IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 June 2025, accepted 21 July 2025, date of publication 23 July 2025, date of current version 31 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3592152

SURVEY

VPN Traffic Analysis: A Survey on Detection and Application Identification

YASAMEEN SAJID RAZOOQI[1] AND ADRIAN PEKAR[1,2,3], (Member, IEEE)

[1]Department of Networked Systems and Services, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, 1111 Budapest, Hungary
[2]HUN-REN-BME Information Systems Research Group, 1117 Budapest, Hungary
[3]CUJO LLC, 1082 Budapest, Hungary

Corresponding author: Yasameen Sajid Razooqi (rsajid@hit.bme.hu)

This work was supported in part by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences; in part by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, through the 2024-1.2.6-EUREKA Funding Scheme under Project 2024-1.2.6-EUREKA-2024-00009 and Project 2024-1.2.6-EUREKA-2024-00009; and in part by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, through the TKP2021-NVA Funding Scheme under Project TKP2021-NVA-02 and Project TKP2021-NVA-02.

ABSTRACT Network traffic analysis is fundamental for cybersecurity, network management, and policy enforcement. The widespread adoption of encryption, particularly through Virtual Private Networks (VPNs), presents a significant challenge by obscuring traditional visibility methods. While VPNs enhance user privacy and security, they also create blind spots for network operators, potentially concealing malicious activities or hindering performance management. Analyzing the characteristics of traffic flowing through encrypted VPN tunnels, without decryption, has become a critical yet difficult task. This survey provides a comprehensive review of the state-of-the-art in VPN traffic analysis research published over the past decade (2016-2025). We specifically focus on three key tasks: detecting the presence of VPN traffic, identifying the specific VPN protocol or service used, and classifying the application traffic encapsulated within VPN tunnels. Based on a systematic review of the literature, we provide an in-depth analysis of the features, methodologies (including traditional and learning-based approaches), and datasets employed in recent studies. We synthesize reported performance results, analyze trends in feature and methodology evolution, and highlight the prevalent use and limitations of benchmark datasets. The survey identifies key technical challenges, discusses the implications of VPN traffic analysis for network security and Quality of Service (QoS), and proposes promising future research directions. This work serves as a vital resource for researchers and practitioners navigating the complexities of analyzing encrypted VPN traffic in modern networks.

INDEX TERMS Application identification, deep learning, encrypted traffic classification, machine learning, network security, survey, VPN detection, VPN traffic analysis.

I. INTRODUCTION

The modern digital landscape is increasingly reliant on network traffic analysis for critical functions spanning network security, Quality of Service (QoS) management, resource planning, and policy enforcement. Traditionally, techniques like Deep Packet Inspection (DPI) provided granular visibility into traffic content and application usage by examining payload data. However, the widespread adoption of encryption, primarily driven by rising privacy concerns and security requirements, has fundamentally altered this paradigm. Protocols like Transport Layer Security (TLS) 1.3, DNS over HTTPS / DNS over TLS (DoH/DoT), and Quick UDP Internet Connections (QUIC) obscure the content layer, rendering traditional DPI ineffective and creating significant blind spots for network operators and security analysts [1], [2].

Virtual Private Networks (VPNs) represent a particularly significant manifestation of this challenge. VPNs are designed to create encrypted tunnels, encapsulating original

The associate editor coordinating the review of this manuscript and approving it for publication was Jose Saldana.

network traffic to provide privacy, anonymity, bypass censorship, or enable secure remote access. While invaluable for users, this encapsulation and encryption process effectively hides the nature of the underlying communication from external observers. Analyzing traffic flowing through a VPN tunnel *without* decrypting it has become a pressing need, yet it is fraught with technical difficulties. This challenge necessitates the development and refinement of sophisticated traffic analysis techniques that can operate on the observable characteristics of encrypted flow patterns rather than their content.

The inability to analyze VPN traffic poses several critical problems. From a security perspective, malicious activities, covert channels, or command-and-control traffic can be effectively masked within a VPN tunnel, bypassing traditional signature-based detection systems. For network operators, the lack of visibility hinders accurate traffic classification, impacting QoS provisioning, load balancing, and capacity planning. Furthermore, in certain regulated environments, understanding VPN usage and the applications within them may be necessary for policy compliance or incident response. Therefore, developing robust capabilities to analyze VPN traffic, including detection, identification, and application classification, is paramount in today's encrypted network environment.

A. CATEGORICAL REVIEW OF EXISTING SURVEYS

To contextualize the contribution of this survey, we first provide a brief review of existing survey literature related to VPNs and encrypted traffic analysis published within the past decades. Based on their primary focus, we categorize these works into two main themes: (1) surveys primarily addressing VPN technology and deployment aspects, and (2) surveys focused on general encrypted traffic classification, where VPNs might be included but are not the central subject of analysis.

1) VPN TECHNOLOGY AND DEPLOYMENT

These surveys address VPN systems as infrastructure components, focusing on aspects such as protocol design, deployment models, mobility handling, compliance, and usage trends. Surveys such as [3], [4], [5], and [6] examine architectural aspects of VPNs, including protocol configurations, mobility support, compliance auditing, and deployment models across wired, wireless, and cloud environments. However, these works do not explore VPN traffic classification or detection tasks such as tunnel identification, protocol inference, or encrypted application analysis. Expanding on deployment diversity, [4] introduces a structured taxonomy of VPN designs, covering multiple protocol types and presenting comparative insights drawn from deployments across industry sectors such as healthcare, utilities, and cloud infrastructure. Security and policy enforcement receive detailed attention in [5], which proposes a compliance assessment framework encompassing encryption strength, tunnel setup, logging, and vulnerability risks like DNS or IPv6 leaks. A mobile-specific perspective is explored by [7], who reviews tunnel establishment modes, OSI-layer support, and commercial VPN solutions under handoff conditions. Lastly, [6] presents a broad survey of VPN use trends, provisioning types (*e.g.*, remote access, site-to-site), platform diversity (hardware vs. software), and the evolution of VPN functionality in modern networking contexts, particularly within 5G and cloud-native architectures.

2) ENCRYPTED TRAFFIC CLASSIFICATION INCLUDING VPN

These surveys examine encrypted traffic classification using Machine Learning (ML) and Deep Learning (DL) techniques, often addressing VPNs alongside other encrypted protocols such as TLS, QUIC, and Tor. In most cases, VPN traffic is one component of a broader dataset rather than a central focus. Surveys like [8], [9], and [10] focus on encrypted traffic classification techniques, often in the context of protocols such as TLS, QUIC, or Tor. While VPN traffic is sometimes included in datasets or model evaluations, it is rarely treated as a standalone classification task. These surveys tend to address broader modeling strategies (*e.g.*, deep learning, pre-training, graph-based inference) without isolating VPN-specific performance or application-level differentiation. Even in earlier surveys like [11] and more comprehensive reviews such as [12], VPNs are either grouped under general encrypted protocols or referenced in the context of network security use cases. These contributions do not provide task-specific evaluations of VPN classification accuracy, nor do they offer detailed comparisons of feature types, model architectures, or datasets used for VPN-related analysis. Specifically, Yang et al. [8] reviews deep learning techniques designed for unidirectional encrypted traffic, including Convolutional Neural Network (CNNs), Recurrent Neural Network (RNNs), and Graph Neural Networks (GNNs). While the analysis introduces a novel taxonomy based on directional feature handling, VPN traffic appears only briefly. A more tunnel-centric approach is taken by Sui et al. [9], who investigates methods for detecting VPNs and similar traffic obfuscation mechanisms across multiple protocol layers, using both traditional fingerprinting and learning-based techniques. Rezaei et al. [13] structures encrypted traffic classification around a deep learning pipeline, discussing model architectures and feature types with minimal attention to VPNs. Similarly, Sharma et al. [14] offers a comprehensive taxonomy of ML and DL techniques for encrypted traffic analysis, including advanced models like GANs and transformers, but considers VPNs only as one of several application domains. Historical contributions such as [11] provide early insights into feature-based classification, noting VPN traffic in the context of IPsec but not emphasizing it in evaluation. More recent efforts like [12] extend encrypted traffic analysis to tasks such as intrusion detection, app usage inference, and privacy leakage, with VPN detection appearing within multi-purpose

TABLE 1. Survey organization and content overview.

Section	Topic Area	Focus	Key Topics & Contributions
Section I	Introduction	Context	Problem definition, literature gap analysis, three core VPN tasks (detection, identification, application classification)
Section II	Analysis	Trends	Dataset usage patterns, temporal evolution of methods (2016-2025), performance comparison across 30 studies
Section III	Implications	Applications	Network security & policy enforcement, QoS management, privacy & ethical considerations
Section IV	Methods	Techniques	7 class. approaches: Traditional ML, ensemble learning, deep learning, graph NNs, language models, semi-supervised, heuristic
Section V	Challenges	Limitations	Technical challenges (encryption, obfuscation), dataset limitations (ISCXVPN2016 issues), benchmarking problems
Section VI	Future	Directions	Advanced methodologies (transformers, self-supervised), novel features, real-world deployment strategies
Section VII	Conclusion	Summary	Key findings synthesis, research contributions, recommendations for future work

frameworks. Shen et al. [15] presents a modular breakdown of encrypted traffic analysis goals, including anomaly and usage detection, where VPNs are discussed in the context of larger behavioral classification. A similar holistic structure is found in [16], which introduces a pipeline that includes traffic representation, modeling, and performance evaluation, but treats VPN as part of a generalized encrypted traffic group. Notably, Dong et al. [10] provides the only survey to explicitly highlight pre-training techniques—such as BERT-style transformer models—for traffic classification that includes VPNs and Tor. Still, the treatment of VPNs remains secondary to the broader focus on model generalization and pre-training architectures.

B. GAP AND CONTRIBUTION

Based on the review of existing literature, there is a clear need for a comprehensive, up-to-date survey specifically dedicated to the analysis of VPN traffic. Existing surveys either focus on VPN infrastructure or treat VPN traffic analysis as a secondary component within broader encrypted traffic classification discussions. None provide a focused, analytical review of the techniques, inputs, and performance specifically for the critical tasks of:

1) *VPN Presence Detection:* Distinguishing VPN traffic from other encrypted or non-encrypted traffic.
2) *Specific VPN Identification:* Identifying the particular VPN protocol (*e.g.*, OpenVPN, WireGuard) or commercial service (*e.g.*, Tinc, ExpressVPN) being used.
3) *Application Identification within VPNs:* Determining the original application (*e.g.*, Youtube, WhatsApp, Discord, etc.) or application category (*e.g.*, web browsing, video streaming, file transfer) whose traffic is encapsulated within the VPN tunnel.

This survey aims to fill this gap by providing a systematic and analytical review of research conducted over the past decade specifically addressing these three core VPN traffic analysis tasks. Beyond merely listing research papers, our key contributions are:

- A comprehensive and current review of the state-of-the-art in VPN presence detection, specific VPN identification, and application identification within VPN tunnels.
- A detailed analysis and comparison of the input features, methodologies (including traditional, machine learning, and deep learning approaches), and datasets employed in the surveyed literature.
- A synthesis of reported performance metrics and accuracies, providing insights into the effectiveness and limitations of different techniques.
- Identification of key trends, prevailing challenges, and promising future research directions specific to VPN traffic analysis in the context of evolving VPN technologies and analysis techniques.

By providing this focused and analytical perspective, this survey serves as a valuable resource for researchers and practitioners seeking to understand the current capabilities, limitations, and potential advancements in gaining visibility into encrypted VPN traffic.

C. SURVEY METHODOLOGY

This survey is based on a systematic review of peer-reviewed literature published between 2016 and 2025. We conducted extensive searches across major academic databases, including IEEE Xplore, Elsevier, ACM Digital Library, SpringerLink, Wiley Online Library, and Google Scholar, using relevant keywords such as "VPN traffic analysis," "VPN detection," "VPN identification," "encrypted traffic classification VPN," "application identification VPN," "machine learning VPN traffic," and combinations thereof. Papers were selected based on their direct relevance to the three core tasks defined in Section I-B and their publication date within the specified timeframe. The selected papers were then analyzed, categorized, and synthesized to extract key information regarding their objectives, methodologies, features, datasets, and reported performance.

D. SURVEY ORGANIZATION AND STRUCTURE

To provide readers with a clear understanding of how this survey is organized, Table 1 presents an overview of the paper's structure and key content areas.

