Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083669

REVIEW

A Survey of AI-Based Encrypted Traffic Detection: Multi-Level Taxonomy and Structural Analysis of Intent–Behavior–Model Coupling

Yeog Kim, Changhoon Lee and Kiwook Sohn*

Cryptographic and Information Security Laboratory, Seoul National University of Science and Technology, Seoul, Republic of Korea

*Corresponding Author: Kiwook Sohn. Email: kiwook@seoultech.ac.kr

Received: 08 April 2026; Accepted: 29 May 2026; Published: 27 July 2026

ABSTRACT: With the widespread adoption of encryption protocols, payload-based traffic analysis has become increasingly infeasible, posing significant challenges for intrusion detection systems (IDS). Consequently, AI-based approaches for encrypted traffic analysis have gained substantial attention. However, existing studies are often evaluated using inconsistent criteria, including heterogeneous attack labels, behavioral representations, and model architectures, making systematic comparison difficult. To address this limitation, this paper proposes a three-level analytical taxonomy for encrypted traffic analysis, structured around attack objectives (Level 1), observable network behaviors (Level 2), and detection models (Level 3). The proposed framework provides a structured perspective for analyzing how detection objectives, behavioral abstractions, and model design interact under encryption constraints. Based on a systematic analysis of 53 representative studies, this survey examines the relationship between attack objectives, behavior patterns, datasets, evaluation metrics, and AI-based detection models. The analysis indicates that behavioral patterns play an important role in connecting attack objectives with detection models, while also revealing imbalances in the coverage of attack objectives across existing studies. In addition, the survey highlights how dataset selection and evaluation criteria influence the interpretation of model performance in encrypted traffic analysis. Overall, the proposed taxonomy provides a behavior-centric analytical framework for organizing existing encrypted traffic analysis studies and offers insights for future IDS research in encrypted network environments.

KEYWORDS: Encrypted traffic analysis; AI-based intrusion detection; attack objective modeling; behavioral representation; multi-level taxonomy; C2 detection; network traffic classification

1 Introduction

1.1 Background and Motivation

As Internet traffic has increasingly shifted toward encrypted communication, most network traffic is now transmitted using protocols such as TLS, HTTP, VPN, and SSH. This change in the network security environment is largely driven by encryption, which ensures the confidentiality and integrity of user data while simultaneously making it difficult for network security devices to directly analyze packet payloads [1–4]. To overcome this limitation, recent studies have actively explored approaches that analyze encrypted traffic using metadata such as packet length, inter-packet time intervals, traffic direction, and session-level statistics [5–7]. In particular, machine learning and deep learning techniques have demonstrated the ability to learn complex traffic patterns and nonlinear attack behaviors, showing the potential to complement or replace traditional rule-based or signature-based detection methods. These approaches are significant in that they extend beyond simple traffic type classification and can support the identification and

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

response to actual intrusion activities. With the increasing number of such studies, various survey papers on AI-based encrypted traffic detection and classification have been conducted. Existing surveys mainly focus on detection models, feature representations, or overall research trends, and typically analyze how proposed models classify encrypted traffic, or examine studies based on data collection methods, feature representation, and performance evaluation. Some surveys also analyze the limitations of existing research and suggest future research directions [8,9]. Research on encrypted traffic attack detection and classification identifies attacks through observed network behaviors and detection techniques that determine whether such behaviors correspond to malicious activities. Based on the interpretation of these behaviors, it is possible to infer the underlying attack objectives [10]. A single attack objective may be realized through one or more categories of behaviors, and these behaviors can also be identified using various detection methods. Encrypted traffic–based attack detection and classification inherently involve multiple conceptual layers. Encrypted traffic–based attack detection begins with the observation of specific communication patterns in network traffic [11]. These behaviors are identified through detection techniques, and the interpretation of such behaviors allows the underlying attack objectives to be inferred [12]. A single attack objective may manifest through multiple types of behavioral patterns, and the same behavior may also be identified by different detection techniques. Therefore, attack objectives, observable network behaviors, and detection techniques form an interrelated hierarchical structure. In other words, the process of encrypted traffic–based attack detection consists of concepts at different levels. In this paper, existing studies are reorganized and analyzed according to three levels: attack objectives, observable network behaviors, and detection techniques. Based on this framework, we systematically compare how specific attack objectives manifest as network behaviors and how these behaviors are effectively identified through corresponding detection techniques.

1.2 Challenges of Intrusion Detection in Encrypted Traffic

The most fundamental challenge in intrusion detection for encrypted traffic is the lack of payload visibility. As a result, the direct contents related to attack intent and attack behavior cannot be observed, forcing IDS to rely on limited header information and statistical characteristics such as the average packet size and its distribution. In addition, encrypted traffic often shows similar patterns between normal traffic and malicious traffic, which makes anomaly detection difficult. In the case of signature-based detection, only limited metadata or behavior-based signatures can be used, which also imposes constraints [13]. Furthermore, packet data collected in real network environments contains overwhelmingly more normal traffic than malicious traffic [14]. This leads to a class imbalance problem in AI model training for encrypted traffic classification, which results in a bias toward normal patterns [15–18]. Consequently, the performance of detecting malicious packets may be degraded. Deep learning-based IDS have shown high performance. However, it is difficult to reinterpret which features were used to determine an attack, which creates limitations in applying them to real operational environments [19,20].

1.3 Limitations of Existing Surveys

As research on AI-based encrypted network traffic analysis and intrusion detection has increased, numerous survey studies have recently been published. These surveys tend to focus on AI-based models rather than specifically on encrypted network environments when organizing existing studies, addressing how model architectures are designed, what features and feature representation methods are used, and which datasets are employed. In addition, by summarizing studies primarily based on the evaluation methods of each model, they provide an overview of the overall technological development trends in AI-based encrypted traffic analysis as well as useful datasets and the features used in these studies. Existing studies have mainly focused on classifying encrypted traffic according to application types or detecting malicious traffic at the

flow level. In the case of flow-level malicious traffic detection, the maliciousness is determined by analyzing an entire network flow, which leads to the use of statistical features and a focus on post-attack detection. From an attack perspective, analyzing encrypted traffic can be useful for identifying what activities were performed on the network for specific attack objectives and for detecting attacks at an early stage. From this perspective, the following limitations were identified when examining existing survey papers. First, most surveys organize encrypted traffic analysis according to the objectives of existing studies, focusing on application identification, service classification, and protocol recognition, while maliciousness detection mainly relies on flow or session-related features. This approach is limited in explaining how detailed changes in packet flows and temporal patterns, rather than overall flow-level statistics, are associated with specific attack objectives for real-time attack detection and early indication of attacks. Second, existing surveys generally summarize AI-based encrypted traffic analysis models from a technical processing perspective, including data collection, feature representation, modeling, and evaluation, while relatively little attention is given to their connection with the operational context of intrusion detection systems. If encrypted traffic analysis techniques were systematically organized in relation to the functional requirements of IDS, such as signature detection, anomaly detection, and real-time detection, they could be highly useful for practical security operations. Third, it remains difficult to systematically explain how anomaly detection results identified in encrypted traffic are related to actual attack stages or attack objectives. For example, when anomalous patterns are detected, a structured connection is required to clarify what observable behaviors correspond to the anomaly, what attacks were carried out through those behaviors, and what the attacker's intent was behind those actions. Such connections could play an important role in managing false positives, responding to zero-day attacks, and supporting operational environments. Fourth, although the importance of the data imbalance problem and explainable AI (XAI) has been mentioned in several surveys, these issues are often presented mainly as future research directions. In real IDS environments, attack traffic is extremely rare compared to normal traffic, which can lead to learning bias and distortion of detection performance. In addition, the explainability requirements necessary for security analysts to trust and respond to AI-based detection results also need to be more systematically organized. In conclusion, existing survey studies have organized the technological developments in encrypted traffic analysis mainly from the perspective of AI models, focusing on datasets, features, model architectures, and evaluation methods. However, there is still room for improvement in approaches that explicitly connect the relationships among attack objectives, observable behaviors, and detection models, and that reorganize the literature from the perspective of AI-based intrusion detection systems. Table 1 summarizes these limitations and outlines how the proposed taxonomy addresses them. Fig. 1 conceptually compares the classification-oriented structures commonly adopted in existing surveys with the IDS-oriented multi-level taxonomy proposed in this study.

Table 1: How the Level 1–2–3 taxonomy addresses limitations of existing surveys.

Limitation	Resolution in Level 1–2–3 Taxonomy
Classification-oriented bias	Level 1 separates attack detection objectives from general traffic classification, while Level 3 distinguishes architectural choices aligned with detection goals.
Weak IDS perspective	The Level 1–Level 2 linkage restores the IDS pipeline by connecting objectives with observable behaviors, and Level 3 aligns detection strategies accordingly.

(Continued)

Table 1 (continued)

Limitation	Resolution in Level 1–2–3 Taxonomy
Limited packet-level focus	Level 2 expands behavioral patterns to packet-, sequence-, session-, and timing-level representations, and Level 3 compares architectures across these abstraction levels.
Unclear mapping between techniques and attack scenarios	Level 3 systematically maps model architectures to specific objectives (Level 1) and behavioral patterns (Level 2).
Gap between anomaly detection and real attacks	The Level 1–Level 2 structure links anomaly signals to concrete attack intents, while Level 3 clarifies how anomaly detection models operationalize this linkage.
Data imbalance not structurally analyzed	Level 3 enables comparison of datasets, labeling strategies, and learning methods within an intent-oriented analytical layer (Level 1).
Fragmented XAI discussion	The Level 1–Level 3 alignment identifies where explainability is required across objectives, behavioral patterns, and model architectures.
Difficulty identifying research gaps	The Level 1–Level 2 combination space reveals underexplored objective–behavior regions, and Level 3 highlights corresponding architectural underutilization.

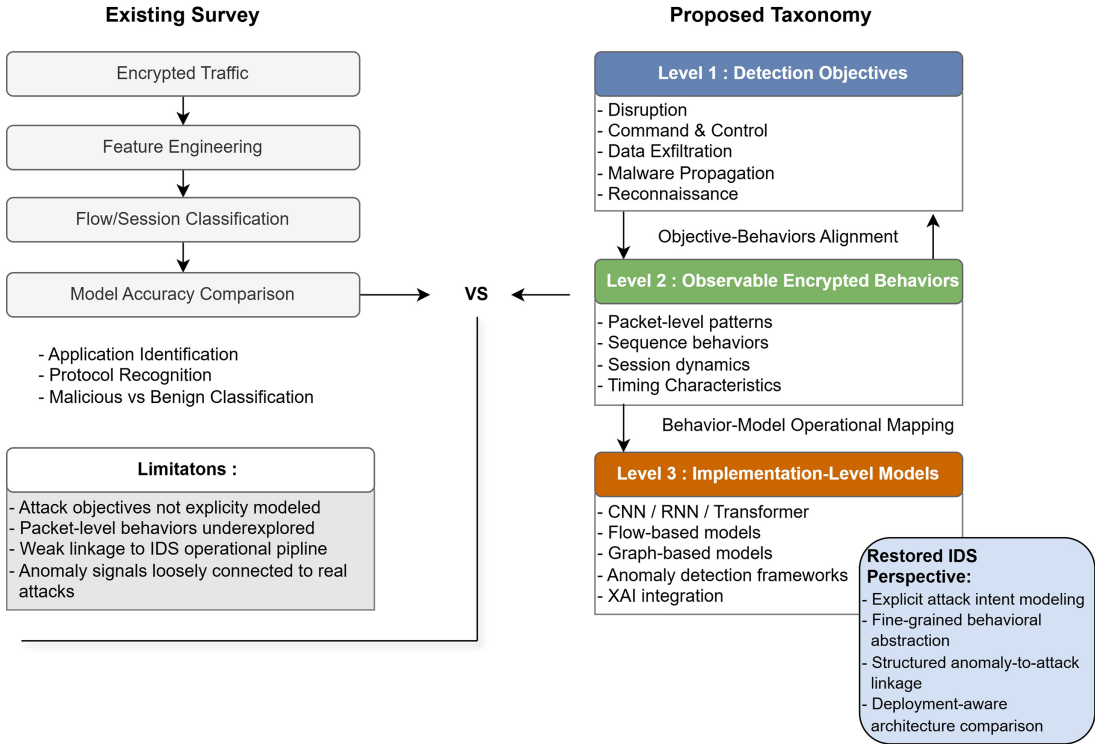

Figure 1: Conceptual comparison between classification-oriented survey-structures and the proposed IDS-oriented multi-level taxonomy.

To address this need, this study introduces a multi-level taxonomy consisting of:

(Level 1) detection objectives,

(Level 2) observable behavioral characteristics in encrypted traffic, and

(Level 3) concrete implementations in prior studies.

Through this three-level-taxonomy, this study aims to systematically reinterpret existing research from the perspective of intrusion detection and attack packets, and to provide an integrated understanding of IDS operational direction in encrypted traffic environments.

1.4 Contribution of This Survey

This paper presents an analytical framework for AI-based encrypted traffic analysis from an attack objective-oriented perspective. It goes beyond classification by revealing how behavioral abstraction mediates the relationship between detection objectives and model design under encryption constraints. The main contributions of this paper are as follows. First, we propose a three-level analytical framework consisting of attack objectives (Level 1), observable network behaviors (Level 2), and detection models (Level 3). This framework formalizes how intrusion detection tasks can be interpreted under encrypted conditions, where payload information is unavailable, and highlights the role of behavioral abstraction in connection detection objectives with model architectures. Second, we identify inconsistencies in how attack labels are defined and used across prior studies, particularly the conflation of attack objectives and behavioral manifestations. To address this, we redefine attack labels based on objectives-oriented categories and systematically map existing studies to these categories, enabling consistent analysis at the level of attack intent.

Third, through a systematic analysis of 53 representative studies, we reveal key structural properties of current research. In particular, we show that detection models are not directly determined by attack objectives, but are mediated by the choice of behavioral representation, which acts as a critical abstraction layer under encryption. Fourth, we provide a comparative analysis of anomaly-based and signature-based approaches in relation to attack objectives and behavioral patterns, highlighting how their effectiveness varies depending on the level of behavioral abstraction. Finally, we discuss practical design implications for real-world IDS deployment in encrypted environments, emphasizing that the selection of behavioral representation is a fundamental factor influencing detection capability. Unlike prior surveys that focus on model performance or feature engineering, this work provides a behavior-centric analytical framework that explains how detection capability is structurally determined under encryption constraints. Based on these insights, we outline a research roadmap for developing more effective AI-based intrusion detection strategies under encryption.

1.5 Organization of the Paper

The remainder of this paper is organized as follows.

Section 2 reviews background concepts related to encrypted traffic analysis, including visibility constraints, detection tasks, and the evolution of AI-based approaches. Section 3 presents the proposed three-level taxonomy, defining attack objectives (Level 1), observable behavioral patterns (Level 2), and their structural relationship, along with implications for intrusion detection system (IDS) design. Section 4 maps existing studies to the proposed taxonomy. After describing the label harmonization and normalization process, we analyze the distribution of studies across Levels 1–3 and identify structural patterns in behavioral representation and model selection. Section 5 provides a comparative structural analysis across the three levels, highlighting cross-level coupling patterns and structural bottlenecks in current encrypted traffic research. Section 6 discusses future research directions toward deeper integration of attack intent,

behavioral abstraction, and detection models. Finally, the paper concludes with a summary of key findings and contributions.

2 Background and Preliminaries

With the widespread adoption of protocols such as TLS, HTTPS, and QUIC, encrypted traffic has become dominant in modern network environments. As a result, traditional payload-based analysis methods used in intrusion detection systems are no longer applicable and research has increasingly focused on inferring threats based on observable network behaviors and metadata under encrypted constraints.

2.1 Encrypted Traffic and Visibility Constraints

As encryption has become widespread, network traffic is increasingly transmitted through encrypted protocols such as TLS/HTTPS, VPN, SSH, and QUIC. This implies that anomaly detection based on traditional payload analysis is no longer feasible. Consequently, network security analysis for encrypted traffic relies on observable metadata and behavioral features, including:

- Packet length and directionality (uplink/downlink)
- Session duration and flow-level statistics
- Inter-packet timing and periodicity
- Destination IP/domain relationships
- Connection frequency and traffic distribution patterns

These features reflect the behavioral characteristics of network traffic rather than its content. Therefore, threat detection in encrypted environments has shifted toward behavior-based analysis, which forms the conceptual foundation of Level 2 (Observable Behavior) in the proposed framework. Fig. 2 presents the conceptual motivation for the proposed taxonomy by illustrating how visibility constraints in encrypted traffic have encouraged model-centric AI research, while the alignment between attack objectives and behavioral representations remains insufficiently explored.

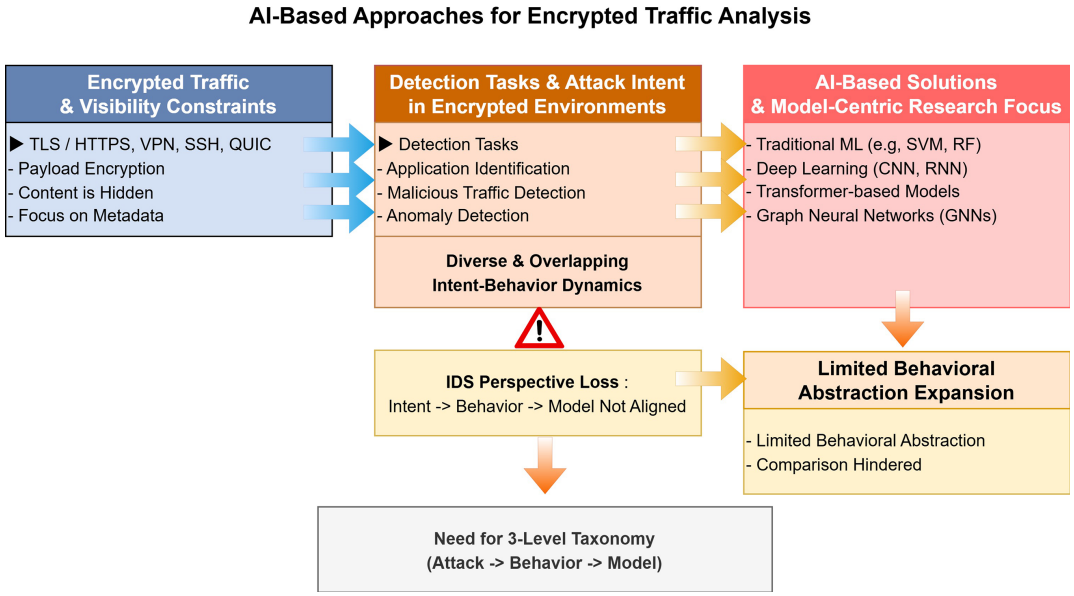

Figure 2: Conceptual framework illustrating how visibility constraints in encrypted traffic have driven model-centric AI research. While architectural sophistication has increased, alignment between attack objectives and behavioral representations remains limited, motivating the need for a three-level IDS-oriented taxonomy.

2.2 Detection Tasks and Intent Modeling in Encrypted Environments

Encrypted traffic analysis encompasses a variety of detection tasks, including:

- Application identification
- Malicious traffic detection (Malware/Botnet/C2 Detection)
- Anomaly Detection
- Circumvention Communication Detection (VPN/Proxy/DoH, etc.)

These tasks do not always correspond directly to attack intent. For Example, application identification does not directly reveal malicious intent, whereas detecting C2 communication corresponds to a clear attack objective [21]. Furthermore, from the perspective of attack intent and network behavior, the same attack objective can manifest through multiple behavioral patterns, while similar behaviors may be observed across different attack objectives. For instance, C2 communication may appear as periodic beaconing or long-lived encrypted sessions, while periodic traffic patterns may also arise in benign applications or other malicious scenarios such as data exfiltration preparation [22–26]. Most existing studies designed detection models based on dataset-specific labels, while systematic classification at the level of attack intent remains l limited. As a result, model-centered comparisons may lead to misleading conclusions, since studies targeting different attack objectives and behavioral patterns are evaluated solely based on model performance. Therefore, it is necessary to explicitly distinguish attack objectives, observable behavior, and detection models, and to analyze the structural relationships among them.

