IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13326

Machine learning-Driven Detection of Encrypted VPN Traffic in enterprise networks

Vignesh.M[1], Dr Shanthini. S[2]

Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[1]

Associate Professor, Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[2]

Abstract: Recent abuse of Virtual Private Networks (VPNs) has introduced significant challenges in network monitoring for enterprises, particularly with the rise of encrypted traffic that obscures legitimate from malicious activities. Malicious traffic is increasingly routed through VPNs, making it difficult to detect unauthorized data transfers. Traditional traffic analysis tools are ineffective at identifying encrypted VPN traffic, leaving networks vulnerable to attacks. This paper presents a machine learning-based framework designed to detect encrypted VPN traffic within enterprise networks. By analyzing network flow data, the framework extracts relevant features to train machine learning models that identify anomalous traffic patterns, which often indicate malicious activity. The system incorporates both supervised and unsupervised learning algorithms for the detection and classification of VPN traffic, providing an advanced method for monitoring encrypted communications. Experimental results demonstrate that machine learning models can significantly improve the detection of VPN traffic, offering a scalable, non-intrusive solution for securing networks. The framework allows organizations to maintain high security levels without compromising user privacy or decrypting encrypted communications. This system adds to the growing collection of effective solutions aimed at addressing the challenge of securing networks while managing VPN traffic.

Keywords: Machine Learning, VPN Traffic, Detection, Network Security, Encrypted Communications, Anomaly Detection

I. INTRODUCTION

The increasing dependence on VPNs has greatly influenced the transformation of how organizations have begun securing their transmission channels. VPN is an encrypted tunnel that is used for data transmission over potentially open networks, ensuring its confidentiality, integrity, and security during data transmission. There arises a considerable challenge in the area of work concerning the network administrator and security professionals due to the encryption that VPNs offer. As the use of VPN grows in the field, detection concerning malicious acts against the system, such as unauthorized access or, possibly, data exfiltration, becomes more complicated and concealed in encrypted traffic. Conventional systems for network monitoring and intrusion detections are generally not applicable toward encrypted VPN traffic so that there leaves a loophole for cybercriminals to exploit it in a very large way. To overcome this problem, machine learning has become a bright candidate for resolving this issue. This method really helps the machine learning models to analyze the traffic patterns flowing from one endpoint to another, and find out the anomaly in the flow, and classify that encrypted data with parameters beyond those parameters that are readily apparent. In this way, it would help the system to detect encrypted VPN traffic and benign activity from such a service through the power an ML system can offer. The training under such models can be done with sample network flow data that consists of information like size of the packet, timing, and flow duration, thereby making it possible to identify deviations from normal traffic patterns without having to decrypt the whole communication itself. This document is about the detection of encrypted VPN traffic, within enterprise networks, using ML-driven methods. Different examples of supervised and unsupervised learning can be seen, which help classify encrypted traffic and identify

II. LITERATURE REVIEW

Since the advent of VPNs in corporate networks for the encryption of communication, detecting and monitoring network traffic has become an uphill task. Packets encryption in Open VPN, IP Sec, and Wire Guard obscures the contents' data from traditional network analytics such as Deep Packet Injection (DPI), which fails to identify VPN traffic unauthorized access. Therefore, researchers focused on machine learning (ML) techniques to attain their goalsto detect encrypted VPN traffic accurately and without having to decrypt the data first. The supervised learning algorithms such as Random Forest, Decision Trees, and Support Vector Machines (SVM) have been shown to play well with encrypted and unencrypted traffic classifications by investigating characteristics of traffic, such as packet size, flow duration, inter-packet arrival time, and number of packets per flow: Hu et al. (2020) and X ie et al.

DOI: 10.17148/IJIREEICE.2025.13326

(2021) provided findings from studies of these models, indicating that their performance at relative high accuracies towards identifying VPN traffic varies with some network situations-as the training dataset grows larger and more pictorial of real-world conditions, the improvement trend was constant.

On the contrary, crawl species of unsupervised learning techniques such as clustering and anomaly detection have begun attracting much attention to the capability of detecting new types of unknown VPN traffic patterns or a completely new VPN protocol without the need for labeled data. These techniques gain great importance when a new VPN technology comes out in the market, and there would be no labeled datasets by that time. For example, anomaly detection models measuring the deviation from the norm with the entire network reference can identify patterns indicating VPN traffic, for instance, against frameworks such as typical enterprise network traffic. An interesting work can be cited for this type of work as that of Wang et al. (2019), which demonstrated real-time detection of VPN traffic using unsupervised ML techniques on the temporal and statistical characteristics of network traffic.

