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 6, Issue 13, April 2026
Impact Factor: 8.2

A Review On VPN-Aware Intrusion Detection Using Machine Learning Techniques

Dr. Sarvesh Warjurkar, Prajwal Khobragade, Pragati Tiwari,
Nutan Shinganjude, Naman Raut, Nayan Matte
Department of Computer Science and Engineering
G. H. Raisoni College of Engineering and Management, Nagpur, India

Abstract: *In this paper we provide a comprehensive survey of current approaches for machine learning based IDSs, focusing on the task of VPN-aware traffic classification. Due to the ever increasing portion of encrypted communication and the presence of Virtual Private Networks (VPNs), traditional IDS techniques that rely on fixed signatures and predefined rules are not no longer sufficient to keep up with the constantly evolving nature of modern cyber threats. To overcome the mentioned limitations, in this research, several machine learning and deep learning approaches including Random Forest, Support Vector Machines, ensemble learning techniques (e.g., AdaBoost, Bagging, Stacking) and neural networks (e.g., multi-layered perceptron, convolutional neural networks) have been evaluated for their ability to detect malicious and encrypted network traffic, with their performance measured using metrics such as accuracy, precision, recall and running time. This paper also identifies key research gaps, particularly the lack of an integrated framework capable simultaneously performing intrusion detection and VPN classification. Lastly, the paper recommends a future research direction for the development of an intelligent, scalable, and adaptive system for IDS..*

Keywords: Intrusion Detection System, Machine Learning, VPN Traffic, Network Security, Deep Learning**.**

I. INTRODUCTION

As Internet technologies continues to grow and more people rely on online systems, network traffic becomes much larger and more complex, making it harder and more expensive for organizations to protect their IT infrastructures. However, the data breach attacks are far from being straightforward: they have evolved from simple Distributed Denial of Service (DDoS) attacks has now evolved into more advanced threats like ransomware, phishing, Advanced Persistent Threats (APTs), and other cunning business email compromises (BECs). Traditional perimeter-based security approaches are no longer enough to effectively defend against today's increasingly sophisticated cyber threats.
Intrusion Detection Systems (IDS) can serve as a valuable component of a network based security infrastructure. IDS can be classified into two categories:

- Signature-based IDS (detect known attacks)
- Anomaly-based IDS (detect deviations from normal behavior)

This paper compares the advantages and limitations of 'Signature Based' detection approaches with those of 'Anomaly Based' detection approaches. While 'Signature Based' techniques are good at detecting known threats with high accuracy, they fall down when faced with a zero-day attack. On the other hand, 'Anomaly Based' detection can spot unknown threats but, typically, with a high false positive rate. Encryption and VPNs (Virtual Private Networks) have, however, become so prevalent that mere payload inspection is insufficient for intrusion detection, shifting attention of researchers and industry experts towards machine learning methods which can detect network intrusions by possibly learning traffic patterns, statistical attributes and behavior from network traffic. This paper is dedicated to a detailed overview of state-of-the-art machine learning based techniques for detection employed by IDS systems as well as for

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 6, Issue 13, April 2026
Impact Factor: 8.2

traffic classification employed by VPNs. In our consideration we try to present advantages and disadvantages of these methods, and point out possible future developments in studied field.

II. LITERATURE SURVEY

In recent years, many studies have focused on the classifying VPN and non-VPN traffic using machine learning techniques, especially as encrypted communication has become more common in modern networks. Razooqi and Pekár [1] presented a detailed survey on VPN traffic analysis, pointing out key challenges such as the inability to inspect payload data due to encryption and the dependence on statistical and flow-based features. Their study highlights the need for intelligent models capable of identifying traffic patterns without accessing sensitive information.

Al-Fayoumi et al. [2] proposed a machine learning-based classification model that makes use of time-related features such as packet inter-arrival time and flow duration. Their approach achieved high accuracy in distinguishing VPN and non-VPN traffic, showing that temporal features can effectively represent encrypted traffic behavior. Similarly, Miller et al. [3] explored the use of behavioral traffic patterns and statistical features for detecting VPN traffic. Their findings shows that machine learning algorithms can accurately classify VPN traffic based on flow-level attributes, even when payload data is unavailable. However, these methods mainly focus on classification and do not include mechanisms to detect malicious activities within the identified traffic.

At the same time , machine learning-based intrusion detection systems (IDS) have been widely studied to enhance network security. Sharlet Alex et al. [2] showed that the Random Forest algorithm is very effective in detecting malicious and benign traffic, achieving an accuracy of around 95%. Its strength come from its ensemble structure, which reduces overfitting and improves ability to generalize across different datasets. Kaushik [3] introduced an optimized Support Vector Machine (SVM) model along with feature selection techniques, which not only improved detection performance but also significantly reduced false positive rates. This highlights the importance of selecting relevant features for efficient intrusion detection.