2.3 AI-Based Approaches for Encrypted Traffic Analysis

Encrypted traffic analysis has employed a variety of AI-based techniques. Early studies relied on traditional machine learning methods (e.g., SVM and Random Forest), which later expanded to deep learning–based models such as CNN and RNN [27]. More recent approaches incorporate advanced architectures, including Transformer-based sequence models, Graph Neural Networks (GNN), multimodal fusion, as well as pre-training and contrastive learning techniques [28]. Despite these advancement in model architectures, the input features used to represent network behavior have remained largely unchanged. In most studies, behavioral features are still primarily based on packet length sequences, session statistics, and temporal patterns [29,30]. This suggests that improvements in detection performance have been driven more by model innovation than by diversification of behavioral representations [31–34]. Recent advances in multimodal deep learning have demonstrated the effectiveness of combining heterogeneous feature representations for encrypted traffic classification [13,35]. In addition, explainable AI (XAI) techniques have been increasingly explored to improve the interpretability of detection models in security-critical applications. Recent survey studies have also systematically reviewed XAI approaches for traffic analysis and intrusion detection [36]. However, these approaches primarily focus on improving model performance and interpretability, rather than analyzing the structural relationships between attack objectives, behavioral representations, and detection models. Therefore, it is necessary to analyze encrypted traffic research by hierarchically distinguishing attack objectives (Level 1), observable behaviors (Level 2), and detection models (Level 3) which forms the basis of the analytical framework proposed in this study.

3 Taxonomy of Malicious Traffic in Encrypted Networks

3.1 Conceptual Foundations of the Proposed Taxonomy

3.1.1 Why Payload-Based Classification Is Infeasible in Encrypted Traffic

End-to-end encryption protocols such as TLS 1.3, QUIC, and VPN maintain encryption keys only at the communication endpoints. As a result, payload visibility is not available to network-based intrusion

detection systems, making techniques such as signature-based detection, Deep Packet Inspection (DPI), and content-based rule detection fundamentally infeasible. In addition, due to privacy protection and legal or regulatory constraints, payload content–based analysis is also not permissible [37–39]. Most existing studies use a largely similar feature space. Accordingly, this study classifies prior work by comparing them according to the level of network behavior pattern. For example, if a communication is represented at a low level—such as packet length sequences, packet inter-arrival times, TLS record sizes, or byte sequences—packet-level characteristics can be identified. In contrast, when it is represented at a higher level—such as flow statistics, session duration, burst patterns, or periodicity—session- or flow-level characteristics can be identified [40,41]. Therefore, rather than comparing studies based on the detection models, this work analyzes and compares prior research according to the level of behavioral representation. The attack objectives are then derived from the attack scenarios or datasets targeted in each study. In other words, the taxonomy proposed in this paper is designed based on the perspective of network-based intrusion detection environments. To formalize the proposed taxonomy, we define the relationships among attack objectives, behavioral representations, and detection models as follows.

$$Let\ O = \{o_1, o_2, \ldots, o_n\}, B = \{b_1, b_2, \ldots, b_m\}, and\ M = \{m_1, m_2, \ldots, m_k\} \tag{1}$$

We define a mapping $f: O \rightarrow 2^B$ that associates each attack objective with a set of behavioral representations, and a mapping $g: B \rightarrow 2^M$ that maps each behavioral representation to a set of detection models. Under this formulation, behavioral abstraction (B) acts as an intermediate layer that mediates the relationship between attack objectives (O) and detection models (M). This formulation captures the inherent many-to-many relationships between attack objectives and behavioral representations, as well as between behavioral representations and detection models. This formulation is empirically supported by the observed cross-level mappings in Section 4.

3.1.2 Why Attack Objectives and Behavioral Patterns Are Necessary

Due to payload invisibility in encrypted environments, intrusion detection must rely on behavioral features extracted from network metadata, such as packet size distributions, transmission periodicity, session duration, connection creation frequency, and communication relationships between hosts [42,43]. Although these features do not directly reveal the content of an attack, they manifest as different network behavior patterns depending on the attack objective [3]. Attack objectives represent the ultimate intent that an attacker seeks to achieve. Fig. 3 illustrates the visibility differences between unencrypted and encrypted network traffic and highlights the types of behavioral information that remain available for IDS analysis in encrypted environments.

From a security perspective, representative attack objectives can be categorized into relatively clear forms such as service disruption, maintenance of remote control, data exfiltration, malware propagation, and reconnaissance. In contrast, with respect to attack activities, malware continuously evolves into numerous variants, and accordingly, indicators such as hash values, utilized domains, and encryption methods change rapidly. However, the attack objective itself tends to remain relatively consistent even when attack variants change. This characteristic explains why intent-based classification can generalize to previously unseen or variant attacks. Meanwhile, network behavior patterns describe how attack objectives manifest within encrypted communications. For example, denial-of-service attacks exhibit traffic burst patterns that generate a large volume of traffic within a short period of time, while command-and-control (C2) communication for remote control often takes the form of periodic and low-rate communication. Data exfiltration attacks tend to maintain long-lived connections and transmit data outward, resulting in abnormal uplink traffic ratios. Reconnaissance or malware propagation often involves repeatedly generating numerous short-lived

connections. These behavioral patterns can be observed through network metadata such as packet size, transmission periodicity, connection duration, and traffic directionality, even without access to payload content. By jointly considering attack objectives and behavioral patterns, the taxonomy proposed in this paper causally connects "why such traffic occurs" with "how the attack manifests in encrypted environments." This provides a foundation for systematically analyzing the relationships among attack types, feature extraction methods, and detection model design.

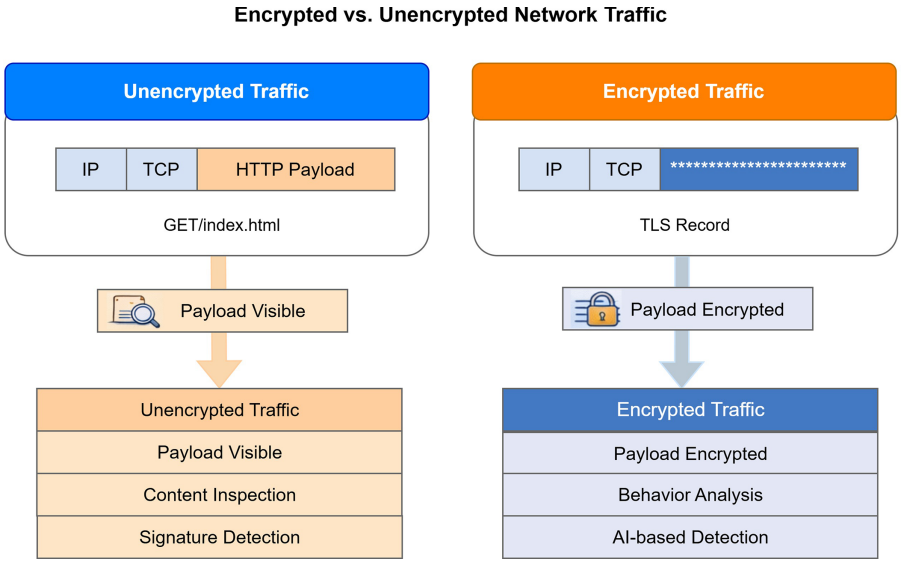

Figure 3: Visibility differences between unencrypted and encrypted network traffic in IDS analysis.

3.1.3 Limitations of Existing Classification Criteria

Existing encrypted traffic analysis studies have primarily classified research based on machine learning or deep learning models. The classification criteria typically organize approaches according to the machine learning or deep learning models used, the feature extraction methods employed, or the encryption protocols targeted for analysis. In addition, a large portion of studies focus on detection performance in terms of binary classification between malicious and benign traffic. However, such classification approaches have limitations in systematically understanding the characteristics of malicious traffic itself. First, model-centered classification provides limited explanatory power regarding how specific attack types are represented through network behaviors. Moreover, in the case of binary classification, different types of attacks are treated as a single malicious traffic category, which restricts the analysis of behavioral differences among attacks and limits the evaluation of generalization capability for previously unseen attacks. In addition, the attack labels used in existing studies vary depending on the datasets or experimental environments. For example, C2-related attacks are labeled as Botnet in the CTU-13 dataset, *Botnet traffic* in the ISCX Botnet dataset, and *Malware communication* in some studies [44,45]. As a result, comparisons across studies may not be properly conducted, and it becomes difficult to derive general conclusions regarding which feature extraction techniques or models are effective under specific attack scenarios. These limitations highlight the need for an integrated classification framework that can reorganize existing studies based on common criteria centered on attack objectives and network behaviors.

3.1.4 Design Principles of the Proposed Taxonomy

Based on the discussions above, the malicious traffic taxonomy proposed in this paper is designed according to the following principles:

1. Encryption-awareness: Assumes payload invisibility and utilizes only network information that can be observed even in encrypted environments [11].
2. Intent-driven abstraction: Adopts attack intent as the highest-level criterion for classifying malicious activities from the structural perspective of *attack objective–attack behavior–detection model* [46,47].
3. Behavioral interpretability: Explicitly defines behavioral patterns through which attack objectives manifest in network traffic, thereby ensuring a clear linkage with feature extraction and detection model design [48–50].
4. Comparability and extensibility: Enables labels and features used across different datasets or studies to be mapped to a common framework while allowing flexible extension of the taxonomy to accommodate new behavioral patterns or detection methods [51–54].

The taxonomy proposed in this paper, designed based on these principles, can serve as a common framework for systematically understanding malicious traffic in encrypted network environments and for comparing and analyzing existing AI-based detection studies.

3.2 Level 1: Attack Objective-Based Classification

When an attacker performs C2 communication using HTTPS and when it is conducted through DNS tunneling, the attack objective remains the same—C2—while the actual communication method differs, resulting in different attack behaviors. Furthermore, in encrypted network environments, the internal contents of packets cannot be directly inspected. Therefore, classifications based on fine-grained technical characteristics, such as specific attack techniques or protocol implementation details, are vulnerable to environmental changes. In this sense, attack methods can vary depending on the tools used, communication channels, packet structures, and encryption mechanisms. In contrast, attack objectives tend to remain relatively stable compared to attack methods. An attack objective represents a higher level of abstraction over attack methods, describing attacks at a conceptual level above individual techniques or observable network behaviors. By organizing attacks at this level, IDS can use attack objectives as a practical basis for assessing the severity of detected threats and determining response priorities.

In general, attack objectives are carried out with goals such as gaining system access, maintaining system control, collecting information, exfiltrating data, or disrupting services. These objectives tend to remain largely unchanged despite technological and environmental evolution. Attack objectives can also be structured according to stages in the cyber attack lifecycle. For example, reconnaissance, intrusion, command and control, data exfiltration, and service disruption represent common objective categories that appear across a wide range of attack scenarios. Accordingly, this paper classifies observable malicious traffic in encrypted network environments into the following five attack objective categories:

- Reconnaissance and Scanning:
- Intrusion and Exploitation
- Command and Control (C2) Communication
- Data Exfiltration
- Disruption and Denial of Service

These categories are defined at a mutually exclusive, high-level objective layer and are designed to comprehensively cover most attack scenarios occurring in encrypted environments. Each category is interpreted in conjunction with observable network behaviors defined in Level 2.

3.2.1 Reconnaissance and Scanning

Reconnaissance and Scanning aim to identify the structure and vulnerabilities of a target system prior to intrusion. It includes attack methods such as port scanning, service discovery, and verification of authentication mechanisms. Since the contents of requests cannot be directly inspected in encrypted environments, such attacks are observed through behaviors such as attempts to connect to multiple destinations, attempts to access multiple ports, connection bursts within a short period of time, an increase in failed connections, and abnormal diversity in access patterns. Although these behaviors may not cause immediate damage, the corresponding attack objective serves as an important early signal indicating the possibility of subsequent intrusion.

3.2.2 Intrusion and Exploitation

Intrusion and Exploitation aims to gain unauthorized access by exploiting system vulnerabilities and to perform actions within the compromised system. This includes remote code execution, authentication bypass, and vulnerability exploitation. In encrypted environments, the exploit payload cannot be directly examined; thus, such attacks are observed through behaviors such as repeated authentication failures, transitions in session states, and changes in connection patterns or connection environments. This stage represents a critical phase in determining the success of a cyberattack. Failure to detect such activities significantly increases the likelihood of subsequent C2 communication and data exfiltration.

3.2.3 Command and Control (C2) Communication

C2 communication refers to the process by which a compromised system communicates with an attacker-controlled server to receive commands or transmit status information. This stage is central to maintaining persistence and enabling further attack propagation. In encrypted environments, C2 traffic often resembles normal HTTPS traffic. Traditionally, it has been observed through behaviors such as periodic communication, repeated connections to a limited set of destinations, and unusually long session durations. However, recent trends include randomization of communication intervals, abuse of legitimate service infrastructures, and intermittent or non-continuous communication patterns to evade detection. Therefore, effective C2 detection requires not reliance on a single indicator but a comprehensive analysis that considers long-term communication patterns, destination relationship structures, and correlations with other attack objectives.

3.2.4 Data Exfiltration

Based on the IDS perspective discussed above, Level 1 classifies encrypted traffic analysis studies according to attack objectives. The definitions of the attack objectives are summarized in Table 2, and the corresponding Level 1 classification framework is presented in Table 3.

Data Exfiltration aims to transfer internally collected information to external entities. In encrypted environments, since the actual content of transmitted data cannot be inspected, inference relies on indirect indicators such as abnormal increases in outbound traffic, changes in session persistence, or concentrated transmission events where large volumes of data are sent within a short time period. Recently, evasion techniques such as low-and-slow exfiltration, masquerading as legitimate services, and blending with C2 communication have become more prevalent. As a result, detection based on single sessions or short-term traffic variations is insufficient. IDS must perform a comprehensive analysis by considering long-term directional changes, communication stability, and correlations with intrusion and C2 stages.

Table 2: Definitions of attack objectives in encrypted traffic environments.

Attack Objective	Intent	Observable Behaviors
Reconnaissance & Scanning	Identify the structure and vulnerabilities of the target system	Port scanning, service discovery, and authentication mechanism verification
Intrusion & Exploitation	Perform actions within the system by gaining unauthorized access through exploiting system vulnerabilities	Remote code execution, authentication bypass, and vulnerability exploitation
Command and Control (C2) Communication	A compromised system communicates with an attacker-controlled server to receive commands or transmit status information	Observable Behaviors: Periodic communication, repeated connections to a limited set of destinations, and unusually long session durations
Data Exfiltration	Exfiltrate internally collected information to external entities	Abnormal increases in outbound traffic, changes in session persistence, and concentrated transmission events where large volumes of data are sent within a short period of time
Disruption and Denial of Service	Degrade the availability of a system or network to disrupt normal service operation	High-rate burst patterns involving a large volume of connection requests or packets within a short period of time

Table 3: Level 1 attack objective-based classification in encrypted environments.

Attack Objective	Primary Security Impact	Typical Observable Indicators (Encrypted)	IDS Implication
Disruption & DoS	Availability degradation	Traffic bursts, resource spikes	Requires rapid mitigation
C2 Communication	Persistent remote control	Periodic connections, stable destinations	Identify compromised hosts
Data Exfiltration	Confidentiality breach	Abnormal outbound trends, long sessions	Detect post-compromise activity
Intrusion & Exploitation	Unauthorized access	Session anomalies, authentication failures	Critical compromise stage
Reconnaissance & Scanning	Information gathering	Repeated short connections, destination diversity	Early-stage warning signal

3.2.5 Disruption and Denial of Service

Disruption and Denial of Service (DoS) aim to degrade system or network availability, thereby preventing normal service operation. Even in encrypted environments, such attacks are observed as high-rate burst patterns involving a large volume of connection requests or packets within a short period. However, recent

attacks increasingly employ low-and-slow or distributed approaches, as well as application-layer techniques, to mimic normal traffic. Therefore, simple increases in traffic volume are insufficient for detection, and it is necessary to analyze long-term resource consumption patterns along with correlations across multiple flows. From an IDS perspective, disruption attacks require immediate response, as delayed detection can directly result in service outages. Collectively, these five attack objectives provide a stable and operationally meaningful abstraction for intrusion detection in encrypted environments. Because identical observable behaviors may correspond to different attack purposes depending on context, objective-level interpretation becomes essential for risk assessment and response prioritization. By abstracting attack intent independently from implementation details, Level 1 establishes a consistent analytical foundation that supports systematic linkage to Level 2 (observable behaviors) and Level 3 (detection models), thereby enhancing both cross-study comparability and practical IDS interpretability.

3.3 Level 2: Observable Network Behavior Patterns

Level 2 provides a classification of network behaviors that can actually be observed by an IDS in encrypted network environments. In environments where the payload is not visible, intrusion detection relies on metadata and behavioral features observable at the network and transport layers, which appear as patterns such as traffic flow, temporal dynamics, and connection behavior. For traffic flow, examples include persistent communication from a specific internal host to a single external IP, or traffic concentrated during specific time periods. Temporal dynamics refer to how traffic changes over time, such as periodically repeated communication or long-lived sessions. Connection behavior describes how connections are established and maintained, with examples including attempts to connect to multiple ports within a short period or a sudden increase in failed connection attempts. In this study, Level 2 is defined not as a classification of individual patterns but as categories that encompass behaviors that can be stably observed in encrypted environments. Specific patterns such as high-rate bursts, long-lived sessions, and periodic low-rate communication can be modified depending on attack strategies or environments; therefore, Level 2 classifies behavioral categories rather than individual patterns. Behavior category–based abstraction allows various variations to be captured while acknowledging that such patterns can change, and relying on a single pattern as an independent criterion limits generalizability. It also enables structural comparison of similarities and differences across different attack objectives (Level 1). The behavioral abstractions used in Level 2 are summarized in Table 4. These categories capture communication characteristics that remain observable despite payload encryption and provide the analytical basis for behavior-driven threat detection.

Accordingly, this study classifies observable behaviors that can be utilized by IDS in encrypted network environments into the following five categories.

- Traffic Rate and Burst Patterns
- Packet Size and Directionality Patterns
- Temporal and Periodicity Patterns
- Connection and Session Behavior Patterns
- Destination and Communication Relationship Patterns

Level 2 categorizes the observable feature space in encrypted environments and provides the practical data dimensions used as input for AI-based detection models. However, these behavioral categories alone cannot uniquely determine attack objectives. The same observable behavior patterns may arise from different malicious intents or even from benign traffic. Therefore, an effective IDS should not rely solely on Level 2 behavioral information, but instead interpret it in conjunction with Level 1 attack objectives through a hierarchical analytical framework. This approach improves detection accuracy in real-world environments and ensures stable applicability even in encrypted network settings.

Table 4: These behavioral abstractions capture communication characteristics that remain observable to network-based intrusion detection systems despite payload encryption and provide the analytical basis for behavior-driven threat detection.

Observable Behavior	Description
Traffic Rate and Burst Patterns	Traffic transmission rate and burst characteristics provide fundamental observable signals in encrypted environments. Sudden high-rate bursts may indicate disruption or DoS activity, while sustained low-rate traffic may appear in stealthy C2 communication or gradual data exfiltration. Detection, therefore, requires analysis of burst persistence, repetition, and temporal distribution rather than instantaneous traffic volume alone.
Packet Size and Directionality Patterns	Packet size distributions and traffic directionality remain observable despite payload encryption. Repeated transmission of similarly sized packets or persistent directional bias, such as abnormal outbound dominance, may indicate malicious communication patterns, including C2 activity or data exfiltration.
Temporal and Periodicity Patterns	Temporal characteristics and communication periodicity provide important cues for identifying automated attacks and persistent communication behaviors. Regular connection intervals, abnormal session durations, or time-of-day concentration patterns may indicate coordinated malicious activity.
Connection and Session Behavior Patterns	Connection establishment patterns and session dynamics provide indirect signals of malicious activity. Indicators may include repeated connection attempts, abrupt session transitions, abnormal session persistence, or unusual authentication or handshake behaviors.

