КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (26), 2024
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2024.26.639
UDC 004.056.5:004.7

Pavlo Pidhornyi
PhD student
Sumy State University, Sumy, Ukraine
ORCID ID: 0009-0008-8604-8051
pashapro49@gmail.com

ANALYTICAL REVIEW OF MODELS AND SYSTEMS FOR NETWORK TRAFFIC CLASSIFICATION

Abstract. This article presents a comprehensive analytical review of network traffic classification models and systems, essential for managing the complexities of modern network environments. The study covers traditional and advanced methods, including statistical approaches, machine learning, and deep learning techniques, highlighting their strengths and limitations. It also explores both commercial and open-source systems, offering insights into their practical applications and effectiveness. The rapid evolution of network technologies has significantly enhanced global data exchange and connectivity but has also introduced new challenges in managing and securing complex network environments. As networks expand and grow more heterogeneous, the ability to classify and manage network traffic efficiently becomes critical for optimizing network performance, ensuring security, and supporting operational continuity. Network traffic classification is an essential function that enables network administrators to apply appropriate policies, detect anomalies, and prevent malicious activities. Traditional classification methods, such as payload-based detection and port-based classification, are increasingly unreliable due to the rise of encrypted traffic and evolving network protocols, which obscure important traffic details. Therefore, it is necessary to explore advanced approaches such as machine learning, deep learning, and transfer learning. These methods can handle large volumes of data and adapt to new traffic types, improving accuracy and efficiency. This paper presents a comprehensive review of existing models and systems for network traffic classification, including both commercial and open-source solutions. The study covers a range of techniques from traditional statistical methods to advanced machine learning and artificial intelligence (AI)-based techniques. Special attention is given to key performance indicators (KPIs) such as latency, packet loss, jitter, throughput, and bandwidth utilization, which are vital for effective network monitoring and optimization. In light of increasing encryption and evolving cyber threats, the paper emphasizes the importance of adaptive systems, real-time analysis, and the integration of AI and machine learning technologies. The review concludes by identifying future research directions and emerging trends, crucial for developing the next generation of traffic classification systems.

Kaywords: network traffic classification; key performance indicators; real-time traffic analysis; transfer learning.

INTRODUCTION

The advancement of network technologies has brought benefits in connectivity and data exchange; on the other hand, it has introduced threats in managing and securing networks. As networks grow increasingly complex and heterogeneous, the ability to effectively identify and categorize network traffic becomes critical for optimizing performance, ensuring security, and maintaining operations. Thus, network traffic classification stands as a critical function based on various parameters such as protocol, source, destination, and data patterns. This process enables network administrators to apply appropriate policies, detect anomalies, and prevent malicious activities.

Traditional methods of traffic classification often fall short in the face of encrypted traffic, evolving protocols, and real-time processing demands. It is necessary to address these challenges by providing a comprehensive review of contemporary models and systems and exploring such innovative approaches as machine learning and big data analytics. Such a study can identify gaps in current methodologies and propose potential avenues for future research.

This paper aims to provide a thorough analytical review of the models and systems employed in network traffic classification.

It covers a range of methods, from traditional statistical approaches to advanced machine learning and deep learning techniques, highlighting their strengths and limitations. Additionally, the paper explores commercial and open-source systems that implement these methods, providing a practical perspective on the effectiveness and efficiency of different traffic classification models. Furthermore, it presents the mathematical foundations of traffic analysis, real-time traffic monitoring techniques, and the security implications.

In light of the ever-evolving network environments and the growing cyber threats, this review tries to identify key research areas and emerging trends that will shape the next generation of network traffic classification systems.

INDICATORS AND CRITERIA OF NETWORK TRAFFIC CLASSIFICATION

Network traffic classification is a process that involves the identification and categorization of traffic flows based on a variety of attributes such as packet headers, flow statistics, and application signatures. This procedure is fundamental for quality of service (QoS) management, traffic shaping, and intrusion detection [18]. Accurate classification allows network administrators to allocate resources more effectively, prioritize critical applications, and protect the network from malicious activities [1].

Key performance indicators (KPIs) like bandwidth utilization, latency, packet loss, and error rates are essential for effective traffic classification [3], [4]. These indicators not only provide insights into the network health but also assist in detecting abnormal traffic patterns that could indicate security breaches [5] or network bottlenecks. Below is a list of traffic KPIs that should be paid continuous attention.

1. Latency

Definition: The time it takes for data to travel from the source to the destination.

Traffic Feature: Identifies the nature of traffic; low latency for real-time applications like VoIP and gaming, higher latency for non-real-time applications like file downloads.

Crucial Change: Sudden increase.

Possible Root: Network congestion, routing issues, or a problem with network hardware.

Possible Effect: Slows down real-time applications leading to poor user experience.

2. Packet Loss

Definition: The percentage of packets lost during transmission across the network.

Traffic Feature: Indicates the reliability and quality of the network; critical for real-time applications.

Crucial Change: Significant increase.

Possible Root: Network congestion, faulty network hardware, or interference.

Possible Effect: Degrades the quality of service, leading to disruptions in voice and video communications.

3. Jitter

Definition: The variability in packet arrival times across the network.

Traffic Feature: Important for applications requiring consistent delivery times, such as VoIP and video conferencing.
Crucial Change: Increase in variability.
Possible Root: Network congestion, routing instability, or inconsistent network paths.
Possible Effect: Causes poor quality in real-time communications, resulting in choppy audio or video.

4. Throughput

Definition: The actual rate at which data is successfully transmitted over the network.
Traffic Feature: Reflects the network's capacity to handle data; higher throughput indicates better network performance.
Crucial Change: Sudden drop.
Possible Root: Network congestion, bandwidth limitations, or hardware failure.
Possible Effect: Slower data transfer rates, affecting the performance of applications and services.

