IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

# Survey on Malicious Packet Classification Techniques

**Pratibha Tambewagh**
Lecturer, Department of Information Technology
Bharati Vidyapeeth Institute of Technology, Kharghar, Navi Mumbai, Maharashtra, India

**Abstract**: *Malicious packet classification is a critical component in safeguarding network infrastructures against evolving cyber threats. Traditional methods, such as port-based and payload-based inspections, have become less effective due to the increasing use of encryption and sophisticated evasion techniques by attackers. Consequently, there has been a significant shift towards leveraging machine learning (ML) and deep learning (DL) approaches to enhance detection capabilities. Recent advancements highlight the efficacy of transformer-based models in analyzing raw payload data for identifying malicious traffic. For instance, a novel approach utilizing transformers demonstrated notable accuracy in distinguishing between benign and malicious packets, even when relying solely on payload bytes . Similarly, the adaptation of natural language processing techniques, such as Word2Vec, to network packet data has shown promise in automatic feature extraction, facilitating improved classification performance .The integration of convolutional neural networks (CNNs) and gated recurrent units (GRUs) has also been explored, particularly in the context of Internet of Things (IoT) networks, achieving high accuracy in both binary and multiclass classification tasks. Furthermore, the application of variational autoencoders (VAEs) has been investigated for detecting anomalies in network traffic, offering potential in identifying denial-of-service (DoS) and distributed denial-of-service (DDoS) attacks. In addition to these methods, the emergence of self-supervised learning and contrastive learning techniques has opened new avenues for enhancing the robustness of malicious traffic classification systems. These approaches aim to improve the generalization capabilities of models, particularly in scenarios involving encrypted or obfuscated traffic. Despite these advancements, challenges persist, including the need for large, labeled datasets, the handling of encrypted traffic, and the development of models resilient to adversarial attacks. Future research directions may focus on addressing these issues, as well as exploring hybrid models that combine multiple learning paradigms to bolster detection efficacy.*

**Keywords**: Packet classification,DD0S attacks, Cyber security

## I. INTRODUCTION

In today's digitally connected world, packet classification plays a vital role in securing network infrastructures by analyzing and categorizing data packets based on their header or payload attributes. When extended to malicious packet classification, this process becomes essential for identifying and mitigating cyber threats such as Distributed Denial of Service (DDoS) attacks, port scans, malware communications, and other forms of unauthorized access or exploitation. With increasing sophistication in cyberattacks, malicious packet classification serves as a foundational mechanism in modern cybersecurity systems, enabling timely threat detection and response.

The significance of this domain is underscored by its central role in intrusion detection systems (IDS), firewalls, and anomaly detection frameworks that protect critical data and infrastructure. However, numerous challenges persist, including the growing use of encryption, which limits visibility into payloads; the imbalance between benign and malicious data in training datasets; the emergence of adversarial evasion techniques; and the need for real-time processing in high-speed networks. This survey aims to explore and compare various approaches—ranging from


DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

classical rule-based and statistical techniques to advanced machine learning and deep learning models—highlighting recent developments from 2023 to 2024 and identifying future research directions in malicious packet classification

## II. TAXONOMY OF MALICIOUS PACKETS

### 1. Based on Intent

Understanding the intent behind malicious packets helps in designing specialized detection and mitigation strategies. Different attack objectives generate distinct traffic patterns and packet behaviors.

**DDoS Packets:**

Distributed Denial of Service (DDoS) attacks aim to overwhelm a target server or network by flooding it with massive volumes of packets, often from distributed sources (botnets). These packets may appear normal individually but are sent in high frequency or volume to exhaust bandwidth, memory, or CPU resources. Common types include TCP SYN floods, UDP floods, and HTTP floods. Their signatures often include patterns like repeated requests from spoofed IPs, malformed headers, or unusual traffic bursts.

**Malware Command-and-Control (C2) Traffic:**

After infecting a system, many types of malware need to communicate with their controller (the attacker). This communication occurs via command-and-control channels, which send instructions and receive data from compromised machines. C2 traffic may use common protocols (like HTTP, HTTPS, or DNS) to evade detection and can be disguised to blend in with normal network traffic. Indicators include persistent low-volume traffic to rare domains or IP addresses, odd timing patterns, or encrypted payloads with consistent packet sizes.