3.4 Relationship between Attack Objectives and Behavior Patterns

In encrypted network environments, the relationship between attack objectives and observable behaviors inherently exhibits a many-to-many structure. A single attack objective may manifest through different behavioral patterns depending on attack strategies, network conditions, and operational environments. Conversely, identical or similar behavioral categories may be observed across different attack objectives. Therefore, there are limitations in deterministically identifying attack objectives based solely on observable behaviors, and probabilistic and interpretive inference is required by considering multiple sources of information and contextual factors. This structure highlights the limitations of intrusion detection approaches that rely on single behavioral indicators or specific patterns. In encrypted environments, since payload information is not accessible, observable behaviors function not as direct indicators of attack intent but as indirect evidence suggesting the possibility of an attack. Accordingly, effective intrusion detection should be understood as a hierarchical inference process in which possible attack objectives are first derived from observed behaviors through the interpretation of behavioral combinations and patterns, and then progressively narrowed down. To systematically explain this relationship, this study separates and defines attack objectives (Level 1) and observable network behavior categories (Level 2), and proposes a taxonomy that interprets the relationship between the two levels from a relation-centric perspective.

3.4.1 Conceptual Relationship between Level 1 and Level 2

In encrypted traffic environments, attack objectives (Level 1) and observable network behavior categories (Level 2) are analytically distinct in intrusion detection, yet they are semantically connected in that they describe the same attack at the objective level and the behavioral level, respectively. Since the payload cannot be directly inspected in encrypted environments, attack objectives cannot be directly observed. Therefore, IDS must infer Level 1 through Level 2 behavioral information such as connection patterns, periodicity, traffic directionality, and session persistence. In addition, a single attack objective may be manifested through multiple behavioral categories depending on attack strategies, network conditions, and operational environments, while the same behavioral category may appear across different attack objectives. Accordingly, Level 1 and Level 2 are mapped in a many-to-many relationship. Fig. 4 illustrates the conceptual relationship between attack objectives (Level 1) and observable network behavior categories (Level 2) in encrypted traffic environments. Behavioral categories highlighted in bold indicate representative characteristics that are generally strongly associated with specific attack objectives. However, the relationship between observable behaviors and attack objectives is inherently probabilistic and interpretive. For example, a sudden increase in traffic rate may strongly indicate a disruption attack. However, without additional contextual information, the same pattern may also appear under certain conditions, making it difficult to distinguish from data exfiltration or C2 communication. Likewise, periodic communication and persistent sessions are typical characteristics of C2 activity, but such behaviors may also occur in benign applications or other attack scenarios. Therefore, observable behaviors do not directly determine attack objectives and must be interpreted by considering additional contextual information together with other behavioral evidence.

3.4.2 Mapping Matrix between Attack Objectives and Observable Behaviors

Table 5 provides a more structured representation of the relationship between attack objectives (Level 1) and observable network behaviors (Level 2) that is conceptually illustrated in Fig. 4. Specifically, Table 5 presents a mapping matrix between Level 1 attack objectives and Level 2 behavioral categories. It is derived from cross-analysis of 53 prior studies and reflects recurring behavioral associations rather than deterministic rules. Each cell distinguishes between cases in which a behavioral category is observed as a relatively dominant characteristic of a specific attack objective and cases in which it may be observed depending on attack strategies or environmental conditions. This matrix does not imply that a single behavioral pattern corresponds deterministically to a particular attack objective. Rather, it functions as an analytical tool for structuring the interpretive space between observable behaviors and attack objectives in encrypted environments. Through this matrix, IDS can more systematically consider the range of possible attack objectives when a particular behavioral category is observed, while avoiding definitive judgments based solely on a single behavioral indicator. In addition, the matrix shows that a single attack objective may be associated with multiple behavioral categories simultaneously, while the same behavioral category may commonly appear across different attack objectives. These characteristics emphasize that intrusion detection in encrypted environments should be based on multidimensional behavioral interpretation.

The symbol • (Strong association) indicates that a behavioral category and an attack objective have been repeatedly observed and described as dominant characteristics in prior studies, while △ (Context-dependent) denotes that the observed behaviors vary depending on attack strategies or environmental conditions across the literature.

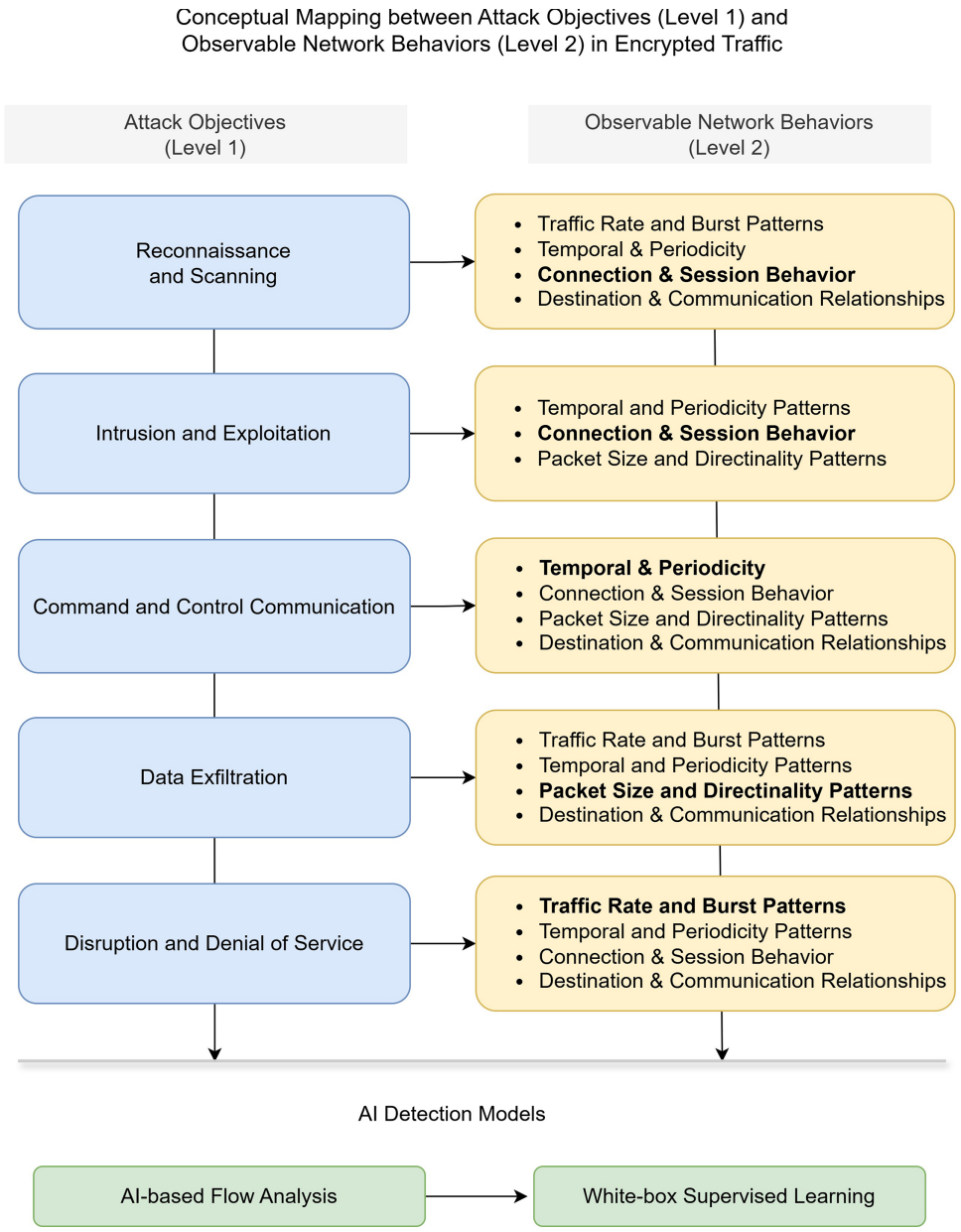

Figure 4: Conceptual taxonomy illustrating the relationship between attack objectives (Level 1) and observable network behavior categories (Level 2) in encrypted traffic environments. Bold categories indicate the most common behaviors for each attack objective.

Table 5: Mapping matrix between attack objectives and observable network behavior categories.

Attack Objective (Level 1)	Traffic Rate & Burst	Packet Size & Directionality	Temporal & Periodicity	Connection & Session Behavior	Destination & Relationship
Reconnaissance & Scanning	△	△	△	●	●
Intrusion & Exploitation	△	△	△	●	△

(Continued)

Table 5 (continued)

Attack Objective (Level 1)	Traffic Rate & Burst	Packet Size & Directionality	Temporal & Periodicity	Connection & Session Behavior	Destination & Relationship
Command & Control (C2)	△	●	●	△	●
Data Exfiltration	△	●	△	△	●
Disruption & DoS	●	△	△	●	△

Note: ● Strong association; △: Context-dependent association.

3.4.3 Implications for IDS Design and Analysis

In encrypted network environments, attack objectives cannot be directly observed and must be inferred from observable behavioral information. Observable behaviors, in turn, can only be meaningfully interpreted when analyzed in relation to attack objectives. For example, periodic communication, long-lived sessions, and burst traffic are, in isolation, merely network phenomena. However, determining whether such behaviors indicate C2 communication, data exfiltration, or benign traffic requires interpreting them in connection with attack objectives. In other words, observable behaviors are ambiguous without the context of attack objectives, while detecting attack objectives without observable behaviors is not feasible. Therefore, Level 1 and Level 2 can be considered closely related. Fig. 4 and Table 5 demonstrate that intrusion detection in encrypted network environments is not a simple rule-based classification problem, but rather an inference problem that requires interpreting multidimensional combinations of behavioral features. Observable network behaviors do not directly reveal attack intent; instead, they serve as indirect indicators suggesting possible attack objectives. Accordingly, effective intrusion detection requires a hierarchical analytical structure that integrates multiple behavioral categories observed at Level 2 and interprets them in relation to Level 1 attack objectives. This approach overcomes the generalization limitations of single-pattern or threshold-based rule detection and motivates the design of hierarchical AI models that capture the relationship between behaviors and attack objectives. Furthermore, the relational structure presented in this section can be extended in future work to model the relative importance of behavioral categories with respect to attack objectives or to analyze how these associations evolve over time. Such extensions can enhance the ability to detect previously unseen or evolving attacks in encrypted network environments.

Finally, the proposed Level 1–Level 2 structure provides a conceptual basis for reinterpreting existing encrypted traffic analysis studies from the perspective of the relationship between attack objectives (Level 1) and observable behaviors (Level 2). This enables prior work to be reorganized not merely in terms of features or models used, but in terms of which attack objectives are primarily addressed and which behavioral categories are repeatedly utilized for their detection. This perspective can be used to identify research concentrations as well as relatively underexplored areas, and to provide a reference framework for future detection strategy development. In the next section, we revisit existing studies based on the proposed taxonomy and analyze them with a focus on the relationship between attack objectives and observable behaviors. Through this analysis, we discuss directions for improving attack detection in encrypted traffic environments.

4 Mapping Existing Studies to the Proposed Taxonomy

4.1 Paper Selection and Analysis Framework

This survey collected AI-based encrypted traffic analysis studies published between 2023 and February 2026 through a systematic search and selection process. The objective of this procedure is to ensure

reproducibility and methodological transparency, while also including a balanced set of studies that demonstrate representativeness and academic rigor in the fields of encrypted traffic analysis and security detection.

4.1.1 Paper Search Strategy

The literature search was conducted using major academic databases and conference proceedings, including IEEE Xplore and Springer. The search period was limited to publications from January 2023 to February 2026. The search was performed by combining the following three groups of keywords using the AND operator:

1. "encrypted traffic" OR "TLS" OR "HTTPS" OR "QUIC."
2. "deep learning" OR "machine learning" OR "transformer" OR "GNN."
3. "classification" OR "detection" OR "malware" OR "botnet."

The search results were initially screened based on abstracts and main content. After removing duplicate papers, the final set of studies was selected according to predefined inclusion and exclusion criteria.

4.1.2 Inclusion and Exclusion Criteria

The criteria used for selecting papers for the final analysis are summarized in Table 6.

Table 6: Criteria for selecting papers to be analyzed.

Inclusion Criteria	Exclusion Criteria
Studies that analyze encrypted network traffic, such as TLS, HTTPS, and QUIC	Studies that focus only on unencrypted (plaintext) traffic
Studies that employ machine learning or deep learning–based detection/classification methods	Studies on pure network optimization or QoS management
Studies that clearly present observable network behavior modeling or model structures	Studies focusing on the design of encryption protocols themselves
Studies that include experimental evaluation	Proposal-based studies without experimental validation

After applying these criteria and completing the screening process, 53 papers were selected for the final analysis. Fig. 5 presents the PRISMA-style workflow used for literature search, screening, and study selection, resulting in the final set of 53 studies included in this survey.

These studies were subsequently mapped to the proposed three-level taxonomy consisting of attack objectives (Level 1), observable behavioral representations (Level 2), and detection models (Level 3). The complete list of the analyzed papers is provided in Appendix A.

Each selected study was analyzed using the three-level taxonomy proposed in Section 3. This framework decomposes encrypted traffic detection research into three analytical dimensions: attack objectives (Level 1), observable behavioral representations (Level 2), and detection techniques (Level 3). The purpose of this structure is to separate detection intent, behavioral evidence, and model implementation so that heterogeneous studies can be compared within a unified analytical framework.

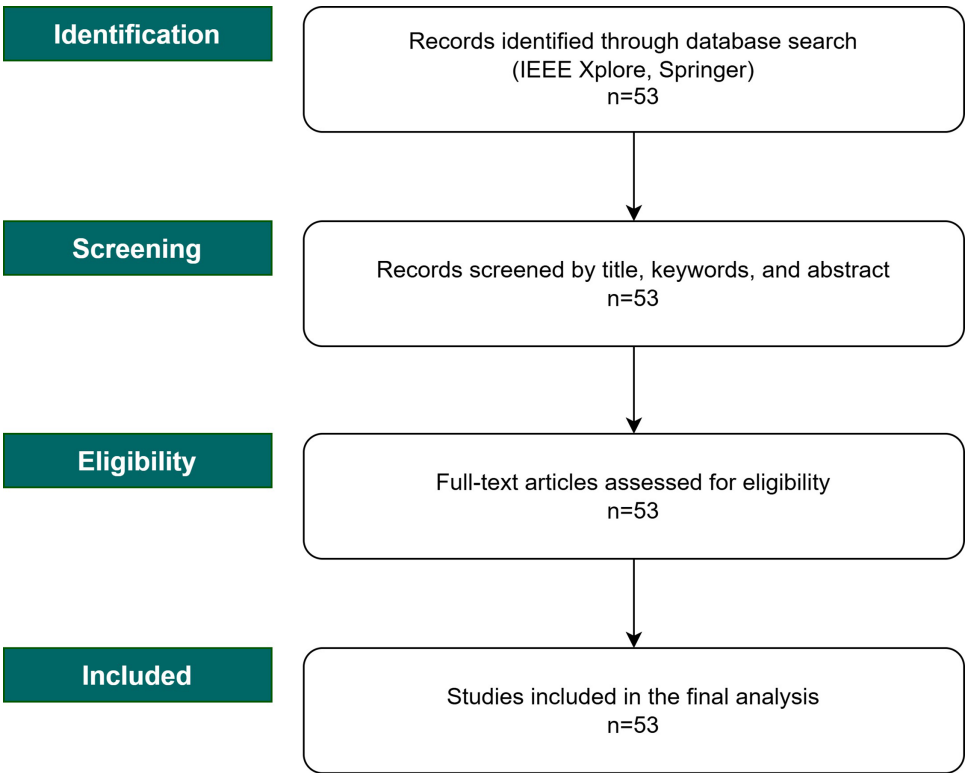

Figure 5: RISMA-style workflow illustrating the literature search and screening process used in this survey. A total of 53 studies were identified and included in the final analysis.

At Level 1, the primary detection or classification objective of each study was normalized into one of the predefined attack objective categories:

- Reconnaissance and Scanning
- Intrusion and Exploitation
- Command and Control (C2) Communication
- Data Exfiltration
- Disruption and Denial of Service

Studies focusing solely on application identification without explicit attack context were categorized as Non-Attack-Oriented, since their objectives are not directly related to attack detection. At Level 2, each study was mapped according to the observable behavioral characteristics represented by its input features rather than the specific model architecture used. The behavioral categories include:

- Traffic Rate and Burst Patterns
- Packet Size and Directionality Patterns
- Temporal and Periodicity Patterns
- Connection and Session Behavior Patterns
- Destination and Communication Relationship Patterns

Importantly, Level 2 classification is based on behavioral representation. For example, packet-length sequences modeled using a Transformer are mapped to the Packet Size and Directionality category because the underlying input representation captures packet-size-related behavior. At Level 3, studies were

categorized according to the AI model architecture and learning strategy employed for detection. The categories include:

- CNN/RNN-based models
- Transformer-based models
- Graph Neural Network-based models
- Pre-trained/Foundation models
- Generative models
- Unsupervised anomaly detection

To ensure consistent cross-study mapping, each selected study was processed using the following structured procedure:

1. Identify the stated detection or classification objective in the study.
2. Extract explicit attack labels and map them to the corresponding Level 1 attack objective category.
3. If no attack objective is specified (e.g., application identification or generic traffic classification), label the study as Non-Attack-Oriented.
4. Analyze the input features and representations and map them to the appropriate Level 2 observable behavior category.
5. Map the model architecture and learning strategy to the corresponding Level 3 detection technique category.

This structured mapping procedure enables systematic cross-level comparison of prior research across attack objectives, observable behaviors, and detection techniques, thereby providing a unified analytical perspective on AI-based encrypted traffic detection studies. Using the 53 studies identified through the PRISMA-style screening process, each study was systematically mapped to the proposed three-level analytical framework described below.

4.2 Harmonization Aper Selection and Analysis Framework

Existing studies on attack detection and classification in encrypted network traffic have been conducted under heterogeneous conditions, including differences in datasets, threat models, and experimental environments. As a result, the naming and classification criteria of attack labels are inconsistently defined across studies. For example, some studies use broad terms such as "Malware Traffic," "Botnet," "DoS," and "Anomaly," while others employ more fine-grained labels such as "C2 Communication," "Data Leakage," "Port Scan," and "Brute Force." These inconsistencies make direct comparison across studies difficult and hinder systematic analysis of how frequently specific attack objectives have been addressed. To address this issue, this study first organizes the attack labels used in prior work and then normalizes them according to the proposed three-level taxonomy, mapping them into higher-level concepts (Level 1: Attack Objectives) and behavior-based categories (Level 2: Observable Network Behavior Categories). This section describes the normalization process and the principles applied in this procedure.

4.2.1 Extraction of Original Attack Labels

First the following information was identified from each selected paper:

1. The attack classes or label names explicitly defined in the paper
2. The attack categories included in the dataset used for experiments
3. The description of the attack definition

During this process, it was found that different terms were often used for the same attack type. For example:

- "Botnet Traffic," "C2 Channel," "Backdoor Communication."
- "Data Leakage," "Exfiltration," "Information Stealing."
- "DoS," "DDoS," and "Service Disruption."

These labels conceptually refer to similar attack objectives, but are expressed differently at the label level (Level 3). In this study, such original labels are defined as "Original Attack Labels" and are used as the starting point for the subsequent normalization process.

4.2.2 Normalization to Level 1: Mapping to Attack Objectives

The first step of normalization is to map the attack labels defined in prior studies to the Level 1 attack objectives proposed in this study. This mapping is performed based on the ultimate security threat objective that the attack aims to achieve. For example, when prior studies use broad labels such as "Malware Traffic" or "Anomaly", the attack objective is inferred from the observed behaviors. If periodic communication patterns are observed, the attack may be interpreted as C2 activity, whereas sustained high-volume outbound traffic may indicate Exfiltration. In this way, Level 1 categorization is performed according to attack intent rather than the original label name itself.

Representative examples of the mapping between original attack labels and the normalized Level 1 attack objective categories are presented in Table 7.