The survey structure follows a deliberate logical progression: the comprehensive analysis of existing literature, trends and applications (Sections II and III), provides the foundation for evaluating current classification techniques (Section IV), which in turn reveals key challenges and limitations (Section V), ultimately guiding the identification of promising future research directions (Section VI).

E. SURVEYED LITERATURE AT A GLANCE

To provide a quick, visual overview of the research landscape covered in this survey, Fig. 1 presents a hierarchical mind map

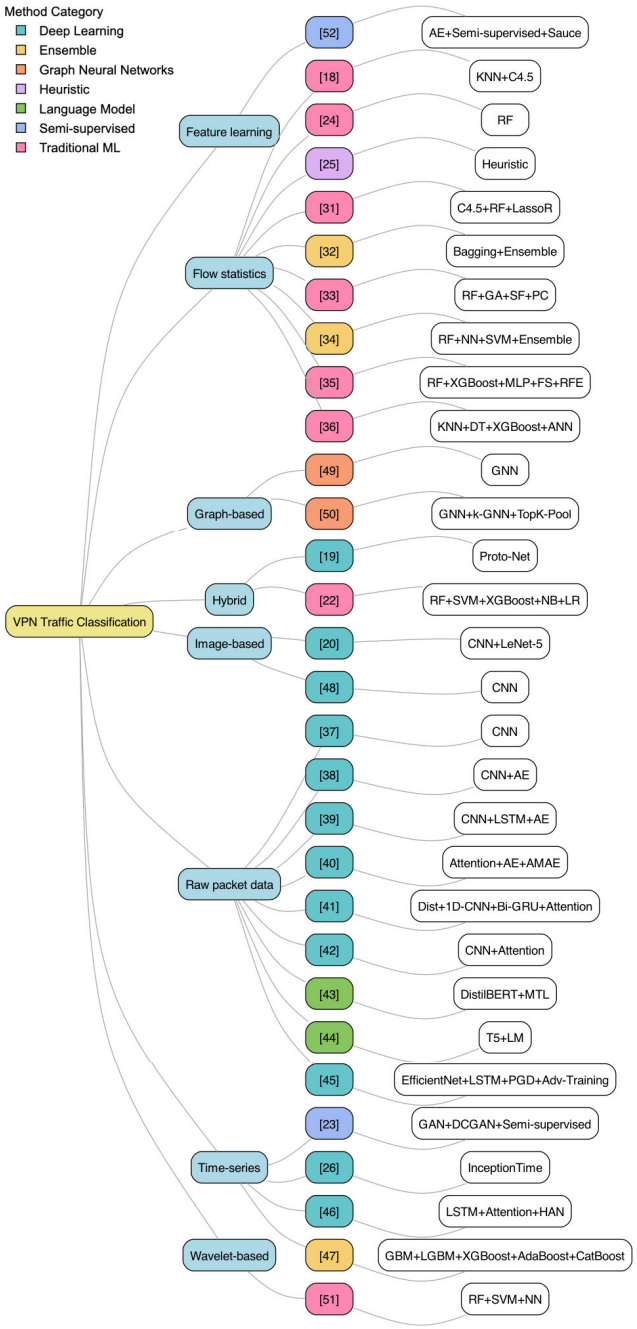


FIGURE 1. Hierarchical mind map summarizing surveyed papers on VPN traffic classification. The structure flows from the central topic to feature categories, then to individual papers (represented by their reference ID), and finally to the specific classification method employed. The color of the paper ID node corresponds to the general method category provided in the legend.

summarizing the VPN traffic analysis techniques discussed in the literature. This visualization organizes the surveyed papers based on the primary feature category employed, further detailing the specific method used and indicating its broader methodological category through color-coding.

The mind map is structured hierarchically, originating from the central theme of ''VPN Traffic Classification''.

- The first level branches out to the primary *Feature Categories* identified across the surveyed literature (*e.g.*, Flow statistics, Raw packet data, Time-series).
- Each feature category then connects to the individual *Papers* (represented by their reference ID, *e.g.*, ' [6]', ' [7]') that utilized features from that category. The color of these ID nodes signifies the *Method Category* (*e.g.*, Traditional ML, Deep Learning, Ensemble) employed in that paper, according to the legend in the figure.
- Finally, each paper ID node links to the specific *Method* (*e.g.*, 'KNN+C4.5', 'CNN', 'LSTM+Attention+HAN') reported in that study.

This visualization serves as a quick reference and structural summary of the works that will be discussed in detail in Section IV. It allows readers to rapidly identify papers associated with specific feature types, observe the methodological categorization (via color) associated with those papers, and see the precise algorithm or technique employed. While the specific placement generated by the layout algorithm primarily aims for readability and reduced edge crossing rather than deep semantic clustering, the map effectively illustrates the breadth of approaches within each feature domain and the diversity of methods applied across the field. It visually complements the subsequent analysis presented in Section II by mapping the individual studies onto a structured overview based on key technical characteristics before their detailed examination.

II. ANALYSIS

This section analyzes the surveyed literature across several key dimensions, including dataset usage, feature engineering trends, employed methodologies, and performance evolution, to identify common practices, challenges, and potential future directions in VPN traffic analysis.

A. DATASET USAGE PATTERNS

The choice of datasets significantly influences the evaluation, comparability, and generalizability of research in VPN traffic analysis. Fig. 2 presents a visualization of dataset usage across the surveyed papers, revealing critical trends and potential biases within the field. The percentages shown represent the proportion of dataset mentions across all surveyed papers, acknowledging that some studies utilize multiple datasets.

The analysis highlights several key points regarding dataset selection and its implications:

1) ISCXVPN2016 DOMINANCE AND LIMITATIONS

The *ISCXVPN2016* dataset, introduced by the University of New Brunswick's Information Security Centre of Excellence (ISCX) [17], [18], is one of the most widely used testbeds for VPN traffic classification. It includes 14 traffic classes, covering seven application types (Browsing, Email, Chat, Streaming, File Transfer, VoIP, P2P), each recorded under both unencrypted and VPN-encrypted conditions. The dataset

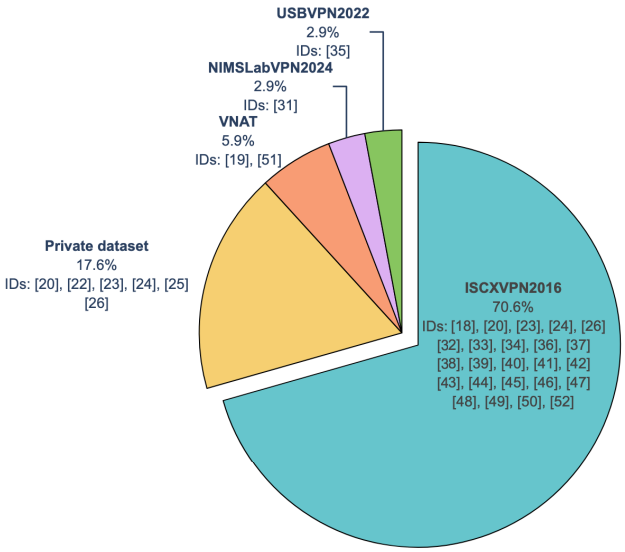


FIGURE 2. Distribution of datasets used in VPN traffic analysis research. Each slice represents the percentage of total dataset mentions across all papers. Paper reference IDs associated with each dataset mention are listed.

was created using real applications such as Skype for Voice over IP (VoIP), BitTorrent for P2P, and YouTube for streaming, with users "Alice" and "Bob" generating traffic both with and without a commercial OpenVPN (UDP mode) connection. The dataset consists of Packet capture files (pcap) (raw packet captures) and CSV files (flow-based features), totaling 28 GB of data.

The ISCXVPN2016 dataset is overwhelmingly the most prevalent benchmark, used by 24 out of 30 unique papers in this survey. This accounts for 70.6% of all dataset mentions across the papers (the 30 unique papers we survey utilize a total number of 34 datasets). Its widespread adoption has established a valuable common evaluation framework, enabling more direct performance comparisons between different classification techniques. However, while fostering comparability, this heavy reliance raises concerns about potential overfitting to its specific traffic patterns and the dataset's age relative to the rapid evolution of VPNs.

Despite its widespread use, researchers have identified several inconsistencies in the ISCXVPN2016 dataset that raise concerns about its suitability for benchmarking VPN detection models. Jorgensen et al. [19] highlighted that the dataset contains unencrypted payload within traffic labeled as VPN. For instance, in the ICQ Chat VPN capture, plaintext chat messages were found, suggesting that either the network tap was placed before encryption or certain packets were not encrypted at all. Additionally, some VPN-labeled captures contained multiple concurrent connections rather than a single VPN tunnel, further affecting data consistency. Similarly, Shapira et al. [20] noted that chat traffic lacks diversity, prompting them to supplement the dataset with additional real-world traffic samples. They also identified categorization inconsistencies, where background traffic was not properly filtered, potentially impacting classification accuracy. Furthermore, the dataset was captured in a controlled lab environment, rather than a live ISP network, limiting its generalizability. The exclusive use of OpenVPN in UDP mode may bias classification models toward recognizing OpenVPN-specific patterns rather than detecting VPNs in a broader context.

These discrepancies have serious implications, methods leveraging payload inspection (*e.g.*, DPI, certain deep packet approaches [21]) may gain an unfair advantage from unencrypted content, while methods relying on flow-level features may be skewed by the presence of multiple connections within a single capture file. Consequently, performance results reported on ISCXVPN2016, especially for encrypted traffic scenarios, must be interpreted with significant caution. Researchers using this dataset should avoid assuming payload encryption and consider post-processing flow data or replaying traffic through a VPN to ensure evaluation validity for tunnel-based analysis.

2) USE OF PRIVATE DATASETS

Representing 17.6% of mentions (used by 6 papers, *e.g.*, [20], [22], [23], [24], [25], [26]), private datasets allow researchers to investigate specific scenarios or newer VPNs. However, their inherent inaccessibility severely hinders reproducibility and cross-study comparison, representing a significant barrier to validating and building upon prior work.

3) EMERGENCE OF NEWER PUBLIC DATASETS

To address the need for more diverse and up-to-date data, several newer datasets have been introduced.

MIT Lincoln Laboratory released the *VPN Non-VPN Network Application Traffic (VNAT)* dataset in 2022 [19], [27]. This dataset consists of traffic from 10 popular applications (Netflix, YouTube, Vimeo for streaming; Zoiper VoIP; Skype chat; SSH/RDP for remote access; SFTP, rsync, SCP for file transfer, etc.), each captured in two conditions: raw (non-VPN) and encrypted through OpenVPN. The collection covers about 33,700 flows (272 hours of traffic) across five broad categories (Streaming, VoIP, Chat, Command & Control, File Transfer). Importantly, MIT's dataset provides not just raw pcaps but also prepared feature sets, including wavelet-based features and TLS metadata features, shared in ready-to-use data frames. It is fully public and aimed at improving reproducible research, which was previously hindered by data scarcity. The MIT dataset is more recent and includes some mobile and web-based apps, reflecting modern traffic mixes. The presence of both encrypted and unencrypted versions of the same traffic allows testing application classifiers in ideal vs encrypted scenarios.

However, the dataset was generated in a synthetic lab environment, which limits its realism. It includes only a single VPN protocol (OpenVPN), failing to represent the broader protocol diversity, and lacks background noise typically present in real-world settings.

The *USBVPN2022* dataset [28], [29] contains network traffic collected from six VPN protocols: PPTP, L2TP, L2TP-IPsec, SSTP, OpenVPN, and WireGuard. It includes both VPN and non-VPN traffic types across a variety of application categories, such as web browsing, email, VoIP, streaming, and SSH. The traffic was generated in a controlled setup using virtual machines and routers, ensuring consistent labeling and environment control. Each traffic flow is stored in JSON format and includes detailed per-flow and per-packet attributes. The dataset captures complete session behavior, including both VPN handshake phases and continuous communication, making it suitable for multi-task classification tasks such as VPN detection, protocol identification, and application-level analysis.

Still, it shares common limitations of synthetic datasets, such as the absence of spontaneous user behavior and network noise. It also includes variable bitrate (VBR) traffic like video streams, which can introduce inconsistencies in flow patterns and affect result reproducibility across models.