**Scanning/Probing Attempts:**

These are reconnaissance activities used by attackers to map the network, identify open ports, detect services, and probe vulnerabilities before launching an attack. Scanning packets may be ICMP (ping sweeps), TCP SYN packets sent to many ports (port scanning), or HTTP requests targeting specific known vulnerabilities. Detection often relies on identifying unusual traffic patterns, such as high connection attempts to multiple destinations or ports in a short time frame.

**Data Exfiltration Packets:**

In data breaches, exfiltration involves transferring sensitive information from inside the network to external destinations. This can be done using DNS tunneling, HTTP POST requests, or encrypted channels like HTTPS or SSH. Exfiltration traffic is often low and stealthy, using protocols and ports that are generally trusted. It may also exhibit compression or encoding techniques to pack large amounts of data into smaller payloads.

### 2. Based on Layer (OSI Model)

Malicious packets can also be categorized by the network stack layer they operate on, which helps determine the detection mechanisms and countermeasures needed.

- Network Layer (Layer 3):This layer handles IP addressing and routing. Attacks here often involve IP spoofing, where attackers forge the source IP address of packets to hide their identity or to bypass access controls. Other techniques include fragmenting packets in unusual ways to evade detection or probing routing behaviors. Tools like NetFlow and packet inspection systems are used to detect irregularities at this layer.
- Transport Layer (Layer 4):This layer governs end-to-end communication using TCP or UDP. A classic attack at this layer is the TCP SYN flood, where an attacker sends a barrage of TCP connection requests without completing the handshake, thereby exhausting server resources. Other examples include UDP floods, which exploit the stateless nature of UDP, and TCP reset (RST) attacks, which terminate active connections. These attacks can often be detected through packet rate analysis, incomplete connections, and abnormal session behavior.
- Application Layer (Layer 7):This is the most targeted and flexible layer, where attackers exploit application-specific protocols like HTTP, DNS, and SMTP. Examples include:
- DNS Tunneling: Using DNS requests/responses to covertly transmit data.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

- HTTP Floods: Bombarding web servers with seemingly legitimate HTTP requests to consume resources.
- SQL Injection payloads embedded in HTTP requests. Since this layer often involves encrypted content (e.g., HTTPS), detection usually relies on traffic patterns, domain analysis, or behavioral anomaly detection rather than payload inspection.

## III. CLASSICAL APPROACHES

### a. Signature-Based Detection

Signature-based detection is one of the earliest and most widely used methods for identifying malicious packets in network traffic. This technique relies on maintaining a database of known attack signatures—predefined patterns or characteristics that are unique to specific types of malicious activity. These signatures may include specific byte sequences in packet payloads, header anomalies, known malicious IP addresses, or sequences of protocol misuse. During packet inspection, the system compares each packet or flow against this database to identify matches that suggest a known threat.

Tools like Snort and Suricata are prominent open-source Intrusion Detection Systems (IDS) that employ signature-based detection. These tools use rule-based configurations that define what constitutes suspicious activity. For example, a rule might specify that a packet with a particular TCP flag combination, originating from a known botnet IP, and targeting a specific port (e.g., 445 for SMB exploits) should trigger an alert. These systems operate efficiently in real-time environments and are commonly deployed at network perimeters to scan large volumes of incoming and outgoing traffic.

The primary strength of signature-based detection lies in its speed and accuracy when dealing with known threats. Because the detection process is essentially a pattern-matching task, it can be optimized for high-throughput environments with minimal resource consumption. Additionally, the false positive rate is typically low, as each signature is designed to match a very specific attack characteristic.

However, a major limitation of this approach is its inability to detect zero-day or previously unseen attacks. If an attacker modifies the structure of a known exploit or introduces a novel method that has no existing signature, the system will not recognize it as malicious. This makes signature-based systems reactive by nature—they can only protect against threats that have already been discovered and analyzed. To stay effective, such systems require constant updates to their signature databases, which can be labor-intensive and time-delayed.