5. Bandwidth Utilization

Definition: The amount of bandwidth being used compared to the total available bandwidth.
Traffic Feature: Indicates the level of network load; helps in assessing network capacity and usage.
Crucial Change: Unusually high utilization.
Possible Root: Excessive data transfers, DDoS attacks, or misconfigured applications.
Possible Effect: Network congestion, reduced performance, and potential service outages.

6. Error Rate

Definition: The frequency of errors occurring in the transmission of data packets.
Traffic Feature: Indicates the health of the network; lower error rates signify more reliable data transmission.
Crucial Change: Increase in error rates.
Possible Root: Hardware issues, poor signal quality, or interference.
Possible Effect: Data corruption, need for retransmissions, and overall decrease in network efficiency.

7. Network Availability/Uptime

Definition: The percentage of time the network is operational and available for use.
Traffic Feature: Measures the reliability and resilience of the network infrastructure.
Crucial Change: Decrease in availability or uptime.
Possible Root: Hardware failures, software bugs, or power outages.
Possible Effect: Loss of productivity, potential data loss, and impact on critical business operations.

8. Round Trip Time (RTT)

Definition: The time it takes for a data packet to travel from the source to the destination and back.
Traffic Feature: Affects the speed and responsiveness of applications, especially interactive ones.
Crucial Change: Increase in RTT.
Possible Root: Network congestion, inefficient routing, or distant servers.
Possible Effect: Slower application performance, especially noticeable in interactive or real-time applications.

9. Connection Establishment Time

Definition: The time required to establish a connection between two network endpoints.

Traffic Feature: Important for services requiring quick connections, such as web browsing and online transactions.
Crucial Change: Prolonged connection times.
Possible Root: DNS resolution issues, network congestion, or authentication problems.
Possible Effect: Delays in accessing services, leading to poor user experience and potential timeouts.

10. Unusual Traffic Volume

Definition: An unexpected increase or decrease in the volume of network traffic.
Traffic Feature: May indicate changes in network usage patterns or potential security issues.
Crucial Change: Sudden spikes or drops in traffic volume.
Possible Root: DDoS attacks, data exfiltration, or changes in user behavior.
Possible Effect: Network congestion, potential security breaches, or system overload.

Network traffic can be classified into various categories based on different criteria, namely:

1. By Application Type

Web Traffic: Includes HTTP and HTTPS traffic, typically generated by web browsers and web applications.
Email Traffic: Traffic related to email services, including protocols like SMTP, IMAP, and POP3.
Streaming Media: Includes traffic from audio and video streaming services, such as YouTube, Netflix, and Spotify.
VoIP Traffic: Voice over IP traffic, often using protocols like SIP, RTP, and H.323.
File Transfer: Traffic involving the transfer of files, using protocols like FTP, SFTP, and SMB.
Peer-to-Peer (P2P): Traffic generated by P2P applications like BitTorrent.
Gaming Traffic: Traffic between online gaming services and platforms.

2. By Protocol

TCP Traffic: Traffic using the Transmission Control Protocol, common for reliable, ordered, and error-checked delivery.
UDP Traffic: Traffic using the User Datagram Protocol, often used for applications that require speed and efficiency over reliability.
ICMP Traffic: Internet Control Message Protocol traffic, used for diagnostic and error messages (e.g., ping).

3. By Service Type

Real-Time Communications: Traffic that requires low latency and jitter, such as VoIP and video conferencing.
Transactional Traffic: Traffic involving transactions, often seen in online banking, e-commerce, and financial services.
Bulk Transfer: Large data transfers, such as backups, software updates, and data replication.

4. By Security Level

Encrypted Traffic: Traffic that is encrypted for security purposes, including HTTPS, VPNs, and SSH.
Unencrypted Traffic: Traffic that is not encrypted, more vulnerable to interception and eavesdropping.

5. By Network Layer

Layer 2 (Data Link Layer): Traffic classified by MAC addresses, VLANs, or Ethernet protocols.
Layer 3 (Network Layer): Traffic classified by IP addresses and routing information.
Layer 4 (Transport Layer): Traffic classified by transport layer ports and protocols (e.g., TCP/UDP).
Layer 7 (Application Layer): Traffic classified by application layer protocols and content (e.g., HTTP, DNS).
6. By Source/Destination
Local Traffic: Traffic that stays within a local network or organization.
External Traffic: Traffic that crosses the boundary of a local network, often interacting with external networks or the internet.
Intra-domain Traffic: Traffic within the same domain or organization.
Inter-domain Traffic: Traffic between different domains or organizations.
7. By Usage Pattern
Interactive Traffic: Traffic that involves real-time user interaction, such as browsing, gaming, or video calls.
Non-Interactive Traffic: Automated or scheduled traffic, such as system updates, backups, or automated scripts.
8. By Purpose
Business-Critical Traffic: Traffic essential for the operation of critical business functions.
Non-Critical Traffic: Traffic not essential to core business functions, such as recreational or personal use traffic.
9. By Traffic Behavior
Constant Bit Rate (CBR) Traffic: Traffic with a consistent flow of data, often seen in streaming or real-time applications.
Variable Bit Rate (VBR) Traffic: Traffic with fluctuating data rates, typical in internet browsing and general application use.
10. By Compliance and Regulation
Regulated Traffic: Traffic that must comply with legal and regulatory requirements, such as data protection regulations (e.g., GDPR, HIPAA).
Non-Regulated Traffic: Traffic not subject to specific legal or regulatory compliance.

Significant challenge for the classification process is the increasing use of encryption, which obscures traffic contents and complicates traditional classification methods. Additionally, the rapid evolution of network applications creates a constantly shifting landscape of traffic types.