Dalhousie *NIMSLabVPN2024* Dataset [30], [31], publicly available on IEEE DataPort. It comprises labeled encrypted traffic generated by five widely used applications, Slack, TikTok, Twitch, Chrome, and Google Drive, spanning four service categories: messaging, streaming, browsing, and cloud storage. The traffic was captured using Cloudflare WARP VPN across two computing platforms (MacBook and Redmi smartphone) and two network environments (campus enterprise Wi-Fi and home personal Wi-Fi). For each application and setup, VPN and non-VPN traffic were collected at five different times throughout the day to capture temporal variability. Rather than providing raw packet capture (pcap) files, the dataset consists of flow-level records extracted using Tranalyzer, which include over 70 statistical metadata features per flow. This flow-based structure makes the dataset lightweight, privacy-conscious, and directly usable for machine learning tasks. It enables reproducible evaluation, comparative benchmarking, and supports further research in encrypted VPN traffic classification across diverse platforms and conditions.

While the NIMSLabVPN2024 dataset is valuable for VPN traffic analysis, it has several limitations that may affect its generalizability. It includes traffic from only five applications,Slack, TikTok, Twitch, Chrome, and Google Drive,across four service categories, limiting the diversity of traffic behaviors. All traffic was generated using scripted interactions, which ensures consistency but may not reflect the complexity of real user behavior. Data was collected on only two devices, a MacBook and a Redmi smartphone, and across two network environments, restricting variation in hardware and network conditions. The dataset provides only flow-level data extracted via Tranalyzer, without raw pcap files, which prevents packet-level inspection and limits certain types of analysis. Moreover, all VPN traffic was routed exclusively through Cloudflare WARP, which means the dataset does not capture performance or traffic pattern differences across other VPN providers. These constraints should be considered when applying or generalizing findings from this dataset to real-world scenarios.

Together, these datasets address several issues found in earlier benchmarks, including lack of protocol diversity and outdated traffic types. They represent a step forward in public dataset availability for VPN research. While current adoption rates remain modest, 5.9% for VNAT, 2.9% for USBVPN2022 and 2.9% for NIMSLabVPN2024, this is likely due to their recent release rather than a reflection of their utility. Their controlled environments ensure consistency but reduce realism, and each presents protocol or behavioral limitations that future work should address. With increased awareness and evaluation in broader contexts, these three datasets are well-positioned to support more rigorous, reproducible experimentation in VPN traffic classification.

Beyond data integrity, consistent and unambiguous categorization of traffic within datasets presents another challenge. For instance, traffic like Skype video calls in ISCXVPN2016 could arguably fit multiple categories (*e.g.*, VoIP, Video Streaming, potentially even Web), highlighting the difficulty in creating universally accepted ground truth labels for application and category identification tasks.

In conclusion, while ISCXVPN2016 has served as a unifying benchmark, critical data integrity issues necessitate extreme caution in interpreting results derived from it, particularly for methods analyzing encrypted payloads or flow characteristics. The field urgently requires thoroughly validated, diverse, and publicly accessible datasets, alongside clear guidelines for their use and interpretation, to ensure the continued progress and reliability of VPN traffic analysis research.

B. TEMPORAL EVOLUTION OF VPN TRAFFIC ANALYSIS

Fig. 3 and Fig. 4 visualize the temporal evolution of feature representation techniques and classification methodologies, respectively, in VPN traffic analysis research surveyed from 2016 to 2025. These stacked bar plots illustrate the proportion of papers employing each category per year, revealing shifts in research focus and highlighting dominant and emerging trends. Note that the number of surveyed papers per year varies, with lower counts in the earlier years (2016-2021), making percentage trends potentially more volatile during that period compared to the more recent years (2022-2025).

1) FEATURE REPRESENTATION TRENDS

Fig. 3 shows a clear diversification in feature engineering over time. *Flow statistics* features [18], [24], [25], [32], [33], [34], [35], [36] represent the most traditional approach and exhibit a sustained presence across the entire period. While consistently used, their relative proportion fluctuates, indicating they remain a foundational but not exclusive choice.

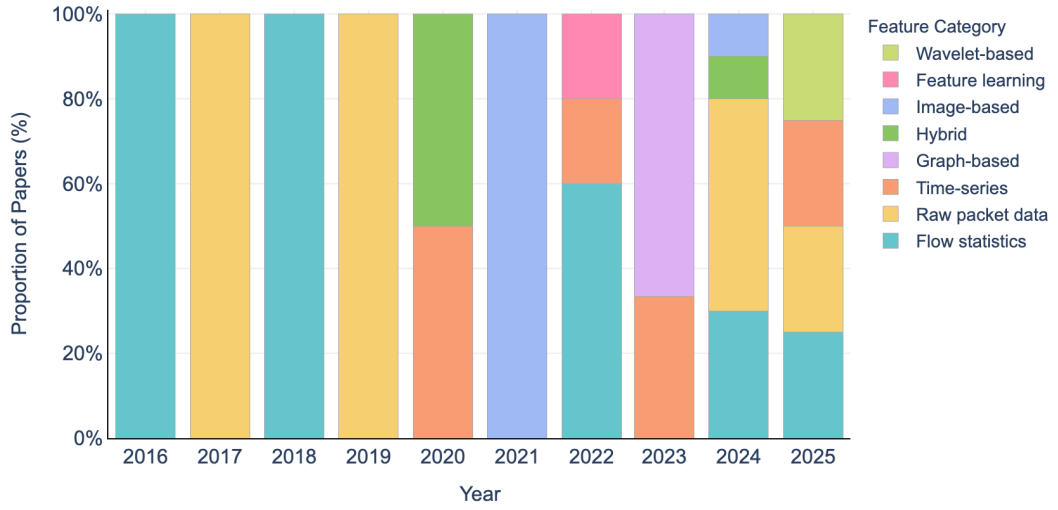


FIGURE 3. Evolution of feature representation approaches in VPN traffic analysis research (2016-2025). The y-axis shows the proportion of papers using each feature category per year, calculated based on the total number of unique papers published in that year.

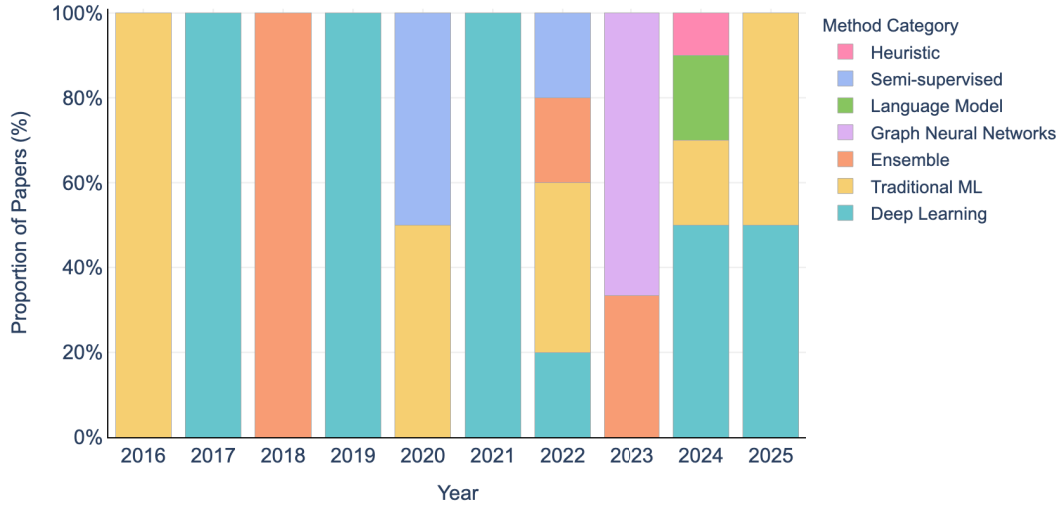


FIGURE 4. Evolution of classification methodologies in VPN traffic analysis research (2016-2025). The y-axis shows the proportion of papers using each method category per year, calculated based on the total number of unique papers published in that year.

Features derived directly from *Raw packet data* [37], [38], [39], [40], [41], [42], [43], [44], [45], often leveraged by deep learning models, gained significant traction starting in 2017 and constitute a major focus in recent years (2024-2025).

Time-series features [23], [26], [46], [47], capturing the temporal dynamics of traffic, emerged around 2020 and appear consistently in subsequent years, often used with recurrent or attention-based neural networks.

More specialized representations like *Image-based* features [20], [48] (treating traffic as images for CNNs), *Graph-based* features [49], [50] (modeling traffic flows or packets as graphs for GNNs), *Wavelet-based* features [51] (for time-frequency analysis), and automated *Feature learning* [52] using techniques like autoencoders have appeared more recently, showcasing ongoing innovation in data representation.

Hybrid approaches [19], [22], combining multiple feature types, also emerge, particularly in later years, suggesting efforts to capture complementary information.

Overall, the trend moves from reliance on statistical summaries towards utilizing richer, often less processed data representations (raw packets, time-series, graphs) that allow more complex models to potentially uncover deeper patterns, though flow statistics remain a viable option.

2) CLASSIFICATION METHODOLOGY TRENDS

Fig. 4 illustrates the adoption of different classification paradigms. *Traditional ML* algorithms (*e.g.*, Support Vector Machines (SVM), Random Fores (RF), k-Nearest Neighbors

(KNN), C4.5 a Decision Tree algorithm (DT), Extreme Gradient Boosting (XGBoost) [18], [22], [24], [31], [33], [35], [36], [51] show continuous usage throughout the surveyed period, indicating their enduring utility, sometimes enhanced by sophisticated feature engineering or feature selection techniques.

Deep Learning [19], [20], [26], [37], [38], [39], [40], [41], [42], [45], [46], [48] experienced a rapid rise starting around 2017 and has become a dominant approach, encompassing various architectures like CNNs, Long Short-Term Memory (LSTMs)/Gated Recurrent Unit (GRUs), autoencoders, and attention mechanisms, often applied to raw packet or time-series data.

Ensemble methods [32], [34], [47], which combine predictions from multiple base classifiers (either traditional ML or DL), appear periodically, often aiming for improved robustness and accuracy.

Newer paradigms have emerged recently, including *Graph Neural Networks* [49], [50] (specifically for graph-based features), *Language model*-based approaches [43], [44] (treating packet sequences like text), *Semi-Supervised* learning [23], [52] (leveraging unlabeled data), and *Heuristic* methods [25]. These represent specialized or cutting-edge techniques applied to VPN analysis.

The trend clearly shows a shift towards Deep Learning, but the persistence of Traditional ML and the emergence of diverse specialized techniques suggest the field employs a range of methodologies tailored to specific data representations and research goals.

3) SYNERGIES AND CO-EVOLUTION

The concurrent evolution of features and methods is not coincidental. The rise of Deep Learning is strongly correlated with the increased use of raw packet data and time-series features, which these models are well-suited to process end-to-end. Similarly, the appearance of Graph Neural Networks directly corresponds to the adoption of graph-based feature representations. Conversely, sophisticated feature engineering (*e.g.*, wavelets) can enhance the performance of Traditional ML models. This interplay underscores that advancements often occur at the intersection of data representation and model architecture. Continued progress likely requires innovation on both fronts simultaneously.

C. PERFORMANCE COMPARISON ANALYSIS

To provide a comparative overview of capabilities across the surveyed studies, Fig. 5 visualizes representative performance scores for different VPN traffic classification tasks. We complement Fig. 5 with Table 2, which presents the same performance data in tabular form for easier reference and direct comparison. Understanding the methodology behind this visualization is key to its interpretation. For each surveyed paper and classification task (binary, application, category, specific VPN), we curated the data as follows: if a study reported results using multiple metrics or configurations, we selected the single highest reported score for inclusion, aiming to represent the peak capability demonstrated by that approach. Furthermore, if a study provided separate scores for VPN and non-VPN traffic using the same metric (*e.g.*, binary precision), these two scores were averaged to yield a single, balanced performance indicator for that metric type within the study; if only one (VPN or non-VPN) or a combined score was reported, that single value was used. This curation approach, while simplifying the visualization and highlighting achieved potential, means the plot represents best-case or averaged scenarios rather than the full range of performance reported within individual studies. Additionally, one study [25] reporting a False Positive Rate significantly outside the range of other metrics was excluded to maintain visual clarity of the high-performance region.