Because of these limitations, signature-based detection is often combined with anomaly-based or behavior-based methods in modern hybrid intrusion detection systems, allowing for more adaptive and comprehensive threat detection.

Certainly! Here's a detailed explanation of **Statistical Methods** for malicious packet classification, suitable for the next section of your survey paper:

### b. Statistical Methods

Statistical methods are a foundational approach to anomaly detection in network traffic, including the classification of malicious packets. Unlike signature-based techniques that rely on known patterns, statistical methods examine the **distribution and statistical properties of packet attributes**—such as packet sizes, inter-arrival times, flow durations, and source/destination IP distributions—to detect deviations from established baselines. The central idea is that normal network behavior follows predictable statistical patterns, and deviations from these patterns may indicate malicious activity.

One widely used technique in this category is **entropy-based detection**, where the entropy (i.e., the randomness or unpredictability) of packet attributes like destination IPs or ports is monitored. A sudden spike or drop in entropy may suggest activities such as scanning (many destinations) or flooding (repeated access to a single destination). Similarly, **threshold-based rules** can flag anomalous flows—such as connections exceeding a certain packet rate or duration—based on predefined statistical limits. For example, if the average number of packets per flow exceeds a threshold, it may indicate a DoS attack or data exfiltration attempt.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

**The advantage** of statistical methods lies in their **ability to detect unknown or zero-day threats**, as they do not depend on prior knowledge of attack signatures. They are particularly useful for identifying subtle or slow attacks (e.g., low-rate DoS, beaconing from C2 servers) that blend in with normal traffic but statistically deviate over time.

However, statistical approaches **suffer from high false positive rates**, especially in dynamic or heterogeneous network environments where normal behavior changes frequently. A legitimate spike in traffic (e.g., software updates, backup operations) may be wrongly flagged as malicious. Additionally, **tuning statistical models**—choosing the right thresholds, windows, and metrics—can be difficult and environment-specific. Poor tuning can either make the system overly sensitive or too lenient, reducing its effectiveness.

Despite these challenges, statistical methods remain a crucial component of hybrid detection systems and are often used to complement signature- and AI-based approaches, especially in real-time monitoring and early-warning systems.

## IV. MACHINE LEARNING APPROACHES

### a. Supervised Learning

Supervised learning is a widely adopted machine learning approach for malicious packet classification, where a model is trained using labeled datasets that clearly distinguish between benign and malicious traffic. This approach assumes the availability of a comprehensive dataset in which each instance (i.e., packet or flow) is tagged with its correct class label. The goal is to learn a function that maps input features—such as packet size or protocol type—to a specific output label (e.g., "normal" or "malicious").

Several algorithms are commonly used in this domain:

- Decision Trees offer interpretable models that split the dataset based on feature thresholds, making them useful for understanding attack logic.
- Support Vector Machines (SVMs) are powerful classifiers, particularly effective in high-dimensional feature spaces, often used for binary classification of attacks.
- Random Forests and XGBoost are ensemble methods that combine multiple decision trees to improve accuracy and reduce overfitting.
- k-Nearest Neighbors (k-NN) classifies new instances based on the majority class of their nearest data points, though it can be computationally expensive for large datasets.

Supervised models rely heavily on labeled datasets for training and validation. Popular datasets include NSL-KDD, a refined version of the older KDD'99 dataset; UNSW-NB15, which offers modern and diverse attack types; and CICIDS2017, which includes realistic traffic scenarios and up-to-date attack patterns like DDoS, brute-force, and infiltration. These datasets provide labeled flows or packets with a wide range of features used during model training.

Common features extracted from traffic include:

- Packet-level features: size, protocol, flag bits, TTL
- Flow-level features: duration, number of packets per flow, bytes transferred
- Network-level metadata: source/destination IP and port, time intervals
- These features help capture both structural and behavioral aspects of network traffic.

Supervised learning models are particularly effective in scenarios where labeled data is abundant and up-to-date, offering high accuracy in identifying known attack patterns. However, they may struggle with concept drift (i.e., evolving attack techniques) and zero-day attacks, which are not represented in the training data. Additionally, their performance heavily depends on the quality and diversity of the training dataset, making dataset curation a critical task.

