取了不同的流量特征作为视频指纹，并建立了适应多种协议的指纹库。研究团队通过捕获 IM 客户端下载视频的流量，针对不同协议类型提取相应特征：在

GQUIC 和 HTTPS 协议下，首先对视频传输长度进行协议偏移修正，得到修正长度作为指纹特征；而在明文流量下，直接使用视频传输的完整长度作为指纹特征。这些特征被用来构建视频指纹库。在识别阶段，研究团队基于传输协议从目标流量中提取相应的视频修正长度或传输长度特征，并采用带有阈值的序列匹配技术在指纹库中进行比对，从而实现加密视频流量的识别。与仅基于传输层数据分组长度提取指纹并未考虑协议附加数据的 Tang-Shrink[61] 和 Du-LSTF[62] 方法不同，Huang-Mix[59]方法在特征提取过程中结合了加密协议信息和相应处理技术，更全面地反映了实际传输特征，显著提高了指纹匹配的准确性，从而实现了在即时通讯软件复杂协议环境下的高精度视频流识别，表现出较强的模型泛化能力、场景适应性和协议适应性。然而，复杂的修正过程导致了计算效率的下降，这限制了其在某些场景中的适用性。

总体而言，基于密文指纹的匹配识别方法具有如下特点：首先，密文指纹的稳定性受到实际网络传输状态变化的显著影响。在复杂网络环境下，同一视频不同会话的流量中提取的流量特征会因网络条件波动而产生较大差异，导致密文指纹本身难以保持稳定性。其次，该类方法在协议泛用性方面存在局限，尤其是在采用多路复用技术的 HTTP/2 和 HTTP/3 协议场景下，由于客户端可同时发起多个请求并并发接收多个响应，流量特征的动态变化

表 5 基于明文指纹的匹配识别法研究汇总

方法 [1]	模型	特征	复杂环境	平台	视频数	封闭世界性能	开放世界性能
Reed-block (2016) [63]	KD-树	WAP 到客户端的吞吐量	从随机位置开始播放	Netflix	63	准确率:>90%	未考虑
Reed-segment (2017) [64]	KD-树	TCP/IP 头部的流量特征	实时检测	Netflix	200	准确率:99.5%	未考虑
Xu-CSI (2020) [65]	迪杰斯特拉算法	视频片段传输长度序列	不同带宽情况	YouTube Facebook etc.	5	准确率:>95%	未考虑
Zhang-RoVIA (2024) [66]	DTW 算法	视频组合块的传输长度序列	时延、限制带宽、不同分辨率	YouTube	180	准确率:99.7% F1 分数:99.7%	准确率:96.7% 召回率:96.7% F1 分数:97.0%
Zhang-Evi (2025) [97]	图匹配	音视频组合块传输长度序列	时延、限制带宽	YouTube	150	准确率:97.3%~99.6% F1 分数:96.2%~99.1%	准确率:96.2%~98.8% F1 分数:95.3%~97.4%
Du-LSTF (2023) [62]	词袋模型	视频片段修正长度序列	未考虑	YouTube	1,000	准确率:96.19%	未考虑
Tang-Zenith (2024) [67]	TLM	视频块的修正长度序列	不同带宽、不同分辨率、VPN	YouTube	62,232	准确率:97.32%	未考虑
Wu-DF (2020) [55]	局部序列比对算法	视频片段传输长度差分指纹	未考虑	Facebook	205,677	准确率:99.8%~100% 精确率:3.3%~100%	未考虑
Björklund-KD (2025) [96]	KD-树	视频片段传输长度序列	实时检测、VPN	Amazon Max SVT Play	116,070 99,478 26,816	召回率:85.6%~99.8%	未考虑
Wu-HHTF (2021) [70]	k 段匹配算法	HTTP/1.1 视频片段修正长度序列	未考虑	Facebook	205,677	准确率:99.99%~100% 精确率:65.7%~100% 假阳率:≈0	未考虑
Wu-H2 (2025) [6]	HMM	HTTP/2 的视频片段修正长度序列	不同分辨率	Facebook Instagram etc.	310,350 98,540 15,000	准确率:97.82%~100% 精确率:97.96%~100% F1 分数: 97.96%~100%	未考虑

Wu-H3 (2025) [51]	HMM	HTTP/3 的视频组合块修正长度序列	未考虑	YouTube Facebook Instagram	362,502 283,895 74,342	准确率:94.35%~100% 精确率:97.95%~100% F1 分数:94.74%~99.75%	未考虑
Hu-QUIC (2024) [71]	HMM	QUIC 流视频片段修正长度序列	未考虑	YouTube Facebook Instagram	50,011 50,009 50,002	准确率:98.03%~99.04% 精确率:98.06%~100% F1 分数:98.49%~99.43%	准确率:97.08%~98.43% 精确率:96.68%~97.16% F1 分数:96.83%~97.65%
Zhao-QUIC (2024) [72]	组合匹配法	QUIC 流视频片段修正长度序列	未考虑	YouTube	500	准确率:>97% 精确率:>95% F1 分数:>96%	未考虑

注：[1]方法名根据作者和方法特点自拟

续表 5 基于明文指纹的匹配识别法研究汇总

方法[1]	模型	特征	复杂环境	平台	视频数	封闭世界性能	开放世界性能
Quan-short (2024) [73]	SVP-DOHMM 匹配算法	视频或视频片段修正长度序列	未考虑	Triller	12,318 111,105	准确率:98.16% 精确率:99.90% 误报率:0	准确率:97.64% 精确率:99.68% 误报率:0.02%
Zhao-software (2024) [74]	HMM	视频片段修正长度	时延、丢包、跳转播放	Twitter Facebook Instagram	302,358 304,198 90,439	准确率:98.81%~99.50% 精确率:99.43%~100% F1 分数:99.06%~99.69%	未考虑
Zhu-asymmetric (2023)[75]	HMM	视频片段修正长度序列	未考虑	YouTube Vimeo	362,502 107,896	准确率:97.43%~98.22% 精确率:99%.08%~100% F1 分数:98.70%~99.10%	未考虑

注：[1]方法名根据作者和方法特点自拟

使基于密文指纹的识别难以适应，因此该类方法适用于 HTTP/1.1 这种单请求、单响应的非流水线传输场景。再次，从场景适应性来看，得益于 HAS 技术固有的 ON/OFF 特征，该类方法的特征相似度较强，能够在 VPN 等特定环境下表现出一定的适用性。然而，在工程应用方面，该类方法面临指纹库规模扩展带来的误差累积问题，密文指纹稳定性的不足会随指纹库扩大而放大，从而影响整体识别准确率。因此，密文指纹识别法目前更适用于小规模场景，对于需要大规模加密视频流量识别的实际应用仍存在明显挑战。

3.2.2 基于明文指纹的匹配识别法

基于明文指纹的匹配识别法，是指通过将带外方式获取能够稳定且唯一标识视频内容的应用层数据特征作为明文指纹，并构建明文指纹库，然后将待检测的加密视频流量提取出的流量序列特征，或者经特征修正算法处理后的流量序列修正特征，与明文指纹库中的指纹进行相似度比对，实现对加密视频流量的内容判别。流量序列特征通常是基于 ON/OFF 模式或等长时间窗口对加密流量进行分割，得到的视频片段传输长度序列或局部字节和序列；流量序列修正特征则依赖领域专家知识，通过设计特定算法对原始流量序列特征进行修正。相比于每次传输都会变化的密文指纹，明文指纹具有更强的稳定性。本文对基于明文指纹的匹配识别法进行了系统的综述，并将相关研究成果归纳总结于表 5 中。

Reed 等人[63]首次提出了基于明文指纹的匹配识别法，并在 DASH 协议场景下实现了加密视频流量的明文指纹匹配识别。该方法通过利用段索引框（sidx）提取 Netflix 视频的片段长度，并将单个视频片段的长度序列中每连续 30 个视频片段作为该视频的一条指纹，以这些指纹为节点构建基于 KD-树（K-Dimension Tree）的指纹库。此外，研究还通过分析无线网络中的 Block Acks，估算无线接入点（Wireless Access Point, WAP）与客户端之间的吞吐量，并将该吞吐量数据按 4 秒时间窗口进行聚合，最终得到视频片段的传输长度序列。通过皮尔逊相关系数和 KD-树结构，研究团队将视频片段的传输长度序列与指纹库中的指纹进行匹配，实现了

在 802.11n 标准的加密 Wi-Fi 环境下对 VBR DASH 视频的识别，并且能够应对波束成形和多输入多输出的复杂环境。相较于 Saponas-DTF[53]方法，Reed-block[63]方法不仅识别时间更短，且识别效果更为精准，展现了良好的场景适应性和协议支持能力。然而，该方法在处理相似视频场景时，可能会出现误识别的情况。

2017 年，Netflix 对其基础设施进行了升级，通过引入 HTTPS 加密技术来增强视频流的隐私保护。针对这一变化，Reed 等人[64]在其之前的 Reed-block[63]方法基础上进行了改进，特别考虑了 HTTP 和 TLS 协议头对加密视频流量特征提取的影响。研究者通过解析 MPEG4 元数据中的段索引框来构建视频的明文指纹，并结合 TCP/IP 协议头部的明文信息提取视频片段的传输长度序列，实现了在无需应用层信息的情况下进行实时视频流量匹配。该方法在 HTTPS 加密环境下基于包含 42,027 个 Netflix 视频指纹的指纹库实现对 200 个 20 分钟的视频流的识别，识别准确率达到 99.5%，并且大多数识别过程在 2 分 30 秒内完成。该方法能够在有限的硬件条件下实现实时运行，并且支持大量并发流的处理。然而，该方法并未充分考虑复杂网络环境和不同平台对识别性能的影响，且缺乏对模型健壮性和特征泛化能力的验证，这使得其在实际部署中的应用仍面临一定的挑战。

Xu 等人[65]提出了一种名为 CSI 的系统，旨在推断在 HTTPS 和 QUIC 加密流量环境下移动自适应比特率（ABR）视频的自适应分辨率行为。CSI 系统通过获取包含目标视频编码和分段信息的 ABR 清单，从中解析出视频片段长度序列，构建视频的明文指纹。系统随后通过分析加密流量中的数据分组大小和时间信息，推断出视频片段的估计长度序列，并使用图搜索算法将视频片段估计长度序列与视频指纹进行匹配，从而实现加密视频流量的识别。实验结果表明，CSI 系统在识别六大主流视频服务时，推断准确率超过 95%，且对 10 分钟的视频分析仅需数秒。然而，CSI 系统也存在一定局限性。例如，由于该系统难以识别 QUIC 协议中的重传分组，因此在 QUIC 加密数据上的估计误差高达 5%，从而增加了基于匹配的视频片段识别算法的误识率。此外，尽管 CSI 系统在实验环境中表现良好，但其在复杂网络环境中的验证较为简单，尚未充分证明其在实际网络中的模型健壮性，也未建立完善的数据采集机制。

YouTube 平台为保护用户隐私，对其视频传输技术进行了改进，采用了组合传输机制。在该机制下，客户端根据当前的网络状况和缓冲区状态灵活选择请求的视频片段数量，从而将多个视频片段合并为视频块进行传输，同时音频也被组合为音频块进行传输。由于加密协议的干扰，使得视频块中的各个片段难以区分，导致依赖视频片段传输长度序列的传统方法失效。为了解决这一问题，Zhang 等人[66]提出了 RoVIA 方法，该方法首次对 YouTube 的组合传输模式进行了研究，并基于其关键特性实现了对 YouTube 加密视频流量的识别。RoVIA 方法通过提取视频索引文件中的视频片段序列作为视频片段指纹，并从加密视频流中提取每个视频块的传输长度序列作为流量特征。在匹配过程中，RoVIA 方法动态地组合视频片段指纹，形成视频块指纹，并使用动态时间规整（DTW）算法对流量特征与视频块指纹之间的相似性进行度量，从而实现加密视频流的识别。随后，YouTube 再次更新其视频传输模式，采用了音频片段和视频片段共同组合成一个传输单元的方式进行传输。为了应对这一挑战，Zhang 等人[97]提出了一种名为 Evi 的高效加密动态 DASH 视频流量识别方法，利用马尔科夫过程从训练样本中学习组合转移概率，生成轻量化的视频指纹库，这些指纹涵盖了多种可能的分段组合。在识别阶段，该方法从加密流量中提取块序列，并构建块转移图，通过动态规划算法计算多源最短距离来评估流量与指纹的匹配度，实现加密视频流量识别。与 Schuster-CNN[52]和 Dubin-BPP[76]方法相比，Zhang 等人的方法[66][97]在面对网络状况不佳、流量模式变化和固定质量播放带来的样本差异时，展现出了独特的识别优势，具有较好的场景适应性。然而，这类方法也存在一些局限性。首先，这类方法仅适用于 HTTP/1.1 协议，未考虑 HTTP/2 和 QUIC 协议的情况，协议泛化能力弱。其次，这类方法中的流量特征未进行深度的长度修正，特征表征能力有限。这使得流量特征与视频块指纹之间可能存在较大误差，从而影响其在不稳定网络环境中的识别能力。