Table 7: Examples of mapping original attack labels used in prior studies to the normalized Level 1 attack objective categories defined in the proposed taxonomy.

Original Attack Label (Prior Studies)	Normalized Level 1 Attack Objective
DoS/DDoS	Disruption and Denial of Service
Botnet/C2 Channel	Command and Control (C2) Communication
Data Leakage/Exfiltration	Data Exfiltration
Exploit/Brute Force	Intrusion and Exploitation
Port Scan/Network Scan	Reconnaissance and Scanning

This process was conducted by comprehensively considering the experimental settings and threat model descriptions of each study, and its purpose is not to modify or reinterpret the original experimental results of prior work, but rather to add a standardized interpretative layer for comparison.

4.2.3 Alignment with Level 2: Mapping to Observable Network Behavior Categories

After Level 1 normalization, the input features and behavioral representations used in each study were analyzed and aligned with the Level 2 observable behavior categories defined in Section 3. Level 2 was constructed not based on attack objectives, but on the behavioral characteristics actually used in the experiments presented in prior studies. Representative features used in prior work to observe behavioral patterns include the following:

- Packet length mean
- Upload/download ratio
- Session duration
- Inter-arrival time
- Repeated destination access

However, simply listing such features makes direct comparison across studies difficult. Therefore, this study realigned these individual features into higher-level behavioral categories. For example:

- Packet length distributions and upload/download ratios → Packet Size and Directionality Patterns
- Session duration and connection frequency → Connection and Session Behavior Patterns
- Periodic communication interval analysis → Temporal and Periodicity Patterns
- Repeated connections to identical destinations → Destination and Communication Relationship

The reconstructed behavioral categories are as follows:

- Traffic Rate and Burst Patterns
- Packet Size and Directionality Patterns
- Temporal and Periodicity Patterns
- Connection and Session Behavior Patterns
- Destination and Communication Relationship Patterns

4.2.4 Ambiguity and Multi-Label Handling Principles

When specific attack objectives were not explicitly indicated, such as in labels like "Malicious Traffic" or "Anomaly", the most appropriate Level 1 attack objective category was assigned by comprehensively reviewing the experimental settings, dataset composition, and threat model descriptions provided in the paper. Only when the attack objective could not be reasonably identified was the label classified as "Generic Malicious Activity". In some studies, a single attack was associated with multiple observable behavioral categories, or a single model was designed to detect multiple attack objectives simultaneously. Cases were also identified in which a single attack was related to multiple Level 2 behavioral categories. In such cases, this study adopted a many-to-many mapping strategy based on the following principles:

- When a single Original Label was related to multiple Level 1 objectives, the label was mapped to all relevant attack objectives.
- When a study used multiple Level 2 behavioral categories, all applicable categories were recorded in order to preserve the multidimensional relationship between attack objectives and observable behaviors.

These processing principles reflect the conceptual framework presented in Section 3, in which the relationship between attack objectives and observable behaviors is inherently non-deterministic. This provides the interpretative basis for the mapping results presented in Section 4.3.

4.3 Distribution of Studies across Attack Objectives (Level 1)

This section presents the distribution of the selected 53 studies across the Level 1 attack objectives defined in Section 3. The purpose of this analysis is not to redefine the Level 1 categories, but to examine how existing research has been structured in terms of attack intent.

4.3.1 Classification Results Based on Level 1

After normalization, the 53 selected studies were categorized according to the Level 1 attack objective taxonomy defined in Section 3. Table 8 summarizes the number of studies in each Level 1 category together with their corresponding paper identifiers (P01–P53). Studies primarily focused on application identification do not explicitly model attack intent and are therefore categorized as Non-Attack-Oriented.

Table 8: Distribution of selected studies across Level 1 attack objectives.

Level 1 Category	Count	Paper IDs
Non-Attack-Oriented	36	P01, P03, P04, P06, P07, P09, P10, P11, P17, P19, P20, P22, P23, P24, P25, P26, P28, P29, P30, P31, P32, P33, P34, P35, P36, P37, P38, P39, P43, P44, P45, P46, P50, P51, P53
Command and Control (C2) Communication	12	P02, P08, P12, P14, P15, P16, P18, P21, P27, P41, P42, P52
Reconnaissance and Scanning	2	P13, P40
Data Exfiltration	1	P05
Disruption and Denial of Service	2	P48, P49

The full mapping between paper identifiers and paper titles is provided in Appendix A. This classification provides the basis for analyzing the distribution of research attention across different attack objectives.

4.3.2 Distribution and Structural Implications of Level 1 Classification

The distribution of the 53 selected studies across Level 1 attack objectives is summarized as follows:

- Non-Attack-Oriented: 67.9% (36 studies)
- Command and Control (C2) Communication: 22.6% (12 studies)
- Reconnaissance and Scanning: 3.8% (2 studies)
- Data Exfiltration: 1.9% (1 study)
- Disruption and Denial of Service: 3.8% (2 studies)

Non-attack-oriented studies constitute the majority of the literature. This indicates that a substantial portion of recent encrypted traffic research focuses on application identification rather than explicit attack intent modeling. In other words, the field appears to have emphasized traffic classification performance and model design ("how to detect") more strongly than structured attack-objective representation ("what to detect"). Among attack-oriented studies, Command and Control (C2) communication detection accounts for the largest share. This may reflect the relatively stable behavioral characteristics of C2 communication in encrypted environments and its central role in post-compromise attack stages.

In contrast, only one study explicitly focused on Data Exfiltration as a primary attack objective, and none of the selected studies explicitly centered on Disruption and Denial of Service within the defined timeframe. This does not imply that these attack objectives lack importance; rather, it suggests that intent-driven modeling of these objectives in encrypted traffic contexts remains comparatively limited within the analyzed corpus. Similarly, studies addressing Reconnaissance or Intrusion and Exploitation constitute a relatively small proportion of the dataset. This may be associated with the inherent difficulty of reliably identifying early-stage or exploit-driven attacks using only observable encrypted traffic behaviors. Overall, the observed distribution provides a structural overview of how recent encrypted traffic research is organized across attack objectives. This distributional perspective establishes the foundation for further examining how attack intent is operationalized through observable behavioral abstractions in encrypted environments. The distribution of attack objectives across the 53 selected studies reveals a clear imbalance in research focus. A substantial proportion of studies fall under the Non-Attack-Oriented category, primarily focusing

on generic traffic classification, malware detection, or encrypted traffic analysis without explicitly modeling attack intent. Although AI-based detection techniques are widely applied, explicit modeling of attack objectives remains comparatively underexplored. This suggests that the research community has prioritized performance optimization and classification accuracy over structured, intent-aware threat modeling. This imbalance further justifies the necessity of the proposed Level 1–2–3 taxonomy framework.

4.3.3 Distinction between Attack-Oriented and Non-Attack-Oriented Studies

Level 1 classification results indicate that the selected studies can be broadly divided into two groups: attack-oriented studies and non-attack-oriented studies. Attack-oriented studies explicitly define and detect specific attack objectives, whereas non-attack-oriented studies focus on application identification, traffic classification, or general anomaly detection without assuming a specific attack intent. Non-attack-oriented studies often emphasize representation learning and model performance in encrypted environments rather than attack objectives themselves. This distinction suggests that prior studies differ not only in their targets, but also in whether their design is driven primarily by attack intent or by feature/model optimization. Such a distinction provides an important basis for interpreting the relationship between research objectives, behavioral patterns, and detection models in the subsequent Level 2 and Level 3 analysis.

4.3.4 Significance of Level 1 Normalization

Level 1 classification is not simply a process of listing studies by attack type, but rather a normalization procedure that reorganizes prior work under a unified attack-objective framework despite differences in datasets and label definitions. Through this process, heterogeneous attack labels used across prior studies are transformed into a common attack-objective structure, making cross-study comparison possible. This allows us to identify which attack objectives have been repeatedly studied in encrypted traffic environments and which remain relatively underexplored.

Furthermore, the results of this section provide the basis for the Level 2 behavioral analysis presented in Section 4.4, enabling structural examination of the relationship between attack objectives and observable network behaviors. This contributes to validating the hierarchical analytical framework proposed in this study for encrypted traffic intrusion detection.

4.4 Distribution of Observable Behavioral Representations (Level 2)

Level 2 refers to the observable behavioral characteristics actually utilized by AI-based models in encrypted environments. In this section, based on the Level 2 categories defined earlier, the 53 selected studies are analyzed to identify which observable network behaviors were used as input features for attack detection.

4.4.1 Overall Distribution of Level 2 Behavioral Categories

The Level 2 taxonomy introduced in this study categorizes observable behavioral representations into five. Based on the Level 2 behavioral categories defined in this study, the input features used in each paper were analyzed and mapped to the corresponding categories. Since many studies employed multiple behavioral categories simultaneously, overlapping assignments were allowed. The results are summarized as follows:

- Packet Size and Directionality Patterns: 6 studies
- Connection and Session Behavior Patterns: 26 studies
- Temporal and Periodicity Patterns: 24 studies
- Destination and Communication Relationship Patterns: 4 studies
- Traffic Rate and Burst Patterns: 3 studies

These results indicate that the behavioral representations used in encrypted environments are concentrated primarily on statistical and temporal characteristics observable at the session and flow level. In particular, Connection and Session Behavior Patterns (26 studies) and Temporal and Periodicity Patterns (24 studies) account for the largest share, showing that most prior work has relied on session duration, the number of connections, packet inter-arrival time, and periodic communication intervals as key behavioral indicators. This suggests that, because payload contents cannot be directly inspected in encrypted environments, attack detection has been largely based on session-level stability and time-dependent behavioral characteristics. In contrast, Packet Size and Directionality Patterns (6 studies), Destination and Communication Relationship Patterns (4 studies), and Traffic Rate and Burst Patterns (3 studies) were used much less frequently. This implies that packet size and directionality features, long-term relationship-based features such as repeated host–destination communication or destination diversity, and traffic rate or burst-based indicators have played a relatively limited role in prior encrypted traffic analysis studies. Such underuse may be attributed to the fact that these features often require more specific attack conditions, longer observation windows, or additional contextual information about communication relationships.

4.4.2 Level 1–Level 2 Cross-Mapping Matrix

To analyze the structural relationship between attack objectives (Level 1) and observable behavioral representations (Level 2), a cross-mapping matrix was constructed based on the 53 analyzed studies. Fig. 6 shows the distribution of the analyzed studies across the Level 1 attack objective categories.

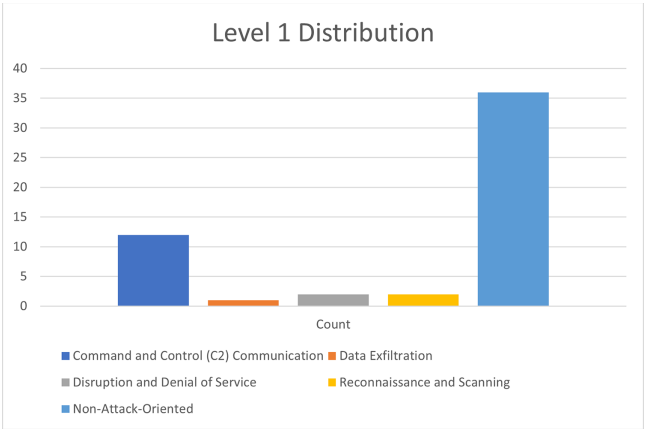

Figure 6: Distribution of studies across Level 1 attack objectives.

The matrix summarizes how frequently specific behavioral representations appear in studies targeting each attack objective category. Table 9 presents the mapping results between Level 1 attack objectives and Level 2 behavioral representations. Fig. 7 provides a heatmap visualization of this relationship, where darker cells indicate a greater number of studies utilizing the corresponding behavioral representation.

Because some studies employ multiple behavioral representations, each representation was counted independently, and therefore a single study may contribute to multiple Level 2 categories. The matrix highlights several notable patterns.

Table 9: Cross-mapping matrix between Level 1 attack objectives and Level 2 observable behavioral representations across the analyzed studies. Multiple representations per study are counted independently.

Level 1	Level 2				
	Packet & Direction	**Traffic Rate & Burst**	**Temporal & Periodicity**	**Connection & Session Behavior**	**Destination & Communication Relationship**
C2 Communication	0	1	9	2	0
Data Exfiltration	0	0	0	0	1
Disruption/DoS	0	0	1	1	0
Reconnaissance	0	1	1	0	0
Non-Attack-Oriented	6	1	9	10	1

Note: 5 Level 2 categories only. Multiple representations per study are counted independently.

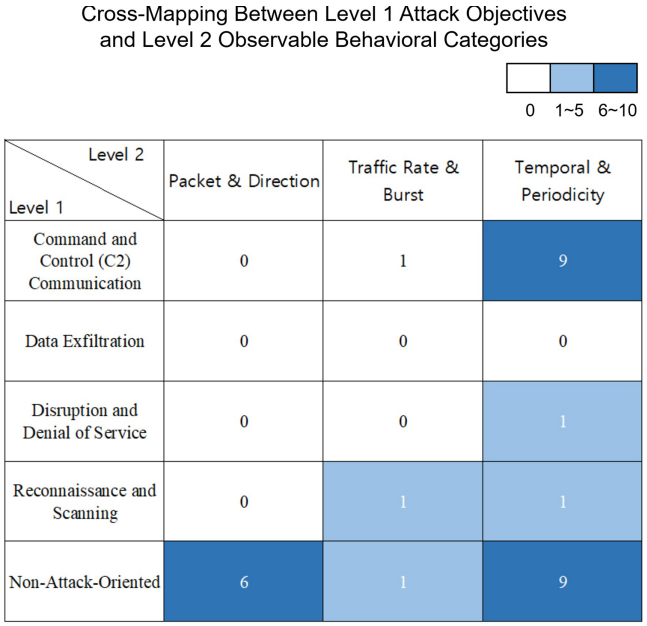

Figure 7: Cross-mapping heatmap between Level 1 attack objectives and Level 2 observable behavioral categories. Darker cells indicate a greater number of studies utilizing the corresponding behavioral representation.

First, Command and Control (C2) communication is strongly associated with temporal and session-level behavioral representations. Many studies rely on periodic communication patterns and persistent connection behavior to identify C2 infrastructures in encrypted environments where payload inspection is not possible. Second, Disruption and Denial-of-Service attacks appear in studies that utilize both traffic burst dynamics and session-level statistics, reflecting the importance of capturing abnormal traffic intensity and connection behavior during service disruption attempts. Third, Reconnaissance activities are primarily associated with rate-based and temporal patterns, which correspond to repeated scanning attempts or distributed probing behavior across network targets. Finally, Non-Attack-Oriented studies, such as application identification or general encrypted traffic classification, are distributed across multiple behavioral categories. These studies most frequently utilize session-level statistics and temporal characteristics, reflecting the dominant

role of aggregated flow behavior in encrypted traffic analysis. Overall, the cross-mapping matrix reveals that encrypted traffic detection research tends to rely on a limited set of behavioral abstractions—particularly session persistence and temporal periodicity—to infer malicious activity in the absence of payload visibility. This observation supports the necessity of separating attack intent and behavioral representation within a structured taxonomy framework.

Furthermore, this matrix empirically supports the many-to-many relationship between attack objectives and observable behavioral categories proposed in Section 3. No attack objective is uniquely associated with a single behavioral abstraction, and no behavioral representation exclusively corresponds to a single attack objective. This indicates that, in prior studies, the alignment between attack intent (Level 1) and behavioral abstraction (Level 2) has largely remained implicit. Most existing studies tend to select behavioral representations primarily to improve model performance rather than to systematically explain how those representations correspond to specific attack objectives. These observations suggest the need for further examination of how behavioral abstractions influence the design of Level 3 detection techniques, which is analyzed in the following section.

4.4.3 Structural Role of Behavioral Abstraction

The Level 2 analysis reveals that encrypted traffic research differs not only in attack categories but also in behavioral abstraction levels. Behavioral abstraction can be broadly categorized into:

- Low-level abstraction: packet length, direction, basic statistics
- Mid-level abstraction: session persistence, flow statistics, temporal dependency
- High-level abstraction: graph-based interactions

Fig. 8 illustrates the behavioral abstraction pyramid of Level 2 observable network behaviors, highlighting the progression from packet-level representations to higher-level relational abstractions.

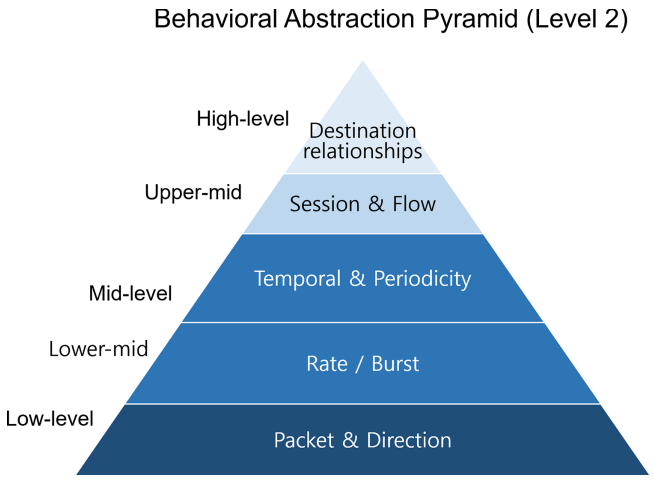

Figure 8: Behavioral abstraction pyramid of Level 2 observable network behaviors.

The majority of studies concentrate on mid-level abstraction, while high-level relational modeling remains relatively limited. This progression from packet-level to structural abstraction reflects ongoing attempts to compensate for payload invisibility through behavioral modeling. However, despite the diversity of model architectures, behavioral representations remain concentrated within a relatively narrow abstraction spectrum. These abstraction levels provide a structural bridge to the Level 3 model analysis presented in Section 4.5.

4.5 Detection Techniques across Behavioral Patterns (Level 3)

This section analyzes the detection models and learning strategies (Level 3) used in the 53 selected studies and examines how they structurally correspond to Level 2 behavioral representations. Level 3 represents the AI model architectures and learning paradigms employed for encrypted traffic detection and reflects the technical evolution of the field.

4.5.1 Distribution of Detection Techniques

The models used in the selected studies were categorized as follows:

- Traditional machine learning (e.g., Random Forest, SVM)
- CNN-based deep learning models
- RNN/LSTM-based sequence models
- Transformer-based models
- Graph Neural Networks (GNN)
- Contrastive learning approaches
- Pre-trained/Foundation model approaches
- Federated learning approaches
- Generative model-based approaches

(Multiple models per study were counted independently.)

The analysis reveals several trends:

- Transformer-based models and pre-training approaches have increased in recent studies.
- CNN and RNN architectures remain widely adopted, particularly in packet-length and temporal sequence modeling.
- GNNs are primarily employed when destination relationships or communication graphs are explicitly modeled.
- Traditional machine learning approaches are more common in early studies or resource-constrained settings.
- Federated and generative approaches appear in specialized contexts such as distributed environments or limited data scenarios.

Overall, the results indicate significant architectural diversity in encrypted traffic detection research.

4.5.2 Structural Coupling between Behavioral Representation and Model Architecture

The analysis reveals a consistent structural coupling between Level 2 behavioral abstractions and Level 3 model selection. For example, temporal behavioral representations such as packet length sequences, periodic intervals, and inter-arrival times are commonly paired with sequence-based models such as RNN, LSTM, GRU, and Transformer architectures. This reflects the natural compatibility between temporal abstractions and sequential modeling capabilities. In particular, C2-oriented studies that rely on periodic communication patterns predominantly employ temporal sequence models, indicating architectural specialization for persistence modeling. In contrast, session- and flow-level statistical features such as session duration, number of flows, and upload/download ratio are typically encoded as fixed-length feature vectors. As a result, these representations are frequently combined with CNN-based architectures or traditional machine learning models such as Random Forest and SVM. Graph-based behavioral representations, including destination relationships and communication structures, are consistently associated with GNNs. Because relational modeling requires explicit structural learning mechanisms, architectural selection in these cases

directly reflects the structural characteristics of the behavioral abstraction. Traffic rate and burst patterns are primarily addressed using traditional machine learning techniques, suggesting that burst-based statistical features can often be effectively handled without requiring deep architectural complexity. In contrast, hybrid or multimodal behavioral representations tend to employ ensemble or composite architectures in order to integrate multiple abstraction levels within a unified detection framework. Overall, these findings indicate that model architecture selection is more strongly determined by the level of behavioral abstraction than by attack objective. In other words, architectural diversity at Level 3 does not necessarily imply diversification of attack-objective modeling at Level 1. Rather, it reflects technical adaptation to specific forms of observable network behavior. This structural dependency highlights the importance of separating attack intent (Level 1) from behavioral abstraction (Level 2) within a unified taxonomy framework. Without such separation, architectural innovation may be misinterpreted as conceptual expansion in attack modeling.