Analysis of these curated performance results reveals several key observations:

- *Metric Heterogeneity*: Despite selecting the highest score per study, the choice of *which* metric was reported as highest varies. Accuracy and F1-score are most common, with some instances of Precision (*e.g.*, [18], [36], [37]). This underlying heterogeneity in reporting still complicates direct comparisons, as the ''best'' score might be measured differently across papers. The lack of universally reported metrics like Recall further limits comprehensive comparison.
- *Performance Variation by Task*: Even considering best/averaged scores, performance varies significantly by task complexity. Binary VPN vs. non-VPN classification consistently achieves very high scores (often $>$ 95%, *e.g.*, [20], [24], [34], [35], [36], [37], [40], [43], [45], [46], [47], [51], [52]), suggesting high efficacy is attainable for this task. Application and Category identification exhibit wider performance ranges ($75\% - 99.9\%$ for Application, *e.g.*, [18], [20], [26], [31], [38], [42], [43], [49], and $78\% - 99.9\%$ for Category, *e.g.*, [19], [20], [23], [26], [32], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [46], [47], [49], [50]), reflecting their inherent difficulty. Specific VPN identification results [22], [35] are limited but show high potential ($94\% - 99\%$).
- *Methodological Impact*: Cross-referencing these top/averaged scores suggests Deep Learning (DL) approaches frequently achieve the highest reported performance, especially for Application and Category tasks (*e.g.*, [20], [26], [37], [38], [39], [40], [41], [42], [43], [44], [46], [47], [49], [50]). However, Traditional ML (*e.g.*, [35], [36], [51]) and Ensemble methods (*e.g.*, [34], [47]) also reach near-optimal performance in several cases, particularly for Binary classification, indicating their continued relevance when potentially combined with effective feature engineering.
- *Influence of Datasets*: The majority of results shown are on ISCXVPN2016. While high peak performance is achieved on other datasets (*e.g.*, [19], [22], [35], [51]), direct comparison remains challenging. Importantly, the

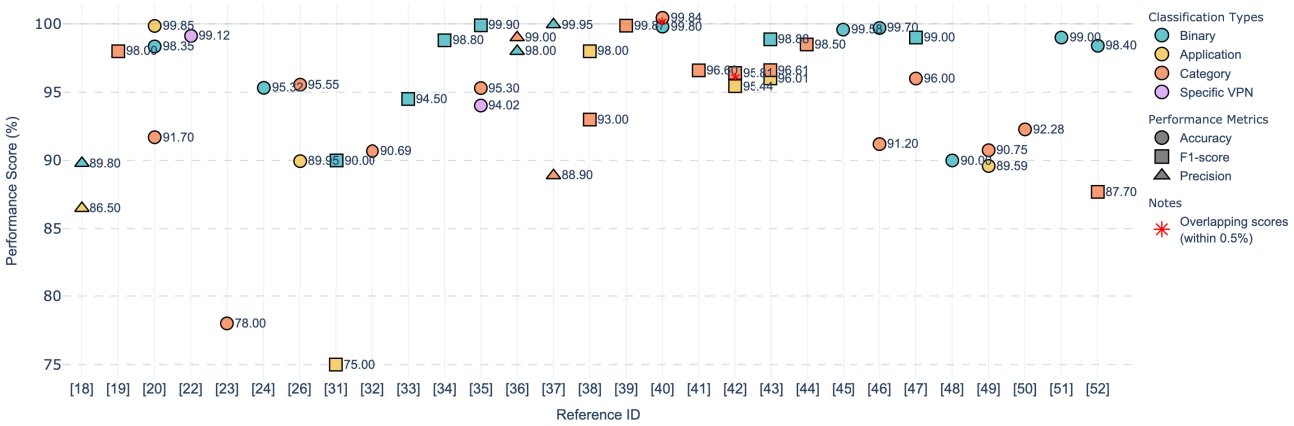


FIGURE 5. Performance comparison of network traffic classification methods (excluding [25]). Points represent the highest reported score or an average of VPN/non-VPN scores for a given task and metric within each study. Colors represent classification types; marker shapes indicate the performance metric (primarily Accuracy and F1-score, some Precision). Higher scores indicate better performance. Note the potential influence of dataset characteristics (Section II-A) and data curation methodology on reported scores.

TABLE 2. Performance comparison of VPN detection methods across different datasets and classification tasks.

Citation	Dataset	Binary		Application		Category		Specific VPN	
		Metric	Score	Metric	Score	Metric	Score	Metric	Score
Draper-Gil *et al.* [18]	ISCXVPN2016	Precision	89.80†	Precision	86.50†	–	–	–	–
Jorgensen *et al.* [19]	VNAT	–	–	–	–	F1-score	98.00	–	–
Shapira *et al.* [20]	Private dataset, ISCXVPN2016	Accuracy	98.35†	Accuracy	99.85†	Accuracy	91.70†	–	–
Gao *et al.* [22]	Private dataset	–	–	–	–	–	–	Accuracy	99.12
Iliyasu *et al.* [23]	Private dataset, ISCXVPN2016	–	–	–	–	Accuracy	78.00	–	–
Li *et al.* [24]	Private dataset, ISCXVPN2016	Accuracy	95.32	–	–	–	–	–	–
Hanlon *et al.* [25]	Private dataset	FPR	29.00	–	–	–	–	–	–
Kotak *et al.* [26]	Private dataset, ISCXVPN2016	–	–	Accuracy	89.95†	Accuracy	95.55†	–	–
Liu *et al.* [31]	NIMSLabVPN2024	F1-score	90.00	F1-score	75.00	–	–	–	–
Caicedo-Muñoz *et al.* [32]	ISCXVPN2016	–	–	–	–	Accuracy	90.70†	–	–
Al-Fayoumi *et al.* [33]	ISCXVPN2016	F1-score	94.50†	–	–	–	–	–	–
Almomani [34]	ISCXVPN2016	F1-score	98.80	–	–	–	–	–	–
Fesl *et al.* [35]	USBVPN2022	F1-score	99.90	–	–	Accuracy	95.30	Accuracy	94.02
Gudla *et al.* [36]	ISCXVPN2016	Precision	98.00	–	–	Precision	99.00	–	–
Wang *et al.* [37]	ISCXVPN2016	Precision	99.95†	–	–	Precision	88.90†	–	–
Lotfollahi *et al.* [38]	ISCXVPN2016	–	–	F1-score	98.00	F1-score	93.00	–	–
Zeng *et al.* [39]	ISCXVPN2016	–	–	–	–	F1-score	99.87	–	–
Cui *et al.* [40]	ISCXVPN2016	Accuracy	99.80†	–	–	Accuracy	99.85†	–	–
Seydali *et al.* [41]	ISCXVPN2016	–	–	–	–	F1-score	96.60	–	–
Chai *et al.* [42]	ISCXVPN2016	–	–	F1-score	95.44	F1-score	95.81	–	–
Park *et al.* [43]	ISCXVPN2016	F1-score	98.88†	F1-score	96.01	F1-score	96.61	–	–
Luo *et al.* [44]	ISCXVPN2016	–	–	–	–	F1-score	98.50	–	–
Tawfeeq *et al.* [45]	ISCXVPN2016	Accuracy	99.58†	–	–	–	–	–	–
Yao *et al.* [46]	ISCXVPN2016	Accuracy	99.70	–	–	Accuracy	91.20	–	–
Abbas *et al.* [47]	ISCXVPN2016	F1-score	99.00	–	–	Accuracy	96.00	–	–
Liu *et al.* [48]	ISCXVPN2016	Accuracy	90.00	–	–	–	–	–	–
Huoh *et al.* [49]	ISCXVPN2016	–	–	Accuracy	89.60†	Accuracy	90.75†	–	–
Okonkwo *et al.* [50]	ISCXVPN2016	–	–	–	–	Accuracy	92.28†	–	–
Razooqi *et al.* [51]	VNAT	Accuracy	99.00	–	–	–	–	–	–
Lin *et al.* [52]	ISCXVPN2016	Accuracy	98.40	–	–	F1-score	87.70†	–	–

† Single balanced score obtained via averaging VPN and non-VPN scores.

high scores reported on ISCXVPN2016 must be viewed critically, considering not only the data curation method used for this plot but also the dataset integrity issues discussed in Section II-A.

In summary, while the curated performance data indicates significant capabilities, particularly in binary VPN detection, the variability in finer-grained tasks persists. The visualization highlights peak or averaged performance, and

interpreting these results requires acknowledging the data selection methodology, the heterogeneity of reported metrics, and the critical caveats surrounding the most commonly used benchmark dataset (ISCXVPN2016).

D. ACCURACY TRENDS ON ISCXVPN2016

Given the widespread use of the ISCXVPN2016 dataset (Section II-A), we examine accuracy trends across three common tasks: binary classification (7 studies), category classification (9 studies), and application classification (3 studies). Accuracy was the most consistently reported metric. No studies reported accuracy for the ''specific VPN identification'' task.

Fig. 6 shows the yearly *median* accuracy for each task from 2016 to 2025. Median values are used to reduce the impact of outliers, especially in years with fewer data points. From the figure, we find that binary classification consistently yields the highest accuracy, often exceeding 98% in recent years (2021-2025). Category classification typically ranges from 78% to 99.8%. Application classification achieved its highest median accuracy of 99.8% in 2021, but lower scores in other years.

Binary classification accuracy remains consistently high throughout the period, with minimal variation. Category classification shows improvement over time, with some fluctuation but generally increasing performance. Application classification exhibits the most variability, with a peak in 2021 and a drop in subsequent years, indicating higher sensitivity to task setup and methodology.

Lastly, high median scores for category and application tasks often coincide with deep learning-based studies (*e.g.*, [20], [26], [40], [46], [49], [50]). At the same time, ensemble and traditional ML models also contribute significantly to binary and category task performance (*e.g.*, [24], [47]), demonstrating the ongoing value of well-engineered non-DL approaches.

These trends suggest binary classification performance on ISCXVPN2016 may be nearing saturation. In contrast, category and application tasks still present variability and opportunities for further improvement.

III. IMPLICATIONS AND APPLICATIONS

Despite the challenges, VPN traffic analysis techniques offer significant implications for various network management and security domains.

A. NETWORK SECURITY AND POLICY ENFORCEMENT

The ability to reliably detect VPN usage is fundamental for enforcing organizational security policies, particularly in environments requiring visibility into network connections (*e.g.*, restricting unsanctioned encrypted tunnels or ensuring traffic flows through mandated security gateways). The high accuracy achieved for binary classification (Section II-D) suggests deployable solutions for basic VPN detection are feasible.

Identifying specific applications or traffic categories within VPN tunnels, while currently more challenging (Section II-C), offers deeper security insights. It enables detection of policy violations (*e.g.*, unauthorized file sharing), identification of potential covert channels used by malware (*e.g.*, C&C communication over seemingly legitimate VPNs), or monitoring for data exfiltration attempts. The varying performance seen in application/category classification highlights the need for context-specific deployment; while not perfect, current methods might still provide valuable signals for threat hunting when correlated with other security data.

The capability to identify specific VPN services or protocols (represented by the ''specific VPN'' task, *e.g.*, [22], [35]) could enable risk-based policy enforcement, differentiating between corporate-sanctioned VPNs and potentially less trusted consumer services. However, the limited research and variable performance in this specific task indicate it requires further development for reliable application.

Furthermore, temporal analysis methods (*e.g.*, using time-series features [26], [46]) could potentially detect anomalous VPN usage patterns deviating from normal user behavior, aiding in the detection of compromised accounts or insider threats, although this remains an emerging application area.

B. QUALITY OF SERVICE (QOS) MANAGEMENT

VPN classification allows network operators to implement more granular QoS policies, moving beyond simple port- or IP-based rules which are ineffective for encrypted tunnels. By identifying VPN traffic and potentially its internal application category, operators can prioritize latency-sensitive applications (VoIP, video conferencing) or manage bandwidth consumption for bulk traffic (file transfers, streaming) even within VPNs.

The improving median accuracy for category classification on ISCXVPN2016 (Section II-D) suggests that this level of granularity is becoming increasingly practical for QoS differentiation. Recognizing broad categories such as ''Streaming'' or ''Chat'' often provides sufficient information for effective traffic prioritization, without requiring precise per-application classification.

However, the computational cost of accurate real-time classification remains a significant consideration, as noted in Section V-A. High-throughput environments or resource-limited network edges may struggle to implement complex DL-based classifiers at line rate. This highlights the need for continued research into efficient yet accurate models or hardware acceleration techniques to make sophisticated VPN-aware QoS feasible in diverse deployment scenarios.