CLASSIFICATION MODELS AND TECHNIQUES

Classification models and techniques used in network traffic classification range from traditional rule-based and statistical methods to advanced machine learning approaches.

Rule-based methods can be divided into two groups:

- Payload-Based Detection [18], [22] with predefined patterns or signatures to classify traffic. Commonly used in intrusion detection systems (IDS) for identifying known malicious traffic [5].
- Port-Based Classification [3], [18] based on well-known port numbers associated with specific services (e.g., HTTP on port 80). This method is straightforward but increasingly unreliable due to port-hopping and encrypted traffic.

- Heuristic Rules use expert knowledge to define rules and heuristics for classifying traffic. While not flexible, they can be effective in specific contexts [12].
- *Statistical methods* build patterns of network behavior [10] based on:
- Frequency Analysis that examines the frequency of certain patterns or features within the traffic to identify and classify it.
- Flow-Based Analysis that uses statistical properties of traffic flows (e.g., flow duration, packet size) to classify the traffic type.
- Statistical Anomaly Detection that identifies deviations from a statistical model of normal network behavior [2].

To overcome the challenges of widespread encryption, new types, and high traffic volumes, modern traffic classification approaches increasingly rely on advanced techniques, including machine learning (Fig. 1), deep learning, and transfer learning.

Machine Learning Models

- Supervised Learning: Requires labeled training data where the model learns to classify traffic based on known categories.
  - Decision Trees: A tree-like model of decisions that splits the data into categories based on feature values.
  - Random Forests: An ensemble method using multiple decision trees to improve classification accuracy.
  - Support Vector Machines (SVM): A method that finds the optimal hyperplane separating different traffic classes in a feature space.
  - Neural Networks: Especially deep learning models like CNNs and RNNs, which can learn complex patterns in traffic data.
- Unsupervised Learning: Does not require labeled training data and is used to find natural groupings in the data.
  - Clustering Techniques: Such as K-Means, DBSCAN, which group traffic into clusters based on similarity.
  - Principal Component Analysis (PCA): A dimensionality reduction technique that transforms data into principal components to highlight variations.

Semi-Supervised Learning: Combines a small amount of unlabeled data to improve learning efficiency.

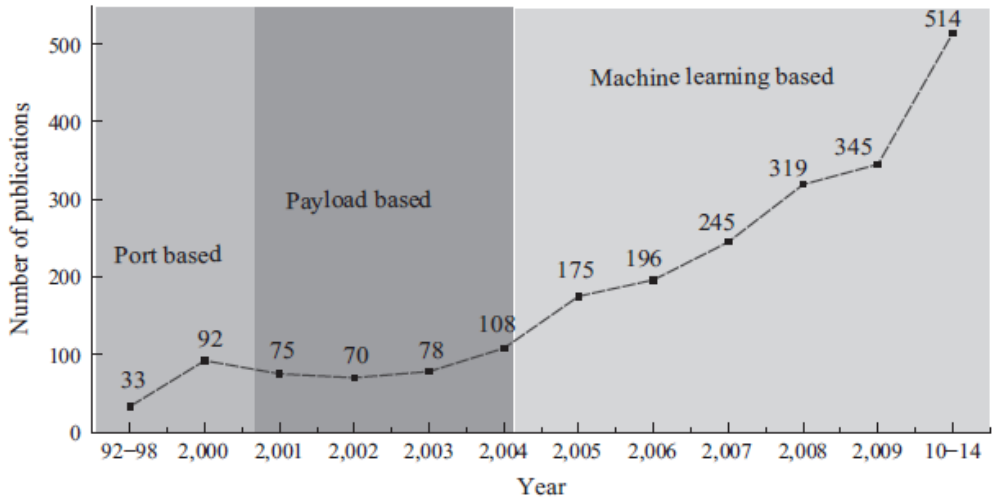

Fig. 1. Introduction of machine learning approaches into evolution of network traffic classification [22]

Deep Learning Techniques

- Convolutional Neural Networks (CNNs): Useful for capturing spatial dependencies in traffic data, often used for classifying encrypted traffic.
- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Suitable for capturing temporal dependencies in sequential data, such as packet flows over time.

Transfer Learning

- Pre-Trained Models: Uses models pre-trained on large datasets and fine-tunes them on specific traffic classification tasks, particularly useful when labeled data is scarce.

Any of these AI techniques is more flexible and scalable in comparison to traditional methods, since they are capable of analyzing vast amounts of traffic data, learning from patterns, and adapting to new traffic types without explicit reprogramming.

Among other approaches, one should mention *Hidden Markov Models (HMM)* based on the probabilistic transitions between states built on traffic sequences and detect patterns [22].

Most of systems combine multiple methods into *Hybrid Techniques* (e.g., rule-based with machine learning, or statistics with HMM) to improve classification accuracy and robustness.

Classification models and techniques are selected and combined based on the specific requirements of the network environment, the nature of the traffic, and the goals of the analysis (e.g., performance monitoring, QoS, security detection, compliance). In practice, the choice can also depend on the availability of labeled data, computational resources, and the need for real-time analysis.

SYSTEMS FOR NETWORK TRAFFIC CLASSIFICATION

Systems for network traffic classification include commercial solutions, open-source applications, and custom tools.

Widely used commercial systems offer a range of features, including real-time monitoring, traffic analysis, and security capabilities, which are essential for managing network performance, security, and compliance to the regulatory requirements (Table 1).