4.5.3 Integrated Structural Summary of the Three-Level Mapping

Integrating the Level 1–Level 2–Level 3 mapping reveals several structural characteristics of contemporary encrypted traffic detection research. First, research attention across attack objectives (Level 1) is uneven. A substantial proportion of studies fall under Non-Attack-Oriented categories, while C2 detection dominates among explicitly attack-oriented works. Objectives such as Data Exfiltration, Reconnaissance, and Disruption/DoS remain comparatively underrepresented. Second, behavioral representations (Level 2) are predominantly centered on mid-level abstractions, particularly session statistics and temporal dependencies. Fine-grained packet-level semantics and high-level relational modeling remain less explored in comparison. Third, model architectures (Level 3) exhibit significant diversity, including CNNs, RNNs, Transformers, GNNs, federated learning, and generative approaches. However, this architectural diversity largely operates within a relatively narrow behavioral abstraction spectrum. Fourth, architectural evolution appears to follow behavioral abstraction complexity. As representations move from statistical flow features toward relational or sequential abstractions, model architectures shift toward structurally expressive forms such as Transformers and GNNs. Collectively, these findings suggest that encrypted traffic detection research has progressed primarily through refinement of behavioral abstraction and architectural sophistication rather than through systematic expansion of attack-intent modeling. Fig. 9 provides an integrated overview of the proposed three-level taxonomy, illustrating how attack objectives (Level 1), observable behavioral representations (Level 2), and detection models (Level 3) are structurally connected through behavioral abstraction.

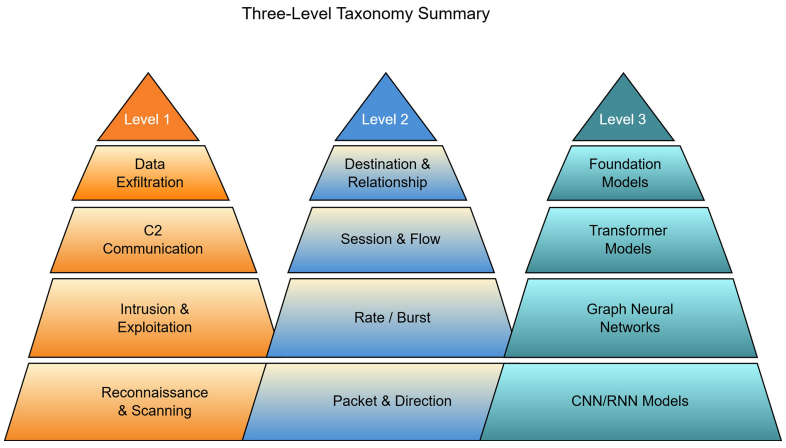

Figure 9: Overview of the proposed three-level taxonomy for AI-based encrypted traffic analysis. Level 1 represents attack objectives, Level 2 observable behaviors, and Level 3 detection models. The framework shows how behavioral representations connect attack intent and model architecture.

From an intrusion detection system (IDS) perspective, this structural configuration reveals two critical gaps:

1. Explicit alignment between attack intent and behavioral abstraction remains insufficiently articulated.
2. Architectural advancement has not been consistently integrated with intent-aware detection objectives.

The proposed three-level taxonomy addresses these gaps by:

- Separating attack objectives (Level 1) from behavioral evidence (Level 2),
- Structuring architectural implementations (Level 3) within this dual-layer framework,
- Restoring a coherent IDS-oriented interpretation of encrypted traffic analysis research.

This integrated view validates the practical utility of the Level 1–Level 2–Level 3 taxonomy beyond statistical categorization. It demonstrates how structured normalization and cross-level mapping enable systematic identification of research imbalances, underexplored objective–behavior regions, and architectural concentration patterns. The implications of this structural configuration for detection robustness, real-time deployment, and future research directions are discussed in the next section. Fig. 10 presents the proposed analytical framework of encrypted traffic analysis. Under encryption constraints, where payload information is not accessible, intrusion detection relies on observable behavioral abstractions. In this framework, behavioral representation (Level 2) acts as a critical mediation layer that connects attack objectives (Level 1) and detection models (Level 3). The relationship between Level 1 and Level 2 is inherently many-to-many, while model selection is largely determined by the characteristics of behavioral representation. This structure explains why architectural diversity does not necessarily correspond to diversification in attack-objective modeling.

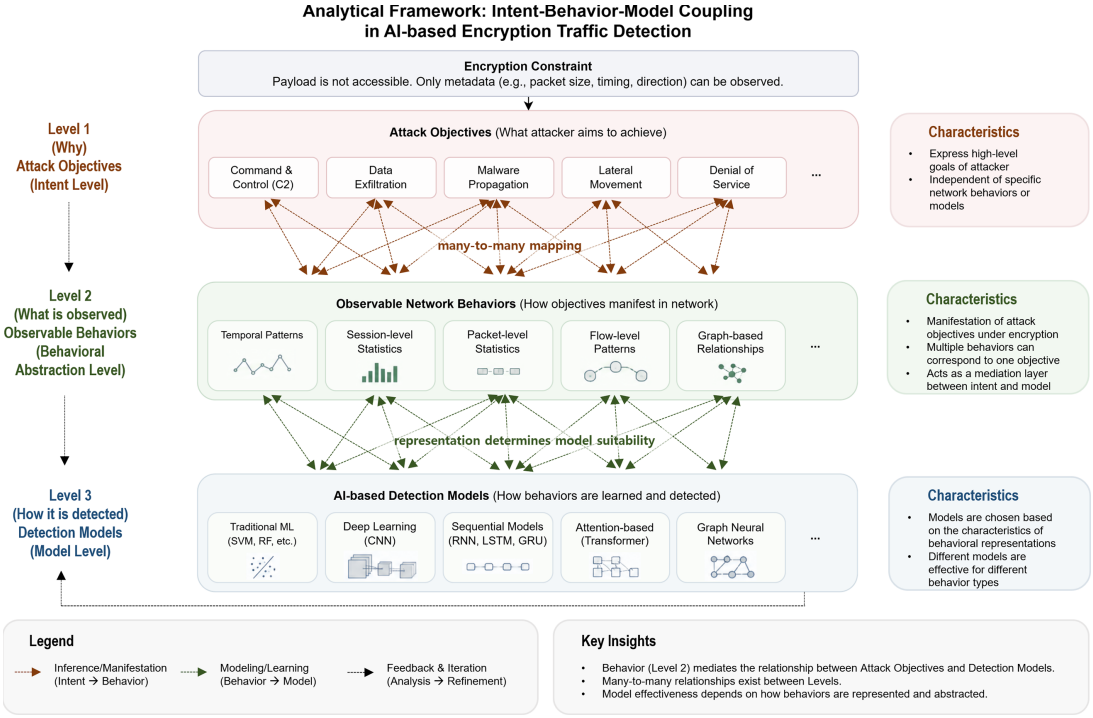

Figure 10: Analytical framework illustrating how behavioral abstraction (Level 2) mediates the relationship between attack objectives (Level 1) and detection models (Level 3) under encryption constraints. The framework highlights the many-to-many relationship between attack objectives and observable behaviors, and the dependency of model selection on behavioral representation.

4.6 Public Datasets and Evaluation Practices in AI-Based Encrypted Traffic Analysis

Experimental evaluation in encrypted traffic analysis is strongly influenced by both benchmark dataset selection and evaluation methodology. Therefore, understanding commonly used datasets and evaluation metrics is essential for interpreting reported detection performance across studies.

4.6.1 Commonly Used Public Datasets

The evaluation of encrypted traffic detection methods heavily relies on publicly available network intrusion detection system (NIDS) datasets. As summarized in Table 10 and Appendix A Table A1, a total of 13 representative public and self-collected datasets were identified across the analyzed studies. Among them, a small subset of benchmark datasets appears frequently, with ISCX VPN-nonVPN, USTC-TFC2016 being the most frequently used datasets, followed by CICIDS2017, and ISCX Tor-nonTor. However, the distribution of dataset usage is highly imbalanced. In particular, ISCX VPN-nonVPN and USTC-TFC2016 dominate the experimental evaluation, each appearing in more than 20 studies within the analyzed corpus. In contrast, many other datasets, such as CTU-13, Bot-IoT, and various self-collected enterprise datasets, are used only once or in a limited number of studies.

Table 10: Public benchmark datasets used in the analyzed studies and their associated attack objectives.

Dataset	Usage Count	Representative Attacks/Traffic	Evaluated Level 1 Objectives
ICS VPN-nonVPN	28	VPN and non-VPN encrypted application traffic	Non-Attack Oriented
USTC-TFC2016	26	Malware traffic, encrypted benign application traffic, and botnet communication	Control & Command (C2), Malware Propagation
CICIDS2017	6	DoS, DDoS, PortScan, Botnet, brute force, infiltration	Disruption/DoS, Reconnaissance, C2
ISCX Tor-nonTor	4	Tor and non-Tor encrypted traffic	Non-Attack-Oriented
UNSW-NB15	4	Generic intrusion behaviors, exploits, reconnaissance, fuzzing	Reconnaissance, Distruption/DoS
AWF Dataset	2	Encrypted web traffic and website fingerprinting traces	Non-Attack-Oriented
DataCon2020	2	Encrypted malicious traffic and abnormal encrypted flows	C2, Generic Malicious Communication
CSTNET-TLS1.3	2	TLS 1.3 encrypted traffic sessions	Non-Attack-Oriented
CIC-AndMal-2017	1	Android malware communication traffic	C2, Malware Propagation
CTU-13	1	Botnet communication scenarios span bot activity	Control & Command (C2)
Bot-IoT	1	IoT botnet traffic, DDoS, reconnaissance, data exfiltration	Disruption/DoS, Reconnaissance, Data Exfiltration
QUIC-related self-collected datasets	1	QUIC-encrypted application traffic	Non-Attack-Oriented
Self-collected enterprise datasets	Multiple isolated cases	Organization-specific encrypted traffic	Varies depending on scenario

Representative datasets used in encrypted traffic analysis include:

- ISCX VPN-nonVPN: A widely used dataset for encrypted traffic classification, distinguishing between VPN and non-VPN traffic

- USTC-TFC2016: A dataset containing labeled traffic for both benign applications and malware communication
- CICIDS2017: A comprehensive dataset covering various attack types, including DoS and infiltration
- ISCX Tor-nonTor: A dataset used for distinguishing Tor and non-Tor encrypted traffic
- UNSW-NB15: A general-purpose intrusion detection dataset containing exploits, reconnaissance, fuzzing, and generic intrusion behaviors

These datasets differ significantly in terms of attack scenarios, traffic composition, labeling granularity, and feature representation. These differences directly affect which attack objectives (Level 1) and observable behavioral patterns (Level 2) are represented during model training and evaluation. This diversity indicates that experimental results are strongly influenced by dataset selection rather than solely by model design. Consequently, direct performance comparison across studies should be interpreted with caution. In particular, datasets emphasizing botnet communication or long-lived encrypted sessions tend to reflect Control and Command (C2)-oriented objectives, whereas datasets dominated by scanning or flooding behaviors are more strongly associated with Reconnaissance or Disruption objectives.

4.6.2 Evaluation Metrics

AI-based encrypted traffic detection studies commonly evaluate model performance using standard classification metrics such as accuracy, precision, recall, F1-score, and AUC. However, the operational requirements of NIDS environments introduce additional considerations, including class imbalance, false alarm reduction, and real-time detection capability.

In encrypted traffic analysis, evaluation metrics may be significantly affected by traffic imbalance, heterogeneous attack distributions, and low-rate attack scenarios. Consequently, conventional classification metrics alone may not fully capture the practical challenges of encrypted traffic detection in operational NIDS environments. Highly imbalanced datasets and low-rate attack scenarios may cause overall accuracy to overestimate practical detection capability. In operational NIDS environments, recall and false positive rates are particularly important because excessive false alarms may overwhelm security analysts, while missed detections may allow malicious encrypted traffic to remain undetected. In addition, real-time detection requirements are not adequately reflected by conventional offline evaluation metrics.

The analyzed studies exhibit a strong concentration on a small number of benchmark datasets, particularly ISCX VPN-nonVPN and USTC-TFC2016. Moreover, different datasets emphasize different attack objectives and observable behavioral representations, indicating that dataset selection implicitly constrains the Level 1 attack objectives and Level 2 behavioral representations learned by AI-based detection models. Similar observable network behaviors may correspond to different attack objectives in encrypted traffic environments. Because payload visibility is limited, behavioral representations such as periodic communication, burst transmission, or long-lived sessions do not uniquely determine attack objectives. As a result, the same benchmark dataset may support the evaluation of multiple Level 1 objectives depending on the selected traffic subsets, attack scenarios, and feature representations. Although some datasets were originally developed for general intrusion detection or encrypted traffic classification, the analyzed studies frequently reused them for different analytical purposes, including encrypted traffic classification, malware communication analysis, botnet detection, and general NIDS evaluation.

4.6.3 Implications for Comparative Analysis

The diversity of datasets and evaluation metrics introduces significant variability in experimental settings across studies. As a result, model performance is often influenced not only by architectural design,

but also by dataset characteristics and evaluation criteria. The diversity of datasets and evaluation metrics introduces significant variability in experimental settings across studies. As a result, model performance is influenced not only by architectural design, but also by dataset composition, attack distribution, feature representation, and evaluation criteria. Consequently, direct performance-based comparison across studies may provide an incomplete understanding of encrypted traffic detection capability. These observations further support the need for a structural analysis framework that focuses on the relationship between attack objectives, behavioral representations, and detection models, rather than relying solely on reported performance metrics.

5 Comparative Structural Analysis of Encrypted Traffic Detection Research

This section synthesizes the Level 1–Level 2–Level 3 mapping results presented in Section 4 to analyze which attack objectives have been most frequently studied, which behavioral representations have been most commonly used, and which models have been repeatedly combined with specific behavioral representations across encrypted traffic analysis research. The purpose of this section is not to compare the performance of individual studies, but rather to identify the hierarchical relationships and recurring research tendencies among attack intent, behavioral representation, and model architecture.

5.1 Structural Distribution across the Three Levels

The distribution of studies across the three levels reveals several important structural characteristics.

First, research attention at the attack-objective level (Level 1) is highly imbalanced. A substantial proportion of studies fall into non-attack-oriented categories, indicating that much of the existing research prioritizes traffic classification or application identification over explicit attack-intent modeling. Among attack-oriented studies, C2 detection dominates, while objectives such as data exfiltration, reconnaissance, and disruption remain comparatively underexplored. This suggests that current research is not evenly aligned with the full spectrum of real-world attack objectives, but instead concentrates on a narrow subset of post-compromise behaviors. Second, behavioral representations (Level 2) are strongly concentrated on mid-level abstractions, particularly session-level statistics and temporal patterns. In contrast, fine-grained packet-level semantics and high-level relational or graph-based representations are relatively underutilized. This indicates that the choice of behavioral abstraction is constrained not only by observability under encryption but also by prevailing feature engineering practices. Third, model architectures (Level 3) exhibit significant diversity, including CNNs, RNNs, Transformers, and GNNs. However, this diversity is largely confined within a limited range of behavioral abstractions. This reveals that architectural innovation has progressed independently of expansion in behavioral representation, leading to a concentration of models operating on similar input spaces.

5.2 Cross-Level Coupling Patterns

Beyond individual distributions, the interaction between levels reveals deeper structural dependencies. As behavioral representations evolve from statistical features to sequential and relational abstractions, model architectures correspondingly shift toward more expressive forms such as Transformers and GNNs. This indicates that model selection is not directly determined by attack objectives, but is instead mediated by the level of behavioral abstraction. Furthermore, the relationship between attack objectives (Level 1) and behavioral patterns (Level 2) is inherently many-to-many. A single attack objective can manifest through multiple behavioral patterns, while similar behavioral patterns may correspond to different attack objectives. This suggests that attempts to directly map attack types to models without considering behavioral abstraction are structurally limited. The mapping presented in Table 11 is derived from a systematic analysis of 53 studies,

rather than a predefined categorization. In addition, Table 12 summarizes the relevance of evaluation metrics from the perspective of practical NIDS deployment and operational requirements.

Table 11: Common evaluation metrics used in AI-based encrypted traffic analysis.

Metric	Definition	Interpretation
Accuracy	Ratio of correctly classified samples	Overall classification performance
Precision	Ratio of correctly predicted positive samples	Reliability of positive alerts
Recall	Ratio of detected positive samples	Ability to detect malicious traffic
F1-score	Harmonic mean of precision and recall	Balanced evaluation under class imbalance
AUC	Area under the ROC curve	Threshold-independent discrimination capability

Table 12: Relevance of evaluation metrics in NIDS environments.

Metric	Relevance in NIDS environments
Accuracy	May be misleading under highly imbalanced traffic distributions
Precision	Important for reducing false alerts and analyst workload
Recall	Critical because missed attacks may remain undetected
F1-score	Useful for balancing false positives and false negatives
AUC	Useful for comparing detection capability across thresholds
False Positive Rate	Operationally important in large-scale monitoring systems

As shown in Table 13:

1. Packet-level representations are primarily associated with non-attack-oriented studies and are typically combined with CNN or Transformer architectures.
2. Temporal abstractions are strongly linked to C2 detection and frequently paired with RNN or Transformer models.
3. Session/flow-based representations are coupled with C2 and certain intrusion studies, often implemented using CNN or traditional ML approaches.
4. Graph-based representations appear in intrusion detection and advanced C2 modeling, exclusively in conjunction with GNN architectures.

These patterns indicate that the relationship between intent and model is not direct. Instead, behavioral representation (Level 2) operates as a structural mediator. Model selection is more strongly influenced by the type of behavioral abstraction than by the attack objective itself.

To further validate these observations, we analyze the distribution of behavioral representations across attack objectives. Temporal behaviors show a higher association with C2-related objectives (7/18, 38.9%), whereas session-based representations are relatively more associated with non-attack-oriented studies (3/10, 30.0%). This distribution empirically supports that behavioral abstraction is closely linked to specific attack objectives. This indicates that model selection is not arbitrary, but strongly conditioned on the level of behavioral abstraction. Furthermore, overlap analysis reveals that each attack objective is associated with multiple behavioral representations, and each behavioral category corresponds to multiple attack objectives. On average, each attack objective is associated with approximately 2–3 behavioral categories, indicating that multi-mapping is an empirically observed structural property rather than an assumption.

Table 13: Structural coupling across levels (based on 53 studies).