### b. Unsupervised and Anomaly Detection

Unsupervised learning and anomaly detection techniques play a critical role in malicious packet classification, especially when labeled data is scarce or when the goal is to detect novel, previously unseen attacks. Unlike supervised learning, these methods do not rely on labeled training data. Instead, they learn the structure or distribution of normal (benign) traffic and flag deviations from this learned behavior as potential anomalies—making them well-suited for zero-day threats and unknown attack vectors.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

Several algorithms are widely used in this category:

- Isolation Forest is an ensemble-based anomaly detection method that isolates anomalies by recursively partitioning data. Malicious packets, which tend to behave differently from the majority, are easier to isolate and thus receive higher anomaly scores.
- One-Class SVM (Support Vector Machine) attempts to find a decision boundary that encloses most of the benign data points. Packets falling outside this boundary are considered suspicious.

Autoencoders, a type of neural network used in deep learning, are trained to reconstruct normal traffic patterns. During inference, packets that result in high reconstruction errors (i.e., those the model cannot reproduce well) are flagged as anomalies, indicating potential malicious behavior.

These approaches are particularly effective for environments where attack types are constantly evolving, and new threats may not resemble historical attack data. They are also useful in scenarios where obtaining large amounts of labeled malicious traffic is impractical.

However, unsupervised methods face several challenges:

- Imbalanced training data: In real-world traffic, malicious packets are rare compared to benign ones. Training models on mostly benign data can lead to poor detection of subtle or sophisticated attacks.
- High sensitivity to noise: Legitimate but unusual user behavior (e.g., a software update or large file transfer) can be misclassified as an anomaly, resulting in high false positive rates.
- Interpretability: Many anomaly detection models, especially deep learning-based ones, lack transparency in their decision-making processes, making it difficult for analysts to validate alerts.

Despite these limitations, unsupervised and anomaly detection techniques are a crucial component of modern hybrid detection systems, where they complement signature and supervised approaches by offering the flexibility to detect new and evolving threats without prior knowledge.

**c. Deep Learning**

Deep learning has revolutionized the field of malicious packet classification by enabling automated feature extraction and modeling of complex patterns in network traffic. Unlike traditional machine learning models that rely heavily on manual feature engineering, deep learning methods can learn hierarchical representations directly from raw or minimally processed data, leading to improved detection accuracy, especially in complex and evolving threat landscapes.

Convolutional Neural Networks (CNNs) are primarily known for their success in image processing, but they have been effectively adapted to network security tasks by representing packets or flows as images. For instance, raw packet bytes can be converted into 2D byte plots or grayscale images, where CNNs excel at extracting spatial patterns indicative of malicious payloads or anomalous byte distributions. This approach captures subtle byte-level correlations that traditional handcrafted features might miss.

Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks, are well-suited for analyzing sequential data, making them ideal for modeling time-dependent behavior in network flows. By processing sequences of packets or flows over time, RNNs can detect patterns such as slow, stealthy command-and-control communications or multi-step attack sequences that manifest over extended periods. LSTMs, in particular, address the vanishing gradient problem, allowing models to learn long-term dependencies in traffic data.

More recently, Transformer architectures, originally developed for natural language processing, have emerged as a promising approach for sequence-based packet inspection. Transformers leverage self-attention mechanisms to capture relationships between distant elements in a sequence without relying on recurrent connections, enabling efficient modeling of complex interactions in packet flows or payloads. This makes them particularly useful for large-scale traffic analysis and real-time threat detection.

Preprocessing techniques play a crucial role in adapting raw network data for deep learning models. Two common strategies are:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

- Packet as Image: Converting packet bytes into 2D arrays or images (byte plots), facilitating CNN-based feature extraction.
- Packet as Sequence: Encoding packet contents or flow features as sequences or embeddings, suitable for RNNs, LSTMs, or Transformers.

While deep learning offers superior detection performance and adaptability, it also comes with challenges such as the need for large labeled datasets, high computational requirements, and interpretability issues. Nonetheless, deep learning remains a cutting-edge approach that continues to advance the state of malicious packet classification.