C. PRIVACY AND ETHICAL CONSIDERATIONS

The development and deployment of VPN traffic analysis techniques inherently intersect with user privacy expectations and ethical considerations. While network operators and organizations have legitimate interests in security, compliance, and performance management, these must be balanced

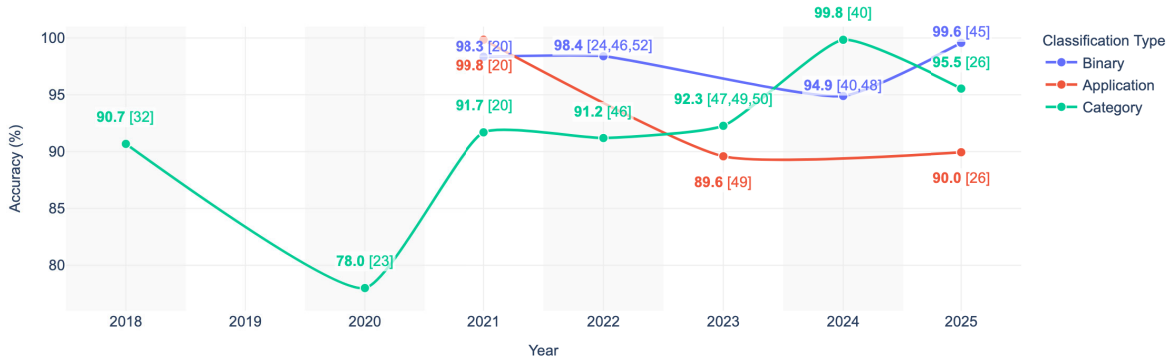


FIGURE 6. Median accuracy on the ISCXVPN2016 dataset over time (2016-2025), separated by classification type. Only types with reported accuracy scores on this dataset are shown. Paper reference IDs contributing to each median point are indicated in the interactive version or supplemental data.

against the privacy afforded by VPNs, which are often used specifically to shield activities from monitoring.

The increasing accuracy of classification methods, particularly fine-grained application/category identification (*e.g.*, achieving high accuracy in studies like [20], [40]), intensifies this tension. The capability to infer specific user activities within an encrypted tunnel raises significant privacy concerns. Ethical deployment necessitates careful consideration of proportionality (is the level of monitoring necessary for the stated goal?), purpose limitation (using the information only for legitimate, predefined purposes), transparency (informing users about monitoring practices), and data minimization (collecting/analyzing only the necessary information).

Legal frameworks like GDPR in Europe impose strict regulations on processing network data that could be linked to individuals, requiring clear legal bases and robust safeguards. Deploying VPN analysis tools, especially those inferring application usage, requires careful legal and ethical review within specific jurisdictional contexts.

Furthermore, widespread or overly aggressive VPN detection/classification could have a chilling effect on the legitimate use of VPNs for security (*e.g.*, on public Wi-Fi), accessing geo-restricted content legally, or protecting sensitive communications. Research and deployment practices should ideally strive for solutions that meet security/operational needs with the minimum possible impact on user privacy and the beneficial uses of VPN technology.

IV. REVIEW OF CLASSIFICATION TECHNIQUES

VPN traffic classification has prompted the development of diverse techniques, ranging from traditional machine learning algorithms to advanced deep learning and semi-supervised methods. This survey reviews the major categories of existing approaches. The categorization adopted here is one of several possible schemes, structured primarily around methodological distinctions,namely, traditional supervised learning, deep learning, and other emerging paradigms. Certain approaches, such as those involving language models, may be considered a subset of deep learning. However, we treat them separately when their architectures or training strategies justify distinct consideration. The goal is to provide a clear, practical organization of the literature without implying a definitive or exhaustive taxonomy.

A. TRADITIONAL SUPERVISED LEARNING

Traditional supervised learning has played a foundational role in early VPN traffic classification research. These methods typically rely on statistical, payload-independent flow features such as packet size distributions, inter-arrival times, and flow durations. Classical models like DT, k-NN, Naïve Bayes, and SVM have been widely used to perform binary or multi-class classification of encrypted traffic. The following studies highlight the breadth of supervised approaches, their evolving feature engineering strategies, and comparative effectiveness in both legacy and modern datasets.

Draper-Gil et al. [18] proposed a supervised learning method based solely on time-related flow features, including inter-arrival time, duration, and idle/active time. Their approach used C4.5 and k-NN classifiers to detect VPN traffic and classify the category of encrypted applications such as VoIP, chat, and streaming. Achieving over 80% accuracy in both detection and classification, their results demonstrated that simple, content-independent features can effectively characterize encrypted flows.

However, Gao et al. [22] identified limitations in relying exclusively on time or entropy-based features, particularly when handling obfuscated VPN traffic. To address this, they introduced a two-stage classification framework using Sample Entropy Fingerprints and Payload Length Sequence (PLS). While entropy captured randomness in packet payloads, it struggled with obfuscation. PLS, which records bidirectional payload length sequences, proved more robust and protocol-specific. Models trained on PLS features,

especially using t RF, achieved up to 98.7% F1-score, outperforming entropy-only methods and traditional DPI tools.

Building on time-based analysis, Al-Fayoumi et al. [33] developed a flow-based classification model using RF to distinguish VPN from non-VPN traffic. Their feature set included flow duration, inter-arrival times in both directions, and flow throughput. To enhance efficiency and model interpretability, they applied feature selection methods such as Pearson Correlation and Genetic Algorithms (GA), achieving over 95% accuracy even after dimensionality reduction. This made their approach practical for real-time deployment.

In contrast, Li et al. [24] focused on tunneling protocol characteristics (TPC) rather than conventional time features. They proposed four statistical feature categories, including two novel entropy-based metrics designed to capture the structural effects of VPN encapsulation. Their model, trained with RF, achieved 99.02% accuracy on real-world VPN traffic and 95.32% on the ISCXVPN2016 dataset. This approach combined high accuracy with interpretability, avoiding the computational demands of deep learning.

To address dataset limitations, Fesl et al. [35] introduced USBVPN2022, a custom dataset comprising six VPN protocols and multiple traffic types. They evaluated several models,RF, XGBoost, MLP, and LSTM,across tasks such as VPN detection, protocol classification, and encrypted traffic identification. Their framework achieved up to 99.9% accuracy in VPN detection and over 94% in encrypted traffic classification. Importantly, they incorporated obfuscation techniques to assess model robustness and inform the development of stealth VPNs resistant to ML-based detection.

Building on the need for richer datasets and reliable feature extraction, Liu et al. [31] introduce a new dataset and use statistical flow features extracted with Tranalyzer, a high-performance network flow analyzer. After removing protocol-specific and constant fields to avoid model bias, they retained 71 features for laptops and 76 for smartphones. These features include flow duration, packet counts, byte statistics, and inter-arrival times,suitable for analyzing encrypted traffic without deep packet inspection. The authors applied three traditional machine learning models: C4.5 DT, RF, and Lasso Regression. These models were trained on their dataset for three classification tasks: binary classification (VPN vs. non-VPN), application classification (*e.g.*, Slack, Chrome), and category-level classification (*e.g.*, messaging, streaming, browsing, cloud storage). RF achieved the best performance, with F1-scores reaching, in some categories, 100% for binary classification and over 90% for application classification.

While prior work focused on accuracy and feature design, Gudla et al. [36] emphasized efficiency and scalability. They proposed TCC, a time-constrained classification framework targeting both VPN and non-VPN encrypted traffic using the ISCXVPN2016 dataset. Their hybrid dimensionality reduction method, combining autoencoders with Linear Discriminant Analysis (LDA), reduced feature space while maintaining accuracy. Among the tested models, a DT with autoencoder preprocessing achieved perfect F1-scores in several classes. Compared to CNN and LSTM, their approach offered faster and more consistent performance, well-suited for real-time systems.

Finally, Razooqi et al. [51] explored signal processing techniques by applying Discrete Wavelet Transform (DWT) to packet size sequences. They extracted statistical features from both traffic directions at decomposition levels 5 and 12. RF models trained on these wavelet-based features achieved up to 99% F1-score. Although higher decomposition levels slightly improved accuracy, they also increased dimensionality. The study showed strong generalization across traffic types like chat, VoIP, and file transfer, highlighting the ability of wavelet features to capture fine-grained and coarse traffic behaviors.

B. ENSEMBLE LEARNING

Ensemble learning methods have gained prominence in VPN traffic classification due to their ability to combine multiple models and improve generalization. These approaches leverage the strengths of individual classifiers while mitigating their weaknesses, often resulting in higher accuracy and robustness. The following studies demonstrate various ensemble strategies,bagging, stacking, and boosting,each tailored to enhance performance in different classification scenarios using time-based and statistical features.

Caicedo-Muñoz et al. [32] proposed a QoS-Classifier that uses time-related features to distinguish VPN from non-VPN traffic. The approach integrates a two-stage process,classification and marking,aligned with Per-Hop Behavior (PHB) in DiffServ architectures. The authors created a QoS-labeled dataset based on known VPN traffic and evaluated several machine learning algorithms. Bagging achieved the highest classification accuracy, reaching 94.42%. Their findings confirm the effectiveness of time-based features in analyzing encrypted traffic and demonstrate how PHB labeling can enhance classification accuracy while supporting network-level QoS policies for real-time traffic management.

While CaicedoMuoz focused on bagging, Almomani [34] introduced a stacking ensemble model combining RF, Artificial Neural Networks (ANNs), and SVM as base classifiers, with Logistic Regression (LR) serving as the meta-learner. Their system was trained using 61 statistical flow-level features such as duration, packet sizes, byte rates, and header flags. Tested on the ISCXVPN2016 dataset, the stacking model achieved 99.3% classification accuracy, outperforming all individual models. Notably, the method operated without payload inspection, relying solely on metadata features, which enhances its suitability for encrypted traffic environments.

Expanding on ensemble diversity, Abbas et al. [47] proposed a two-strategy ensemble framework using boosting-based models such as Light Gradient Boosting Machine (LGBM), XGBoost, and Categorical Boosting (CatBoost), combined with thorough feature engineering and

partitioning techniques. Their evaluation covered both binary and multiclass classification tasks on the ISCXVPN2016 dataset. For multiclass application category-level classification, XGBoost achieved 96.0% accuracy using 15-second non-VPN traffic segments. In binary classification, AdaBoost reached 93.1% accuracy. Their second strategy, which included dataset merging, feature scaling via min–max normalization, and stratified k-fold cross-validation, resulted in a peak accuracy of 99.7% using LGBM, outperforming prior approaches.

C. DEEP LEARNING

Deep learning has become a key enabler in encrypted VPN traffic classification, offering automatic feature extraction and high classification accuracy without the need for manual engineering. Unlike traditional approaches that rely on statistical or handcrafted features, deep neural networks learn complex spatial, temporal, and hierarchical patterns directly from raw or minimally processed traffic data. Various architectures,including CNNs, RNNs, attention mechanisms, and hybrid models,have been applied to traffic sequences, images, and signal representations to enhance performance, generalizability, and robustness in real-world scenarios.

Wang et al. [37] proposed an end-to-end method using one-dimensional convolutional neural networks (1D-CNN) for classifying encrypted traffic, including VPN and non-VPN flows. This model processes raw packet data directly, removing the need for handcrafted features. Validated on the ISCXVPN2016 dataset, the approach achieved significant gains in accuracy, precision, and recall over traditional machine learning baselines. The study found 1D-CNN to outperform 2D-CNN due to its better alignment with one-dimensional traffic sequences, confirming the effectiveness of CNN-based models in encrypted traffic classification.

Building on this, Lotfollahi et al. [38] developed Deep Packet, an end-to-end deep learning framework that integrates stacked autoencoders and 1D-CNN for both application and traffic category classification. Trained on the same ISCXVPN2016 dataset, Deep Packet achieved a weighted F1-score of 98% for application identification and 0.93 for traffic category classification. Like the previous work, it bypasses manual feature design, demonstrating scalability and speed in processing raw packet-level data.

While Deep Packet focused on single-architecture models, Zeng et al. [39] introduced Deep-Full-Range (DFR), a hybrid framework that combines CNN, LSTM, and stacked autoencoders (SAE). DFR jointly captures spatial, temporal, and coding-level features from raw traffic, enabling precise classification without using payload data or sensitive information. Evaluated on the ISCXVPN2016 dataset, DFR achieved up to 99.85% accuracy, outperforming previous deep learning and machine learning baselines in terms of F1-score and storage efficiency.