Furthermore, Samriya and Kumar developed a lightweight intrusion detection system using statistical feature selection methods like Chi-square and ANOVA. Their model achieved accuracy levels of up to 99%, making it well-suited environments with limited resources. In addition, hybrid and ensemble learning approaches [4] have been proposed to combine multiple classifiers, which helps improves both detection accuracy and overall robustness. However despite these advancements, most existing IDS models focus only on identifying attacks and do not consider whether the traffic originates from VPN or non-VPN sources, limiting their applicability in encrypted environments.

As encryption protocols continue to be widely adopted, identifying malicious activities within encrypted traffic has become a major challenge for researchers. Wang et al. [5] introduced a machine learning-based method for encrypted traffic classification that relies on statistical features rather than payload inspection. Their approach shows that useful patterns can still be extracted from encrypted data, allowing for effective classification. Similarly, Zeng et al. [8] proposed a deep learning-based framework combined with Bayesian data fusion, which improves classification accuracy for complex and high-dimensional traffic data. Deep learning models, such as neural networks, are especially effective in capturing hidden patterns and nonlinear relationships in encrypted traffic.

However, despite their high performance, these approaches often require substantial computational resources and large training datasets, which makes them less practical for real-time deployment. In addition, they do not integration with VPN-aware intrusion detection systems, which is crucial for achieving comprehensive network security.

Overall, the shows that although considerable progress has been made in VPN traffic classification, intrusion detection, and encrypted traffic analysis, these areas are mostly treated independently. There is a clear need for integrated frameworks that can simultaneously classify VPN traffic, detect malicious activities, and adapt to changing network conditions.

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 6, Issue 13, April 2026
Impact Factor: 8.2

Comparative Analysis

Table 1: Comparative Analysis of Existing Techniques

Ref	Method Used	Focus Area	Accuracy	Limitation
[1]	Survey	VPN Traffic	—	No implementation
[2]	Random Forest	IDS	~95%	No VPN detection
[3]	SVM + Optimization	IDS	High	High complexity
[4]	Ensemble ML	IDS	High	Computational cost
[5]	ML	Encrypted Traffic	Good	Limited feature scope
[6]	Statistical + ML	IDS	~99%	Dataset dependent
[7]	Time-based ML	VPN	High	Limited features
[8]	ML	VPN Detection	Moderate	Low scalability
[9]	Deep Learning	Traffic Classification	High	High computation

Research Gap

Even though there has been noticeable progress in this field, existing research still has some important limitations. Current machine learning-based intrusion detection systems, as well as VPN traffic classification methods, each come with their own set of challenges and shortcomings. However, these two research areas are often explored separately, without combining their strengths into a single, unified system. Intrusion Detection Systems (IDS), whether traditional or based on machine learning, are generally designed to detect malicious activities in network traffic without considering whether the traffic originates from a VPN or a non-VPN source.However, with more than 50% of all network traffic currently. Since VPN traffic is encrypted, IDS models often find it difficult to accurately interpret its underlying patterns. This limitation makes it even harder for them to detect sophisticated and hidden attacks. On the other hand, VPN traffic classification techniques focus on distinguishing between VPN and non-VPN traffic by analyzing statistical and behavioral features such as packet size, flow duration, and timing patterns. Although these approaches can achieve high accuracy, they do not include an intrusion detection mechanism. As a result, traffic may be correctly identified as VPN, yet still contain malicious activity that goes unnoticed. Despite advancements in both IDS and VPN classification models, there is still a strong need for practical solutions that can effectively handle network traffic classification while also ensuring security. One of the key limitations of current approaches is their inability to adapt to constantly changing network traffic. As data flows through a network, patterns evolve over time due to variations in user behavior, running applications, and even ongoing malicious activities. This continuous change in traffic patterns is referred to as *concept drift*, and it can significantly reduce the accuracy of machine learning models that are trained on static datasets and cannot adapt over time. While many modern neural network architectures achieve state-of-the-art performance, especially in tasks like image classification, they often require substantial computational power and large volumes of labeled data, making them less practical for real-world deployment.

This makes it difficult to deploy such models in real time, especially on edge devices and embedded systems commonly used in small enterprises. Additionally, many existing approaches do not adequately address scalability or the need for high-speed, real-time processing required to effectively handle modern network traffic.

A major research gap in addressing VPN-related intrusion misuse is the lack of standardized datasets that include both labeled attack traffic and clear indications of whether the traffic originates from VPN or non-VPN sources. Most existing studies rely on separate datasets—one for training intrusion detection models and another for VPN classification—making it difficult to build a unified system. Although there is a substantial amount of research in this area, there is still no comprehensive framework that effectively integrates these approaches and provides a consistent way to interpret the results.