## V. ENCRYPTED PACKET CLASSIFICATION

With the widespread adoption of encryption protocols such as SSL/TLS, traditional packet inspection methods that rely on analyzing packet payloads have become largely ineffective. This shift presents a significant challenge for malicious packet classification, as payload-based detection techniques, including deep packet inspection (DPI), cannot access the encrypted contents. To address this, encrypted packet classification leverages **flow metadata and statistical features** rather than payload data to identify malicious activity.

Key metadata features include **packet sizes**, **inter-arrival times** between packets, and **directionality** (i.e., whether packets are incoming or outgoing). These features can reveal patterns indicative of specific applications or attack types. For example, certain malware command-and-control channels may exhibit characteristic timing intervals or packet size distributions that differ from benign encrypted traffic.

Since DPI is ineffective against encrypted streams, alternative approaches such as **SSL/TLS fingerprinting** have been developed. SSL fingerprinting analyzes parameters from the handshake phase—like cipher suites, certificate details, and protocol versions—to identify anomalous or suspicious encrypted sessions. Moreover, **flow-based behavior detection** uses machine learning models trained on metadata sequences to detect deviations from normal encrypted traffic behavior, helping to uncover stealthy attacks such as encrypted tunneling or data exfiltration.

Encrypted packet classification thus represents a crucial area of research for modern network security, balancing privacy preservation with effective threat detection in an increasingly encrypted internet landscape.

## VI. REAL-TIME AND EDGE-BASED DETECTION

With the exponential growth of network traffic and the increasing complexity of cyber threats, real-time malicious packet classification has become essential for timely detection and mitigation. This urgency has driven the development of **lightweight models** optimized for deployment on edge devices such as routers, IoT gateways, and network switches, where computational resources and power are limited.

To meet these constraints, techniques from **Tiny Machine Learning (TinyML)**—which focus on creating compact, efficient models—are employed. Model optimization strategies such as **pruning** (removing redundant neurons or connections) and **quantization** (reducing the precision of model parameters) help shrink model size and speed up inference without significantly sacrificing accuracy. These adaptations allow models to run directly on edge hardware, enabling faster threat response and reducing the need to send large volumes of data to centralized servers for analysis.

Hardware acceleration further enhances real-time performance. Implementations on **Field-Programmable Gate Arrays (FPGAs)** and **Graphics Processing Units (GPUs)** provide parallel processing capabilities tailored to the demands of high-throughput packet inspection. FPGAs, in particular, offer energy-efficient and customizable architectures suited for fixed-function network security tasks, while GPUs excel at accelerating deep learning inference for complex models.

However, deploying detection systems in real-time and on edge devices requires balancing the trade-off between **accuracy and latency**. Highly accurate, deep learning models often demand substantial computation, increasing detection delay. Conversely, lighter models optimized for speed may miss subtle attack signatures. Designing effective real-time systems involves carefully tuning this balance to maintain acceptable detection rates while ensuring timely alerts.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

Real-time and edge-based detection represents a critical frontier in malicious packet classification, enabling scalable, responsive defenses across distributed network environments.

## VII. PUBLIC DATASETS FOR RESEARCH

Effective research in malicious packet classification heavily depends on the availability of quality datasets that represent diverse and realistic network traffic, including a wide range of attack scenarios. Several public datasets have been widely used by researchers to train, test, and benchmark classification models:

- **CICIDS2017:** Developed by the Canadian Institute for Cybersecurity, this dataset contains highly realistic network traffic combining both normal activities and various types of attacks such as DDoS, brute force, infiltration, and web attacks. It provides detailed flow-based features along with precise labels, making it suitable for supervised learning and anomaly detection studies.
- **UNSW-NB15**: This dataset offers a more modern and diverse set of attack types, including reconnaissance, exploits, and malware behaviors. It includes both raw packets and extracted features, which help researchers develop models that address contemporary threats. Its realistic traffic generation and comprehensive labeling make it a valuable resource for evaluating classification algorithms.
- **NSL-KDD**: An improved version of the older KDD'99 dataset, NSL-KDD resolves many of the latter's issues such as redundant records and imbalanced classes. Although it reflects attack types from an earlier era, it remains widely used due to its accessibility and extensive documentation. It provides a solid benchmark for evaluating baseline classification performance.
- **TON_IoT:** Targeting the emerging threat landscape of the Internet of Things (IoT), TON_IoT includes telemetry data from IoT devices, network traffic, and operating system logs. This dataset captures attack scenarios specific to IoT ecosystems, such as Mirai botnet activities and other IoT-specific exploits, making it crucial for research focused on securing IoT networks.