与 Reed[63],[64]、Xu[65]和 Zhang[66]等人提出的使用视频片段传输长度序列作为特征的识别方法不同，Du 等人[62]通过修正视频片段传输长度，显著降低了流量特征与视频指纹之间的长度误差。他们提出了一种名为长短期频率（Long-Short Terms Frequency, LSTF）的方法，通过解密视频流获取视

频片段长度并构建视频指纹，再使用线性回归模型对视频片段的传输长度进行修正。视频片段修正长度和指纹被输入到词袋模型中，通过长短期频率匹配实现 TLS 场景下的加密视频流识别，能够适应如视频块丢失或重复出现的复杂情况，展现出较好的模型健壮性和场景适应能力。与 Gu-PDTW[54]、Wu-DF[55]和 Yang-markov[57]等方法相比，LSTF 在准确率、抗干扰能力以及计算效率方面均具有显著优势，使其在实际应用中表现出更高的可行性和实用性。然而，LSTF 方法也有一定的局限性，特别是对 TLS 协议的依赖，使其难以扩展到 QUIC 协议的场景。

为了实现对多种传输协议的泛用性，Du[62]团队中的 Tang 等人[67]对 Du-LSTF[62]方法进行了改进，提出了 Zenith 方法，旨在 TLS 场景和 QUIC 场景下实时识别带有失真的 DASH 加密视频流。Zenith 方法通过提取视频索引文件中的稳定视频片段序列作为视频指纹，构建指纹数据库。该方法针对 TLS、QUIC 以及 VPN 等不同环境，提出了相应的视频中间指纹提取方案，并采用随机游走模型将视频指纹转换为中间指纹。同时，Zenith 通过引入流量语言模型（Traffic Language Model, TLM）来模拟序列匹配问题，以应对视频数据丢失和重传的问题。此外，Zenith 还引入了频率字典结构，从而提升识别速度。Zenith 方法能够在使用半分钟的视频流量的情况下实现 97.32%的识别准确率，能够应对低带宽、自适应分辨率和 VPN 场景。然而，Zenith 方法在处理视频指纹时，采用随机游走模型将指纹转化为中间指纹，尽管这一方法显著减小了指纹库的大小并提高了计算效率，但也导致了大量视频指纹的丢失，进而影响了其在实际应用中的效果。

为降低因网络层、传输层及加密协议引入的特征波动带来的影响，Wu 等人[55]提出了一种基于视频片段明文长度序列的差分指纹识别方法，该方法基于视频片段明文长度构造明文差分指纹，将相邻视频片段传输长度的差值序列作为流量特征，并采用局部序列比对算法实现对加密视频流量的识别。研究团队对 HTTP/1.1 和 TLS 协议对视频片段传输长度的影响进行了详细分析，并通过回归模型对加密后的视频片段传输长度偏差进行修正，降低特征波动的影响。为验证方法有效性，Wu 等人[55]构建了包含约 20 万个指纹的大规模开放世界数据集，并在该数据集上进行评估。实验结果显示，该方法在连续截获 12 秒加密视频流的情况下，识别准确率可超过 98%，误报率接近 0。与使用密文差分指纹的 Gu-PDTW 方法[54]相比，该方法在所需视频播放时长更短、数据规模更大、识别准确率更高，具有明显优势。然而，该方法实验对象主要为 Facebook 平台视频，尚未在其他平台上验证其模型泛用性；此外，该方法依赖于对 HTTP/1.1 和 TLS 协议的深入解析，导致其对 HTTP/2 和 QUIC 等新协议的适应能力有限，缺乏协议适应性。

Björklund 等人[96]继承了 Reed 等人[63]的研究思路，设计了一套基于自适应码率分段特性的加密视频识别系统，在不依赖传输层明文信息的条件下，通过分析分段请求模式完成视频内容推断。该系统构建了覆盖多平台的大规模指纹库，并结合 KD-树检索与时间序列匹配实现高效指纹比对，在真实网络环境中对主流视频服务的识别精度超过 99.5%，且在 VPN 和 Wi-Fi 嗅探等复杂场景下仍保持较高健壮性。该方法采用的归一化与差分的特征处理方法一定程度上实现了跨传输协议的通用性识别，但是不可避免地存在特征表征能力弱的问题，无法应对复杂波动的真实网络环境。

随着网络传输技术的持续演进，HTTP/1.1、HTTP/2 与 HTTP/3 等多版本协议在实际应用中并行发展，形成了多协议共存的网络生态。针对不同协议版本所呈现的特征差异，学术界提出了多种面向加密视频流量识别的技术方案，力图在不同协议环境下实现对加密视频流量的有效识别。

HTTP/1.1 协议传输视频时采用非流水线通信模式，客户端与服务器之间严格按照请求-响应顺序进行交互，且每个请求的所有响应均具备相同的响应序列号，从而可据此将视频流划分为独立的视频片段。针对这一协议特性，Wu 等人[70]提出了 HHTF 方法，通过汇总相同响应序列号下的数据分组载荷，获得视频片段传输长度序列，并依次对该序列进行精确修正，逐步消除 HTTP 头部与 TLS 头部引入的传输长度偏移，最终生成视频片段的修正长度序列。随后，HHTF 方法采用 k 段序列匹配算法，实现了修正长度序列与明文指纹的匹配。实验结果表明，该方法使用连续三个视频片段传输长度在包含 205,677 个指纹的数据库中实现有效识别。相比于 Reed-segment[64]方法，HHTF 在视频片段传输长度的修正与特征表达能力上均有较大提升。然而，该方法依赖于模拟数据指纹库，尚未在真实网络环境中验证其场景适应性，并且主要针对 HTTP/1.1 协议下的非流水线加密视频流，对多路复用特性的

HTTP/2与HTTP/3协议的适应性和模型健壮性仍待提升。

HTTP/2协议在继承HTTP/1.1基本语义的基础上，新增了二进制分帧、多路复用和首部压缩等关键技术。具体而言，二进制分帧机制对数据封装和传输方式进行了重新设计，所有数据以帧为最小单元，并统一采用二进制编码；首部压缩利用HPACK技术将HTTP头部信息压缩成HEADERS帧进行独立传输；多路复用则允许每个请求或响应通过独立的流标识符在同一TCP连接中交错传输，实现了音视频等多种数据的高效并行传递。针对HTTP/2的协议特性，Wu等人[6]进一步开展了加密视频流量识别研究，提出了一种面向HTTP/2多路复用特征的识别方法。该方法首先利用特定长度的TLS记录进行音视频片段划分，通过片段长度过滤获取视频片段序列，并详细分析了HTTP/2传输过程中音视频数据长度偏移的机制，针对TLS头部、HTTP/2帧头和HTTP头部建模修正，实现了对加密音视频片段长度的逐层校正。随后，采用隐马尔科夫模型（Hidden Markov Model, HMM）和维特比算法实现加密视频的识别。实验结果显示，该方法在包含五个平台共423,890条指纹的数据集上能够有效识别固定分辨率与自适应分辨率视频。在与基于HTTP/1.1的Wu-HHTF[70]方法的对比实验中，无论在6,090条固定分辨率视频流还是766条自适应分辨率视频流的数据集上，Wu-HHTF[70]方法在5,000指纹库和400,000指纹库条件下的准确率和召回率均低于75%，而Wu-H2[6]在相同条件下均保持在97%以上。这一结果表明，基于HTTP/1.1的识别方法难以适应HTTP/2的多路复用机制。

HTTP/3协议以QUIC作为底层传输协议，具备流控制、基于流的多路复用以及全加密等关键特性。具体而言，HTTP/3通过在基于UDP的QUIC数据单元中嵌入专用的传输控制信息，实现了可靠性保障；流级多路复用机制允许客户端在同一连接内同时请求和接收多个音视频片段，服务器也可将多个片段组合后并行传输。然而，QUIC协议的全加密特性显著增加了分析难度，传输控制信息难以准确剥离，音视频块的组合数据也难以单独分割，从而对精确修正视频片段特征带来了巨大挑战。针对上述问题，Wu等人[51]提出了面向HTTP/3协议的加密视频流量识别方法。该方法基于对HTTP/3协议机制的深入剖析，直接提取组合后的视频块的传输长度特征，并逐数据分组修正由Packet Number、Offset、Length和Stream ID等字段引入的长度偏移，从而得到视频块修正长度序列。研究团队以视频片段明文指纹库为基础，生成视频块明文指纹库，结合隐马尔科夫模型和维特比算法，实现在YouTube、Facebook和Instagram这三大主流平台共72万条指纹的指纹库场景下对加密视频流量的识别。值得注意的是，该方法只需少量样本即可训练修正参数，对训练样本数量不敏感，展现出优异的特征表达能力与跨平台场景适应性。

随着针对YouTube组合传输模式的加密视频流量识别方法出现[66],[51]，YouTube平台对其传输机制进行了对抗性升级，引入了更加复杂的音视频块组合传输模式。客户端在不同网络状态和缓冲区条件下，可灵活选择请求的音视频片段数量，将多个视频片段与音频片段合并为单一的音视频块进行传输。这一变化极大增加了流量结构的复杂性，使得原有的识别方法难以适应新的流量模式。为应对上述挑战，Wu等人[71],[72]提出了面向音视频组合传输模式的加密QUIC视频流识别方法。该方法以音视频片段为基本单元，依据实际观测到的组合规律，预先对音视频片段进行组合生成音视频块长度，并以此为指纹存储于键值结构的内存指纹数据库中。利用QUIC协议特性，从加密流量中精准提取音视频块的传输长度序列，并通过基于协议特性的长度修正和基于线性回归的长度修正，消除由Packet Number、Offset、Length及Stream ID等字段带来的长度偏移，获得修正后的音视频块长度序列。结合隐马尔科夫模型和动态剪枝算法，实现了对音视频组合块的识别。实验结果表明，在开放世界环境下，该方法的准确率超过97%，展现出卓越的特征表征能力。然而，由于QUIC协议全加密特性，该方法在面对丢包、漏包等弱网环境时，仍存在一定局限，模型的健壮性和场景泛化能力有待提升。此外，针对QUIC协议的精细化处理，使得该方法难以迁移至其他类型传输协议场景，协议适应性有限。

随着短视频应用的快速发展，面向短视频平台的加密视频流量识别逐渐成为学术界关注的新热点。现有方法多聚焦于长视频的HAS传输模式，对短视频平台普遍采用的混合传输与预加载机制适应性不足。为解决该难题，Quan等人[73]提出了一种专门针对短视频混合传输与预加载特性的加密视频流量识别方法。研究团队首先构建了一个包含HTTP渐进下载（HTTP Progressive Download,

HPD）模式下视频文件长度和 HAS 模式下视频片段长度序列的混合视频指纹库。随后，研究团队从加密流量中提取视频资源长度特征，并采用线性回归模型依次修正 HTTP 头部、TLS 头部以及预加载启动数据引入的长度偏移，消除因数据缺失和协议封装导致的误差，从而获得精确的修正长度序列。最终，Quan 等人[73]基于短视频观看行为和平台预加载机制的特点，设计了以短视频模式驱动的优化隐马尔科夫模型（SVP-DOHMM），结合阈值与优化策略，实现了对 HTTP/1.1 协议传输的加密短视频流的识别。实验结果表明，该方法在含有超过十万条短视频指纹的真实数据集上，达到了 98.16%的识别准确率，在开放世界环境下，达到了 97.64%的识别准确率。该方法不仅适用于长视频场景，更能有效覆盖具备混合传输与预加载机制的主流短视频平台，展现了良好的模型健壮性和广泛的应用适应能力。