Simultaneously detect cyberattacks and classify VPN/non-VPN traffic

Operate efficiently in real-time environments

Scale to large and high-speed network infrastructures

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 6, Issue 13, April 2026 Impact Factor: 8.2

This paper addresses key research gaps and offers new insights into effectively understanding and characterizing encrypted network traffic, supporting the development of next-generation intrusion detection systems in increasingly complex network environments.

III. PROPOSED DIRECTION

To overcome the limitations identified in existing research, this paper proposes an integrated and adaptive framework that brings together intrusion detection, VPN traffic classification, and concept drift handling using machine learning techniques. The system is designed to work with network traffic collected from multiple sources and analyzes it using flow-based and statistical features—such as packet size, flow duration, and inter-arrival time—allowing it to effectively examine encrypted communication without relying on payload inspection.

Initially, the data is preprocessed through steps such as cleaning, normalization, and feature selection to improve the model's efficiency and accuracy. The refined data is then fed into a VPN classification module, where machine learning algorithms like Random Forest or Support Vector Machines are used to differentiate between VPN and non-VPN traffic based on their behavioral patterns. Subsequently, the classified traffic is passed to an intrusion detection module, which uses a hybrid approach that combines machine learning and deep learning models to accurately identify malicious activities, even within encrypted or VPN-based traffic. To maintain effectiveness in dynamic network environments, the framework also includes a concept drift detection mechanism that continuously monitors changes in traffic patterns. Whenever a drop in performance is observed, the system adapts by updating the model through incremental learning techniques. Additionally, a feedback-driven model update process is incorporated to continuously improve detection accuracy by retraining the system with newly observed data. This integrated approach offers a scalable, efficient, and real-time solution that can simultaneously detect cyberattacks, classify VPN traffic, and adapt to evolving network conditions—ultimately strengthening overall network security in modern, encrypted environments.

IV. CONCLUSION

Machine learning algorithms are widely used in network traffic analysis for both attack detection and VPN traffic classification. In addition to traditional methods, techniques such as Random Forest, Support Vector Machines, and deep learning models have recently gained significant attention for their ability to improve the accuracy of intrusion detection systems and traffic classification. Most existing approaches focus on maximizing accuracy for two separate problems, without attempting to integrate them into a unified solution. However, the growing use of encryption and the ever-changing nature of network traffic introduce significant challenges, such as reduced visibility into data and the impact of concept drift on model performance. Current solutions do not adequately address these challenges. To overcome these limitations, there is a strong need for an integrated framework that can perform both intrusion detection and VPN traffic classification in real time, while also adapting to concept drift in encrypted network environments. In this paper, we emphasize the importance of a unified and adaptive approach that can handle all these tasks simultaneously. Furthermore, achieving high accuracy while minimizing false alarms is essential for the effectiveness of such a framework. With this motivation, we also explore future directions for designing unified models that can efficiently and scalably address the challenges posed by encrypted networks, ultimately strengthening cyber defense capabilities.

REFERENCES

[1] Y. S. Razooqi and A. Pekár, "VPN Traffic Analysis: A Survey on Detection and Application Identification," IEEE Access, vol. 13, pp. 132830–132848, 2025.

[2] Sharlet Alex et al., "Machine Learning Based Network Traffic Analyser for Malicious and Benign Traffic Detection," IEEE ICCTDC, 2025.

[3] Sunil Kaushik, "Robust Machine Learning-Based Intrusion Detection System Using Simple Statistical Techniques in Feature Selection," Scientific Reports (Springer Nature), 2025.

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 6, Issue 13, April 2026
Impact Factor: 8.2

[4] "Hybrid and Ensemble Machine Learning Techniques for Intrusion Detection," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 2024.
[5] Z. Wang, J. Li, and Y. Chen, "Encrypted Network Traffic Classification Based on Machine Learning," ICT Express, vol. 10, no. 1, pp. 112–118, 2024, Elsevier.
[6] J. K. Samriya and S. Kumar, "Machine Learning-Based Network Intrusion Detection Optimization for Cloud Computing Environments," IEEE Transactions on Consumer Electronics (TCE), 2024.
[7] M. Al-Fayoumi, M. Al-Fawa'reh, and S. Nashwan, "VPN and Non-VPN Network Traffic Classification Using Time-Related Features," Computers, Materials & Continua, vol. 72, no. 2, pp. 3091–3111, 2022.
[8] S. Miller, K. Curran, and T. Lunney, "Detection of Virtual Private Network Traffic Using Machine Learning," International Journal of Wireless Networks and Broadband Technologies, vol. 9, no. 3, pp. 1–15, 2020.
[9] Y. Zeng, H. Gu, W. Wei, and Y. Guo, "Network Traffic Classification Using Deep Learning Networks and Bayesian Data Fusion," IEEE Access, vol. 8, pp. 12831–12845, 2020.