Table 1

Commercial Systems for Network Traffic Monitoring & Classification

Name	Description	Key Features	Use Cases
Cisco NetFlow	A feature on Cisco routers and switches that provides detailed visibility into network traffic flows. It captures and analyzes IP traffic data as it enters or exits an interface.	• Flow-based traffic monitoring and analysis. • Identification of applications, users, and devices consuming network bandwidth. • Security monitoring and threat detection.	Network performance management, capacity planning, anomaly detection
Palo Alto Networks	A suite of products, including next-generation firewalls (NGFW) and advanced threat protection systems, designed to provide comprehensive	• Application identification and traffic classification using App-ID technology. • User and device visibility with User-ID and Device-ID. • Integrated threat prevention, including antivirus, anti-	Network security, threat detection and prevention, traffic visibility

	network security and traffic analysis	spyware, and intrusion prevention.	
Fortinet FortiGate	Series of firewalls provide high-performance network security, including advanced traffic classification and inspection capabilities.	• Deep packet inspection (DPI) for detailed traffic analysis. • Application control and web filtering. • Integration with FortiGuard Labs for threat intelligence and updates.	Comprehensive security, application control, secure SD-WAN.
SolarWinds Network Performance Monitor (NPM)	A comprehensive network monitoring tool that helps in detecting, diagnosing, and resolving network performance issues.	• Real-time network traffic analysis and monitoring. • Customizable dashboards and alerts. • Integration with NetFlow Traffic Analyzer for detailed traffic insights.	Network performance monitoring, capacity planning, troubleshooting
Darktrace	Cybersecurity solutions powered by AI, focusing on threat detection and autonomous response	• Uses machine learning and AI for real-time traffic analysis and anomaly detection. • Self-learning technology that adapts to the unique patterns of network behavior. • Autonomous response capabilities to mitigate threats.	Threat detection and response, anomaly detection, cybersecurity
Check Point Next-Generation Firewalls	Next-generation firewall solutions that include advanced traffic inspection and threat prevention capabilities	• Advanced traffic classification and application control. • Integrated threat prevention with firewall, IPS, and antivirus. • Comprehensive threat intelligence and reporting.	Network security, traffic management, threat prevention
Zscaler Internet Access	A cloud-based security platform that provides secure internet and web access, with comprehensive traffic analysis capabilities	• Secure web gateway with traffic classification and control. • SSL inspection and advanced threat protection. • User and application visibility.	Cloud security, secure web access, compliance

Open-source applications and tools offer reliable solutions for traffic monitoring, analysis, and security, able to match and excel the capabilities of commercial solutions [23].

Wireshark is the most widely used open-source network protocol analyzer that allows users to capture and interactively browse the network traffic for troubleshooting, analysis, and software development [15]. The key aspects and features of Wireshark are listed below:

- Packet Capture: capture live packet data from a network interface or read packets from a previously saved capture file. It supports various file formats, including pcap (used by tcpdump), pcapng, and others.
- Deep Packet Inspection: detailed visibility into network protocols at different layers, from the physical to the application layer. This includes the ability to see the entire packet contents, including headers and payloads.
- Protocol Analysis: dissect hundreds of network protocols, providing detailed analysis and display of protocol-specific information. It helps in understanding protocol operations, detecting anomalies, and debugging implementations.

- Filtering and Searching: powerful filtering capabilities allow users to focus on specific packets of interest based on criteria like IP address, protocol type, port number, or specific field values.
- Statistics and Summaries: a variety of statistical analyses and summaries, including protocol hierarchy statistics, conversations, endpoint statistics, and IO graphs. These features help users understand traffic distribution and identify trends or issues.
- Visualization and Analysis: graphical tools for visualizing various aspects of traffic data, such as flow graphs, IO graphs, and round-trip time (RTT) measurements. Good for identifying patterns, bottlenecks, and anomalies (Fig. 2).
- Customization and Extensibility: supports user-created plugins and dissectors, which allow for the extension of its capabilities to recognize and dissect new or proprietary protocols.
- Multi-Platform Support: Windows, macOS, Linux, and other Unix-like systems.
- Security Features: tools for inspecting secure protocols, such as SSL/TLS, and can be configured to decrypt captured traffic if the necessary encryption keys are available.

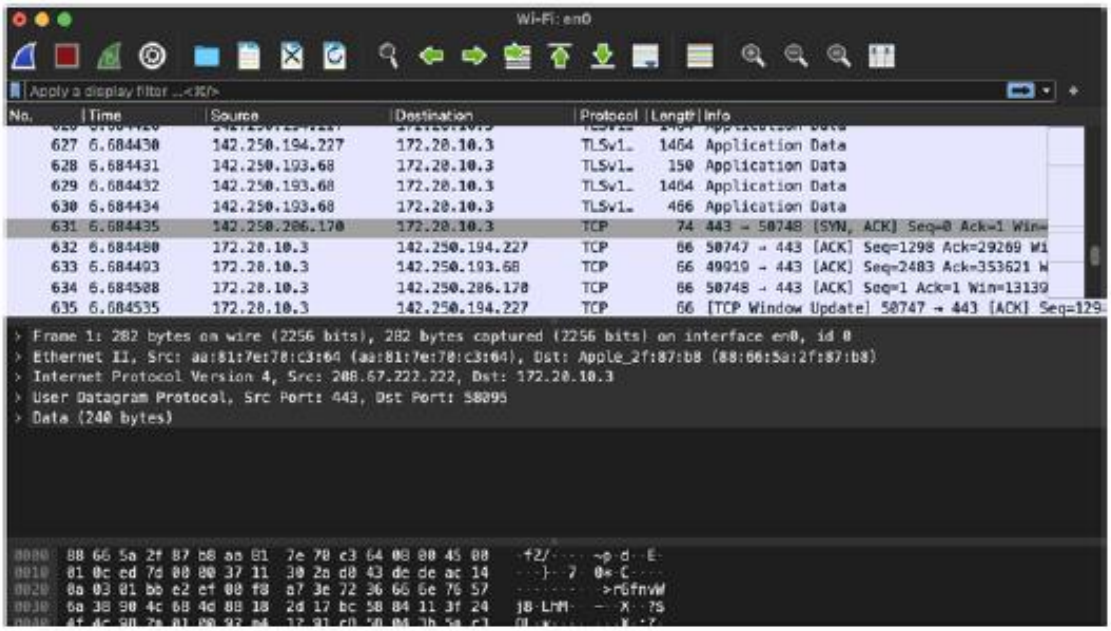

Fig. 2. Wireshark user interface [15]