In contrast, Shapira et al. [20] presented FlowPic, which transforms network flows into 2D histograms representing packet size and inter-arrival time. These visual representations are classified using CNNs, capturing both structural and temporal aspects of VPN traffic. FlowPic achieved over 99% accuracy on ISCXVPN2016 and ISCX Tor datasets, generalizing well to unseen applications and preserving privacy by avoiding payload analysis. The method outperformed conventional feature-based and deep learning models across multiple encrypted traffic scenarios.

Expanding on temporal modeling, Yao et al. [46] proposed an attention-based deep learning framework using LSTM and Hierarchical Attention Networks (HAN). By treating flows as time-series and applying attention mechanisms, the model identified critical packets and byte regions in the sequence. On the ISCXVPN2016 dataset, the system achieved 91.2% accuracy for a 12-class task, outperforming CNN-based and traditional models. The results emphasized the benefit of attention in capturing temporal relevance for encrypted traffic.

Focusing on IoT-specific challenges, Liu et al. [48] introduced a CNN-based classification system that incorporates both spatial features (*e.g.*, IPs, ports, flags) and temporal metrics (*e.g.*, delays, durations). These were encoded into structured image representations and processed using CNNs. The framework achieved 90% accuracy and 93.3% recall while maintaining low latency and high throughput, making it suitable for real-time IoT environments.

Meanwhile, Cui et al. [40] introduced AMAE, an interpretable deep learning model combining attention mechanisms and autoencoders. Encrypted session-level traffic was converted into 28×28 grayscale images and processed through spatial and channel attention layers to enhance important byte-level patterns. The model achieved 99.69% accuracy for VPN traffic and 100% for non-VPN traffic on the ISCXVPN2016 dataset, outperforming CNN+RNN and standard autoencoders. AMAE also provided interpretable visualizations, addressing the lack of transparency in most deep learning classifiers.

In a different direction, Jorgensen et al. [19] proposed an extensible classification framework using wavelet-transformed features and traditional flow statistics. Stationary Wavelet Transform (SWT) was applied to extract frequency-domain features like wavelet energy and Shannon entropy. These were input into a prototypical neural network with uncertainty estimation via Mahalanobis distance, enabling detection of out-of-distribution samples. The approach enhanced adaptability and interpretability in dynamic network environments.

To address scalability and deployment challenges, Seydali et al. [41] developed a distributed deep learning pipeline using Apache Kafka, Apache Spark Streaming, and TensorFlow. The model combined 1D-CNN with an attention-based Bidirectional Gated Recurrent Unit (Bi-GRU) to capture both spatial and temporal features. Evaluated on the ISCXVPN2016 dataset, the system demonstrated up to 25% improvement in F1-score under scaling, showing effective performance in high-throughput, real-time network scenarios.

Focusing on packet-level representation, Chai et al. [42] introduced Combo Packet, which uses CNNs with spatial and channel attention to extract byte-level and contextitual features from packet sequences. Validated on the ISCXVPN2016 dataset, the model achieved 97.04% and 97.13% accuracy in category and application classification, respectively. Its emphasis on contextitual packet relationships improved robustness to transmission variability, outperforming state-of-the-art models.

For adversarial robustness, Tawfeeq et al. [45] proposed a hybrid model combining EfficientNet-B0 and BiLSTM, trained with Projected Gradient Descent (PGD). This architecture processes variable-length raw packet sequences into fixed-size vectors and resists network anomalies such as delays and congestion. On the ISCXVPN2016 dataset, the model reached 99.81% accuracy for normal and 99.35% for adversarial traffic, showing high robustness and generalization.

Finally, Kotak et al. [26] introduced a time-series classification framework based on InceptionTime. Their model used 1-second interval statistical features with a 60-second sliding window, trained on a custom dataset from three geographic regions. It achieved 93.8% accuracy for VPN category-level and 86.5% for application-level classification. The model also performed well on non-VPN and IoT traffic, outperforming feature-based methods in handling tunnel-encapsulated VPN flows.

D. GRAPH NEURAL NETWORKS

Graph Neural Networks (GNNs) have emerged as a promising direction for encrypted VPN traffic classification, offering a distinct shift from traditional sequential or grid-based models. Instead of treating flows or packets as isolated instances, GNNs model traffic as graphs where nodes represent packets or flows and edges encode temporal, structural, or contextitual relationships. This graph-based perspective allows GNNs to learn complex dependencies and interactions across traffic elements, making them well-suited for handling encrypted and variable-length traffic in dynamic network environments.

Huoh et al. [49] proposed a GNN-based framework that classifies encrypted VPN and non-VPN traffic using flow-level input. Unlike conventional CNN or RNN architectures, their model processes raw bytes, temporal relationships, and metadata within a non-Euclidean graph structure. The framework employs an encoder-core-decoder design to learn feature representations from the graph topology. Trained on the ISCXVPN2016 dataset, the GNN outperformed CNN and LSTM baselines in both category and application-type classification, demonstrating higher accuracy and F1-scores. The study highlights GNN's flexibility in dealing with variable-length flows and its robustness across both encrypted and unencrypted data.

Building on this direction, Okonkwo et al. [50] introduced a higher-order GNN that models individual traffic sessions as graphs with packets as nodes and sequential dependencies as edges. This setup enables the model to capture both local packet features and global structural patterns. Their architecture includes five GNN layers with GraphConv operations, batch normalization, and top-K pooling for effective representation learning and dimensionality reduction. Evaluated on the ISCXVPN2016 and USTC-TFC datasets, the model achieved 97.48% accuracy for VPN traffic and 87.07% for non-VPN traffic in category classification, outperforming existing techniques. The study also analyzed the impact of input truncation and adopted fixed-length graph representations with stratified cross-validation to ensure generalization.

E. LANGUAGE MODEL

Recent research has explored the application of pre-trained language models to encrypted VPN traffic classification, motivated by their strong performance in sequence modeling and transfer learning. These models treat network traffic as a form of structured textit, enabling the use of architectures originally designed for natural language processing. By leveraging models like Bidirectional Encoder Representations from Transformers (BERT) and Text-to-Text Transfer Transformer (T5), these approaches can perform multi-task learning (MTL), eliminate the need for manual feature engineering, and generalize well with limited labeled data.

Park et al. [43] introduced a Multi-Task Learning (MTL) framework based on DistilBERT for encrypted traffic classification. The model simultaneously addresses three tasks: detecting encryption status (VPN vs. non-VPN), identifying the application category (*e.g.*, VoIP, streaming), and classifying specific applications (*e.g.*, Skype, Gmail). Trained on the ISCXVPN2016 dataset, the model uses hard parameter sharing to optimize training efficiency and incorporates class and task weights to mitigate imbalances. It achieved 99.29% accuracy for encryption detection, 97.38% for category classification, and 96.89% for application-level classification. The results demonstrate the effectiveness of MTL in capturing interrelated traffic patterns while improving training speed and robustness.

Building on the idea of traffic-as-text, Luo et al. [44] proposed an approach using a fine-tuned T5 model for encrypted traffic classification. Their method converts traffic data into a textual format suitable for language modeling and applies transfer learning to fine-tune the T5 architecture. Validated on the ISCXVPN2016 dataset, the model achieved a weighted F1-score of 98%, outperforming baseline models such as CNN, stacked autoencoders (SAE), and DeNeT-Lang. It showed strong performance in classifying VPN traffic types, including chat, VoIP, email, and streaming, even with limited training data. The study highlights the potential of language models for scalable and accurate encrypted traffic classification without manual feature extraction.

F. SEMI-SUPERVISED LEARNING

Semi-supervised learning offers a practical solution for encrypted VPN traffic classification in scenarios where labeled data are scarce. These approaches combine the strengths of supervised and unsupervised learning by leveraging a small labeled dataset alongside a larger pool of unlabeled traffic. This is especially useful in encrypted environments, where manual labeling is costly and limited in scope. Recent work in this area has focused on generative models, clustering-based frameworks, and self-training techniques to enhance classification performance with minimal supervision.

Iliyasu et al. [23] proposed a semi-supervised approach using Deep Convolutional Generative Adversarial Networks (DCGAN) to address the challenge of limited labeled data in VPN traffic classification. Their model combines a small number of labeled samples with synthetic data generated by a GAN. Time-series features such as inter-arrival time, packet length, and direction are transformed into pseudo-image matrices for input to the classifier. Evaluated on the ISCXVPN2016 and QUIC datasets, the model achieved 78% and 89% accuracy, respectively, using only 10% labeled data. It outperformed CNN and Multilayer Perceptron (MLP) baselines on the QUIC dataset, showing strong potential for real-world encrypted traffic analysis under data-constrained conditions.

Complementing this, Lin et al. [52] introduced Sauce, a semi-supervised model designed to reduce dependence on labeled data while maintaining high performance in VPN detection. The framework integrates an autoencoder for unsupervised feature extraction with an auxiliary network that uses pseudo-labels to guide clustering. Dimensionality reduction is achieved through t-distributed Stochastic Neighbor Embedding (t-SNE), followed by k-means clustering. Sauce attained clustering accuracy of 98.4% on the VPN dataset and 98.0% on the Tor dataset, using only 5% labeled data. Compared to models requiring more supervision, Sauce improves generalization and clustering stability through self-training, offering a reliable alternative to fully supervised classification.

G. HEURISTIC METHODS

Heuristic methods offer a rule-based alternative to data-driven approaches for VPN traffic detection. These techniques apply manually defined rules based on protocol behavior, timing patterns, or statistical properties, enabling interpretable and computationally efficient detection. Although often less flexible than learning-based models, heuristic approaches can deliver high precision in specific use cases, especially when targeting well-defined protocol anomalies or behaviors.

Hanlon et al. [25] proposed a protocol-agnostic VPN detection method based on heuristic analysis of encapsulated Transmission Control Protocol (TCP) behavior in UDP tunnels. Instead of using machine learning, the method relies on Request for Comments (RFC)-defined TCP traits, including the presence of a Three-Way Handshake (3WHS), timely acknowledgments within 500 milliseconds (500msACK), and acknowledgments after receiving twice the Remote Maximum Segment size (2×RMSS). These traits serve as indicators of TCP encapsulation. Evaluated on real-world traffic, the approach achieved a false positive rate as low as 0.11%, outperforming many machine learning models. It proved particularly effective for identifying file transfer and SSH traffic, though it was less accurate for multiplexed web traffic. The study underscores the potential of protocol behavior analysis as a lightweight and effective VPN detection strategy, revealing weaknesses in existing tunneling practices.

V. CHALLENGES AND LIMITATIONS

Despite significant advancements, VPN traffic analysis faces persistent technical hurdles, data limitations, and methodological inconsistencies that impede further progress and reliable deployment.

A. TECHNICAL CHALLENGES

The core technical challenges stem primarily from encryption, obfuscation, and the operational constraints of real-world networks.

- *Encryption and Protocol Evolution:* The fundamental difficulty lies in classifying traffic whose content is intentionally obscured by encryption. While current methods successfully exploit metadata (flow statistics, packet timing/size), the continuous evolution of VPN protocols (*e.g.*, WireGuard gaining traction) and encryption standards (like TLS 1.3 reducing handshake information) constantly threatens the longevity of existing feature-based approaches. Methods relying on specific protocol artifacts are particularly vulnerable.
- *Obfuscation and Adversarial Tactics:* Recognizing the effectiveness of traffic analysis, VPN services and privacy tools increasingly employ obfuscation techniques like traffic shaping, padding, and randomized timing to mimic benign traffic (*e.g.*, standard HTTPS). These techniques directly attack the statistical and temporal features used by many classifiers. Our survey reveals a significant gap in addressing these adversarial tactics, with only limited work [45] explicitly incorporating adversarial robustness into model training. This arms race necessitates research into features and models inherently resilient to common obfuscation strategies.
- *Scalability and Real-time Performance:* The computational demands of sophisticated methods, especially Deep Learning models operating on raw packet or fine-grained temporal data, pose practical challenges for real-time, line-rate deployment on high-speed networks or resource-constrained edge devices. Most surveyed studies focus primarily on classification accuracy, often neglecting crucial performance metrics like

inference latency, throughput, and resource utilization (CPU/memory). This accuracy-efficiency trade-off remains a critical, underexplored area essential for translating research into deployable solutions.
- *Distinguishing Encrypted Traffic Types:* Differentiating VPN traffic from other ubiquitous encrypted flows (HTTPS, DoH/DoT, QUIC, SSH, secure messaging) is increasingly complex. As more internet traffic becomes encrypted by default, relying solely on encryption presence is insufficient. Classifiers need robust features that specifically fingerprint VPN tunneling behavior amidst a sea of other encrypted protocols, a challenge acknowledged but not comprehensively solved in the current literature.