These publicly available datasets serve as foundational resources that support reproducibility and comparative evaluation in malicious packet classification research, enabling advancements in detection methodologies.

## VIII. EVALUATION METRICS

Evaluating the performance of malicious packet classification systems requires a set of well-defined metrics that quantify their effectiveness in accurately detecting threats while minimizing errors. The following metrics are commonly used in research to assess classification models:

| Metric | Description |
|---|---|
| **Accuracy** | Measures the overall rate of correctly classified packets, including both malicious and benign traffic. It provides a general sense of model performance but can be misleading in imbalanced datasets. |
| **Precision** | Indicates the proportion of packets classified as malicious that are actually malicious. High precision means fewer false alarms (false positives). |
| **Recall (Sensitivity)** | Reflects the ability of the model to identify all malicious packets correctly. High recall indicates fewer missed attacks (false negatives). |
| **F1-Score** | The harmonic mean of precision and recall, providing a balanced measure when there is an uneven class distribution or when both false positives and false negatives are critical. |
| **False Positive Rate (FPR)** | The rate at which benign packets are incorrectly classified as malicious, impacting the usability and trustworthiness of the detection system. |
| **Detection Latency** | The time taken by the system to analyze and classify packets, which is critical for real-time applications where delays can lead to missed prevention opportunities. |

Selecting appropriate evaluation metrics is essential to ensure that malicious packet classification models meet the specific requirements of security environments, balancing detection accuracy with operational efficiency.


DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

## IX. CHALLENGES AND FUTURE DIRECTIONS

The field of malicious packet classification faces several ongoing challenges that researchers must address to improve detection accuracy, robustness, and deployment feasibility.

Encrypted Traffic: With the increasing use of encryption protocols like TLS/SSL, traditional payload-based inspection techniques are rendered ineffective. Future research must focus on innovative methods that rely on flow metadata, timing patterns, and statistical features to accurately classify malicious behavior without direct access to payload data.

Adversarial Machine Learning: Attackers are actively developing sophisticated methods to evade machine learning (ML) models by crafting malicious packets that can bypass detection. This arms race necessitates the development of robust ML models that are resilient against adversarial attacks and can detect subtle evasion tactics.

Zero-Day Attacks: Detecting previously unseen threats remains a critical challenge. Improving the generalization capabilities of classifiers through transfer learning, unsupervised anomaly detection, and continual learning can help systems identify novel attacks without prior labeled examples.

Model Interpretability: Many advanced ML and deep learning models operate as "black boxes," providing little insight into their decision-making processes. Enhancing model interpretability is essential for building trust in security-critical applications, enabling analysts to understand, validate, and act upon model predictions effectively.

Edge Deployment: As real-time detection increasingly moves toward edge devices with limited computational resources, adapting models for these constrained environments is paramount. Techniques such as model compression, pruning, quantization, and hardware acceleration will be vital to maintain detection accuracy while meeting latency and power consumption requirements.

Addressing these challenges will drive the next generation of malicious packet classification systems, making them more adaptive, secure, and practical for deployment in diverse and evolving network environments.

## X. CONCLUSION

This survey reviewed key techniques in malicious packet classification, spanning classical methods such as signature- and statistical-based detection, as well as modern approaches leveraging supervised, unsupervised, and deep learning algorithms. Each methodology offers unique strengths and faces distinct challenges, highlighting the complexity of detecting malicious packets in today's dynamic threat landscape.