社交软件中的加密视频流量识别受到复杂的用户行为、频繁更新的内容以及动态波动的网络环境的影响，面临着极大的技术挑战。针对上述难题，Zhao 等人[74]提出了一种适用于复杂网络环境下社交软件加密视频流量的识别方法。该方案设计并实现了视频数据快速采集系统 VDRAS，能够根据网页标签和资源特征自动精准定位视频内容，并动态更新视频指纹库。在特征提取阶段，方法通过对 HTTP/2 和 TLS 协议的深入分析，系统性修复了因丢包和重传导致的数据分组乱序与缺失，得到了视频片段传输长度序列，借助协议逆向，逐层消除 TLS 头部、HTTP/2 帧头部及 HTTP 头部对长度特征的偏移，最终获得准确的修正长度序列。随后，利用隐马尔科夫模型实现修正长度序列与视频指纹的动态匹配。实验结果显示，该方法在高延迟、高丢包和跳转播放等复杂网络环境下，识别准确率超过 90.21%，并在 Twitter、Facebook、Instagram 等主流平台上实现了加密视频流量识别，具有极强的健壮性和平台泛化能力。需要指出的是，该方法尚未验证带宽受限下的识别问题，也无法对未知视频进行识别。

现有的加密视频流量识别方法通常依赖双向视频流量特征，这使得它们在广泛存在的非对称路由环境中难以有效适应。为了解决这一问题，Zhu 等人[75]提出了一种基于单向上行流量特征提取和隐马尔科夫模型的加密视频识别方法，用于在非对称路由场景中识别加密视频流量。该方法通过分析客户端 TCP 报文中的 ACK 信息，提取视频片段的传输长度，并利用回归模型将 HTTP 协议、TLS 协议对视频片段传输长度的影响进行修正。实验结果表明，该方法在包含 470,398 个视频指纹的数据库上，使用 25 到 60 秒的流量数据实现了 HTTP/1.1 协议传输的加密视频流量识别，准确率达到 98.21%，具有较强的特征表达能力。与传统方法不同，该方法仅依赖单向上行流量进行加密视频流量识别，因此特别适用于普遍存在的非对称路由环境，具有卓越的场景泛用性。然而，该方法在实际应用中也面临一些限制。它对领域知识的依赖较高，并且仅能应用于 HTTP/1.1 协议。虽然通过一定的改进可以将其应用于 HTTP/2 协议，但该方法无法应对全加密的 HTTP/3 协议。

总体而言，基于明文指纹的匹配识别方法具有如下特点：首先，明文指纹具有高度稳定性，不会受到网络环境和传输协议的影响，只与基于 HAS 技术生成的视频片段有关。其次，该类方法具有极强的协议适应性和较差的协议泛用性，该类方法可以通过领域专家知识和协议逆向，针对特定的传输协议进行特化，从而实现针对特定传输协议的加密视频流量识别，如：HTTP/1.1、HTTP/2 和 HTTP/3 等，但是也因为该类方法针对性的特化，导致其协议泛化能力受限。再次，由于明文指纹的高度稳定性和环境无关性，因此该类方法通过延迟补偿和丢包补偿等措施，能够很好的适应复杂的网络状态，如：高延迟、高丢包、低带宽等，并且能够在一定程度上应对 VPN 场景。最后，在工程应用方面，明文指纹具有唯一性，因此该类方法不会受到指纹库规模扩展带来的误差累计效应影响，能够实现大规模加密视频流量识别，具有更好的工程实践意义。

3.3 基于机器学习的分类识别法

基于机器学习的分类识别法是一类依赖特征提取和分类模型的加密视频流量识别技术。其核心思想是通过对加密视频流量在传输过程中的统计特征进行量化分析，构建特征向量，利用分类模型实现对流量所属视频内容的自动判别。不同于基于特征工程的匹配识别法侧重于与指纹库的直接特征匹配，基于机器学习的分类识别法关注于从大量流量数据中挖掘具有判别能力的流量特征，并通过训练分类器对未知流量进行归属推断。根据所用特征维度的不同，基于机器学习的分类识别法主要可分为基于一维统计特征和基于多维统计特征两类。

接下来，将分别对这两类方法的代表性研究进行系统综述与比较。

3.3.1 基于一维统计特征的分类识别法

基于一维统计特征的分类识别法，是指在领域专家知识指导下，对加密视频流量进行分析，提取精简的一维统计特征，以刻画视频传输过程中的应用层数据特征稳定性，然后结合具备良好健壮性的机器学习或深度学习分类模型，对提取的一维统计特征进行建模与判别，实现对加密视频流量内容的自动分类与识别。本文对基于一维统计特征的分类识别法进行了综述，并在表 6 中总结了相关研究成果。

Dubin 等人[76]首次提出了一种基于峰值比特率序列的加密视频流量识别方法，利用了 HAS 协议下视频流周期性 ON/OFF 模式的特性。具体而言，该方法将每两个 OFF 阶段之间的数据传输部分视为一个峰值段，通过累加每个峰值段内的总比特数，构建峰值比特率序列作为流量特征。该特征解决了因 HAS 技术迭代导致的定长时间窗口划分方法[63],[64]在分割视频片段时失效的问题。研究团队采用基于峰值的流量分割方式，提取峰值比特率序列，结合改进的最近邻、最近邻类、最近邻类唯一以及基于相似性特征的支持向量机（Similarities as Features Support Vector Machine, SFSVM）等四种机器学习算法进行加密视频流量分类。研究团队采集 HTTP/1.1 协议传输的 100 个视频的 100 轮流量样本作为数据集，并公开了数据集（Dubin Dataset[76]）。

表 6 基于一维统计特征的分类识别法研究汇总

方法[1]	模型	特征	复杂环境	平台	视频数	封闭世界性能	开放世界性能
Dubin-BPP (2017) [76]	NN、NNC、NNCU、SFSVM	峰值比特率	时延、丢包	YouTube	100	准确率:>90%	准确率:97.8%
Dvir-NLP (2020) [78]	K-Means、GMM、AHC、SC	峰值比特率	未考虑	YouTube	100	CP:74% NPC:84% NSCB:68% CCS:5724	未考虑
Schuster-CNN (2017) [52]	CNN	突发流量序列	带宽限制	Netflix YouTube Amazon Vimeo	100 18 10 10	准确率:79%~98.4% 召回率:93%~98.8%	准确率:99.4% 精确率:99.5% 召回率:99% BDR:≈50%
Li-MLP (2018) [84]	MLP、LSTM、CNN	下行非数据帧数量	未考虑	YouTube	10	准确率:97.5%	未考虑
Li-CNN (2022) [85]	CNN	下行非数据帧数量	未考虑	Netflix Stan YouTube	30	准确率：55.4%~99.3% 精确率：61.68%~99.33% 召回率：55.4%~99.3% F1 分数：52.64%~99.3%	准确率：95.3%~100%
Bae-LTE (2022) [79]	CNN	滑动窗口字节	网络拥塞	YouTube Netflix Amazon	100 53 32	准确率:87%~98.5%	准确率:>90%
Yang-siamese (2023) [81]	k-NN、SVM、RF、XGBoost、DT、岭回归	字节率	加入瑞利分布的噪声	YouTube	100	准确率:73.31%~97.39%	未考虑
Yang-triplet (2024) **Error! Reference source not**	逻辑回归、SVM、k-NN、DT、RF、XGBoost	字节率	加入瑞利分布的噪声	YouTube	100	准确率:82.87%~94.08%	准确率:96.89%

found.							
Xu-CNN-LSTM (2024) [83]	CNN-LSTM	视频片段传输长度序列	VPN，海量背景流量	Twitter Dailymotion	48 49	准确率：91%~95% 精确率：91%~96% 召回率：90%~95% F1 分数：90%~95%	未考虑

注：[1]方法名根据作者和方法特点自拟

在该数据集上，该方法在良好网络条件下的分类准确率超过 90%，在 600ms 延迟和 6%丢包率下，基于 SFSVM 的分类模型在封闭世界的准确率保持在 80%以上，在开放世界也超过 70%，具有良好的协议适应性、场景适应性。然而，该方法的缺点在于需要 90 轮样本数据进行模型训练，模型训练成本较高，模型的泛用性差。

针对实际应用场景中在无已知标签的情况下进行加密视频流量识别的需求，Dubin[76]团队的 Dvir 等人[78]在 Dubin-BPP[76]方法基础上进行了改进，将研究范式从有监督分类拓展到无监督聚类。以峰值比特率序列为核心特征，在特征表达上融合了自然语言处理（NLP）方法，分别利用 Word2Vec、词袋模型、深度自编码器和交集相似度函数将峰值比特率序列转化为特征向量。随后，研究团队分别采用 K-Means、高斯混合模型、谱聚类和层次聚类等多种聚类算法，实现了加密视频流在无标签条件下的自动分组。实验结果表明，该方法在 Dubin Dataset[76]上获得了 0.74 的聚类纯度，证明了无监督聚类技术在加密视频流量识别中的可行性。这一工作有效缓解了实际应用中标签稀缺问题，提升了方法在开放世界和大规模真实环境下的场景适用性，为未知视频内容的自动识别提供了新的思路。

Schuster 等人[52]深入研究了 MPEG-DASH 流媒体协议中，由 VBR 编码引发的加密视频流量信息泄漏问题，并提出了一种基于视频流突发流量特征的加密视频识别方法。该方法将时间间隔小于 0.5 秒的连续数据分组定义为突发流量（ON 阶段），并通过将每个突发流量内的数据分组载荷进行累加，形成突发流量字节特征。此外，研究还提取了上下行流的总字节数、数据分组数等其他统计特征，并分别使用 CNN 对这些特征进行模型训练和分类。实验结果表明，该方法在 YouTube 数据集上的召回率达 98.8%，误报率为零，并通过数据计算证明了应用层数据特征的唯一性。此外，研究还验证了该方法在 Netflix、Amazon 和 Vimeo 数据集上适应性。然而，该方法假设流量足够纯净，难以应对背景流量的干扰，也无法处理自适应分辨率场景，模型的健壮性有限。并且，该方法只能应用于非多路复用的 HTTP/1.1 协议，无法应对多路复用带来的干扰，协议适应性有限。

随着移动互联网的普及，Wi-Fi 应用场景变得越来越广泛。在此背景下，Li 等人[84]首次实现了通过被动监听分析 WPA2 加密 Wi-Fi 流量，并基于 MAC 层帧的统计特征开展加密视频流量识别。该方法通过将采集到的 YouTube DASH 加密视频流量按 0.36 秒的时间窗口进行分割，统计了上行与下行数据帧及非数据帧（如控制帧和管理帧）的数量、字节总数、最大/最小分组长度、均值与方差等八类统计特征，并分别将这些特征输入多层感知器(Multilayer Perceptron, MLP)、循环神经网络(Recurrent Neural Network, RNN)和 CNN 进行模型训练和分类识别。实验结果表明，该方法所用特征中下行非数据帧数量与应用层数据特征的关联度最高，采用该特征的 MLP 模型在包含 10 个不同视频的 3198 条流量样本的数据集上实现了 97.5%的识别准确率。通过自身研究以及与 Schuster 等人[52]基于一维 CNN 的加密视频流量识别研究对比，他们发现 MLP 模型不仅结构更为简洁，资源消耗更低，仅需两层隐藏层即可获得优异识别效果，并表现出更强的模型健壮性。

为了进一步推动加密视频流量识别的自动化，Li 等人[85]在 Li-MLP[84]方法的基础上提出了分层识别架构。该方法通过三个专用的 CNN 模型，依次实现流量类型、流量平台和具体内容的三级分类，分层架构采用模块化设计并复用参数，减少了 42%的参数量，同时增加了对音频和网页等新类型流量的识别能力。基于下行非数据帧数量特征的模型在加密 Wi-Fi 流量的流量类型识别任务中达到了 99.4%的准确率，对流量平台的识别准确率也达到了 98.1%。然而，尽管该模型能够高效识别流量类型，但是其轻量化设计使得其在精细度上的表现有所妥协：静态网页的识别准确率超过 98%，但对加密视频流量的识别准确率下降至 80.5%；并且对如 Spotify、Xiami 等音频内容的场景泛化能力较弱，准确率仅为 30%左右，表明该方法在特征泛化能力