Using Wireshark, network administrators can diagnose network issues (such as latency problems, dropped packets, and protocol errors) and identify the root cause by providing detailed packet-level analysis [12]. Security professionals can analyze network traffic for signs of malicious activity, such as unusual protocols, data exfiltration, or unauthorized access attempts. Developers working on networked applications or protocols can verify the correctness of their software, getting information of how their data is transmitted and received. Wireshark is also used for educational purposes to teach students about networking and protocol behavior. It has a large and active community, contributing to its development and providing support through forums, mailing lists, and tutorials.

Many open-source systems incorporate AI techniques to provide advanced capabilities such as anomaly detection, predictive analytics, and automated decision-making.

Zeek (formerly Bro) is a powerful network analysis framework that includes extensive capabilities for security monitoring and traffic analysis [24]. It can be extended with user's AI modules for anomaly detection and pattern recognition. By supporting scripts and custom plugins, it enables users to integrate machine learning models for detecting suspicious activities or network anomalies.

ELK Stack is a set of open-source tools for searching, analyzing, and visualizing log data in real time. The core of the package provides built-in machine learning features for anomaly detection, predictive analytics, and time-series forecasting. Users can create custom AI tasks to analyze data patterns and detect unusual behaviors.

Apache Spot is an open-source cybersecurity project focused on providing a comprehensive platform for anomaly detection, threat intelligence, and data analysis. It uses AI, supports data from various sources, and provides tools for analyzing network flows, DNS logs, and other data.

In addition to commercial and open-source systems, businesses and organizations often develop custom solutions to meet specific needs of network traffic classification and security [9]. They include unique features such as proprietary protocols analysis, in-house expertise, industry-specific traffic patterns recognition, or enhanced data privacy measures. Sectors like finance, healthcare, and critical infrastructure generally opt for custom solutions to address unique regulatory requirements and protect sensitive data.

The landscape of network traffic classification and security systems is diverse, encompassing a range of commercial, open-source, and custom solutions. Effective integration of these systems requires a deep understanding of the available technologies, together with a strategic alignment with organizational goals and security policies [7]. Ultimately, the right combination of tools can significantly enhance network performance, security, and compliance, supporting the dynamic needs in modern network environment.

REAL-TIME TRAFFIC ANALYSIS

Real-time traffic analysis refers to the continuous monitoring and evaluation of network traffic as it flows through the network, providing immediate insights and actionable intelligence.

The importance of real-time traffic analysis lies in its ability to provide timely visibility into network conditions, which is essential for various applications. For example, critical real-time services like VoIP or online gaming require maintaining low latency and jitter; otherwise, they can be disrupted [18].

However, real-time traffic analysis presents significant challenges:

1. the volume and velocity of data generated in modern networks overwhelm traditional monitoring systems and call for advanced solutions capable of handling high-throughput data streams [6];
2. the increasing use of encryption obscures payload data, making it difficult to apply classic deep packet techniques;
3. the need for real-time processing demands low latency and rapid response capabilities [10].

To address these challenges, the following tools and technologies are employed:

- Stream Processing Frameworks. Platforms like Apache Kafka, Apache Flink, and Spark Streaming are designed to handle real-time data streams. They provide

robust mechanisms for ingesting, processing, and analyzing high-volume data in real-time, enabling the detection of patterns and anomalies as data arrives.

- Time-Series Databases. Databases like InfluxDB and Prometheus are optimized for storing and querying time-series data, which is crucial for tracking metrics such as bandwidth usage, packet loss, and latency over time [3], [7]. These databases support high-speed read and write operations, making them suitable for real-time monitoring.
- Intrusion Detection and Prevention Systems (IDPS). Tools like Suricata and Zeek offer real-time network traffic analysis with a focus on security. They can detect malicious activity, anomalous behavior, and policy violations, providing immediate alerts and response mechanisms [19].
- Visualization and Alerting Platforms. Visualization tools like Grafana (Fig. 3) and Kibana, often used in conjunction with time-series databases and IDPS, provide real-time dashboards and alerting systems [12]. These platforms enable network administrators to visualize traffic patterns, set thresholds, and receive alerts when metrics exceed predefined limits.