Given the evolving nature of cyberattacks and the increasing use of encryption, there is a growing need for hybrid and adaptive models that combine the precision of signature-based methods with the flexibility of AI-driven anomaly detection. Furthermore, improving the interpretability of these models is essential to foster trust and enable actionable insights in real-world security operations.

Finally, advancing research in this domain requires access to more diverse, open-source datasets and the development of standardized benchmarks. These resources will facilitate reproducibility, fair comparison of approaches, and continuous innovation to tackle emerging threats effectively.

## REFERENCES

**[1].** Zhang, Y., Liu, J., & Wang, X. (2023). Deep Learning for Malicious Traffic Classification: A Comprehensive Survey. IEEE Communications Surveys & Tutorials, 25(1), 500-529.

**[2].** Singh, A., & Kumar, V. (2024). A Review on Anomaly Detection Techniques in Encrypted Network Traffic. Journal of Network and Computer Applications, 210, 103547.

**[3].** Chen, H., Zhao, R., & Li, M. (2023). Hybrid Models for Real-Time Detection of DDoS Attacks Using Edge Computing. IEEE Access, 11, 23456-23468.

**[4].** Gupta, P., & Sharma, S. (2024). Adversarial Machine Learning in Network Security: Challenges and Solutions. ACM Computing Surveys, 56(3), Article 58.

**[5].** Kim, D., Park, J., & Lee, S. (2023). Transformer-Based Models for Network Traffic Analysis and Malicious Packet Detection. Proceedings of the 2023 IEEE International Conference on Communications (ICC), 1123-1128.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 8, May 2025
Impact Factor: 7.67

**[6].** Alqahtani, H., & Alshammari, R. (2024). Lightweight Deep Learning Models for IoT Security: TinyML Approaches and Applications. Sensors, 24(3), 1258.

**[7].** Rivera, A., & Diaz, J. (2023). Anomaly Detection in Encrypted Network Traffic Using Autoencoders and Flow Metadata. IEEE Transactions on Information Forensics and Security, 18, 407-420.

**[8].** Kumar, S., & Bhatia, A. (2023). A Survey on Public Datasets for Network Intrusion Detection and Classification. Journal of Cybersecurity and Privacy, 3(1), 89-108.

**[9].** Hassan, M., & Tariq, U. (2024). Challenges and Future Directions in Malicious Packet Classification Using Machine Learning. Future Generation Computer Systems, 145, 238-252.

**[10].** Mahesh Datta Sai Ponnuru, Likhitha Amasala, Tanu Sree Bhimavarapu, Guna Chaitanya Garikipati(2023) A Malware Classification Survey on Adversarial Attacks and Defences

**[11].** Huang H., Zhang X., Lu Y., Li Z., Zhou S. 2024 BSTFNet: An Encrypted Malicious Traffic Classification Method Integrating Global Semantic and Spatiotemporal Features

**[12].** Pakmehr, A., Aßmuth, A., Taheri, N., et al. July 2024 DDoS Attack Detection Techniques in IoT Networks: A Survey

**[13].** Liu, Y.; Wang, Z.; Pang, S.; Ju, L. 2024 Distributed Malicious Traffic Detection

**[14].** ACM Digital Library,2025 A Survey on Encrypted Network Traffic: A Comprehensive Survey of Identification/Classification Techniques, Challenges, and Future Directions

**[15].** Anish Singh Shekhawat, Fabio Di Troia, Mark Stamp December 2023 Feature Analysis of Encrypted Malicious Traffic

**[16].** Kyle Stein, Arash Mahyari, Guillermo Francia III, Eman El-Sheikh March 2024

**[17].** A Transformer-Based Framework for Payload Malware Detection and Classification

**[18].** Soumyadeep Hore, Jalal Ghadermazi, Diwas Paudel, et al. May 2023 Deep PackGen: A Deep Reinforcement Learning Framework for Adversarial Network Packet Generation

**[19].** MDPI May 2024 Encrypted Network Traffic Analysis and Classification Utilizing Machine Learning

**[20].** Raj Davis, Jinsheng Xu, Kaushik Roy January 2024 Classifying Malware Traffic Using Images and Deep Convolutional Neural Network