和模型健壮性方面仍存在一定的不足。

Bae 等人[79]提出了一种基于长期演进网络（Long Term Evolution， LTE）无线信号的加密视频流量识别方法，该方法通过提取 LTE 广播控制信道中的下行控制信息（Downlink Control Information，DCI）并估算时间窗口载荷特征来实现视频流量识别。研究者利用身份映射技术和非密信息解析技术，克服用户标识符频繁变更和多信道传输等挑战，将 DCI 中的下行数据分组载荷按照 0.2 秒的等长时间窗口进行划分，并聚合每个窗口内的数据分组载荷，形成时间窗口载荷特征。通过训练和使用一维 CNN 分类模型进行加密视频流量的识别。实验结果表明，在包含 46,810 条真实 LTE 流量实例的数据集上，该方法对 YouTube 视频的识别准确率达到了 98.5%。虽然，该方法通过精细化的特征提取实现了高效的识别，但是，基于 LTE 协议的精细化提取方式不可避免的导致其协议泛用性受到限制，无法扩展至其他协议场景。

为探索加密视频流量识别中通用特征的构建可行性，Yang 等人[81]提出了一种基于视频流字节率序列的低维嵌入方法 EVS2vec，首次将低维特征嵌入技术应用于该领域。他们利用结合了长短期记忆网络（Long Short-Term Memory Network，LSTM）与自注意力机制的嵌入模型，将预处理得到的一维变长字节率序列转换为 8 维定长特征向量，实现了特征的标准化，降低了存储成本，提升了相似性计算的效率。该方法采用孪生网络结构和对比损失函数进行训练，使得同一视频流在嵌入空间内距离较近，而不同视频流距离较远。实验结果表明，在 Dubin Dataset[76]上，EVS2vec 在相似度阈值判别任务中达到 96.71%的准确率，分类准确率为 97.39%，聚类准确率为 73.31%，并解释了其在指纹识别场景下的理论可行性。为了进一步提升特征表征能力和场景适应性，Yang 等人 **Error! Reference source not found.**将孪生网络扩展为三元组网络（triplet），增强了嵌入空间中类内聚合与类间区分的能力，最终实现了 93.60%的指纹识别准确率和 86.46%的聚类准确率，并验证了方法在瑞利噪声环境下的健壮性。进一步在包含 100,000 对相同与不同视频流样本的开放世界测试中，与 Yang-levenshtein[58] 方法对比，Yang-triplet**Error! Reference source not found.**的识别准确率提升了 7.89%，达到 96.89%，展现出更优的开放场景适用性。值得注意的是，该方法未引入组流操作，尚不能有效抵抗背景流量干扰。

随着互联网的持续发展，VPN 环境下的加密视频流量识别需求日益突出。针对该场景，Xu 等人[83]提出了一种基于视频片段传输长度特征的 VPN 加密视频流量识别方法，将 HAS 协议下客户端请求所形成的视频片段传输长度序列作为主要特征。具体而言，研究团队首先通过分析加密流量中的有效负载长度和分组到达间隔，采用 K-means 聚类方法构建标准分布模板，并借助欧氏距离阈值过滤掉背景流量。随后，对筛选出的 VPN 流按照客户端请求分组进行分割，提取视频片段传输长度的分布序列，并将其输入 CNN-LSTM 混合神经网络进行加密视频流量识别。该方法在融合了 57,000 条主干网背景流量的 Twitter（48 部短视频）和 Dailymotion（49 部 3 分钟视频）的 VPN 加密视频流量数据上进行了实验。实验结果显示，该方法在 Twitter 和 Dailymotion 上的识别准确率均可达到 95%; 在采用 20 轮训练样本的情况下，准确率保持在 90%以上。该方法通过深度学习模型充分挖掘了视频片段传输长度特征的表征能力，有效克服了传统 ON/OFF 峰值特征在 VPN 场景下因分组混淆和 TLS 二次加密失效的问题，实现了在隐私增强环境下的加密视频流量识别，并展现出良好的协议泛用性和应用场景适应性。然而，该方法需要预先采集 40 轮训练样本训练 CNN-LSTM 模型，且对 CDN 缓存策略的敏感性较高，导致训练与更新成本较大，模型健壮性尚需提升。

总体而言，基于一维统计特征的分类识别方法具有如下特点：首先，该类方法依赖于领域专家知识，通过对 HAS 传输机制及相关协议的深入理解，能够有效提取反映应用层数据特性的一维统计特征，但其稳定性易受到网络状态波动的影响，导致流量特征表现出一定的不确定性。其次，由于一维统计特征对流量模式变化敏感，该类方法难以适应多路复用场景下并发请求与响应导致的流量特征剧烈变化，但对于非流水线传输模式下特征较为稳定的加密视频流量，仍具有较强的识别能力。再次，依托于机器学习和深度学习模型的健壮性，该类方法能够在高延迟、低带宽、瑞利噪声等复杂网络环境中容忍一定程度的特征扰动，支持 VPN 等隐私增强场景下的识别任务。然而，在实际工程应用中，由于多数方法不具备有效的组流能力，对大规模背景流量的抑制和处理能力有限，难以满足真实应用中高背景噪声场景的需求。此外，机器学习和深度学习模型对大规模训练数据的依赖以及分类规模

的限制，也影响了模型的训练效率和可扩展性，使得该类方法在实际中可识别的视频数量和模型更新能力存在一定瓶颈。

3.3.2 基于多维统计特征的分类识别法

基于多维统计特征的分类识别法，是指通过对加密视频流量中各类数据分组传输参数进行量化分析，充分挖掘并利用反映流量传输行为的多维统计特征，进而构建特征向量，结合具有健壮性的机器学习或深度学习分类模型，实现对加密视频流量的归属判别或内容预测，而无需对流量进行解密。即便在实际应用中，用户采用如 Tor、VPN 等隐私增强技术以提升通信的匿名性和不可识别性，加密视频流量在传输过程中的多维统计特征依然可能呈现出一定的规律性，为建模与识别提供了可能。此类统计特征既包括分组总数、最大/最小分组长度、平均分组长度等原始统计量，也涵盖了如分组长度方差、标准差等经数学变换得到的衍生特征。这些特征能够有效揭示视频传输的规律性和结构性。本文对基于多维统计特征的分类识别法进行了系统梳理，并在表 7 中进行了归纳总结。

随着用户对隐私保护需求的提升，匿名通信网络的应用日益广泛，其中 Tor 网络是典型代表，有大量研究人员针对 Tor 进行网站指纹识别（Website Fingerprinting, WF）研究[98]-[106]。针对隐私增强场景下的加密视频流量识别问题，Rahman 等人[86]首次将 WF 策略引入 Tor 环境下的加密视频识别研究，提出了一种结合数据分组到达时间与方向的二维特征方法。该方法充分利用了 Tor 网络传输采用固定长度加密单元的特性，仅提取并保留数据分组的方向和到达时间信息，通过二者相乘生成 Tik-Tok 特征向量，有效规避了 Tor 的链路加密与填充机制带来的信息丢失。具体实验中，研究团队将每条流量样本按定长截断后转化为 Tik-Tok 特征向量，并输入基于深度指纹识别的 CNN 模型实现封闭世界下的分类识别。该方法在包含 50 个 3 分钟音乐视频和 32,000 条真实 Tor 加密视频流的大规模数据集上，实现了 55%的识别准确率。需要指出的是，由于 DASH 协议下视频流比特率和分段结构随网络环境波动显著，同一视频在不同会话中的流量表现差异较大，而现有深度指纹模型尚未针对视频流特性进行专门优化，导致特征表征与泛化能力有限，整体准确率受限。

为了检验基于机器学习的分类识别法在 Tor 匿名网络开放世界场景下的适用性，Walsh 等人[87]在 Rahman-DF[86]方法的基础上，提出了一种融合统计时序特征的加密视频流量识别方法。该方法通过联合建模数据分组的方向、到达时间和字节大小特征，支持对 YouTube、Facebook、Vimeo 和 Rumble 等主流平台的加密视频流量识别。针对 HAS 协议

表 7 基于多维统计特征的分类识别法研究汇总

方法[1]	模型	特征	复杂环境	平台	视频数	封闭世界性能	开放世界性能
Rahman-DF (2019) [86]	DF	DT 序列和 Tik-Tok 序列	Tor	YouTube	50	准确率:54.7%	未考虑
Walsh-DF (2024) [87]	DF	分组级 D+T+S 组合特征	Tor	YouTube Facebook Vimeo Rumble	60 60 60 60	准确率:75.35%~99.91%	假阳率:≈0.1% 召回率:51.67%~81.67%
Kattadige-SETA++ (2021) [89]	XGBoost	分组级时序统计特征	VPN	YouTube Netflix Stan	20 20 20	准确率: 94.4%~100%	未考虑
Liu-ITP-KNN (2020) [88]	KNN	ITP 特征集	VPN、Tor	YouTube	100	精确率:98.18% 召回率:97.82% F1 分数:98%	未考虑
Amjad-ENCVIDC (2022) [90]	RF、K-NN、SVM	统计特征与信息熵特征	未考虑	YouTube Netflix Dailymotion	12 12 12	准确率:98.13%	未考虑

注：[1]方法名根据作者和方法特点自拟

自适应码率导致的流量长度波动，研究团队将每条四分钟视频流按 1/8 秒滑动窗口划分，统计每个窗口内上行和下行字节数和累计到达时间，从而生成高维统计指纹。为贴近实际应用，该团队在全球 10 个地理节点采集加密视频流，构建了一个开放世界加密视频流数据集，涵盖 HTTPS 与 Tor 两种类型（包括未监控视频的 78,694 个 HTTPS 流量和 76,692 个 Tor 流量）。在模型优化上，采用了改进的深度指纹识别网络以适配不同特征表达形式。实验结果显示，在 HTTPS 环境下，该方法对封闭世界 YouTube 加密视频流量识别的准确率达到 99.19%，在 Tor 场景下准确率为 97.12%；而在 60 部监控视频的 931 条流量与 64,000 部非监控视频流的开放世界测试中，在召回率为 0.5 时，误报率低于 0.002%，体现了良好的场景泛用性。然而，当监控集规模扩展时，统计特征易遭遇“基数灾难”，误报风险显著提升。此外，方法对每个目标视频需采集约 850 条流量轨迹，造成数据收集与存储成本较高，且对不同平台的缓冲机制较为敏感，影响其跨平台泛化能力。

针对隧道场景下加密视频流量识别的实际需求，Kattadige 等人[89]提出了 SETA++架构。该方法利用分组级统计特征，面向 VPN 加密环境下的视频流量分类，将 0.36 秒滑动时间窗口内的数据分组总数、总字节数、平均分组长度等基础统计特征，与 10.8 秒时间窗口内的均值、最大值、标准差等高阶统计特征相结合，构建出包含 63 维的高维流量特征向量。借助 Flow Sampling Sketch 特征压缩与 XGBoost 分层分类模型，SETA++实现了在千兆级高速网络下对 VPN 加密视频流的有效识别。实验结果显示，在涵盖 YouTube、Netflix 和 Stan 三大平台、共 60 个视频、6000 条 VPN 加密流量的测试集上，SETA++平台分类准确率达到 99%，内容分类准确率为 94.4%，即使在 VPN/SSL 加密下依然具备高水平识别能力。这说明，底层统计特征能够有效表征应用层数据特征，并支持在 VPN 环境下的识别应用。需要指出的是，该方法无法应对剧烈网络波动、复杂背景流量和高噪声条件，场景适应性和模型健壮性有待提升。