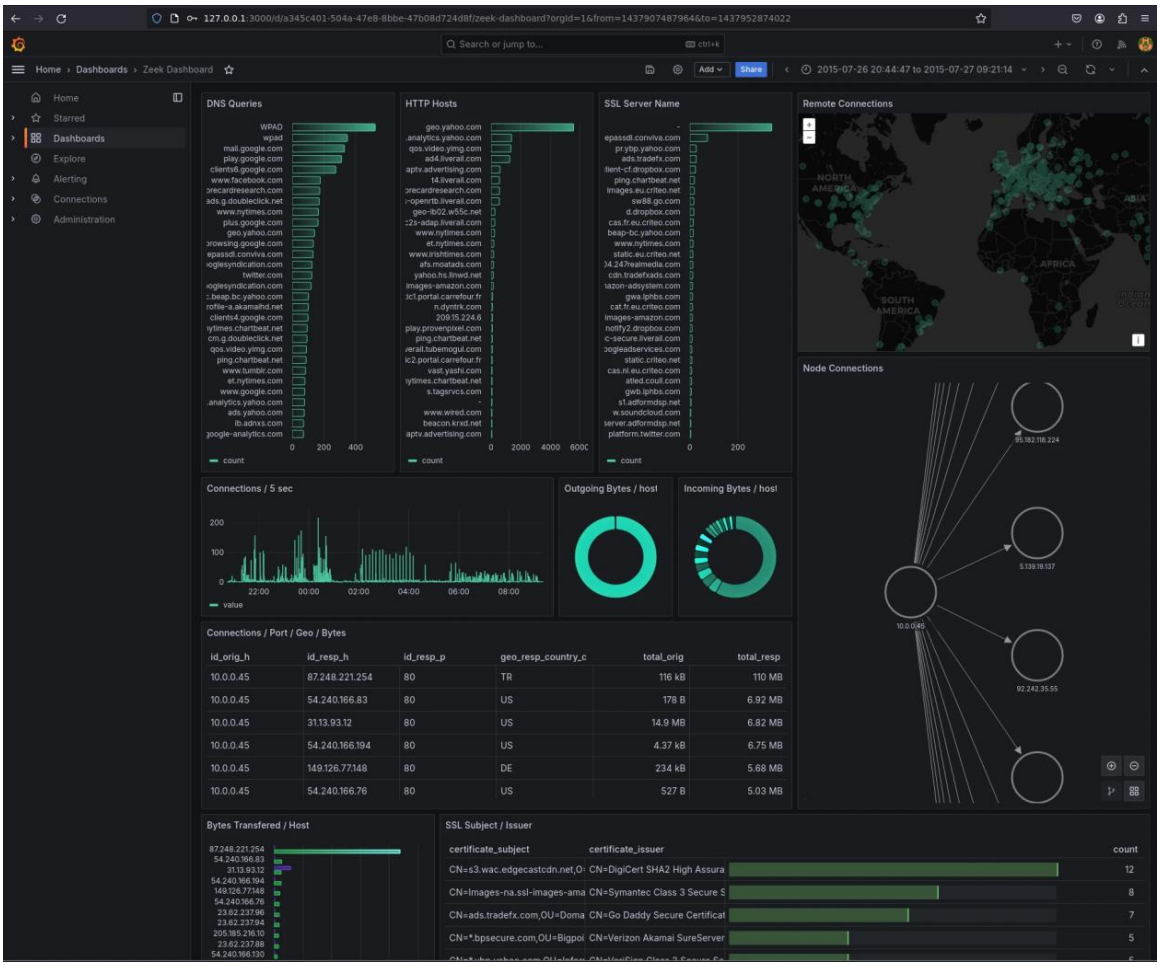

Fig. 3. Zeek dashboard using Graphana [24]

By implementing robust real-time analysis systems, organizations and businesses ensure optimal performance, secure infrastructure, and quick respond to emerging issues, safeguarding both their operations and their users.

FUTURE DIRECTIONS AND TRENDS

Due to advancements in technology, increasing security threats, and the growing complexity of network environments, the field of network traffic classification is rapidly evolving.

Artificial intelligence (AI) and machine learning (ML) automate the classification process, improve accuracy, and adapt to new types of traffic and threats. Future developments are likely to focus on enhancing robustness, interpretability, and efficiency of these models, learning them to handle encrypted traffic and sophisticated threats such as advanced persistent threats (APTs) and zero-day exploits.

Deep learning allows for the analysis of complex traffic patterns and temporal dependencies with the use of convolutional neural networks (CNNs) [17], [21] and recurrent neural networks (RNNs). Future trends may include the development of more specialized neural network architectures tailored to network traffic data [8], [14].

Transfer learning is a machine learning technique where a model pre-trained for a particular task is reused as the starting point for another model that solves a different, but related task. Possible applications of pre-trained models in network traffic classification are:

- Feature Extraction. Pre-trained models, used as feature extractors, can feed the features into other machine learning algorithms for classification.
- Fine-Tuning pre-trained models on specific datasets related to network traffic, improving their ability to classify different types of traffic or detect anomalies.
- Cross-Domain Learning. Applying models trained on one type of network data (e.g., corporate network traffic) to another domain (e.g., cloud traffic), leveraging shared patterns and structures.

Transformers are a type of neural network architecture originally developed for natural language processing (NLP) tasks. In the context of network traffic classification, transformers are valuable for their ability to capture long-range dependencies and contextual information; they are well-suited for analyzing sequential data, such as network traffic flows. The self-attention mechanism allows a transformer to focus on different parts of the input sequence to capture complex dependencies between packets or flows, which are crucial in network traffic classification. Unlike classic ANNs, transformers process input data in parallel, making them more efficient and scalable, especially for large datasets; much more layers and parameters can be added to create “a big picture”.

The integration of transfer learning and transformers offers a powerful approach to network traffic classification. By leveraging pre-trained transformer models, organizations can quickly adapt to new traffic patterns and threats. This combination allows for efficient model training and deployment, particularly in environments with rapidly evolving traffic characteristics.

CONCLUSION

In this review, we have explored the diverse models and systems for network traffic classification used to analyze and secure network environments.