Level 2 Behavioral Representation	Associated Level 1 Objectives	Typical Level 3 Techniques/Models	Paper IDs
Packet & Direction	Non-Attack-Oriented	Transformer-based models, CNN variants, and hybrid deep learning models	P01, P24, P29, P30, P35, P39
Rate/Burst	Reconnaissance, C2 (partial), Non-Attack-Oriented (partial)	Traditional ML (SVM, RF, XGBoost), Hybrid models	P13, P31, P52
Temporal	C2, Disruption/DoS, Reconnaissance, Non-Attack	RNN, LSTM, GRU, Transformer, and Autoencoder-based models	P02, P03, P06, P07, P08, P11, P12, P14, P15, P16, P19, P22, P27, P28, P32, P34, P38, P40, P42, P49
Session & Flow	C2, Disruption/DoS, Non-Attack	CNN, Transformer, Traditional ML, GNN, Hybrid architectures	P09, P17, P18, P20, P23, P25, P26, P27, P33, P36, P37, P43, P44
Destination & Relationship	Data Exfiltration, Non-Attack	Graph Neural Networks (GNN), Autoencoder, Sequence-based models	P05, P53

5.3 Structural Implications for Encrypted Traffic Detection

The observed structural patterns lead to several key implications. First, the lack of explicit alignment between attack objectives and behavioral representations highlights a fundamental gap in current research. This implies that detection strategies are often designed around available features rather than the underlying attack intent. Second, the concentration of studies in specific objective–behavior regions indicates the existence of underexplored areas within the taxonomy. This reveals opportunities for future research to target combinations of attack objectives and behavioral representations that have received limited attention. Third, the dependence of model selection on behavioral abstraction suggests that improvements in detection performance may be more effectively achieved through diversification of behavioral representations rather than solely through architectural innovation. This indicates that advancing feature abstraction is a critical factor for improving detection capability in encrypted environments.

5.4 Summary of Findings

Overall, the analysis demonstrates that encrypted traffic detection research is structurally organized around a hierarchical relationship in which behavioral abstraction mediates the connection between attack intent and model architecture. This reveals that behavioral representation plays a central role in determining both the scope of detectable attack objectives and the selection of appropriate learning models.

5.5 Case Study: Behavior Representation in C2 Detection

Command-and-Control (C2) communication is one of the most extensively studied attack objectives in encrypted traffic analysis, as it represents a critical stage of post-compromise persistence. To demonstrate the practical implications of the proposed framework, we present a case study on C2 detection. In this case study, we analyze how different behavioral representations (Level 2) influence the selection of detection models (Level 3) for the same attack objective (Level 1), demonstrating the practical implications of the proposed

framework. Under encryption constraints, where payload inspection is not feasible, C2 activities manifest through observable network behaviors such as periodic communication patterns, long-lived sessions, and low-rate beaconing traffic. These manifestations can be captured at different levels of abstraction, including packet-level, session-level, and temporal representations. At the packet-level, C2 traffic is represented using features such as packet sizes, inter-arrival times, and direction sequences. This representation provides fine-grained information but often lacks sufficient contextual continuity to capture long-term communication patterns. As a result, models such as convolutional neural networks (CNNs) or sequence-based models (e.g., RNNs) are typically employed to extract local patterns from packet sequences. However, packet-level representations alone may struggle to distinguish C2 traffic from benign periodic applications due to limited temporal context. At the session-level, C2 behavior is modeled using aggregated statistics such as flow duration, total bytes, packet counts, and bidirectional interaction patterns. This level of abstraction provides a more holistic view of communication sessions, enabling traditional machine learning models (e.g., SVM, Random Forest) as well as deep learning models to identify anomalous patterns. While session-level representations improve robustness, they may obscure fine-grained temporal dynamics that are critical for identifying stealthy C2 channels. At the temporal level, C2 detection focuses on long-term behavioral patterns, such as periodic beaconing, burst intervals, and time-series characteristics of communication. This representation captures the persistence and regularity of C2 activities, making it particularly effective for detecting low-rate or stealthy communication channels. Consequently, temporal models such as RNNs, LSTMs, GRUs, and Transformer-based architectures are more suitable for learning sequential dependencies over time. Temporal representations are especially effective in distinguishing C2 traffic from normal application traffic that lacks consistent periodicity. This comparison highlights that, for the same attack objective (C2 detection), the choice of detection model is not directly determined by the objective itself, but rather by the selected behavioral representation. Fig. 11 illustrates this relationship by showing how different behavioral representations (Level 2) lead to different model selections (Level 3) for the same attack objective (Level 1), demonstrating that model selection is mediated by behavioral abstraction rather than directly determined by attack objectives.

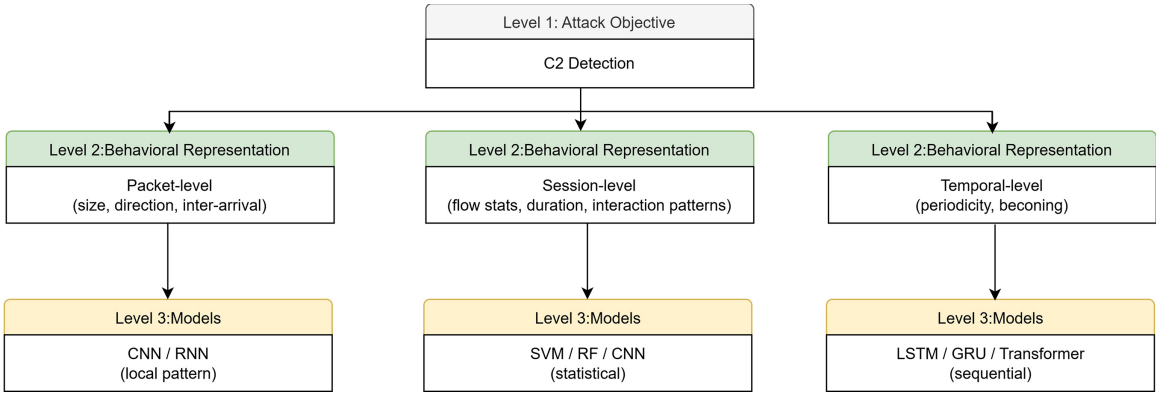

Figure 11: Case study illustrating how different behavioral representations (Level 2) influence the selection of detection models (Level 3) for the same attack objective (C2 detection, Level 1). The figure shows that model selection is mediated by behavioral abstraction rather than directly determined by attack objectives.

Packet-level representations favor local pattern extraction models, session-level representations enable statistical learning approaches, and temporal representations require sequence modeling architectures. From the perspective of the proposed framework, this case study demonstrates that behavioral abstraction (Level 2) acts as a critical mediation layer that determines how attack objectives are translated into learnable

representations and, consequently, which models are most effective. This observation provides practical guidance for IDS design: selecting an appropriate behavioral representation is a prerequisite for effective model selection and directly impacts detection performance.

6 Future Research Directions: Toward Intent–Behavior–Model Integration

The structural analysis in Section 5 shows that existing encrypted traffic research has concentrated primarily on particular behavioral representations (Level 2) and model architectures (Level 3), while attack-objective-driven modeling (Level 1) remains relatively limited. This chapter presents future research directions for addressing this structural imbalance from three perspectives.

6.1 Expanding the Intent Space (Level 1 Expansion)

The results of this study show that current encrypted traffic research is relatively concentrated on C2-oriented modeling, while other attack objectives, such as Data Exfiltration, Reconnaissance, and Disruption/DoS, are addressed much less frequently. In particular, the cross-mapping matrix indicates that C2 is primarily associated with Temporal & Periodicity and Connection & Session Behavior Patterns, whereas Data Exfiltration appears only in connection with Destination & Communication Relationship representations. Reconnaissance and Disruption/DoS are also linked to only a limited subset of behavioral representations. However, recent threat environments exhibit new intent-oriented characteristics that cannot be sufficiently explained by traditional C2-centered detection alone. Examples include stealthy data exfiltration through legitimate service channels such as DNS over HTTPS, QUIC, and CDNs; intelligent C2 mechanisms using content-level obfuscation; AI-driven adaptive communication attacks; and intent-oriented intrusions that exploit behaviors inside specific application protocols. These changes require expansion of intent-level modeling through more fine-grained and hierarchical attack-objective structures. Future research should move beyond single attack-type classification and develop attack-objective frameworks with hierarchical structures, while also modeling the structural relationships among attack objectives.

6.2 Expanding Behavioral Abstraction (Level 2 Expansion)

As observed in this study, encrypted traffic analysis tends to concentrate on a limited set of behavioral representations, including Packet & Direction, Connection & Session Behavior, and Temporal & Periodicity. In particular, most Non-Attack-Oriented studies and C2 studies are centered on session statistics and temporal features, while Packet & Direction-based representations are primarily used in Non-Attack-Oriented research. In contrast, Destination & Communication Relationship-based representations are rarely used overall and appear mainly in Data Exfiltration studies. This indicates that current research mainly relies on features that can be collected relatively reliably in encrypted environments, such as session duration, number of connections, inter-arrival times, and periodic communication intervals. However, emerging threats require more advanced behavioral abstractions. Examples of such future directions include:

1. Long-term consistency analysis of user and endpoint behavior beyond individual sessions
2. Semantic inference using implicit protocol structures and handshake metadata
3. Context-aware behavioral modeling that reflects network environments and service contexts
4. Concept-drift-based detection to address dynamic changes in attacker behavior
5. Relationship-based representations that capture repeated interactions and evolving relationships among multiple entities

These observations suggest that Level 2 behavioral representations should expand beyond simple session- and time-based features toward representations incorporating long-term relationships, interactions, and contextual information.

6.3 Re-Aligning Intent–Behavior–Model

One of the most important structural findings of this study is that model selection tends to be determined more directly by behavioral representation (Level 2) than by attack objective (Level 1). For example, Transformers and LSTMs are repeatedly combined with sequence-based representations regardless of attack objective, while CNNs and traditional machine learning models are repeatedly associated with session-statistics-based features. Similarly, GNNs appear only when relationship-based representations are used. However, these three layers have largely evolved separately. In current research, observable behavioral representations (Level 2) are typically selected first, suitable model architectures (Level 3) are then applied, and attack objectives (Level 1) are inferred from this combination. Future research therefore requires more integrated approaches, including:

- Intent-aware Behavioral Design: explicitly designing behavioral representations according to attack objectives
- Hierarchical Intent Modeling: structuring attack objectives hierarchically and linking them to behavioral representations
- Intent-Conditioned Representation Learning: incorporating attack intent as a conditioning factor during representation learning
- Multi-Intent Learning Frameworks: simultaneously modeling multiple attack objectives
- Behavior–Model Decoupling: reducing the conventional coupling between particular behavioral representations and particular model architectures

Ultimately, an integrated framework that co-designs Intent–Behavior–Model is required. This would represent a transition beyond simple model performance improvement toward a structurally aligned encrypted traffic detection framework in which attack objectives, behavioral representations, and model architectures are explicitly connected.

7 Conclusion

This paper presents an analytical framework for encrypted traffic analysis based on a three-level structure consisting of attack objectives (Level 1), observable behaviors (Level 2), and detection models (Level 3), and systematically analyzes 53 representative studies within this framework. The analysis reveals three major structural characteristics of existing encrypted traffic research. First, there is a clear imbalance at the attack-objective level. A substantial proportion of studies fall into the Non-Attack-Oriented category, focusing on application identification or general classification rather than explicit attack-intent modeling. Among attack-oriented studies, research is heavily concentrated on C2 detection, while other attack objectives remain comparatively underexplored. Second, behavioral representations show a strong convergence toward a limited set of abstractions. Although model architectures continue to evolve rapidly, most studies rely on packet length, session statistics, and temporal periodicity. This indicates that the diversity of behavioral abstraction has not expanded at the same pace as model innovation. Third, cross-level analysis reveals that model selection is more strongly associated with behavioral representations than with attack objectives. Rather than being directly connected, attack objectives and model architectures are structurally mediated through behavioral representations. These findings suggest that encrypted traffic research has evolved more around "how to detect" than "what to detect." Based on these observations, future research should explicitly consider the hierarchical integration of Intent–Behavior–Model, with particular emphasis on intent-aware

detection and the expansion of behavioral representations. By identifying structural imbalance and cross-level coupling, this work provides a behavior-centric analytical framework that explains how detection capability is structurally determined under encryption constraints. Rather than proposing a new detection model, this study provides a structured analytical framework that enables consistent interpretation of heterogeneous studies, which is difficult to achieve through model-centric evaluation alone. The proposed framework can support practical IDS design by guiding the selection of appropriate behavioral representations based on detection objectives. This enables more interpretable and systematically designed detection strategies in encrypted network environments.

Despite its contributions, this study has several limitations. The proposed taxonomy was developed based on the current body of encrypted traffic detection literature and may require further refinement as new attack types, behavioral representations, and foundation-model-based detection approaches continues to emerge. In addition, the structural relationships identified in this study were derived through qualitative literature analysis rather than large-scale experimental validation. Future work should investigate how the proposed Level 1-Level 2-Level 3 framework can be quantitatively validated using benchmark datasets and real-world encrypted traffic environments.

Acknowledgement: This research was supported by Seoul National University of Science and Technology.

Funding Statement: This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2023-00235509, Development of Security Monitoring Technology Based on Network Behavior Against Encrypted Cyber Threats in ICT Convergence Environment).

Author Contributions: Yeog Kim conducted the main research, performed the analysis, and drafted the manuscript. Changhoon Lee reviewed and edited the manuscript. Kiwook Sohn supervised the overall research process, reviewed the manuscript, and served as the corresponding author. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: This study is based on previously published literature and does not involve any newly generated data or materials.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Mapping of surveyed studies to the proposed Level 1–2–3 taxonomy: Each paper is categorized according to attack objective (Level 1), observable network behavior pattern (Level 2), and detection technique (Level 3). The table summarizes the datasets and feature representations used in each study.

Table A1: Summary of the analyzed studies, including datasets, feature representations, and their mapping to the proposed three-level taxonomy (Level 1–Level 3).

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P01 [55]	A Cascaded Broad Learning Network Embedded Image Features for Malware Traffic Classification	Encrypted malware traffic classification	USTC-TFC2016; ISCXVPN2016	Traffic-to-image representation (28 × 28 grayscale) with image processing features	Generic Malicious traffic	Packet Size & Directionality Patterns	Broad Learning System (BLS)—IFCBLN
P02 [56]	A Detection Method for Malware Communication Traffic via Encrypted Traffic Analysis	Encrypted malware communication traffic detection and classification	DataCon2020; CIC-AndMal-2017	Session-based packet byte sequences; temporal correlations; contextual correlations across packet sequences	Command & Control	Temporal Behavior Patterns; Session-level communication patterns	Transformer-based model (Session-Transformer with DNN classifier)
P03 [57]	A Novel Multimodal Deep Learning Framework for Encrypted Traffic Classification	Encrypted traffic application classification	Real data center encrypted traffic dataset	Raw packet bytes + packet-length sequences	Non-Attack-Oriented	Packet size & Directionality Patterns; Connection & Session Behavior Patterns	Transformer-based multimodal model (PEAN)
P04 [58]	Adversarial Attacks on Pre-Trained Deep Learning Models for Encrypted Traffic Analysis	Encrypted traffic classification and robustness evaluation under adversarial attacks	USTC-TFC; ISCX-VPN-Service; ISCX-VPN-App; CSTNET-TLS 1.3	Tokenized packet payload bytes using bigram encoding; BURST-based traffic corpus; packet sequence representation	Non-Attack-Oriented	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Pre-trained Transformer model (ET-BERT)
P05 [59]	Anomaly Detection in the Key-Management Interoperability Protocol Using Metadata	Anomaly detection in encrypted KMIP traffic	Simulated enterprise KMIP metadata dataset (EKM system logs)	KMIP protocol metadata sequences (request types, attributes, operation frequency)	Data Exfiltration/Intrusion	Temporal Behavior Patterns; Statistical flow anomalies	LSTM autoencoder
P06 [60]	ATVITSC: A Novel Encrypted Traffic Classification Method Based on Deep Learning	Encrypted traffic classification (application/malware categories)	- USTC-TFS2016; ISCX-Tor (Tor-nonTor); ISCX-VPN; Cross-Platform Android App dataset	Session images generated from packet payload bytes (first packets/bytes)	Non-Attack-Oriented; Malware Propagation (partial)	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Transformer-based hybrid model (ATVITSC: PVT + Bi-LSTM)

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P07 [61]	AutoML4ETC: Automated NAS for Encrypted Traffic Classification	Encrypted traffic application classification	Real-world ISP encrypted traffic dataset	Flow-based packet length and inter-arrival time sequences	Non-Attack-Oriented	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	AutoML-based neural architecture search (AutoML4ETC)
P08 [62]	BoNC: Discovering and Classifying Novel Encrypted Botnet Traffic	Encrypted botnet traffic detection and classification	ISCX Botnet dataset; CTU-13 botnet dataset	Flow-level statistical features (packet counts, bytes, timing statistics)	Command & Control	Periodic low-rate communication; Long-lived C2 sessions	Clustering-based botnet discovery (BoNC)
P09 [63]	BPF-DAG: Byte-Packet-Flow Feature Fusion via Dynamic Attributed Graph for reliable encrypted traffic classification	Encrypted traffic classification	USTC-TFC2016; ISCX-VPN-Service; ISCX-VPN-App	Byte-, packet-, and flow-level feature fusion with dynamic traffic graph representation	Non-Attack-Oriented; Malware Propagation (partial)	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Graph neural network model (Dynamic Attributed Graph—BPF-DAG)
P10 [64]	Cactus: Obfuscating Bidirectional Encrypted TCP Traffic at Client Side	Defense against encrypted traffic analysis attacks (traffic obfuscation)	Website fingerprinting datasets; Application identification traffic traces	Packet size, direction, and timing patterns manipulated via TCP traffic reshaping	Non-Attack-Oriented (privacy protection)	Non-Attack-Oriented (privacy protection)	TCP traffic obfuscation system (Cactus, eBPF-based client-side mechanism)
P11 [65]	CBS: A Deep Learning Approach for Encrypted Traffic Classification With Mixed Spatio-Temporal and Statistical Features	Encrypted traffic classification	ISCX VPN-nonVPN; ISCX Tor-nonTor; USTC-TFC2016	Mixed spatio-temporal packet features and statistical flow features	Non-Attack-Oriented; Malware propagation (partial)	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	CNN–BiLSTM hybrid deep learning model (CBS)
P12 [66]	CBSeq: A Channel-Level Behavior Sequence for Encrypted Malware Traffic Detection	Encrypted malware traffic detection	ISCX VPN-nonVPN; USTC-TFC2016	Channel-level behavior sequences derived from packet size, direction, and timing	Command & Control; Malware Propagation	Temporal Behavior Patterns; Connection/session behavior patterns	Behavior-sequence deep learning model (CBSeq; CNN + BiLSTM)

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P14 [67]	CL-ViME: Contrastive Learning and Vision Mixture of Experts for Encrypted Traffic Classification	Website fingerprinting attack on encrypted traffic	DF dataset (Deep Fingerprinting dataset); AWF dataset	Packet direction and burst sequences derived from Tor traffic traces	Reconnaissance	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Class-incremental deep learning model (Dynamic Expansion Architecture; CNN-based)
P15 [68]	CMFTC: Cross Modality Fusion Efficient Multitask Encrypt Traffic Classification in IIoT Environment	Encrypted traffic classification in IIoT environment	ISCX VPN-nonVPN; ISCX Tor-nonTor; USTC-TFC2016	Cross-modality features combining packet byte sequences and flow statistical features	Non-Attack-Oriented; Malware Propagation (partial)	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns; Temporal Behavior Patterns	Cross-modality fusion multitask deep learning model (CMFTC; CNN-based)
P16 [69]	CMTSNN: A Deep Learning Model for Multiclassification of Abnormal and Encrypted Traffic of Internet of Things	IoT encrypted traffic anomaly detection and classification	IoT network traffic dataset (normal and attack traffic)	Flow statistical features and packet sequence characteristics	Disruption (DoS/DDoS); Malware propagation; Reconnaissance	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Statistical flow anomalies	CNN-based deep learning model (CMTSNN)
P17 [70]	Data Augmentation-Based Enhancement for Efficient Network Traffic Classification	Encrypted traffic application classification	ISCX VPN-nonVPN 2016	Tabular packet-header features derived from standardized packet fields	Non-Attack-Oriented	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Data-augmentation-enhanced ML/DL traffic classifier
P18 [71]	DNS Over HTTPS Detection Using Standard Flow Telemetry	DNS over HTTPS (DoH) traffic detection	CIRA-CIC-DoHBrw-2020 dataset; DoH resolver traffic traces	Bidirectional flow telemetry features (packet/byte ratios, packet counts, timing statistics)	Command & Control; Data Exfiltration (potential misuse of DoH)	Connection/session behavior patterns; Statistical flow anomalies	Flow-telemetry-based ML classifier
P19 [72]	Encrypted TLS Traffic Classification on Cloud Platforms (NeuTic)	TLS encrypted traffic classification	ISCX VPN-nonVPN dataset; cloud platform TLS traffic traces	TLS flow statistical features and packet length distributions	Non-Attack-Oriented	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Machine learning–based TLS traffic classifier