Liu 等人[88]提出了“间歇性流量模式”（Intermittent Traffic Pattern, ITP）概念，用于刻画加密视频流量中的应用层数据特征，并针对由视频分段传输产生的 ITP 特征开展了识别研究。在此基础上，研究团队设计了 ITP-KNN 方法，构建了一套 70 维的统计特征集以全面描述 ITP 特性。该特征集分为两类：一类为分组到达间隔（Packet Arrival Interval, PAI）的量化特征，针对上行和下行流量分别利用 23 个从 0 至 5 微秒起始、以倍增方式划分的时间窗口统计分组，拼接后形成整体描述；另一类为整体统计特征，分别对上、下行流量的时间序列计算多项描述性及高阶统计量，如最大值、最小值、均值、偏度和峰度等，从不同角度刻画时序特性。该方法将上述特征与 K-近邻（K-Nearest Neighbors, KNN）分类算法结合实现加密视频流量识别。实验结果表明，在涵盖 HTTPS、QUIC、VPN 及 Tor 等多类加密流量、共 30.2 万条流的数据集上，ITP-KNN 方法在 K=9 时召回率达 97.82%，精确率为 98.18%。然而，该方法目前主要针对具备 ITP 特征的视频平台，平台泛用性有待拓展，并且尚未充分验证在大规模高速网络环境下的表现，场景适应性仍需加强。

为促进加密视频流量识别方法向平台通用性和协议通用性的方向迈进，Amjad 等人[90]提出了一种低数据量特征提取方法，该方法仅利用加密视频流前 50 个数据分组的统计特征和信息熵特征，实现了跨平台加密视频流量的分类。基于此，研究者设计了 ENCVIDC 系统，并首次将柯尔莫哥洛夫熵（Kolmogorov Entropy）引入加密视频流量识别领域，用于度量压缩加密数据的随机性。此外，结合香农熵（Shannon Entropy）构建了特征集，增强了特征表达的能力。在特征提取方面，团队从前 50 个数据分组中提取了 51 维统计特征，包括分组大小、标准差、流平均持续时间、分组间时间间隔以及双向数据量比值等，并结合 6 维熵特征，共同描述加密负载的随机性特征。为了提高模型的效率，研究者提出了基于贪心搜索的特征降维策略和动态阈值更新机制，最终构建了基于随机森林、KNN 和支持向量机的加密视频流量分类器，并引入基于接收者操作特征曲线下面积（AUC-ROC）的自适应阈值优化模块，实现了模型的在线动态更新。实验结果表明，该方法在包含 YouTube、Netflix、Dailymotion 等平台的 12 类 3036 条视频流的真实数据集上，RF 分类器的识别准确率达到了 98.13%，每条流量的处理时延为 0.005 毫秒，展现了优异的平台泛化能力和协议适应性。Amjad 等人[90]引入的熵特征在分类低码率视频（如 144p）方面具有明显优势，因为熵特征能够量化数据的不确定性和压缩复杂度，从而稳定地捕捉不同内容类别的统计规律，展示了强大的特征表征能力。但是，该方法仅

能实现视频类别级的粗粒度识别，未能在视频内容级别进行细粒度识别。

总体而言，基于多维统计特征的分类识别方法具有如下特点：首先，多维统计特征之间能够实现一定程度的相互补偿和容错，从而提升了流量特征在复杂网络环境中的稳定性，对高延迟、丢包等网络波动具备较强的抗干扰能力。其次，依托于多维特征的丰富表达和分类模型的健壮性，该类方法能够有效应对多路复用、定长填充及隧道加密等复杂协议场景，在协议泛用性和场景适应性方面表现突出，能够适用于如 Tor、VPN 等多种隐私增强环境。然而，与基于一维统计特征的方法类似，多维统计特征方法在实际应用中也受到分类模型对大规模训练数据的高度依赖以及分类类别扩展能力有限的约束，因此，难以高效应对实际工程场景中大规模加密视频流量的实时识别与模型动态更新需求。

3.4 两类方法的对比分析

在加密视频流量识别领域近二十年的研究中形成的基于特征工程的匹配识别法和基于机器学习的分类识别法两大技术体系具有各自的优势与缺陷。

基于特征工程的匹配识别方法依托领域专家知识与匹配算法，将加密视频流量的序列特征与视频指纹进行相似度计算，从而实现加密视频流量的识别。该类方法的显著特点是对训练数据需求量较低，通常仅需一次或少量几次指纹采集即可构建指纹库并完成识别。例如，Huang-Mix[59]、Wu-H3[51]和 Zhao-software[74]等方法均在单轮采集条件下实现了有效的匹配识别。此外，由于识别过程依赖指纹库而非大规模模型训练，该类方法在应对新增视频时只需扩展或替换指纹库即可完成识别，展现出较强的实用性，Wu-H2[6]、Wu-H3[51]和 Hu-QUIC[71]等方法均可通过替换指纹库实现新增视频的识别。依托低数据需求、指纹库可扩展性和匹配算法高效性，该类方法能够在大规模场景下实现高效识别，例如，Tang-Zenith[67]、Quan-short[73]和 Zhu-asymmetric[75]等方法分别支持 62,232、111,105 和 362,502 个视频的识别。然而，该方法依赖领域专家知识进行特征设计，并且在协议和平台适配性方面存在局限，其泛化能力不足，例如，Zhang-RoVIA[66]、Du-LSTF[62]和 Zhao-QUIC[72]等方法局限于 YouTube 平台，Wu-DF[55]、Wu-HHTF[70]和 Zhang-RoVIA[66]等方法则主要针对 HTTP/1.1 协议，难以直接推广至不同平台和多样化的协议环境。

基于机器学习的分类识别方法依托对加密视频流量统计特征的量化分析，借助机器学习或深度学习模型的特征学习能力，训练分类器以实现加密视频流量的识别。该类方法的显著优势在于泛化能力较强。例如，Bae-LTE[79]、Schuster-CNN[52]和 Li-CNN[85]等方法均在多个视频平台上验证了其跨平台的适应性；Xu[83]、Wu[6]和 Hu[71]等人的研究分别针对 HTTP/1.1、HTTP/2 和 HTTP/3 协议场景，并均与 Bae-LTE[79]方法进行对比，结果表明 Bae-LTE[79]方法具备较好的协议通用性。此外，凭借机器学习和深度学习模型的健壮性，该类方法能够在隐私增强场景中实现有效识别，如 Xu-CNN-LSTM[83]、Liu-ITP-KNN[88]和 Rahman-DF[86]等方法在 Tor、VPN 等场景下均展现出较高的识别准确率。然而，该类方法在实际应用中仍存在局限。一方面，模型训练依赖于大量标注数据，例如 Dubin-BPP[76]、Li-MLP[84]和 Walsh-DF[87]等方法分别需要 90 轮、200 余轮和 800 余轮训练数据，增加了成本和复杂度。另一方面，视频数目增加将导致分类模型需要学习更复杂的决策边界和更细微的特征差异，如果模型容量和数据量未相应增加，则准确性必然下降，为了保证准确性，现有使用此类方法进行视频识别研究的实验中视频规模有限，Dvir-NLP[78]、Yang-triplet**[Error! Reference source not found.]**和 Bae-LTE[79]等方法仅支持 100 个视频的识别，难以满足大规模应用需求。

总体而言，基于特征工程的匹配识别方法与基于机器学习的分类识别方法各具优势：前者对数据依赖低、扩展性强，适用于大规模识别；后者泛化能力突出，尤其在隐私增强场景中具有更强的适应性。

3.5 加密视频流量识别防御对策

加密视频流量识别防御对策主要针对主流视频服务提供商，尽管部分防御策略已在实践中应用，但大多数研究方案仍以理论可行性为主，缺乏实际部署经验。

最常见的防御策略之一是数据填充，主要包括固定大小填充[52],**[Error! Reference source not found.]**,[86]和随机填充[71],[73],[79]两种方式。固定大小填充通过将传输数据填充到最大传输单元的倍数来增强防御效果，但会导致显著的带宽消耗，特别是在服务器资源受限的情况下，效率降低较为严重。相比之下，随机填充虽然减少了填充量，但同样会增加带宽占用，

且对视频服务商的核心目标——用户体验产生负面影响。目前，Facebook 和 Instagram 已采用小规模的随机填充策略来对抗加密视频流量识别，但由于其填充量较小，防御效果仍有限。

另一种常见并且非常有效的防御策略是组合传输[64],[74]。此策略通过将音频和视频片段随机组合成音视频块进行传输，从而模糊视频片段的边界，增加流量模式的随机性，有效对抗加密视频流量的识别。YouTube、Facebook 和 Instagram 等平台已采纳此防御方法，通过将音频片段随机组合成音频块，视频片段随机组合成视频块进行传输[51],[66],[67]。YouTube 在 2023 年底进一步改进了这一模式，实现了音频片段与视频片段的共同随机组合传输[71],[72]。组合传输策略能够有效抵抗基于机器学习的分类识别法，并显著降低基于特征工程的匹配识别法对视频的识别能力[51],[66],[67],[71],[72],[97]。这一方法通过增强数据流的随机性，为视频内容的保护提供了更强的防护机制。

一些防御对策仍处于理论阶段，例如可变缓冲区[52],[63],[73]、固定比特率传输[52],[53],[79]、固定比特率编码[52],[63]和视频可变切分[58],[66],[79]等。可变缓冲区策略要求客户端实施相关操作，随机调整缓冲区容量，以使视频片段的请求间隔变得随机化。然而，这种策略会影响用户体验，且无法有效防止基于特征工程的匹配识别法，因此在实际应用中的价值较低。固定比特率传输策略通过严格控制视频片段的传输速率破坏视频流量的统计特征，但会导致带宽占用过大或信息密度较高的视频片段无法及时传输，从而影响用户体验，这对于视频服务商来说是不可接受的。固定比特率编码通过对视频进行固定比特率编码，使每个视频片段大小相同，但在低信息量的场景下浪费码率，而在高信息量的场景中则可能导致质量下降，视频服务商同样无法接受。视频可变切分涉及服务器资源成本，要求服务器对视频进行实时随机分段，并将分段的索引文件发送给客户端，客户端再根据请求播放视频。这种方法会导致用户初始缓冲时间显著增加，影响用户体验，同时，服务器需要为每个用户存储随机分割的视频片段，占用大量存储空间，因此视频服务商无法承受。尽管这些防御对策在理论上可行，但在实际应用中面临较大挑战。

4 研究挑战与展望

经过近二十年的发展，加密视频流量识别领域取得了诸多具有代表性的研究进展。然而，目前大部分成果仍集中在理论方法与模型验证层面，在实际部署和大规模应用过程中仍面临诸多挑战和亟需攻克的关键问题，主要表现在以下几个方面。

4.1 问题与挑战

4.1.1 理论研究体系薄弱

目前，无论是基于特征工程的匹配识别法还是基于机器学习的分类识别法，加密视频流量识别研究都是以实验验证为主要研究方法。研究通常依赖于现有数据集，先对流量特征进行选择或提取，再采用分类或匹配模型进行识别，并最终通过相关性能指标对方法效果进行评估。该研究范式存在两方面的主要不足。

首先，在特征选择方面缺乏系统性的理论指导。尽管早期研究已确定应用层数据特征具有唯一性与较强的稳定性，并且能够通过 ON/OFF 模式提取出具有一定表征能力的流量特征，但如何进行最优特征表征的理论依据仍不充分。例如，视频片段修正长度序列虽能精确表征应用层数据特征，但其获取过程对协议知识和专家经验要求较高，且需针对不同平台和协议进行定制，难以推广至通用场景。其他统计特征在表征能力上又存在不足，且特征种类与数量的确定往往依赖于多轮实验，缺少明确的理论依据。此外，当前的特征分析方法通常与特定数据集强相关，尚未建立起具有通用性的特征分析体系

其次，在识别模型的选取方面同样缺乏系统性的理论支撑，现有工作大多仍停留在经验驱动和效果对比的层面。具体来看，基于机器学习的分类识别方法往往通过研究者对若干常见分类模型进行实验对比，或依据个人经验与偏好确定最终模型配置，却较少从理论上分析模型结构与加密视频流量特征之间的契合机理；基于特征工程的匹配识别方法亦主要依赖领域专家的直观判断，在众多匹配策略中选择“看似合理”的方案，而缺乏对不同匹配算法优劣及其在加密视频流量识别任务中适用边界的系统性论证。从理论建设的角度看，特征设计与模型选择应当形成统一的分析框架，通过揭示特征空间性质与模型表达能力之间的对应关系，建立可解释的“特征－模型匹配”理论体系，从而为识