B. DATASET LIMITATIONS AND BENCHMARKING ISSUES

The quality, diversity, and accessibility of datasets remain significant bottlenecks, directly impacting the validity and generalizability of research findings.

- *Benchmark Staleness and Integrity:* As highlighted in Section II-A, the field heavily relies on the ISCXVPN2016 dataset (used in 70.6% of surveyed papers). This dataset, captured in 2016, may not represent current VPN protocols, applications, or background traffic accurately. More critically, identified integrity issues (unencrypted payloads in VPN captures, multiple connections per capture file) undermine its validity for evaluating methods designed for encrypted traffic, potentially leading to inflated performance metrics and misleading conclusions, particularly for payload-aware or flow-based methods [25].
- *Lack of Diverse Public Datasets and Reproducibility:* The scarcity of modern, large-scale, and diverse public benchmarks hinders progress. The reliance on private datasets (17.6% of papers) prevents independent verification and reproduction of results. Even among public datasets (ISCXVPN2016, VNAT, USB-VPN2022, NIMSLabVPN2024), variations in capture environments, traffic types, labeling granularity, and preprocessing steps make direct cross-dataset comparisons challenging. The slow adoption of newer datasets like VNAT [19], [51], USBVPN2022 [35], and NIMSLabVPN2024 [31] suggests a need for community consensus or more compelling benchmark options.
- *Ecological Validity and Labeling Ambiguity:* Many datasets are generated in controlled laboratory environments, potentially lacking the noise, scale, and heterogeneity of real-world network traffic (''ecological validity''). Furthermore, datasets often cover a limited range of VPN services, applications, and user behaviors. Labeling ground truth, especially for application or category classification, can also be ambiguous (as noted with Skype traffic) or lack sufficient granularity, impacting the training and evaluation of fine-grained classifiers.
- *Temporal Relevance and Concept Drift:* Internet traffic is non-stationary; application behaviors, protocol usage, and VPN obfuscation techniques evolve. Most existing datasets represent static snapshots in time. The lack of longitudinal datasets capturing this evolution makes it difficult to assess the long-term robustness of classification methods or study the phenomenon of concept drift, where model performance degrades as real-world traffic patterns diverge from the training data.

VI. FUTURE RESEARCH DIRECTIONS

Addressing the identified challenges and leveraging emerging technologies points towards several promising avenues for future research in VPN traffic analysis.

A. ADVANCED METHODOLOGIES AND LEARNING PARADIGMS

- *Transformer Architectures:* Largely untapped in this domain, Transformers' ability to model long-range dependencies could be highly effective for capturing subtle sequential patterns in packet sizes, timings, or flow metadata characteristic of specific VPN protocols or encapsulated applications. Adapting these models for network traffic data is a key direction.
- *Self-Supervised and Contrastive Learning:* Given the difficulty in obtaining large, accurately labeled VPN datasets, self-supervised approaches that learn representations from abundant unlabeled traffic data are highly appealing. Contrastive methods, learning to distinguish similar vs. dissimilar traffic flows/sessions, could generate robust features resistant to minor variations.
- *Multimodal Learning:* Integrating diverse feature types (*e.g.*, flow statistics, packet sequences, DNS queries, TLS handshake details where available) into unified models could provide a more holistic view and improve robustness compared to single-modality approaches. Exploring effective fusion techniques is crucial.
- *Continual Learning and Concept Drift Adaptation:* Developing models that can incrementally adapt to evolving VPN protocols, application behaviors, and obfuscation techniques without catastrophic forgetting is essential for long-term deployment. Online learning and explicit concept drift detection mechanisms are needed.
- *Adversarial Robustness:* Systematically designing and evaluating classifiers against realistic adversarial obfuscation techniques (padding, timing manipulation, traffic morphing) is critical. Incorporating adversarial training or utilizing inherently robust features/models should become standard practice.

B. NOVEL FEATURE REPRESENTATIONS

- *Advanced Temporal/Spectral Analysis:* Beyond basic time-series, exploring sophisticated techniques like wavelet transforms [51] or other spectral methods could better capture multi-scale temporal characteristics or

frequency-domain fingerprints potentially resistant to simple padding/timing obfuscation.

- *Graph-based Representations:* Further investigation into optimal graph construction (*e.g.*, flow graphs, packet graphs, host communication graphs) and tailored GNN architectures holds significant potential, building on promising initial results [49], [50]. Capturing relational information seems key for certain classification tasks.
- *Behavioral Fingerprinting:* Moving beyond passive statistical features, identifying characteristic protocol-level behaviors or responses of different VPN clients/servers to network events (*e.g.*, connection setup, retransmissions, handling congestion) could provide robust fingerprints.
- *Learned Embeddings:* Developing compact, efficient, yet discriminative embeddings for flows or packets using techniques like autoencoders, variational autoencoders, or language model pre-training (*e.g.*, [43], [44]) could improve both accuracy and computational efficiency.

C. REAL-WORLD DEPLOYMENT AND EVALUATION

- *Efficiency and Scalability:* Focused research on model compression, quantization, knowledge distillation, and efficient architectures suitable for high-throughput network devices and edge computing is paramount for practical adoption. Performance should be reported not just in accuracy but also in throughput, latency, and resource usage.
- *Longitudinal Evaluation and Robustness:* Assessing model performance over extended periods using longitudinal datasets (if available) or through realistic simulations of concept drift is needed to understand real-world robustness and retraining requirements.
- *Standardized Benchmarking and Datasets:* The community urgently needs new, large-scale, diverse, validated, and publicly accessible benchmark datasets reflecting contemporary VPN usage and internet traffic. Establishing standardized evaluation protocols, including common task definitions and metrics, is equally crucial for meaningful progress tracking. Efforts should also focus on generating datasets with reliable ground truth that capture various obfuscation techniques.
- *Integration and Explainability (XAI):* Research on integrating VPN classifiers into existing security frameworks (SIEM, SOAR, Firewalls) and providing interpretable explanations for model predictions using XAI techniques will increase operator trust and enable more effective response actions.
- *Privacy-Preserving Techniques:* Exploring methods like federated learning (training models without centralizing raw traffic), differential privacy (adding noise to protect individual flows), or secure multi-party computation could enable effective VPN analysis while mitigating privacy risks, potentially reconciling the tension between monitoring and confidentiality.

VII. CONCLUSION

The increasing adoption of encryption, particularly through VPNs, has fundamentally altered the landscape of network traffic analysis. While providing essential privacy and security benefits, VPNs obscure traditional visibility methods, creating significant blind spots for network management, security monitoring, and policy enforcement. This challenge necessitates the development of sophisticated techniques to analyze encrypted VPN traffic without relying on payload inspection.

This survey conducted a comprehensive review of research in VPN traffic analysis over the past decade (2016-2025), focusing on three core tasks: VPN presence detection, specific VPN identification, and application identification within VPN tunnels. Our analysis of the surveyed literature revealed significant progress driven by the application of various methodologies, from traditional supervised learning and ensemble methods to advanced deep learning architectures, graph neural networks, and language models. These techniques leverage diverse features, primarily derived from flow statistics, packet sequences, and temporal patterns.

Key findings from our analysis highlight that binary VPN detection is generally a well-addressed problem, with many methods achieving high reported accuracies, suggesting feasibility for deployment in certain contexts. However, the more granular tasks of identifying specific VPN protocols/services and, particularly, classifying the application traffic encapsulated within VPN tunnels, remain substantially more challenging, exhibiting greater performance variability across studies.

Despite methodological advancements, the field faces critical limitations, significantly impacted by dataset availability and quality. The heavy reliance on the aging and publicly documented integrity issues of the ISCXVPN2016 dataset raises concerns about the generalizability and true performance of many reported methods. The scarcity of modern, diverse, and ecologically valid public datasets hinders rigorous benchmarking, reproducibility, and the study of dynamic network conditions and concept drift.

The implications of effective VPN traffic analysis are substantial, offering capabilities to enhance network security through improved policy enforcement and threat detection, and enabling more granular Quality of Service management. However, these capabilities must be balanced against the critical privacy and ethical considerations inherent in analyzing encrypted communications.

Addressing the identified challenges points towards vital future research directions. Methodological advancements leveraging paradigms like Transformers and Self-Supervised Learning, coupled with the exploration of novel features such as sophisticated temporal/spectral properties, graph-based representations, and learned embeddings, are crucial. Equally important is a strong focus on real-world deployment

challenges, including improving efficiency and scalability for high-throughput networks, developing robust models resilient to obfuscation and concept drift, establishing standardized benchmarking, and integrating privacy-preserving techniques to navigate the ethical landscape responsibly.

In conclusion, while considerable strides have been made in analyzing VPN traffic despite the fundamental challenges posed by encryption, significant work remains. Future research must prioritize the development of robust, efficient, and ethically conscious techniques supported by realistic and diverse datasets to provide reliable visibility into the ever-evolving encrypted network environment.

REFERENCES

[1] A. Langley, "The QUIC transport protocol: Design and internet-scale deployment," in *Proc. Conf. ACM Special Interest Group Data Commun.*, Aug. 2017, pp. 183–196, doi: 10.1145/3098822.3098842.

[2] K. Jerabek, K. Hynek, and O. Rysavy, "Comparative analysis of DNS over HTTPS detectors," *Comput. Netw.*, vol. 247, Jun. 2024, Art. no. 110452, doi: 10.1016/j.comnet.2024.110452.

[3] A. F. Gentile, P. Fazio, and G. Miceli, "A survey on the implementation and management of secure virtual private networks (VPNs) and virtual LANs (VLANs) in static and mobile scenarios," *Telecom*, vol. 2, no. 4, pp. 430–445, Nov. 2021, doi: 10.3390/telecom2040025.

[4] I. D. Irsyad and E. Mulyana, "A survey on designs and implementations of virtual private network (VPN)," in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Oct. 2023, pp. 1–6, doi: 10.1109/iceei59426.2023.10346627.

[5] H. Abbas, N. Emmanuel, M. F. Amjad, T. Yaqoob, M. Atiquzzaman, Z. Iqbal, N. Shafqat, W. B. Shahid, A. Tanveer, and U. Ashfaq, "Security assessment and evaluation of VPNs: A comprehensive survey," *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–47, Jul. 2023, doi: 10.1145/3579162.

[6] J. Li, B. Feng, and H. Zheng, "A survey on VPN: Taxonomy, roles, trends and future directions," *Comput. Netw.*, vol. 257, Feb. 2025, Art. no. 110964, doi: 10.1016/j.comnet.2024.110964.

[7] A. Alshalan, S. Pisharody, and D. Huang, "A survey of mobile VPN technologies," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1177–1196, 2nd Quart., 2016, doi: 10.1109/COMST.2015.2496624.

[8] C. Yang, Z. Gu, Y. Wei, G. Xiong, G. Gou, S. Yao, and Y. Yu, "Unidirectional encrypted traffic classification: A survey," in *Proc. Asia–Pacific Conf. Image Process., Electron. Comput. (IPEC)*, Apr. 2024, pp. 702–708, doi: 10.1109/ipec61310.2024.00125.

[9] Z. Sui, H. Shu, F. Kang, Y. Huang, and G. Huo, "A comprehensive review of tunnel detection on multilayer protocols: From traditional to machine learning approaches," *Appl. Sci.*, vol. 13, no. 3, p. 1974, Feb. 2023, doi: 10.3390/app13031974.

[10] W. Dong, J. Yu, X. Lin, G. Gou, and G. Xiong, "Deep learning and pre-training technology for encrypted traffic classification: A comprehensive review," *Neurocomputing*, vol. 617, Feb. 2025, Art. no. 128444, doi: 10.1016/j.neucom.2024.128444.

[11] P. Velan, M. Čermák, P. Čeleda, and M. Drašar, "A survey of methods for encrypted traffic classification and analysis," *Int. J. Netw. Manage.*, vol. 25, no. 5, pp. 355–374, Jul. 2015, doi: 10.1002/nem.1901.