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P20 [73]	Encrypted Traffic Classification Through Deep Domain Adaptation Network With Smooth Characteristic Function	Encrypted traffic application classification with domain adaptation	ISCX VPN-nonVPN; USTC-TFC2016	Packet byte sequences and flow statistical features	Non-Attack-Oriented; Malware propagation (partial)	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Deep domain adaptation network (CNN-based)
P21 [74]	Encrypted Traffic Detection in Resource Constrained IoT Networks: A Diffusion Model and LLM Integrated Framework	Encrypted malicious traffic detection in IoT networks	IoT network traffic datasets (normal and attack traffic)	Flow statistical features and temporal traffic attributes	Disruption (DoS/DDoS); Malware propagation; Command & Control	Disruption (DoS/DDoS); Malware Propagation; Command & Control	Diffusion-model–based data generation + LLM-assisted detection framework
P22 [75]	End-to-End Open-Set Semi-Supervised Learning for Fine-Grained Encrypted Traffic Classification	Fine-grained encrypted traffic classification with open-set recognition	ISCX VPN-nonVPN; USTC-TFC2016	Packet byte sequences and packet length/time features	Non-Attack-Oriented; Malware propagation (partial)	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Open-set semi-supervised deep learning model
P23 [76]	Enhancing Accountability, Resilience, and Privacy of Intelligent Networks With XAI	Explainable encrypted traffic classification/intrusion detection	CICIDS2017; USTC-TFC2016	Flow statistical features extracted from network telemetry	Disruption (DoS/DDoS); Malware propagation; Command & Control	Statistical flow anomalies; Connection & Session Behavior Patterns	XAI-enhanced machine learning IDS framework
P24 [77]	EO-EPTC: End-to-End Original Traffic-Based Encrypted Proxy Traffic Classification Framework	Encrypted proxy traffic classification	Original TCP traffic dataset; proxied traffic datasets (Shadowsocks, VMess, Trojan, VLESS)	Traffic sequence features (packet length, timing, direction) extracted from bidirectional flows	Command & Control; Data Exfiltration (proxy-based tunneling)	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	EO-EPTC framework (Seq2Seq-based proxy traffic transformation + classifier)
P25 [78]	Flow-Based Encrypted Network Traffic Classification With Graph Neural Networks	Encrypted traffic application classification	ISCX VPN-nonVPN	Graph representation of flow statistics	Non-Attack-Oriented	Connection & Session Behavior Patterns; Statistical flow anomaly	GNN-based classifier

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P26 [79]	From Packets to Semantics: Transformer-Driven Privacy-Preserving ETA for OTT App Classification	Encrypted traffic analysis for OTT application classification	OTT application traffic dataset (encrypted network traffic traces)	Packet sequence features (packet length, direction, timing) encoded as Transformer inputs	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Transformer-based ETA model
P27 [80]	Identification of Malicious Encrypted Traffic Through Feature Fusion	Malicious encrypted traffic detection	ISCX VPN-nonVPN; USTC-TFC2016	Fusion of packet byte sequences and flow statistical features	Malware propagation; Command & Control	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Statistical flow anomalies	Feature-fusion deep learning classifier
P28 [81]	Machine Learning-Powered Encrypted Network Traffic Analysis: A Comprehensive Survey	Survey of machine learning techniques for encrypted traffic analysis	Survey of 108 studies; commonly discussed datasets include NSL-KDD, UNSW-NB15, ISCX VPN-nonVPN, CICIDS, USTC-TFC2016	Flow statistical features, packet-based features, and raw traffic representations	Multiple objectives: network identification, traffic classification, anomaly detection, privacy leakage detection	Statistical Flow Anomaly; packet sequence patterns	Survey of ML/DL models for encrypted traffic analysis
P29 [82]	MT-DEGCL: Multi-Task Encrypted Traffic Classification With Dual Embedding and Graph Contrastive Learning	Encrypted traffic classification (multi-task learning)	ISCX VPN-nonVPN; USTC-TFC2016	Dual embeddings of packet byte sequences and flow statistical features with graph representation	Non-Attack-Oriented; Malware propagation (partial)	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Graph contrastive learning with dual-embedding multi-task model (MT-DEGCL)
P30 [83]	Multi-Level Pre-Training for Encrypted Network Traffic Classification	Encrypted traffic classification with multi-level pre-training	ISCX VPN-nonVPN; USTC-TFC2016	Packet byte sequences and flow statistical features with hierarchical representations	Non-Attack-Oriented; Malware propagation (partial)	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Multi-level pre-trained deep learning model

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P31 [84]	NFTracker: Fine-Grained NFT Behavior Traffic Identification Over Encrypted Tunnel	Fine-grained NFT behavior identification over encrypted tunnels	Self-collected NFT behavior traffic datasets (OpenSea, LooksRare over V2Ray/VMess)	Packet length, direction, and inter-packet time sequences with sliding-window spatio-temporal features	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Hybrid CNN-Transformer framework (NFTracker)
P32 [85]	Resource-Efficient Spectrum-Based Traffic Classification on Constrained Devices	Resource-efficient encrypted traffic classification on constrained devices	ISCX VPN-nonVPN dataset	Frequency-domain packet length features	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Spectrum-based classifier
P33 [86]	Robust Network Traffic Classification Based on Information Bottleneck Neural Network	Encrypted network traffic classification	ISCX VPN-nonVPN; ISCX Tor2016	Packet-level statistics, transport-layer features, and flow duration features	Non-Attack-Oriented	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns; Statistical flow anomalies	Information Bottleneck Neural Network (IBNN)
P34 [87]	Robustness Matters: Pre-Training Can Enhance the Performance of Encrypted Traffic Analysis	Encrypted traffic analysis (traffic classification/application identification)	DataCon2020, DataCon2021-p2, CSTNET-TLS1.3, EBSNN-dataset, CSTNET-packet	Packet length sequence with direction encoded as tokens	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Pre-trained Transformer model (BERT-ps)
P35 [88]	SHAPE: A Simultaneous Header and Payload Encoding Model for Encrypted Traffic Classification	Encrypted traffic classification	ISCX VPN-nonVPN; USTC-TFC2016	Raw packet header bytes + truncated payload bytes	Non-Attack-Oriented; Malware propagation	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	SHAPE header–payload encoding model
P36 [89]	Two-Stage Distillation-Aware Compressed Models for Traffic Classification	Encrypted traffic classification (application & malware)	ISCX VPN-nonVPN; USTC-TFC2016	Fixed-length packet byte representation	Non-Attack-Oriented; Malware propagation	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns	Distillation-aware compressed DL model (NIN-based)

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P37 [90]	Universal Embedding Function for Traffic Classification via QUIC Domain Recognition Pretraining	Encrypted traffic classification via transfer learning	CESNET-QUIC22 (pretraining); ISCXVPN2016; MIRAGE19; MIRAGE22; UTMOBILENET21; UCDAVIS19; CESNET-TLS22; AppClassNet	Packet size + direction + inter-packet time sequences	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Transfer-learning traffic embedding model
P38 [91]	VPN-Encrypted Network Traffic Classification Using a Time-Series Approach	VPN encrypted traffic classification	ISCX VPN-nonVPN dataset	Packet length time-series with packet direction	Non-Attack-Oriented	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Time-series deep learning model (CNN/LSTM-based)
P39 [92]	WFF-EGNN: Encrypted Traffic Classification Based on Weaved Flow Fragment via Ensemble Graph Neural Networks	Encrypted traffic classification	ISCX VPN-nonVPN; USTC-TFC2016	Weaved flow fragments from packet sequences (packet length and direction)	Non-Attack-Oriented; Malware propagation	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns; Temporal Behavior Patterns	Ensemble Graph Neural Network (WFF-EGNN)
P40 [93]	WF-Transformer: Learning Temporal Features for Accurate Anonymous Traffic Identification by Using Transformer Networks	Anonymous/encrypted traffic identification (website fingerprinting)	AWF dataset; DF dataset (Tor traffic datasets)	Packet direction sequence (±1) with packet timing/ordering	Reconnaissance	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Transformer-based website fingerprinting model (WF-Transformer)
P41 [94]	Detecting Unknown Encrypted Malicious Traffic in Real Time via Flow Interaction Graph Analysis	Unknown encrypted malicious traffic detection	USTC-TFC2016; CICIDS2017	Flow interaction graph features	Command & Control; Malware propagation	Connection & Session Behavior Patterns; Statistical Flow Anomaly	Flow interaction graph GNN

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P42 [95]	Research on network traffic classification method based on dual-branch spatial-temporal attention	Encrypted malware traffic classification	USTC-TFC2016 dataset	Bidirectional flow representation; packet sequences; spatial features from packet matrices; temporal dependencies across packet flows	Command & Control; Malware propagation	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Dual-branch spatio-temporal attention model (2D-CNN + LSTM + attention)
P43 [96]	Integrating Explainable AI for Effective Malware Detection in Encrypted Network Traffic	Encrypted malware traffic detection with XAI	CICIDS2017; USTC-TFC2016	Flow statistical features	Command & Control; Malware propagation	Statistical flow anomaly; Connection & Session Behavior Patterns	Explainable ML-based IDS (XAI)
P44 [97]	Language of Network: A Generative Pre-trained Model for Encrypted Traffic Comprehension	Encrypted traffic understanding and classification via generative pre-trained model	USTC-TFC2016; ISCX VPN-nonVPN; CICIDS2017	Raw packet byte sequences (header + payload bytes) as tokenized traffic sequences	Non-Attack-Oriented; Command & Control; Malware propagation	Packet Size & Directionality Patterns; Temporal behavior patterns	Generative pre-trained traffic model (LoN)
P45 [98]	Deep Learning and Pre-training Technology for Encrypted Traffic Classification: A Comprehensive Review	Survey of deep learning and pre-training methods for encrypted traffic classification	Survey (multiple datasets)	Packet-, flow-, and byte-level representations	Non-Attack-Oriented; Command & Control; Malware propagation	Packet Size & Directionality Patterns; Temporal Behavior Patterns; Connection & Session Behavior Patterns	Survey of deep learning and pre-trained ETA models
P46 [99]	Adversarial Robustness of Traffic Classification under Resource Constraints	Adversarial robustness evaluation for encrypted traffic classification under resource constraints	USTC-TFC2016	Flattened byte sequences; 2D packet-wise time-series representations; packet order information; adversarial perturbation patterns	Non-Attack-Oriented	Packet byte sequence patterns; Packet-wise temporal behavior patterns	Hardware-aware lightweight deep learning models with adversarial robustness evaluation

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P47 [100]	Flow Interaction Graph Analysis for Unknown Encrypted Malicious Traffic Detection (Extended Version)	Unknown encrypted malicious traffic detection	USTC-TFC2016; CICIDS2017	Flow interaction graph derived from flow statistics and host communication relations	Generic Malicious traffic	Connection & Session Behavior Patterns; Statistical flow anomaly	Flow interaction graph-based malicious traffic detection
P48 [101]	Application of EM-transformer hybrid model in real-time detection of directed DDoS attacks on botnet devices	DDoS attack detection in network traffic	Bot-IoT; UNSW-NB15	Flow statistical features combined with EM-GMM posterior probability features	Disruption (DoS/DDoS)	High-rate burst traffic; Statistical Flow Anomaly	EM-Transformer hybrid detection model
P49 [102]	Machine learning-based hybrid technique to enhance cyber-attack perspective	Network intrusion detection using hybrid machine learning technique	UNSW-NB15	Flow statistical traffic features extracted from network flows	Disruption (DoS/DDoS); Reconnaissance;	Statistical Flow Anomaly; Connection & Session Behavior Patterns	Hybrid machine learning intrusion detection model
P50 [103]	Advances in deep learning intrusion detection over encrypted data with privacy preservation: a systematic review	Survey of deep learning-based intrusion detection over encrypted traffic	Survey (multiple datasets: ISCX VPN-nonVPN, CICIDS2017, UNSW-NB15, USTC-TFC2016, etc.)	Packet-, flow-, and statistical traffic representations	Disruption (DoS/DDoS); Command & Control; Malware propagation; Reconnaissance	Packet Size & Directionality Patterns; Connection & Session Behavior Patterns; Statistical Flow Anomaly	Survey of deep learning IDS models for encrypted traffic
P51 [104]	Adaptive detection of encrypted malware traffic via fully convolutional masked autoencoders	Encrypted malware traffic detection	USTC-TFC2016	Packet byte sequences and packet length–direction sequences	Command & Control; Malware propagation	Packet Size & Directionality Patterns; Temporal Behavior Patterns	Fully convolutional neural network (FCN)-based malware traffic detection
P52 [105]	Detecting malware evidences through static and dynamic information using extreme machine learning for forensic analysis	Encrypted malware traffic detection	USTC-TFC2016	Flow statistical features extracted from encrypted traffic metadata	Command & Control; Malware propagation	Connection & Session Behavior Patterns; Statistical Flow Anomaly	Static feature–based machine learning malware traffic detection

(Continued)

Table A1 (continued)

ID	Title	Task/Scope	Datasets	Feature	Level 1	Level 2	Level 3
P53 [106]	An enhanced mechanism for malicious URL detection using deep learning and DistilBERT-based feature extraction	Malicious URL detection	Kaggle Phishing Site URLs; Kaggle Malicious URLs	Machine learning–based malicious URL detection	Malware propagation; Command & Control	Application-Layer Pattern	Machine learning–based malicious URL detection

References

1. Google LLC. HTTPS encryption on the web [Internet]. Mountain View, CA, USA: Google; 2024 [cited 2026 Mar 2]. Available from: https://transparencyreport.google.com/https/overview.
2. Rescorla E. The transport layer security (TLS) protocol version 1.3 [Internet]. RFC 8446. Fremont, CA, USA: Internet Engineering Task Force (IETF); 2018 [cited 2026 Mar 2]. Available from: https://www.rfc-editor.org/rfc/rfc8446.
3. Velan P, Cech T, Roman R, Garcia-Rego JA. A survey of methods for encrypted traffic classification and analysis. Int J Netw Manag. 2015;25(5):355–74. doi:10.1002/nem.1901.
4. Khan JA, Khan MA, Saeed N, Cayrel PL, Hahn C. Intrusion detection systems for in-vehicle networks: protocol, applications, and challenges. IEEE Access. 2025;13(6):215219–50. doi:10.1109/ACCESS.2025.3645058.
5. Di Monda D, Nascita A, Carillo R, Pescapé A. Analyzing the impact of encryption on traffic classification through explainable AI. In: Proceedings of the IFIP Networking 2025; 2025 May 26–30; Limassol, Cyprus. p. 134–9.
6. Abbasi M, Shahraki A, Taherkordi A. Deep learning for network traffic monitoring and analysis (NTMA): a survey. Comput Commun. 2021;170(3):19–41. doi:10.1016/j.comcom.2021.01.021.
7. Papadogiannaki E, Ioannidis S. A survey on encrypted network traffic analysis applications, techniques, and countermeasures. ACM Comput Surv. 2021;54(6):1–35. doi:10.1145/3457904.
8. Kovanen T, David G, Hämäläinen T. Survey: intrusion detection systems in encrypted traffic. In: Internet of things, smart spaces, and next generation networks and systems. Cham, Switzerland: Springer; 2016. p. 281–93. doi:10.1007/978-3-319-46301-8_23.
9. European Union Agency for Cybersecurity. Encrypted traffic analysis. Heraklion, Greece: ENISA; 2020.
10. Wright C, Monrose F, Masson G. On inferring application protocol behaviors in encrypted network traffic. J Mach Learn Res. 2006;7:2745–69. doi:10.5555/1248547.1248647.
11. Luh R, Marschalek S, Kaiser M, Janicke H, Schrittwieser S. Semantics-aware detection of targeted attacks: a survey. J Comput Virol Hack Tech. 2017;13(1):47–85. doi:10.1007/s11416-016-0273-3.
12. Studer A, McLain CD, Lippmann RP. Tuning intrusion detection to work with a two encryption key version of IPsec. In: Proceedings of the IEEE Military Communications Conference (MILCOM 2007); 2007 Oct 29–31; Orlando, FL, USA. Piscataway, NJ, USA: IEEE; 2007. p. 3977–83.
13. Aceto G, Ciuonzo D, Montieri A, Pescape A. MIMETIC: mobile encrypted traffic classification using multimodal deep learning. Comput Netw. 2019;165(1):106944. doi:10.1016/j.comnet.2019.106944.
14. Sattar S, Khan S, Khan MI, Akhmediyarova A, Mamyrbayev O, Kassymova D, et al. Anomaly detection in encrypted network traffic using self-supervised learning. Sci Rep. 2025;15(1):26585. doi:10.1038/s41598-025-08568-0.
15. Jallad KA, Aljnidi M, Desouki MS. Anomaly detection optimization using big data and deep learning to reduce false-positive. arXiv:2209.13965. 2022.
16. Singh A, Jang-Jaccard J. Autoencoder-based unsupervised intrusion detection using multi-scale convolutional recurrent networks. arXiv:2204.03779. 2022.
17. Zhang Q, Xu Y, Yin Z, Zhou C, Jiang Y. Automatic policy synthesis and enforcement for protecting untrusted deserialization. In: Proceedings of the Network and Distributed System Security Symposium (NDSS); 2024 Feb 26–Mar 1; San Diego, CA, USA. doi:10.14722/ndss.2024.24053.
18. Liu L, Wang P, Lin J, Liu L. Intrusion detection of imbalanced network traffic based on machine learning and deep learning. IEEE Access. 2021;9:7550–63. doi:10.1109/ACCESS.2020.3048198.
19. Bedi P, Gupta N, Jindal V. Siam-IDS: handling class imbalance problem in intrusion detection systems using Siamese neural network. Procedia Comput Sci. 2020;171(6):780–9. doi:10.1016/j.procs.2020.04.085.
20. Bedi P, Gupta N, Jindal V. I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. arXiv:2009.10940. 2020.
21. Shanmugam V, Ribeiro B, Silva C, Moniz H. Addressing class imbalance in intrusion detection: a comprehensive evaluation of machine learning approaches. Electronics. 2025;14(1):69. doi:10.3390/electronics14010069.