From traditional rule-based methods to advanced machine learning and AI-driven approaches, each has its unique strengths and applications. As networks become more complex and cyber threats more sophisticated, the continued development of adaptive and scalable systems is essential. The adoption of AI and machine learning, particularly techniques like transfer

learning and transformers, promises significant improvements in the field. Looking ahead, the focus will be on refining these technologies to handle the dynamic nature of network traffic, ensure robust and scalable solutions, protect and optimize network operations in real time.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Alkenani, J., & Nassar Khulood A. (2022). Network Monitoring Measurements for Quality of Service: A Review. *Iraqi Journal for Electrical and Electronic Engineering, 18(2),* 33–42. https://doi.org/10.37917/ijeee.18.2.5
2. Bhattacharyya, D, & Kalita, J. (2016). *DDoS Attacks: Evolution, Detection, Prevention, Reaction, and Tolerance.* CRC Press.
3. Bhuyan, M., Bhattacharyya, D., Kalita, J. (2017). *Network Traffic Anomaly Detection and Prevention: Concepts, Techniques, and Tools.* Springer.
4. Biersack, E., Callegari, C., & Matijasevic, M. (2013). *Data Traffic Monitoring and Analysis.* Springer.
5. Chauhan, Aj. S. (2018). *Practical Network Scanning: Capture Network Vulnerabilities Using Standard Tools Such As Nmap and Nessus*. Packt Publishing.
6. Chen, S., Chen, M., & Xiao, Q. (2017). *Measurement for Big Network Data.* Springer.
7. Collins, M. (2017). Network Security through Data Analysis: From Data to Action. O'Reilly Media.
8. Cui, J., Bai, L., Zhang, X., Lin, Z., & Liu, Q. (2024). The Attention-Based Autoencoder for Network Traffic Classification with Interpretable Feature Representation. *Symmetry, 16(5)* 589. https://doi.org/10.3390/sym16050589
9. Kim, E. (n. d.). *A Survey on Network Measurement: Concepts, Techniques, and Tools*. https://www.cs.helsinki.fi/u/kervasti/projects/A%20Survey%20on%20Network%20Measurement%20-%20Concepts,%20Techniques,%20and%20Tools%20-%20Kim%20Ervasti%20-%2031-12-2016.pdf
10. Stênio, F. (2017). *Performance Evaluation for Network Services, Systems and Protocols.* Springer.
11. Fichera, J., & Bolt, S. (2012). *Network Intrusion Analysis: Methodologies, Tools, and Techniques for Incident Analysis and Response.* Syngress.
12. Forshaw, J. (2018). *Attacking Network Protocols: A Hacker's Guide to Capture, Analysis, and Exploitation.* No Starch Press.
13. He, T., Ma, L., Swami, A., & Towsley, D. (2021). *Network Tomography: Identifiability, Measurement Design, and Network State Inference.* Cambridge University Press.
14. Hu, F. et al. (2023). Network Traffic Classification Model Based on Attention Mechanism and Spatiotemporal Features. *EURASIP Journal on Information Security, 6(2023).* https://doi.org/10.1186/s13635-023-00141-4
15. Vinit, J. (2022). *Wireshark Fundamentals: A Network Engineer's Handbook to Analyzing Network Traffic.* Apress.
16. Wang, L., & Lu, Y. (2018). A Survey of Network Measurement in Software-Defined Networking. *Advances in Intelligent Systems Research, 147*.
17. Nie, Sh., et al. (2020). Network Traffic Classification Model Based on Multi-Task Learning. *Journal of Physics: Conference Series, 1693*. https://doi:10.1088/1742-6596/1693/1/012097
18. Nucci, A., & Papagiannaki, K. (2009). *Design Measurement and Management of Large Scale IP Networks. Bridging the Gap Between Theory and Practice.* Cambridge University Press.
19. Özçelik, I., Brooks, R. R. (2020). *Distributed Denial of Service Attacks: Real-world Detection and Mitigation.* CRC Press.
20. Alkenani, J., & Nassar Khulood A. (2022). Network Monitoring Measurements for Quality of Service: A Review. *Iraqi Journal for Electrical and Electronic Engineering, 18(2),* 33–42. https://doi.org/10.37917/ijeee.18.2.5
21. Sharif, M., & Moein, M. (2021). An Effective Cost-Sensitive Convolutional Neural Network for Network Traffic Classification. In: *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*. https://doi: 10.1109/3ICT53449.2021.9581789
22. Tari, Z., Fahad, A., Almalawi, A., & Yi, X. (2020). *Network Classification for Traffic Management: Anomaly Detection, Feature Selection, Clustering and Classification.* The Institution of Engineering and Technology.
23. Mohan, V., et al. (2011). Active and Passive Network Measurements: A Survey. *International Journal of Computer Science and Information Technologies, 2(4),* 1372–1385.
24. Zeek. (n. d.). *An Open Source Network Security Monitoring Tool*. https://zeek.org/

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (26), 2024
ISSN 2663 - 4023

Підгорний Павло Володимирович
аспірант
Сумський державний університет, Суми, Україна
ORCID ID: 0009-0008-8604-8051
pashapro49@gmail.com

АНАЛІТИЧНИЙ ОГЛЯД МОДЕЛЕЙ І СИСТЕМ КЛАСИФІКАЦІЇ МЕРЕЖЕВОГО ТРАФІКУ

Анотація. У цій статті представлено всеосяжний аналітичний огляд моделей та систем класифікації мережевого трафіку, які є важливими для управління складністю сучасних мережевих середовищ. Дослідження охоплює традиційні та сучасні методи, включаючи статистичні підходи, машинне навчання та глибинне навчання, висвітлюючи їхні сильні сторони та обмеження. Також розглядаються як комерційні, так і відкриті системи з відкритим кодом, надаючи уявлення про їх практичне застосування та ефективність. Швидка еволюція мережевих технологій значно покращила глобальний обмін даними та підключення, але також спричинила нові виклики в управлінні та забезпеченні безпеки складних мережевих середовищ. З розширенням мереж та збільшенням їхньої гетерогенності здатність ефективно класифікувати та керувати мережевим трафіком стає критично важливою для оптимізації продуктивності мережі, забезпечення безпеки та підтримки безперервності роботи. Класифікація мережевого трафіку є необхідною функцією, яка дозволяє адміністраторам мереж застосовувати відповідні політики, виявляти аномалії та запобігати шкідливим діям. Традиційні методи класифікації, такі як детекція на основі аналізу вмісту пакету та класифікація за портами, стають все менш надійними через зростання зашифрованого трафіку та розвиток мережевих протоколів, які приховують важливі деталі трафіку. Тому необхідно досліджувати сучасні підходи, такі як машинне навчання, глибинне навчання та трансферне навчання. Ці методи здатні обробляти великі обсяги даних і адаптуватися до нових типів трафіку, покращуючи точність і ефективність. У цій статті представлено всеосяжний огляд існуючих моделей і систем для класифікації мережевого трафіку, включаючи як комерційні, так і рішення з відкритим кодом. Дослідження охоплює низку технік від традиційних статистичних методів до передових методів на основі машинного навчання та штучного інтелекту (ШІ). Особлива увага приділяється ключовим показникам ефективності (KPI), таким як затримка, втрата пакетів, джиттер, пропускна здатність і використання ширини смуги, які є життєво важливими для ефективного моніторингу та оптимізації мережі. З огляду на зростання кількості шифрування та еволюцію кіберзагроз, у статті наголошується на важливості адаптивних систем, аналізу в реальному часі та інтеграції технологій ШІ та машинного навчання. Огляд завершується визначенням майбутніх напрямів досліджень і нових тенденцій, що мають вирішальне значення для розробки наступного покоління систем класифікації трафіку.