[12] E. Papadogiannaki and S. Ioannidis, "A survey on encrypted network traffic analysis applications, techniques, and countermeasures," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, Jul. 2021, doi: 10.1145/3457904.

[13] S. Rezaei and X. Liu, "Deep learning for encrypted traffic classification: An overview," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 76–81, May 2019, doi: 10.1109/MCOM.2019.1800819.

[14] A. Sharma and A. H. Lashkari, "A survey on encrypted network traffic: A comprehensive survey of identification/classification techniques, challenges, and future directions," *Comput. Netw.*, vol. 257, Feb. 2025, Art. no. 110984, doi: 10.1016/j.comnet.2024.110984.

[15] M. Shen, K. Ye, X. Liu, L. Zhu, J. Kang, S. Yu, Q. Li, and K. Xu, "Machine learning-powered encrypted network traffic analysis: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 791–824, 1st Quart., 2023, doi: 10.1109/COMST.2022.3208196.

[16] Z. Wang, Y. Yang, and Y. Wang, "A survey of encrypted traffic classification: Datasets, representation, approaches and future thinking," in *Proc. IEEE/ACIS 24th Int. Conf. Comput. Inf. Sci. (ICIS)*, Sep. 2024, pp. 113–120, doi: 10.1109/icis61260.2024.10778376.

[17] *ISCXVPN 2016 Datasets Research Canadian Institute for Cybersecurity | UNB— Unb.ca*. Accessed: Mar. 13, 2025. [Online]. Available: https://www.unb.ca/cic/datasets/vpn.html

[18] G. Draper-Gil, A. H. Lashkari, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," in *Proc. 2nd Int. Conf. Inf. Syst. Secur. Privacy*, 2016, pp. 407–414, doi: 10.5220/0005740704070414.

[19] S. Jorgensen, J. Holodnak, J. Dempsey, K. de Souza, A. Raghunath, V. Rivet, N. DeMoes, A. Alejos, and A. Wollaber, "Extensible machine learning for encrypted network traffic application labeling via uncertainty quantification," *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 420–433, Jan. 2024, doi: 10.1109/TAI.2023.3244168.

[20] T. Shapira and Y. Shavitt, "FlowPic: A generic representation for encrypted traffic classification and applications identification," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 2, pp. 1218–1232, Jun. 2021, doi: 10.1109/TNSM.2021.3071441.

[21] Z. Nazari, M. Noferesti, and R. Jalili, "DSCA: An inline and adaptive application identification approach in encrypted network traffic," in *Proc. 3rd Int. Conf. Cryptography, Secur. Privacy*, Jan. 2019, pp. 39–43, doi: 10.1145/3309074.3309102.

[22] P. Gao, G. Li, Y. Shi, and Y. Wang, "VPN traffic classification based on payload length sequence," in *Proc. Int. Conf. Netw. Netw. Appl. (NaNA)*, Dec. 2020, pp. 241–247, doi: 10.1109/NaNA51271.2020.00048.

[23] A. S. Iliyasu and H. Deng, "Semi-supervised encrypted traffic classification with deep convolutional generative adversarial networks," *IEEE Access*, vol. 8, pp. 118–126, 2020, doi: 10.1109/ACCESS.2019.2962106.

[24] Y. Li, F. Wang, and S. Chen, "VPN traffic identification based on tunneling protocol characteristics," in *Proc. IEEE 5th Int. Conf. Comput. Commun. Eng. Technol. (CCET)*, Aug. 2022, pp. 150–156, doi: 10.1109/CCET55412.2022.9906397.

[25] M. Hanlon, G. Wan, A. Ascheman, and Z. Durumeric, "Detecting VPN traffic through encapsulated TCP behavior," in *Free and Open Communications on the Internet*, vol. 2. Stanford, CA, USA: Stanford Univ., 2024, pp. 77–82. [Online]. Available: https://www.petsymposium.org/foci/2024/foci-2024-0016.pdf

[26] J. Kotak, I. Yankelev, I. Bibi, Y. Elovici, and A. Shabtai, "VPN-encrypted network traffic classification using a time-series approach," *IEEE Trans. Netw. Service Manage.*, vol. 22, no. 2, pp. 2225–2242, Apr. 2025, doi: 10.1109/TNSM.2025.3543903.

[27] *VPN/Non-VPN Network Application Traffic Dataset (VNAT) | MIT Lincoln Laboratory— Ll.mit.edu*. Accessed: Mar. 13, 2025. [Online]. Available: https://www.ll.mit.edu/r-d/datasets/vpnnonvpn-network-application-traffic-dataset-vnat

[28] *Encrypted VPN Dataset—zenodo.org*. Accessed: Apr. 10, 2025. [Online]. Available: https://zenodo.org/records/7301756

[29] M. Naas and J. Fesl, "A novel dataset for encrypted virtual private network traffic analysis," *Data Brief*, vol. 47, Apr. 2023, Art. no. 108945, doi: 10.1016/j.dib.2023.108945.

[30] *Dalhousie NIMS Lab VPN 2024 Dataset—IEEE-dataport.org*. Accessed: Apr. 23, 2025. [Online]. Available: https://ieee-dataport.org/documents/dalhousie-nims-lab-vpn-2024-dataset#files

[31] H. Liu, R. Alshammari, and N. Zincir-Heywood, "Edge-cloud VPN traffic analysis over cross platforms," in *Proc. IEEE 10th World Forum Internet Things (WF-IoT)*, Nov. 2024, pp. 1–6, doi: 10.1109/WF-IOT62078.2024.10811218.

[32] J. A. Caicedo-Muñoz, A. L. Espino, J. C. Corrales, and A. Rendón, "QoS-classifier for VPN and non-VPN traffic based on time-related features," *Comput. Netw.*, vol. 144, pp. 271–279, Oct. 2018, doi: 10.1016/j.comnet.2018.08.008.

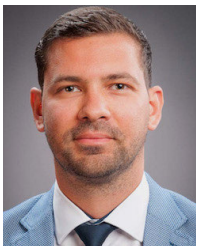
[33] M. Al-Fayoumi, M. Al-Fawa'reh, and S. Nashwan, "VPN and non-VPN network traffic classification using time-related features," *Comput., Mater. Continua*, vol. 72, no. 2, pp. 3091–3111, 2022, doi: 10.32604/cmc.2022.025103.

[34] A. Almomani, "Classification of virtual private networks encrypted traffic using ensemble learning algorithms," *Egyptian Informat. J.*, vol. 23, no. 4, pp. 57–68, Dec. 2022, doi: 10.1016/j.eij.2022.06.006.

[35] J. Fesl and M. Naas, "A complex ML-based approach for virtual private network traffic detection and identification," *KeAi: Int. J. Intell. Netw.*, pp. 1–12, May 2024, doi: 10.2139/ssrn.4828611.

[36] R. Gudla, S. Vollala, K. G. Srinivasa, and R. Amin, "TCC: Time constrained classification of VPN and non-VPN traffic using machine learning algorithms," *Wireless Netw.*, vol. 31, pp. 3415–3429, Mar. 2025, doi: 10.1007/s11276-025-03946-y.
[37] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, "End-to-end encrypted traffic classification with one-dimensional convolution neural networks," in *Proc. IEEE Int. Conf. Intell. Secur. Informat. (ISI)*, Jul. 2017, pp. 43–48, doi: 10.1109/ISI.2017.8004872.
[38] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, "Deep packet: A novel approach for encrypted traffic classification using deep learning," *Soft Comput.*, vol. 24, no. 3, pp. 1999–2012, May 2019, doi: 10.1007/s00500-019-04030-2.
[39] Y. Zeng, H. Gu, W. Wei, and Y. Guo, "*Deep − full − range* : A deep learning based network encrypted traffic classification and intrusion detection framework," *IEEE Access*, vol. 7, pp. 45182–45190, 2019, doi: 10.1109/ACCESS.2019.2908225.
[40] J. Cui, L. Bai, X. Zhang, Z. Lin, and Q. Liu, "The attention-based autoencoder for network traffic classification with interpretable feature representation," *Symmetry*, vol. 16, no. 5, p. 589, May 2024, doi: 10.3390/sym16050589.
[41] M. Seydali, F. Khunjush, and J. Dogani, "Streaming traffic classification: A hybrid deep learning and big data approach," *Cluster Comput.*, vol. 27, no. 4, pp. 5165–5193, Jan. 2024, doi: 10.1007/s10586-023-04234-0.
[42] Y. Chai, Y. Zhu, W. Lin, and D. Li, "Combo packet: An encryption traffic classification method based on contextual information," *Comput., Mater. Continua*, vol. 79, no. 1, pp. 1223–1243, 2024, doi: 10.32604/cmc.2024.049904.
[43] J.-T. Park, C.-Y. Shin, U.-J. Baek, and M.-S. Kim, "Fast and accurate multi-task learning for encrypted network traffic classification," *Appl. Sci.*, vol. 14, no. 7, p. 3073, Apr. 2024, doi: 10.3390/app14073073.
[44] J. Luo, Z. Chen, W. Chen, H. Lu, and F. Lyu, "A study on the application of the T5 large language model in encrypted traffic classification," *Peer Peer Netw. Appl.*, vol. 18, no. 1, pp. 1–13, Nov. 2024, doi: 10.1007/s12083-024-01817-5.
[45] T. M. Tawfeeq and M. Nickray, "Adversarial training for improved vpn traffic classification using efficientnet-b0 and projected gradient descent," *Int. J. Intell. Eng. Syst.*, vol. 18, no. 1, pp. 1200–1215, Feb. 2025, doi: 10.22266/ijies2025.0229.87.
[46] H. Yao, C. Liu, P. Zhang, S. Wu, C. Jiang, and S. Yu, "Identification of encrypted traffic through attention mechanism based long short term memory," *IEEE Trans. Big Data*, vol. 8, no. 1, pp. 241–252, Feb. 2022, doi: 10.1109/TBDATA.2019.2940675.
[47] G. Abbas, U. Farooq, P. Singh, S. S. Khurana, and P. Singh, "Feature engineering and ensemble learning-based classification of VPN and Non-VPN-based network traffic over temporal features," *Social Netw. Comput. Sci.*, vol. 4, no. 5, pp. 1–16, Jul. 2023, doi: 10.1007/s42979-023-01944-5.
[48] B. Liu, S. Kim, and M. Kim, "Adaptive classification of VPN and non-VPN IoT traffic data using CNN with spatio-temporal features," *Asia–Pacific J. Convergent Res. Interchange*, vol. 10, no. 10, pp. 507–519, Oct. 2024, doi: 10.47116/apjcri.2024.10.39.
[49] T.-L. Huoh, Y. Luo, P. Li, and T. Zhang, "Flow-based encrypted network traffic classification with graph neural networks," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 2, pp. 1224–1237, Jun. 2023, doi: 10.1109/TNSM.2022.3227500.
[50] Z. Okonkwo, E. Foo, Z. Hou, Q. Li, and Z. Jadidi, *Encrypted Network Traffic Classification With Higher Order Graph Neural Network*. Cham, Switzerland: Springer, 2023, pp. 630–650, doi: 10.1007/978-3-031-35486-1_27.
[51] Y. S. Razooqi and A. Pekar, "Binary VPN traffic detection using wavelet features and machine learning," 2025, *arXiv:2502.13804*.
[52] K. Lin, X. Xu, and Y. Jiang, "A new semi-supervised approach for network encrypted traffic clustering and classification," in *Proc. IEEE 25th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2022, pp. 41–46, doi: 10.1109/CSCWD54268.2022.9776310.

YASAMEEN SAJID RAZOOQI received the B.Sc. degree from the College of Science, Al-Qadisiyah University, Iraq, in 2011, and the M.Sc. degree from the College of Computer Science and Information Technology, Al-Qadisiyah University, in 2021. She is currently pursuing the Ph.D. degree with the Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary. Her research interests include network event classification, traffic optimization strategies, network routing protocols, energy consumption in wireless sensor networks, and the application of machine learning in networking.



ADRIAN PEKAR (Member, IEEE) received the Ph.D. degree from the Technical University of Kosice, Slovakia, in 2014. He is currently an Associate Professor with the Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary. Before joining academia in Hungary, he gained valuable experience through research, teaching, and engineering roles across Slovakia and New Zealand. His research interests include a wide array of topics, including network and service management, software-defined networking, network function virtualization, network programmability, machine learning, and data science.