别模型的结构设计、参数配置及性能优化提供更加科学、可推演的依据。

4.1.2 研究假设理想化

为了简化学术研究，加密视频流量识别领域通常设定了场景假设、模板假设和识别者能力假设等前提条件。然而，这些假设的设定在一定程度上限制了该研究方法的实际应用，导致其与实际部署存在显著差距。

在场景假设方面，现有研究普遍依赖较为理想化的前提假设，难以充分刻画真实网络环境的复杂性。对于基于特征工程的匹配识别方法而言，通常要求对加密流量进行精细的组流、分片与长度修正，以保证流量特征与视频指纹在片段级别严格对齐。然而，在真实网络中普遍存在的丢包、漏包、重传、乱序等现象，会直接破坏这一精确对齐前提，使得特征修正过程面临较大不确定性，进而削弱方法的实际识别能力。相对地，基于机器学习的分类识别方法通常通过引入覆盖多种网络状态的大规模训练数据来提升模型对不同场景的适应性，但受限于数据采集成本和可观测场景边界，训练集难以穷尽真实网络中的多样化场景，模型在未见场景下的表现仍存在显著不确定性。为贴近实际应用需求，部分工作尝试突破传统封闭世界和开放世界的理想化假设框架，例如 Zhao 等人[74]面向高时延、高丢包环境进行了加密视频流量识别研究，Tang 等人[67]则考察了不同带宽条件下的识别性能。然而，这类研究在实验设计中对网络环境的建模仍相对简化，与复杂现网中多因素耦合、动态变化的实际运行环境仍存在明显差距，尚不足以全面反映复杂网络条件下加密视频流量识别所面临的真实挑战。

在模板设定方面，现有大量工作往往隐含地假定视频流在较短时间窗口内呈现出相对固定的特征模式，即所有样本均可由某一统一模板加以刻画。随着主流平台在编码策略、分段时长、自适应分辨率等方面持续演进，这一前提假设逐渐难以成立：在同一时间尺度下，部分视频流可能仍大致服从既有模板，而另一些则表现出明显偏离。如果仍将“所有视频流均满足统一模板”作为基础前提，势必引入系统性建模偏差。对于基于特征工程的匹配识别方法而言，传输模式的变更会直接改变特征还原的粒度和结构，例如组合传输模式下，流量特征往往只能对应到音视频合并后的块级数据，而无法准确回溯到单个视频片段，使得流量侧特征与指纹库中片段级指纹之间存在一对多甚至多对多的错配，显著放大匹配误差。对于基于机器学习的分类识别方法，传输机制迭代则会从整体上改变流量的时序形态和统计分布，使得不同时期采集的数据样本在特征空间中相互脱节，从而导致基于旧传输模式训练的模型难以迁移到新技术栈下的流媒体场景，影响模型的长期可用性与泛化性。

在识别者能力假设方面，现有研究通常假定识别者拥有与用户完全相同的先验知识，这一假设显然过于理想化。此外，识别者能力假设假定识别者具备精确的流量分割能力，然而，实际情况中并非所有识别者都能具备此能力。尽管有研究试图通过仅采集纯净流量来弥补这一能力的不足，但这种方法可能会削弱识别方法对背景噪声的抗干扰能力，导致其在实际应用中的表现与理论假设存在较大偏差。

4.1.3 实际部署难度大

尽管加密视频流量识别领域已取得一定的研究进展，但距离实际部署和应用仍存在较大的差距。除了研究中存在假设过于理想化的问题外，主要有两个核心问题尚未得到有效解决，即数据过时和增量式视频识别问题。

数据过时问题指的是研究所用数据的采集时间与实际应用场景之间存在明显滞后，使得加密视频流量识别方法在当前网络环境下性能显著下降甚至失效。对于基于特征工程的匹配识别方法，这一问题通常由传输协议演进直接触发，例如原本围绕 HTTP/1.1 设计的特征提取与指纹构建流程，难以在全面迁移至 HTTP/3 的传输环境中继续适用，导致指纹含义与流量语义出现脱节。相较之下，基于机器学习的分类识别方法在协议升级方面具有一定的容错空间，但同样会受到视频切分策略调整的显著影响。当平台改变片段划分粒度或自适应调度策略时，视频传输过程中的时序模式与统计分布将随之整体迁移，使得历史训练样本中学习到的流量特征与当前实际流量之间出现分布偏移，从而削弱模型的判别能力。因此，如何在充分利用既有“过时”数据的前提下，构建易于在线迭代更新并具备跨协议、跨平台通用性的加密视频流量识别框架，已成为推动相关技术走向长期可用与工程落地的关键方向。值得注意的是，Tang 等人[67]已围绕跨协议通用性开展初步探索并取得一定进展，但此类工作仍然相对有限，距离形成系统化、可广泛推广的解决方案尚有较大提升空间。

另一个亟待解决的问题是如何实现增量式视

频识别。增量式识别可以被视为缓解数据过时问题的一条重要技术路径，其核心思想是在系统运行过程中持续自动采集新增流量样本，并对加密视频流量识别模型进行在线或准在线的迭代更新，从而维持模型对最新网络状态与内容演化的适配能力。这一过程涉及若干尚未充分解决的关键环节，包括：如何高效、低干扰地持续采集新数据，如何合理设计旧数据的淘汰策略与更新周期，以及如何在保证识别性能稳定的前提下实现模型的动态迭代。对于基于特征工程的匹配识别方法，尤其是基于明文指纹的匹配识别方法而言，在传输协议与传输模式保持稳定的前提下，仅需通过自动化爬虫周期性获取最新的视频明文指纹并更新指纹库，即可在较小代价下支持视频集合的增量扩展。然而，目前关于指纹库在线更新机制与增量维护策略的系统性研究仍较为匮乏，有关更新频率、版本控制与一致性保障等问题尚未形成成熟方法体系。相比之下，基于机器学习的分类识别方法在执行增量式更新时通常需要为新增视频构建训练样本并对模型进行重新训练或微调，导致更新开销较大、在线部署困难。

4.1.4 隐私增强场景中识别研究不足

当前，加密视频流量识别研究多集中于常规网络环境，尽管已有部分工作[54],[56],[67],[83],[86]-[89],[94],[95]对隐私增强场景进行了探索，但整体仍处于起步阶段，存在明显不足。其主要问题体现在两个方面：一是缺乏覆盖多类隐私增强场景的大规模数据集，二是尚未形成系统化的研究方法。

在数据集层面，现有研究多聚焦于单一隐私增强协议场景，如 Tor 或 Vmess，数据覆盖范围有限。这种局限性导致研究成果在多种隐私增强场景下的泛化能力不足，难以全面验证不同协议条件下的识别性能。

在方法层面，现有研究大多以常规网络环境为主，将隐私增强场景作为验证方法适用性的补充实验，而非系统研究对象。因此，尚未形成面向隐私增强环境的完整技术框架和方法体系。这种缺口制约了加密视频流量识别在复杂隐私增强环境下的进一步发展与应用。

4.2 未来研究方向

4.2.1 发展加密视频流量识别理论研究体系

针对加密视频流量识别领域中理论研究体系薄弱的问题，为了使该领域的研究更加具备指导性和实用性，未来的研究可以在以下几个方向展开：首先，应强化特征理论的研究。这包括特征生成和扩展方法的探索，深入研究流量特征与应用层数据特征之间的关联性，分析特征的独立性，并排除冗余特征。此外，还需要对特征之间可能的相互干扰进行系统分析。其次，跨领域的研究也应得到加强，借鉴其他领域的相关研究成果。从加密视频流量识别模型本身出发，深入探讨各类分类模型与匹配模型对不同特征的敏感性，尝试揭示特征与模型之间固定的匹配关系及其原因。例如，分类模型对统计特征更为敏感，而匹配模型则在处理字节特征时表现得更为有效。这些研究将有助于提升加密视频流量识别方法的理论深度和应用价值。

4.2.2 弱化加密视频流量识别研究假设

弱化加密视频流量识别研究假设是推动该技术实用化的关键步骤。针对这一问题，研究可以从多个方向展开，整体思路是从简到难、从单一条件逐步过渡到多个条件的组合，以此逐渐接近实际应用场景。在场景假设方面，研究不仅应关注丢包、延迟、带宽限制、移动用户等实际因素，还应探索这些因素组合下的影响，例如，研究移动用户在丢包、延迟和带宽限制等不稳定网络条件下的视频流量识别问题。在模板假设方面，研究应着重考虑不同时间段内变化的视频流量数据，挑战传统假设的适用性。在识别者能力假设方面，探索非完全先验知识条件下的加密视频流量识别，尝试研究在先验知识不足的情况下如何保持高效的识别能力，这将有助于更加贴近真实的应用环境。

4.2.3 推动加密视频流量识别研究的应用部署

针对加密视频流量识别应用部署过程中面临的挑战，除了在弱化加密视频流量识别假设方面开展深入研究外，还应从迭代式识别模型的构建与通用性识别算法的设计两大方向同时进行攻关。具体而言，应尽可能多地探索并设计适用于不同场景的通用识别模型，以提升模型的适用范围和准确性。同时，研究应更加注重自适应的增量式迭代识别模型的构建，通过不断迭代和优化，确保识别模型能够在实际应用中应对动态变化的流量特征和网络环境。最后，下一阶段的研究应着重推动加密视频流量识别原型系统的实际落地应用，在实际部署中发现新的问题，并通过不断的调整和优化，逐步解决这些问题，以促进该技术的广泛应用。

4.2.4 强化隐私增强场景加密视频流量识别研究

鉴于当前隐私增强场景下的加密视频流量识别研究仍存在不足，未来需要在数据集构建与方法体系两个方面加以强化。具体而言，在数据集方面，

应面向多维度、多协议场景开展系统性采集，构建覆盖匿名网络、加密代理和隧道协议三大类隐私增强技术体系的数据资源，涵盖 Tor、Vmess、Shadowsocks、OpenVPN 等多种主流协议，以满足多样化实验与验证需求。在方法论层面，应超越以往零散式的适配验证，系统探索隐私增强场景下的通用化技术框架与方法体系，形成完整的研究范式，从而为不同隐私增强环境中的加密视频流量识别提供稳健且可迁移的技术支撑。

4.3 加密视频流量识别技术未来应用场景

加密视频流量识别面向的核心应用，是在缺乏视频服务提供商协作的条件下，由监管方部署于网络中间节点，对端到端加密传输的流量进行被动分析，从而从指纹库或训练集中定位并识别已知的公害视频，实现面向内容的网络监管。由于实际网络环境远比实验条件复杂，识别过程将不可避免地受到主动对抗、协议栈叠加、传输扰动与动态均衡等因素影响，导致流量特征稳定性下降、片段边界难以恢复或可观测信息不足。基于这一现实需求，面向复杂网络场景的加密视频流量识别将构成未来最具代表性的应用场景。

复杂网络场景可概括为四类典型形态。（1）协议嵌套场景，用户可能借助 VPN 或加密代理对原有加密视频流进行再次封装与转发，以改变可观测的传输行为并规避中间节点监管。（2）主动对抗场景，部分用户会采用以洋葱路由为代表的匿名通信网络来隐藏访问关系与行为轨迹，使得识别所依赖的统计模式被链路加密、填充与多跳转发进一步弱化。（3）网络性能波动场景，现实网络中的拥塞、带宽约束与接入切换会引发丢包、重传、乱序等现象，造成流量时序结构与统计分布显著漂移，增加特征重整与稳定表征的难度。（4）动态负载均衡场景，骨干网中的非对称路由、CDN 调度等机制可能导致路径与观测点频繁变化，甚至出现流量片段缺失或视角不完整，从而削弱识别所需的关键证据链。综合而言，上述四类场景覆盖了现实监管中最常见、最具挑战性的部署条件，也构成了未来加密视频流量识别技术面向应用落地时需要重点适配与系统验证的主要方向。

5 结束语

加密视频流量识别技术表明，即便在流量加密的场景下，仍有可能对公害视频实现高效、精准的监测，从而在保障用户隐私安全的同时，助力网络环境的健康有序发展。近年来，随着端到端加密技术的快速普及，加密视频流量识别领域迎来了蓬勃发展，研究人员已开展了较为系统的探索并积累了大量成果，为后续的理论创新和工程应用奠定了坚实基础。本文首次对加密视频流量识别领域进行了系统性的综述，从加密视频流量识别的基本概念出发，系统梳理了其定义、识别假设、威胁模型以及研究意义等核心内容。在此基础上，对该领域的研究进展进行了分类综述和比较分析。最后，针对当前面临的主要挑战，本文提出了四大研究难题，并对未来的发展方向进行了展望。

参 考 文 献

[1] Ni B, Peng H, Chen M, et al. Expanding language-image pretrained models for general video recognition//Proceedings of the European conference on computer vision. Tel Aviv, Israel, 2022: 1-18.