Ключові слова: класифікація мережевого трафіку; ключові показники ефективності; аналіз трафіку в реальному часі; навчання передачі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Alkenani, J., & Nassar Khulood A. (2022). Network Monitoring Measurements for Quality of Service: A Review. *Iraqi Journal for Electrical and Electronic Engineering, 18(2),* 33–42. https://doi.org/10.37917/ijeee.18.2.5
2. Bhattacharyya, D, & Kalita, J. (2016). *DDoS Attacks: Evolution, Detection, Prevention, Reaction, and Tolerance*. CRC Press.
3. Bhuyan, M., Bhattacharyya, D., Kalita, J. (2017). *Network Traffic Anomaly Detection and Prevention: Concepts, Techniques, and Tools.* Springer.
4. Biersack, E., Callegari, C., & Matijasevic, M. (2013). *Data Traffic Monitoring and Analysis.* Springer.
5. Chauhan, Aj. S. (2018). *Practical Network Scanning: Capture Network Vulnerabilities Using Standard Tools Such As Nmap and Nessus*. Packt Publishing.

6. Chen, S., Chen, M., & Xiao, Q. (2017). *Measurement for Big Network Data.* Springer.
7. Collins, M. (2017). Network Security through Data Analysis: From Data to Action. O'Reilly Media.
8. Cui, J., Bai, L., Zhang, X., Lin, Z., & Liu, Q. (2024). The Attention-Based Autoencoder for Network Traffic Classification with Interpretable Feature Representation. *Symmetry, 16(5)* 589. https://doi.org/10.3390/sym16050589
9. Kim, E. (n. d.). *A Survey on Network Measurement: Concepts, Techniques, and Tools.* https://www.cs.helsinki.fi/u/kervasti/projects/A%20Survey%20on%20Network%20Measurement%20-%20Concepts,%20Techniques,%20and%20Tools%20-%20Kim%20Ervasti%20-%2031-12-2016.pdf
10. Stênio, F. (2017). *Performance Evaluation for Network Services, Systems and Protocols.* Springer.
11. Fichera, J., & Bolt, S. (2012). *Network Intrusion Analysis: Methodologies, Tools, and Techniques for Incident Analysis and Response.* Syngress.
12. Forshaw, J. (2018). *Attacking Network Protocols: A Hacker's Guide to Capture, Analysis, and Exploitation.* No Starch Press.
13. He, T., Ma, L., Swami, A., & Towsley, D. (2021). *Network Tomography: Identifiability, Measurement Design, and Network State Inference.* Cambridge University Press.
14. Hu, F. et al. (2023). Network Traffic Classification Model Based on Attention Mechanism and Spatiotemporal Features. *EURASIP Journal on Information Security, 6(2023).* https://doi.org/10.1186/s13635-023-00141-4
15. Vinit, J. (2022). *Wireshark Fundamentals: A Network Engineer's Handbook to Analyzing Network Traffic.* Apress.
16. Wang, L., & Lu, Y. (2018). A Survey of Network Measurement in Software-Defined Networking. *Advances in Intelligent Systems Research, 147.*
17. Nie, Sh., et al. (2020). Network Traffic Classification Model Based on Multi-Task Learning. *Journal of Physics: Conference Series, 1693.* https://doi:10.1088/1742-6596/1693/1/012097
18. Nucci, A., & Papagiannaki, K. (2009). *Design Measurement and Management of Large Scale IP Networks. Bridging the Gap Between Theory and Practice.* Cambridge University Press.
19. Özçelik, I., Brooks, R. R. (2020). *Distributed Denial of Service Attacks: Real-world Detection and Mitigation.* CRC Press.
20. Alkenani, J., & Nassar Khulood A. (2022). Network Monitoring Measurements for Quality of Service: A Review. *Iraqi Journal for Electrical and Electronic Engineering, 18(2),* 33–42. https://doi.org/10.37917/ijeee.18.2.5
21. Sharif, M., & Moein, M. (2021). An Effective Cost-Sensitive Convolutional Neural Network for Network Traffic Classification. In: *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT).* https://doi: 10.1109/3ICT53449.2021.9581789
22. Tari, Z., Fahad, A., Almalawi, A., & Yi, X. (2020). *Network Classification for Traffic Management: Anomaly Detection, Feature Selection, Clustering and Classification.* The Institution of Engineering and Technology.
23. Mohan, V., et al. (2011). Active and Passive Network Measurements: A Survey. *International Journal of Computer Science and Information Technologies, 2(4),* 1372–1385.
24. Zeek. (n. d.). *An Open Source Network Security Monitoring Tool*. https://zeek.org/