22. Neupane S, Ables J, Anderson W, Mittal S, Rahimi S, Banicescu I, et al. Explainable intrusion detection systems (X-IDS): a survey of current methods, challenges, and opportunities. In: Proceedings of the 2022 IEEE International Conference on Big Data (Big Data); 2022 Dec 17–20; Osaka, Japan. doi:10.1109/ACCESS.2022.3216617.
23. Keshk M, Koroniotis N, Pham N, Moustafa N, Turnbull B, Zomaya AY. An explainable deep learning-enabled intrusion detection framework in IoT networks. Inf Sci. 2023;639(7):119000. doi:10.1016/j.ins.2023.119000.
24. Chen X, Wang Y, Li Z, Zhang H. Explainable deep learning-based feature selection and intrusion detection for encrypted traffic. Sensors. 2024;24(16):5223. doi:10.3390/s24165223.
25. Dai H, Sun C, Sun X, Lv B. Malware detection based on periodic behavior. In: Proceedings of the 2023 IEEE International Conference on Big Data (BigData); 2023 Dec 15–18; Sorrento, Italy. Piscataway, NJ, USA: IEEE; 2023. p. 6325–7. doi:10.1109/ICAICA58456.2023.10405439.
26. van der Eijk V. Detecting Cobalt Strike beacons in NetFlow data. Amsterdam, Netherlands: University of Amsterdam; 2020.
27. Zhang Y, Dong H, Nottingham A, Buchanan M, Brown DE, Sun Y. Global analysis with aggregation-based beaconing detection across large campus networks. In: Proceedings of the 39th Annual Computer Security Applications Conference (ACSAC '23); 2023 Dec 4–8; Austin, TX, USA. New York, NY, USA: ACM; 2023. p. 565–79. doi:10.1145/3627106.3627126.
28. Sheridan S, Kechadi T, Le-Khac NA. Detection of DNS-based covert channel beacon signals. J Inf Warf. 2015;14(4):46–58.
29. Sachintha S, Ahmed S, Madurapperuma S. Data exfiltration through electromagnetic covert channels in industrial control systems. Appl Sci. 2023;13(5):2928. doi:10.3390/app13052928.
30. Moore AW, Zuev D. Internet traffic classification using Bayesian analysis techniques. In: Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems; 2005 Jun 6–10; Banff, Canada. New York, NY, USA: ACM; 2005. p. 50–60.
31. Sheridan S, Keane A. Improving the stealthiness of DNS-based covert communication. In: Proceedings of the 16th European Conference on Cyber Warfare and Security (ECCWS 2017); 2017 Jun 29–30; Dublin, Ireland. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 433–41.
32. Sharma A, Lashkari AH. A survey on encrypted network traffic: a comprehensive survey of identification/classification techniques, challenges, and future directions. Comput Netw. 2025;257(3):110984. doi:10.1016/j.comnet.2024.110984.
33. Rezaei S, Liu X. Deep learning for encrypted traffic classification: an overview. IEEE Commun Mag. 2019;57(5):76–81. doi:10.1109/mcom.2019.1800819.
34. Liu Z, Xie Y, Luo Y, Wang Y, Ji X. TransECA-Net: a transformer-based model for encrypted network traffic classification. Appl Sci. 2025;15(6):2977. doi:10.3390/app15062977.
35. Wang X, Yuan Q, Wang Y, Gou G, Gu C, Yu G, et al. Combine intra- and inter-flow: a multimodal encrypted traffic classification model driven by diverse features. Comput Netw. 2024;245(4):110403. doi:10.1016/j.comnet.2024.110403.
36. Nascita A, Aceto G, Ciuonzo D, Montieri A, Persico V, Pescapé A. A survey on explainable artificial intelligence for internet traffic classification and prediction, and intrusion detection. IEEE Commun Surv Tutor. 2024;27(5):3165–98. doi:10.1109/COMST.2024.3504955.
37. Lin X, Xiong G, Gou G, Li Z, Shi J, Yu J. ET-BERT: a contextualized datagram representation with pre-training transformers for encrypted traffic classification. In: Proceedings of the ACM Web Conference 2022; 2022 Apr 25–29; Lyon, France. New York, NY, USA: ACM; 2022. p. 633–42. doi:10.1145/3485447.3512217.
38. Zhang H, Yu L, Xiao X, Li Q, Mercaldo F, Luo X, et al. TFE-GNN: a temporal fusion encoder using graph neural networks for fine-grained encrypted traffic classification. In: Proceedings of the ACM Web Conference 2023; 2023 Apr 30–May 4; Austin, TX, USA. New York, NY, USA: ACM; 2023. p. 2066–75. doi:10.1145/3543507.3583227.
39. Masukawa R, Yun S, Jeong S, Huang W, Ni Y, Bryant I, et al. PacketCLIP: multi-modal embedding of network traffic and natural language semantics through contrastive pre-training and hierarchical graph reasoning. Front Artif Intell. 2025;8:1593944. doi:10.3389/frai.2025.1593944.

40. Gu J, Zhong Y, Yu X. Bridging packet and session: cross-level dual-attention networks for encrypted traffic classification. J King Saud Univ Comput Inf Sci. 2026;38(3):79. doi:10.1007/s44443-026-00470-7.
41. Anderson B, McGrew D. Identifying encrypted malware traffic with contextual flow data. In: Proceedings of the 9th ACM Workshop on Artificial Intelligence and Security (AISec '16); 2016 Oct 28; Vienna, Austria. New York, NY, USA: ACM; 2016. p. 35–46.
42. Zhou J, Fu W, Hu W, Sun Z, He T, Zhang Z. Challenges and advances in analyzing TLS 1.3-encrypted traffic: a comprehensive survey. Electronics. 2024;13(19):3900. doi:10.3390/electronics13204000.
43. Al Homsi M. Investigating the adoption of QUIC and its impact on network monitoring and cybersecurity. [Master's thesis]. Stockholm, Sweden: KTH Royal Institute of Technology; 2025.
44. Azab A, Khasawneh M, Alrabaee S, Choo KR, Sarsour M. Network traffic classification: techniques, datasets, and challenges. Digit Commun Netw. 2024;10(3):676–92. doi:10.1016/j.dcan.2022.09.009.
45. Wang Z, Zhu X, Hu Y, An M, Zhao J. Malware traffic classification using convolutional neural network for representation learning. In: Proceedings of the International Conference on Information Networking (ICOIN 2017); 2017 Jan 11–13; Da Nang, Vietnam. Piscataway, NJ, USA: IEEE; 2017. p. 712–7.
46. Meira J. Structured behavior analysis on encrypted traffic. [Master's thesis]. Lisbon, Portugal: Instituto Superior Técnico; 2020.
47. García S, Grill M, Stiborek J, Zunino A. An empirical comparison of botnet detection methods. Comput Secur. 2014;45:100–23. doi:10.1016/j.cose.2014.05.011.
48. Lashkari AH, Draper-Gil G, Mamun MSA, Ghorbani AA. Characterization of Tor traffic using time-based features. In: Proceedings of the 3rd International Conference on Information Systems Security and Privacy (ICISSP 2017); 2017 Feb 19–21; Porto, Portugal. Setúbal, Portugal: SciTePress; 2017. p. 253–62.
49. Fortinet. What is encryption? Definition, types & benefits [Internet]. Sunnyvale, CA, USA: Fortinet; 2025 [cited 2026 Apr 7]. Available from: https://www.fortinet.com/resources/cyberglossary/encryption.
50. Moskal S, Yang SJ. Cyberattack action-intent framework for mapping intrusion observables. arXiv:2002.07838. 2020.
51. Drašar M, Moskal S, Yang S, Zat'ko P. Session-level adversary intent-driven cyberattack simulator. In: Proceedings of the 2020 Winter Simulation Conference (WSC); 2020 Dec 14–18; Orlando, FL, USA. Piscataway, NJ, USA: IEEE; 2020. p. 1497–508. doi:10.1109/DS-RT50469.2020.9213690.
52. Singh K, Kashyap A, Cherukuri AK. Interpretable anomaly detection in encrypted traffic using SHAP with machine learning models. arXiv:2505.16261. 2025.
53. Li M, Pi D, Chen Y. Adversarial malicious encrypted traffic detection based on interpretable analysis. Symmetry. 2022;14(11):2329. doi:10.3390/sym14112329.
54. Mavroeidis V, Bromander S. Cyber threat intelligence model: an evaluation of taxonomies, sharing standards, and ontologies in cyber threat intelligence. J Cybersecur. 2021;7(1):tyab018. doi:10.1093/cybsec/tyab018.
55. Xu J, Zhang Y, Zhou K, Wang Q, Hua M, Shan L, et al. A cascaded broad learning network embedded image features for malware traffic classification. IEEE Trans Cogn Commun Netw. 2025;11(4):2426–39. doi:10.1109/TCCN.2024.3522143.
56. Wei L, Wang Y, Li X, Li J, Huang Y, Liu Z. A detection method for malware communication traffic via encrypted traffic analysis. IEEE Internet Things J. 2025;12(16):33482–92. doi:10.1109/JIOT.2025.3577245.
57. Lin P, Ye K, Xu CZ, Yang J. A novel multimodal deep learning framework for encrypted traffic classification. IEEE/ACM Trans Netw. 2023;31(3):1369–84. doi:10.1109/TNET.2022.3215507.
58. Seok B, Sohn K. Adversarial attacks on pre-trained deep learning models for encrypted traffic analysis. J Web Eng. 2024;23(6):749–68. doi:10.13052/jwe1540-9589.2361.
59. Baee MAR, Simpson L, Armstrong W. Anomaly detection in the key-management interoperability protocol using metadata. IEEE Open J Commun Soc. 2024;5:156–70. doi:10.1109/OJCS.2024.3386715.
60. Liu Y, Wang X, Qu B, Zhao F. ATVITSC: a novel encrypted traffic classification method based on deep learning. IEEE Trans Inf Forensics Secur. 2024;19(2):9374–89. doi:10.1109/TIFS.2024.3433446.

61. Malekghaini N, Akbari E, Salahuddin MA, Limam N, Boutaba R, Mathieu B, et al. AutoML4ETC: automated neural architecture search for real-world encrypted traffic classification. IEEE Trans Netw Serv Manag. 2024;21(3):2715–30. doi:10.1109/TNSM.2023.3324936.
62. Hu H, Li Z, Doerr C. BoNC: discovering and classifying novel encrypted botnet traffic. IEEE Open J Commun Soc. 2025;6(1):10845–60. doi:10.1109/OJCOMS.2025.3638985.
63. Shi Y, Li G, Wu J, Li J, Fang H. BPF-DAG: byte-packet-flow feature fusion via dynamic attributed graph for reliable encrypted traffic classification. IEEE Trans Inf Forensics Secur. 2026;21(1):197–211. doi:10.1109/TIFS.2025.3643127.
64. Xie R, Cao J, Zhu Y, Zhang Y, He Y, Peng H, et al. Cactus: obfuscating bidirectional encrypted TCP traffic at client side. IEEE Trans Inf Forensics Secur. 2024;19(11):7659–73. doi:10.1109/TIFS.2024.3442530.
65. Seydali M, Khunjush F, Akbari B, Dogani J. CBS: a deep learning approach for encrypted traffic classification with mixed spatio-temporal and statistical features. IEEE Access. 2023;11:141674–702. doi:10.1109/ACCESS.2023.3343189.
66. Liu M, Zhang Y, Liu Y, Zhang Z, Wang X. CBSeq: a channel-level behavior sequence for encrypted malware traffic detection. IEEE Trans Inf Forensics Secur. 2023;18:5011–25. doi:10.1109/TIFS.2023.3300521.
67. Wang Z, Chen X, Li Y, Zhang J. CL-ViME: contrastive learning and vision mixture of experts for encrypted traffic classification. IEEE Trans Netw Serv Manag. 2025;23:1422–34. doi:10.1109/TNSM.2025.3650038.
68. Elayan H, Aloqaily M, Guizani M. CMFTC: cross-modality fusion efficient multitask encrypted traffic classification in IIoT environment. IEEE Internet Things J. 2024;11(9):15984–96. doi:10.1109/TNSE.2023.3279427.
69. Bakhshi T, Ghita B, Manaf NAA. CMTSNN: a deep learning model for multiclassification of abnormal and encrypted traffic of internet of things. IEEE Internet Things J. 2023;10(18):16074–88. doi:10.1109/JIOT.2023.3244544.
70. Wang C, Xu J, Chen Y, Li Z. Data augmentation-based enhancement for efficient network traffic classification. IEEE Access. 2024;12(3):42135–49. doi:10.1109/ACCESS.2024.3525000.
71. Jerabek K, Hynek K, Rysavy O, Burgetova I. DNS over HTTPS detection using standard flow telemetry. IEEE Access. 2023;11:50000–12. doi:10.1109/ACCESS.2023.3275744.
72. Yun X, Wang Y, Zhang H, Li Z, Chen M. Encrypted TLS traffic classification on cloud platforms. IEEE/ACM Trans Netw. 2023;31(2):876–89. doi:10.1109/TNET.2022.3191312.
73. Tong V, Dao C, Tran HA, Tran DD, Binh HTT, Hoang-Nam T, et al. Encrypted traffic classification through deep domain adaptation network with smooth characteristic function. IEEE Trans Netw Serv Manag. 2025;22(1):331–43. doi:10.1109/TNSM.2025.3534791.
74. Li H, Kang H, Liu C, Wang R, Li J, Sun G, et al. Encrypted traffic detection in resource constrained IoT networks: a diffusion model and LLM integrated framework. IEEE Trans Netw Sci Eng. 2026;13(5):5324–44. doi:10.1109/TNSE.2025.3649259.
75. Yang Q, He W, Chen M, Du H, Shao S, Wu F, et al. End-to-end open-set semi-supervised learning for fine-grained encrypted traffic classification. IEEE Trans Inform Forensic Secur. 2026;21:1347–61. doi:10.1109/TIFS.2026.3653575.
76. Conti M, Gangwal A, Lal C, Losiouk E. Enhancing accountability, resilience, and privacy of intelligent networks with XAI. IEEE Commun Surv Tutor. 2024;26(2):1193–221. doi:10.1109/OJCOMS.2025.3608784.
77. Zhang J, Li Y, Wang X, Liu H. EO-EPTC: end-to-end original traffic-based encrypted proxy traffic classification framework. IEEE Access. 2024;12:84577–91. doi:10.1109/TIFS.2025.3646874.
78. Lopez-Martin M, Carro B, Sanchez-Esguevillas A, Lloret J. Flow-based encrypted network traffic classification with graph neural networks. Comput Netw. 2024;245(1514):110398. doi:10.1016/j.comnet.2024.110398.
79. Hassan A, Rauf A, Latif R, Baig AI, Kanso H. From packets to semantics: transformer-driven privacy-preserving ETA for OTT app classification. IEEE Access. 2025;13(11):207694–709. doi:10.1109/ACCESS.2025.3640294.
80. Wang H, Zhang Q, Liu Y, Li Z. Identification of malicious encrypted traffic through feature fusion. IEEE Access. 2024;12:102345–58. doi:10.1109/ACCESS.2023.3279120.
81. Shen M, Ye K, Liu X, Zhu L, Kang J, Yu S, et al. Machine learning-powered encrypted network traffic analysis: a comprehensive survey. IEEE Commun Surv Tutor. 2023;25(1):791–830. doi:10.1109/COMST.2022.3208196.

82. Zhao Y, Chen X, Li P, Wang J. MT-DEGCL: multi-task encrypted traffic classification with dual embedding and graph contrastive learning. IEEE Trans Netw Serv Manag. 2025;22(3):1987–2001. doi:10.1109/TIFS.2026.3664007.
83. Park JT, Choi Y, Cho BS, Kim SH, Kim MS. Multi-level pre-training for encrypted network traffic classification. IEEE Access. 2025;13(8):68643–59. doi:10.1109/ACCESS.2025.3559068.
84. Ding K, Hu X, Shu Z, Wu H. NFTracker: fine-grained NFT behavior traffic identification over encrypted tunnel. Comput Netw. 2025;22(2):1455–68. doi:10.1109/TCCN.2026.3658756.
85. Goez Sanchez GD, Beyazit EA, Fletscher LA, Botero JF, Gaviria N, Latré S, et al. Resource-efficient spectrum-based traffic classification on constrained devices. IEEE Open J Commun Soc. 2024;5:3066–88. doi:10.1109/OJCOMS.2024.3396077.
86. Lin W, Chen Y. Robust network traffic classification based on information bottleneck neural network. IEEE Access. 2024;12(5):150169–79. doi:10.1109/ACCESS.2024.3477466.
87. Yang L, Liu L, Huang JJ, Shi J, Fu S, Wang Y, et al. Robustness matters: pre-training can enhance the performance of encrypted traffic analysis. IEEE Trans Inf Forensics Secur. 2025;20:10588–603. doi:10.1109/TIFS.2025.3613970.
88. Dai J, Xu X, Gao H, Wang X, Xiao F. SHAPE: a simultaneous header and payload encoding model for encrypted traffic classification. IEEE Trans Netw Serv Manag. 2023;20(2):1993–2007. doi:10.1109/TNSM.2022.3213758.
89. Lu M, Zhou B, Bu Z. Two-stage distillation-aware compressed models for traffic classification. IEEE Internet Things J. 2023;10(16):14152–66. doi:10.1109/JIOT.2023.3263487.
90. Luxemburk J, Hynek K, Plný R, Čejka T. Universal embedding function for traffic classification via QUIC domain recognition pretraining: a transfer learning success. arXiv:2502.12930. 2025.
91. Kotak J, Yankelev I, Bibi I, Elovici Y, Shabtai A. VPN-encrypted network traffic classification using a time-series approach. IEEE Trans Netw Serv Manag. 2025;22(2):2225–42. doi:10.1109/TNSM.2025.3543903.
92. Chen Z, Cheng G, Niu D, Qiu X, Zhao Y, Zhou Y. WFF-EGNN: encrypted traffic classification based on weaved flow fragment via ensemble graph neural networks. IEEE Trans Mach Learn Commun Netw. 2023;1:389–411. doi:10.1109/TMLCN.2023.3323915.
93. Li J, Zhang Y, Chen X, Wang H. WF-Transformer: learning temporal features for accurate anonymous traffic identification by using transformer networks. IEEE Trans Inf Forensics Secur. 2024;19:1123–38. doi:10.1109/TIFS.2023.3318966.
94. Fu C, Li Q, Xu K. Detecting unknown encrypted malicious traffic in real time via flow interaction graph analysis. In: Proceedings of the Network and Distributed System Security Symposium (NDSS); 2023 Feb 27–Mar 3; San Diego, CA, USA. doi:10.14722/ndss.2023.23080.
95. Wang Q, Xu T. Research on network traffic classification method based on dual-branch spatial-temporal attention. Acad J Comput Inf Sci. 2024;7(9):1–10. doi:10.25236/AJCIS.2024.070901.
96. Zeleke SN, Jember AF, Bochicchio MA. Integrating explainable AI for effective malware detection in encrypted network traffic. arXiv:2501.05387. 2025.
97. Zhao D, Jiang B, Liu S, Cui S, Shen M, Han D, et al. Language of network: a generative pre-trained model for encrypted traffic comprehension. arXiv:2505.19482. 2025.
98. Dong W, Jing Y, Lin X, Gou G, Xiong G. Deep learning and pre-training technology for encrypted traffic classification: a comprehensive review. Neurocomputing. 2025;617(5):128444. doi:10.1016/j.neucom.2024.128444.
99. Chehade A, Ragusa E, Gastaldo P, Zunino R. Adversarial robustness of traffic classification under resource constraints: input structure matters. In: Proceedings of the International Symposium on Networks, Computers and Communications (ISNCC); 2025 Jul 15–17; Rome, Italy. Piscataway, NJ, USA: IEEE; 2025. p. 1–6. doi:10.48550/arXiv.2512.02276.
100. Fu C, Li Q, Xu K. Flow interaction graph analysis for unknown encrypted malicious traffic detection. IEEE/ACM Trans Netw. 2024;32(4):2785–99. doi:10.1109/TNET.2024.3370851.
101. Sahu S, Mohapatra S, Pattanayak BK. Application of EM-transformer hybrid model in real-time detection of directed DDoS attacks on botnet devices. J King Saud Univ Comput Inf Sci. 2024;36(3):101912. doi:10.1016/j.jksuci.2024.101912.
102. Abbas A, Salahuddin M, Khan MZ, Khan AA, Zaman FU, Inam SA, et al. Machine learning-based hybrid technique to enhance cyber-attack perspective. J Cloud Comput. 2025;14(1):57. doi:10.1186/s13677-025-00782-5.

103. Hendaoui F, Ferchichi A, Trabelsi L, Meddeb R, Ahmed R, Khelifi MK. Advances in deep learning intrusion detection over encrypted data with privacy preservation: a systematic review. Cluster Comput. 2024;27(7):8683–724. doi:10.1007/s10586-024-04424-4.
104. Jia J, Shen M, Yuan Q, Liu Y, Wang J, Kong J, et al. Adaptive detection of encrypted malware traffic via fully convolutional masked autoencoders. Front Comput Sci. 2026;20(4):2004804. doi:10.1007/s11704-025-41273-9.
105. Beg R, Pateriya RK, Tomar DS, Solanki S. Detecting malware evidences through static and dynamic information using extreme machine learning for forensic analysis. Discov Comput. 2025;28(1):329. doi:10.1007/s10791-025-09842-5.
106. Zaimi R, Safi Eljil K, Hafidi M, Lamia M, Nait-Abdesselam F. An enhanced mechanism for malicious URL detection using deep learning and DistilBERT-based feature extraction. J Supercomput. 2025;81(2):438. doi:10.1007/s11227-024-06908-x.