[2] Kumar K, Shrimankar D D, Singh N. Eratosthenes sieve based key-frame extraction technique for event summarization in videos. Multimedia Tools and Applications, 2018, 77: 7383-7404.

[3] Zhu X, Xiong Y, Dai J, et al. Deep feature flow for video recognition//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, USA. 2017: 2349-2358.

[4] Feichtenhofer C, Fan H, Malik J, et al. Slowfast networks for video recognition//Proceedings of the IEEE/CVF international conference on computer vision. Seoul, Republic of Korea. 2019: 6202-6211.

[5] PU Zhan-Xing, GE Yong-Xin. Few-Shot Action Recognition in Video Based on Multi-Feature Fusion. Chinese Journal of Computers, 2023, 46(3): 594-608.
(蒲瞻星, 葛永新. 基于多特征融合的小样本视频行为识别算法. 计算机学报, 2023, 46(3): 594-608.)

[6] Wu H, Luo H, Zhao SS, Liu ST, Cheng G, Hu XY. Encrypted Video Identification Method for HTTP/2 Traffic Multiplexing Features. Journal of Software, 2025,36(7):3375-3404
(吴桦, 罗浩, 赵士顺, 刘嵩涛, 程光, 胡晓艳. 面向HTTP/2 流量多路复用特征的加密视频识别方法. 软件学报, 2025, 36(7):3375-3404)

[7] Dingledine R, Mathewson N, Syverson P. Tor: the second-generation onion router//Proceedings of the 13th conference on USENIX Security Symposium. San Diego, USA, 2004.

[8] Xu Y, Wang L, Chen J, et al. Tor trace in images: a novel multi-tab website fingerprinting attack with object detection. Computer Journal, 2024, 67(8): 2690-2701.

[9] Yuan X, Li T, Li L, et al. HSWF: enhancing website fingerprinting attacks on tor to address real world distribution mismatch. Computer Networks, 2024, 241: 110217.

[10] Zhou Q, Wang L, Zhu H, et al. WF-transformer: learning temporal features for accurate anonymous traffic identification by using transformer networks. IEEE Transactions on Information Forensics and Security, 2023, 19: 30-43.

[11] Zou H, Wei Z, Su J, et al. Relation-CNN: enhancing website fingerprinting attack with relation features and NFS-CNN. Expert Systems with Applications, 2024, 247: 123236.

[12] Deng X, Yin Q, Liu Z, et al. Robust multi-tab website fingerprinting attacks in the wild//Proceedings of the IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2023: 1005-1022.

[13] Zhou Q, Wang L, Zhu H, et al. Few-shot website fingerprinting attack with cluster adaptation. Computer Networks, 2023, 229: 109780.

[14] Ha J, Roh H. QUIC website fingerprinting based on automated machine learning. ICT Express, 2024, 10(3): 594-599.

[15] Qian Y, Huang G, Ma C, et al. Enhancing resilience in website fingerprinting: novel adversary strategies for noisy traffic environments. IEEE Transactions on Information Forensics and Security, 2024, 19: 7216-7231.

[16] Mathews N, Holland J K, Hopper N, et al. Laserbeak: evolving website fingerprinting attacks with attention and multi-channel feature representation. IEEE Transactions on Information Forensics and Security, 2024, 19: 9285-9300.

[17] Deng X, Li Q, Xu K. Robust and reliable early-stage website fingerprinting attacks via spatial-temporal distribution analysis//Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. Salt Lake City, USA, 2024: 1997-2011.

[18] Bahramali A, Bozorgi A, Houmansadr A. Realistic website fingerprinting by augmenting network traces//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. Copenhagen, Denmark, 2023: 1035-1049.

[19] Xie L, Li Y, Yang H, et al. A multi-scene webpage fingerprinting method based on multi-head attention and data enhancement// Proceedings of the International Conference on Information Security and Cryptology (Inscrypt). Hangzhou, China, 2024: 418-432.

[20] Wang Z, Li T, Yin M, et al. WF3A: a N-shot website fingerprinting with effective fusion feature attention. Computers & Security, 2024, 140: 103796.

[21] Xie Y, Feng J, Huang W, et al. Contrastive fingerprinting: a novel website fingerprinting attack over few-shot traces//Proceedings of the ACM Web Conference 2024. Singapore, 2024: 1203-1214.

[22] Tan J, Wang H, Han S, et al. An adaptability-enhanced few-shot website fingerprinting attack based on collusion. IEEE Transactions on Information Forensics and Security, 2024, 19: 8220-8235.

[23] Shen M, Ji K, Gao Z, et al. Subverting website fingerprinting defenses with robust traffic representation//Proceedings of the 32nd USENIX Security Symposium. Anaheim, USA, 2023: 607-624.

[24] Attarian R, Keshavarz-Haddad A. Effective website fingerprinting attack based on the first packet direction only. Computer Networks, 2023, 231: 109811.

[25] Luo C, Tang W, Wang Q, et al. Few-shot website fingerprinting with distribution calibration. IEEE Transactions on Dependable and Secure Computing, 2024, 22(1): 632-648.

[26] Cherubin G, Jansen R, Troncoso C. Online website fingerprinting: evaluating website fingerprinting attacks on tor in the real world//Proceedings of the 31st USENIX Security Symposium. Boston, USA, 2022: 753-770.

[27] Xiao X, Zhou X, Yang Z, et al. A comprehensive analysis of website fingerprinting defenses on tor. Computers & Security, 2024, 136: 103577.

[28] Xu H, Cheng Z, Li S, et al. ProxyKiller: an anonymous proxy traffic attack model based on traffic behavior graphs//Proceedings of the European Symposium on Research in Computer Security (ESORICS). Bydgoszcz, Poland, 2024: 162-181.

[29] Xue D, Kallitsis M, Houmansadr A, et al. Fingerprinting obfuscated proxy traffic with encapsulated TLS handshakes//Proceedings of the 33rd USENIX Security Symposium. Philadelphia, USA, 2024: 2689-2706.

[30] Wu M, Sippe J, Sivakumar D, et al. How the great firewall of China detects and blocks fully encrypted traffic//Proceedings of the 32nd USENIX Security Symposium. Anaheim, USA, 2023: 2653-2670.

[31] Alice, Bob, Carol, et al. How china detects and blocks shadowsocks//Proceedings of the ACM Internet Measurement Conference. Virtual Event, USA, 2020: 111-124.

[32] Fenske E, Johnson A. Bytes to schlep? Use a FEP: hiding protocol metadata with fully encrypted protocols//Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. Salt Lake City, USA, 2024: 1982-1996.

[33] Gu Z, Liu C, Zhang X, et al. DecETT: accurate app fingerprinting under encrypted tunnels via dual decouple-based semantic enhancement//Proceedings of the ACM on Web Conference 2025. Sydney, Australia, 2025: 2413-2423.

[34] Ma X, Qu J, Shi M, et al. Website fingerprinting on encrypted proxies: a flow-context-aware approach and countermeasures. IEEE/ACM Transactions on Networking, 2024, 32(3): 1904-1919.

[35] Günther F, Stebila D, Veitch S. Obfuscated key exchange//Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. New York, USA, 2024: 2385-2399.

[36] Ramesh R, Winter P, Korman S, et al. CalcuLatency: leveraging cross-layer network latency measurements to detect proxy-enabled abuse//Proceedings of the 33rd USENIX Security Symposium. Philadelphia, USA, 2024: 2263-2280.

[37] Wang C, Yin J, Li Z, et al. Identifying VPN servers through graph-represented behaviors//Proceedings of the ACM Web Conference 2024. New York, USA, 2024: 1790-1799.

[38] Xue D, Stanley R, Kumar P, et al. The discriminative power of cross-layer RTTs in fingerprinting proxy traffic//Proceedings of the 32nd Annual Network and Distributed System Security Symposium (NDSS). San Diego, USA, 2025.

[39] Xue D, Ramesh R, Jain A, et al. OpenVPN is open to VPN fingerprinting. Communications of the ACM, 2025, 68(1): 79-87.

[40] Ramesh R, Evdokimov L, Xue D, et al. VPNalyzer: systematic investigation of the VPN ecosystem//Proceedings of the Network and Distributed System Security. San Diego, USA, 2022: 1-16.

[41] Liu M, Gou G, Xiong G, et al. Enhanced detection of obfuscated HTTPS tunnel traffic using heterogeneous information network. Computer Networks, 2025, 257: 110975.

[42] Fu C, Li Q, Shen M, et al. Detecting tunneled flooding traffic via deep semantic analysis of packet length patterns//Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security. Salt Lake City, USA, 2024: 3659-3673.

[43] Zillien S, Schmidbauer T, Kubek M, et al. Look what's there! Utilizing the internet's existing data for censorship circumvention with OPPRESSION//Proceedings of the 19th ACM Asia Conference on Computer and Communications Security. Singapore, Singapore, 2024: 80-95.

[44] Gu Z, Gou G, Hou C, et al. LFETT2021: a large-scale fine-grained encrypted tunnel traffic dataset//Proceedings of the IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Shenyang, China, 2021: 240-249.

[45] Ahn J, Hussain R, Kang K, et al. Exploring encryption algorithms and network protocols: a comprehensive survey of threats and vulnerabilities. IEEE Communications Surveys & Tutorials, 2025, 27(6): 3587-3614.

[46] Zhou Hong-Cheng, Su Jin-Shu, Wei Zi-Ling, et al. A Review of the Research of Website Fingerprinting Identification and Defense. Chinese Journal of Computers, 2022, 45(10): 2243-2278 (in Chinese).
(邹鸿程, 苏金树, 魏子令, 等. 网站指纹识别与防御研究综述. 计算机学报, 2022, 45(10): 2243-2278)

[47] Shen M, Ye K, Liu X, et al. Machine learning-powered encrypted network traffic analysis: a comprehensive survey. IEEE Communications Surveys and Tutorials, 2023, 25(1): 791-824.

[48] Sharma A, Lashkari A H. A survey on encrypted network traffic: a comprehensive survey of identification/classification techniques, challenges, and future directions. Computer Networks, 2025, 257: 110984.

[49] Feng Y, Li J, Mirkovic J, et al. Unmasking the internet: a survey of fine-grained network traffic analysis. IEEE Communications Surveys & Tutorials, 2025, 27(6): 3672-3709.

[50] Papadogiannaki E, Ioannidis S. A survey on encrypted network traffic analysis applications, techniques, and countermeasures. ACM Computing Surveys, 2021, 54(6): 123:1-123:35.

[51] Wu Hua, Ni Shan-Shan, Luo Hao, et al. An Encrypted Video Recognition Method Based on the Transmission Char-acteristics of HTTP/3. Chinese Journal of Computers, 2024, 47(7): 1640-1664.
(吴桦, 倪珊珊, 罗浩,等. 一种基于HTTP/3 传输特性的加密视频识别方法. 计算机学报, 2024, 47(7): 1640-1664)

[52] Schuster R, Shmatikov V, Tromer E. Beauty and the burst: remote identification of encrypted video streams//Proceedings of the 26th USENIX Security Symposium. Vancouver, Canada, 2017: 1357-1374.

[53] Saponas T S, Lester J, Hartung C, et al. Devices that tell on you: privacy trends in consumer ubiquitous computing//Proceedings of the USENIX Security Symposium. Boston, USA, 2007: 55-70.

[54] Gu J, Wang J, Yu Z, et al. Walls have ears: Traffic-based side-channel attack in video streaming//Proceedings of the IEEE International Conference on Computer Communications. Honolulu, USA, 2018: 1538-1546.

[55] Wu H, Yu Z, Cheng G, et al. Identification of encrypted video streaming based on differential fingerprints//Proceedings of the IEEE International Conference on Computer Communications Workshops (INFOCOM WKSHPS). Toronto, Canada, 2020: 74-79.

[56] Afandi W, Bukhari S M A H, Khan M U, et al. Fingerprinting technique for youtube videos identification in network traffic. IEEE Access, 2022, 10: 76731-76741.

[57] Yang L, Fu S, Luo Y, et al. Markov probability fingerprints: a method for identifying encrypted video traffic//Proceedings of the International Conference on Mobility, Sensing and Networking (MSN). Tokyo, Japan, 2020: 283-290.

[58] Yang L, Fu S, Luo Y, et al. A clustering method of encrypted video traffic based on levenshtein distance//Proceedings of the International Conference on Mobility, Sensing and Networking (MSN). Exeter, United Kingdom, 2021: 1-8.

[59] Huang R, Wu H, Wang X, et al. The secret behind instant messaging: video identification attack against complex protocols. Cybersecurity, 2025, 8(1): 13.

[60] Xu M, Du M, Li Z, et al. Rtd: the road to encrypted video traffic identification in the heterogeneous network environments//Proceedings of the IEEE International Conference on High Performance Computing and Communications (HPCC). Hainan, China, 2022: 1697-1704.

[61] Tang W, Du M, Li Z, et al. Shrink: identification of encrypted video traffic based on QUIC//Proceedings of the IEEE International Performance, Computing, and Communications Conference (IPCCC). Anaheim, USA, 2023: 140-149.

[62] Du M, Xu M, Liu K, et al. Long-short terms frequency: a method for encrypted video streaming identification//Proceedings of the International Conference on Computer Supported Cooperative Work

in Design (CSCWD). Rio de Janeiro, Brazil, 2023: 1581-1588.

[63] Reed A, Klimkowski B. Leaky streams: identifying variable bitrate DASH videos streamed over encrypted 802.11n connections//Proceedings of the IEEE Annual Consumer Communications & Networking Conference (CCNC). Las Vegas, USA, 2016: 1107-1112.

[64] Reed A, Kranch M. Identifying HTTPS-protected netflix videos in real-time//Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy. Scottsdale, USA, 2017: 361-368.

[65] Xu S, Sen S, Mao Z M. CSI: inferring mobile ABR video adaptation behavior under HTTPS and QUIC//Proceedings of the Fifteenth European Conference on Computer Systems. Heraklion, Greece, 2020: 1-16.

[66] Zhang X, Xiong G, Li Z, et al. Traffic spills the beans: a robust video identification attack against YouTube. Computers & Security, 2024, 137:103623.

[67] Tang W, Li J, Du M, et al. Zenith: real-time identification of DASH encrypted video traffic with distortion//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia, 2024: 7200-7209.

[68] Wu H, Li X, Cheng G, et al. Monitoring video resolution of adaptive encrypted video traffic based on HTTP/2 features//Proceedings of the IEEE International Conference on Computer Communications Workshops (INFOCOM WKSHPS). Vancouver, Canada, 2021: 1-6.

[69] Wu H, Li X, Wang G, et al. Resolution identification of encrypted video streaming based on HTTP/2 features. ACM Transactions on Multimedia Computing Communications and Applications, 2023, 19(2): 1-23.

[70] Wu Hua, Yu Zhen-Hua, Cheng Guang, et al. Encrypted Video Recognition in Large-scale Fingerprint Database. Journal of Software, 2021, 32(10): 3310-3330(in Chinese).
(吴桦, 于振华, 程光等. 大型指纹库场景中加密视频识别方法. 软件学报, 2021, 32(10): 3310-3330)

[71] Hu N, Wu H, Zhao H, et al. Breaking through the diversity: encrypted video identification attack based on QUIC features//Proceedings of the European Symposium on Research in Computer Security (ESORICS). Bydgoszcz, Poland, 2024: 166-186.

[72] Zhao Y, Wu H, Chen L, et al. Identifying video resolution from encrypted QUIC streams in segment-combined transmission scenarios//Proceedings of the 34th Workshop on Network and Operating System Support for Digital Audio and Video. Bari, Italy, 2024: 50-56.

[73] Quan J, Du J, Wu H, et al. Efficient short video identification attack for scenarios with hybrid transmission modes and preloading mechanism//Proceedings of the 20th International Conference on Mobility, Sensing and Networking (MSN). Harbin, China, 2024: 504-511.

[74] Zhao H, Wu H, Bian X, et al. Unveiling the unseen: video recognition attacks on social software//Proceedings of the 29th Australasian Conference on Information Security and Privacy (ACISP). Sydney, Australia, 2024: 412-432.

[75] Zhu W, Wu H, Quan J, et al. Accurate identification of encrypted videos in asymmetric routing scenarios//Proceedings of the 24st Asia-pacific Network Operations and Management Symposium (APNOMS). Sejong, Korea, 2023: 310-313.

[76] Dubin R, Dvir A, Pele O, et al. I know what you saw last minute-encrypted http adaptive video streaming title classification. IEEE Transactions on Information Forensics and Security, 2017, 12(12): 3039-3049.

[77] Dubin R, Hadar O, Dvir A, et al. Video quality representation classification of encrypted http adaptive video streaming. KSII Transactions on Internet and Information Systems (TIIS), 2018, 12(8): 3804-3819.

[78] Dvir A, Marnerides A K, Dubin R, et al. Encrypted video traffic clustering demystified. Computers & Security, 2020, 96: 101917.

[79] Bae S, Son M, Kim D, et al. Watching the watchers: practical video identification attack in LTE networks//Proceedings of the 31st USENIX Security Symposium. Boston, USA, 2022: 1307-1324.

[80] Khan M U, Bukhari S M, Khan S A, et al. ISP can identify YouTube videos that you just watched//Proceedings of the International Conference on Frontiers of Information Technology (FIT). Islamabad, Pakistan, 2021: 1-6.

[81] Yang L, Wang Y, Fu S, et al. Evs2vec: a low-dimensional embedding method for encrypted video stream analysis//Proceedings of the 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). Madrid, Spain, 2023: 537-545.

[82] Yang L, Fu S, Wang Y, et al. The analysis of encrypted video stream based on low-dimensional embedding method. IEEE Transactions on Information Forensics and Security, 2025, 20: 8280-8295.

[83] Xu Z, Ren X, Zhang Y, et al. Peering through the veil: a segment-based approach for VPN encapsulated video title identification//Proceedings of the IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Sanya, China, 2024: 1500-1505.

[84] Li Y, Huang Y, Xu R, et al. Deep content: unveiling video streaming content from encrypted wifi traffic//Proceedings of the IEEE 17th International Symposium on Network Computing and Applications (NCA). Cambridge, USA, 2018: 1-8.

[85] Li Y, Huang Y, Seneviratne S, et al. From traffic classes to content: a hierarchical approach for encrypted traffic classification. Computer Networks, 2022, 212: 109017.

[86] Rahman M S, Matthews N, Wright M. Poster: video fingerprinting in tor//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York, USA, 2019:

2629-2631.

[87] Walsh T, Thomas T, Barton A. Exploring the capabilities and limitations of video stream fingerprinting//Proceedings of the IEEE Security and Privacy Workshops (SPW). San Francisco, USA, 2024: 28-39.

[88] Liu Y, Li S, Zhang C, et al. ITP-KNN: encrypted video flow identification based on the intermittent traffic pattern of video and K-nearest neighbors classification//Proceedings of the International Conference on Computational Science (ICCS). Amsterdam, Netherlands, 2020: 279-293.

[89] Kattadige C, Choi K N, Wijesinghe A, et al. SETA++: real-time scalable encrypted traffic analytics in multi-gbps networks. IEEE Transactions on Network and Service Management, 2021, 18(3): 3244-3259.

[90] Amjad F, Khan F, Tahir S, et al. ENCVIDC: an innovative approach for encoded video content classification. Neural Computing and Applications, 2022, 34(21): 18685-18702.

[91] Liu Y, Ou C, Li Z, et al. Wavelet-based traffic analysis for identifying video streams over broadband networks//Proceedings of the IEEE Conference on Global Communications (GLOBECOM). New Orleans, USA, 2008: 1-6.

[92] Liu Y, Sadeghi A-R, Ghosal D, et al. Video streaming forensic-content identification with traffic snooping//Proceedings of the 13th Information Security. Boca Raton, USA, 2011: 129-135.

[93] Dahanayaka T, Jourjon G, Seneviratne S. Understanding traffic fingerprinting CNNs//Proceedings of the IEEE 45th Conference on Local Computer Networks (LCN). Sydney, Australia, 2020: 65-76.

[94] Zhang Y, Xu Z, Ren X, et al. SCEP-TI: a side-channel attack on encrypted proxy video streams for video title identification. Computer Networks, 2025, 271: 111630.

[95] Lu J, Zhou Z, Wu H, et al. TorVIA: a novel encrypted video identification method based on tor transmission characteristics. Computer Networks, 2025, 271: 111591.

[96] Björklund M, Duvignau R. Endangered privacy: large-scale monitoring of video streaming services//Proceedings of the 34th USENIX Security Symposium (USENIX Security 25). Seattle, USA, 2025: 6581-6597.

[97] Zhang X, Xiong G, Gou G, et al. Dive into streaming: efficient identification of encrypted dynamic DASH video traffic. Science China Information Sciences, 2025, 69(1): 112304.

[98] Karunanayake I, Jiang J, Ahmed N, et al. Exploring uncharted waters of website fingerprinting. IEEE Transactions on Information Forensics and Security, 2023, 19: 1840-1854.

[99] Gong J, Zhang W, Zhang C, et al. Wfdefproxy: real world implementation and evaluation of website fingerprinting defenses. IEEE Transactions on Information Forensics and Security, 2023, 19: 1357-1371.

[100] Mathews N, Holland J K, Oh S E, et al. Sok: a critical evaluation of efficient website fingerprinting defenses//Proceedings of the IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2023: 969-986.

[101] Shen M, Ji K, Wu J, et al. Real-time website fingerprinting defense via traffic cluster anonymization//Proceedings of the IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2024: 3238-3256.

[102] Yin Q, Liu Z, Li Q, et al. An automated multi-tab website fingerprinting attack. IEEE Transactions on Dependable and Secure Computing, 2021, 19(6): 3656-3670.

[103] Gong J, Zhang W, Zhang C, et al. Surakav: generating realistic traces for a strong website fingerprinting defense//Proceedings of the IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2022: 1558-1573.

[104] Gulmezoglu B. XAI-based microarchitectural side-channel analysis for website fingerprinting attacks and defenses. IEEE Transactions on Dependable and Secure Computing, 2021, 19(6): 4039-4051.

[105] Jiang M, Cui B, Fu J, et al. RUDOLF: an efficient and adaptive defense approach against website fingerprinting attacks based on soft actor-critic algorithm. IEEE Transactions on Information Forensics and Security, 2024, 19: 7794-7809.

[106] Zhao X, Deng X, Li Q, et al. Towards fine-grained webpage fingerprinting at scale//Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security. Salt Lake City, USA, 2024: 423-436.

ZHAO Hang-Yu, Ph.D., candidate. His research interests include encrypted traffic analysis, encrypted video recognition.

WU Hua, Ph.D., associate professor. Her research interests include encrypted traffic analysis, cyberspace security, and network situational awareness.

CHENG Guo, M.S., candidate. His research interests include encrypted traffic analysis, encrypted video recognition.

YIN Xiao-Yue, M.S., candidate. Her research interests include encrypted traffic analysis, encrypted video recognition.

LIU Song-Tao, Ph.D., candidate. His research interests include encrypted traffic analysis, fine-grained webpage fingerprint.

CHENG Guang, Ph.D., professor. His research interests

include cyberspace security monitoring and protection, network traffic big data analysis, botnet and APT attack detection, etc.

HU Xiao-Yan, Ph.D., associate professor. Her research interests include encrypted traffic analysis, cyberspace security, future network architecture, etc.