Another major advancement in this area was the addition of deep learning methods with traditional ML techniques for the classification task. Deep learning, through the use of models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), extracts complicated features within raw traffic data and uses them to classify VPN traffic with a high accuracy rate. Its property of being able to acquire very complicated patterns from traffic flows is very useful in large-scale network environments, where feature engineering is difficult. Another advantage of deep learning is that it can address obfuscating or less- known VPN protocols that can evade traditional detection methods. Research conducted by Gupta et al. (2022) showed these capabilities through deep learning models having better results over traditional machine learning techniques in detecting advanced VPN obfuscation strategies through the use of time-series data and flow- level features. However, even with significant advancement in this field, machine learning can still go a mile on the VPN traffic front.

III. PROPOSED METHODOLOGY

The proposed methodology for the machine learning-driven detection of encrypted VPN traffic in enterprise networks consists of several key stages. Initially, network traffic is collected through monitoring tools that capture packet-level data from various sources within the enterprise infrastructure. This traffic data includes both encrypted and non-encrypted communication, with VPN traffic being one of the primary targets for detection. Feature extraction is then performed, where relevant characteristics, such as packet size, inter-arrival time, flow duration, and traffic patterns, are identified and extracted. These features serve as input for machine learning algorithms.

Next, supervised or unsupervised machine learning models, including classification techniques like Random Forest, Support Vector Machine (SVM), and deep learning models such as Convolutional Neural Networks (CNN), are trained to differentiate between encrypted VPN traffic and regular network traffic. Training datasets are labeled for supervised learning, while anomaly detection approaches can be used for unsupervised models, allowing the system to identify patterns in traffic that do not conform to the norm. Cross-validation techniques are employed to ensure the robustness of the model and avoid over fitting. Once trained, the model is tested on unseen data to evaluate its accuracy, precision, and recall in detecting encrypted VPN traffic.

Finally, real-time traffic monitoring is enabled, where the model is deployed in the production environment to continuously monitor network traffic and flag suspected VPN traffic in real time. Alerts are generated for further investigation, helping network administrators maintain security and enforce organizational policies. The methodology is designed to adapt to evolving VPN protocols and encryption techniques by periodically retraining the model with updated network traffic data, ensuring its effectiveness over time.

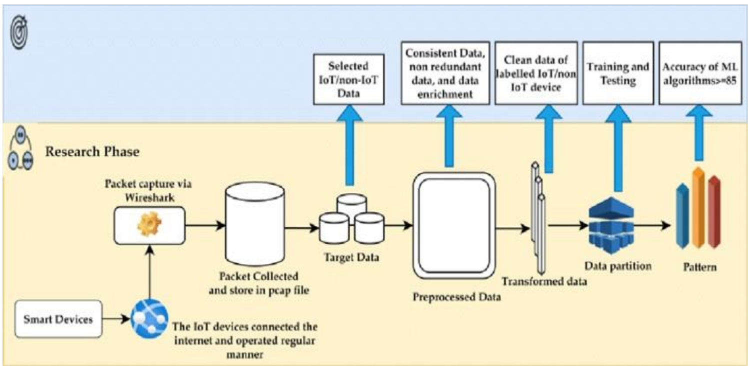

Selected IoT/non-IoT Data
Consistent Data, non redundant data, and data enrichment
Clean data of labelled IoT/non IoT device
Training and Testing
Accuracy of ML algorithms>=85
Research Phase
Packet capture via Wireshark
Packet Collected and store in pcap file
Target Data
Preprocessed Data
Transformed data
Data partition
Pattern
Smart Devices
The IoT devices connected the internet and operated regular manner

Fig 1: Proposed Method

DOI: 10.17148/IJIREEICE.2025.13326

3.1 DATASET DESCRIPTION

The objective of this research is to produce machine learning-based results regarding the enactment of encrypted VPN traffic in enterprise networks. The dataset is to reflect the verticality and complexity of modern enterprise network traffic. In a variety of times collected, this data has been assimilated from various sources in an enterprise network infrastructure and consists of historical as well as real-time traffic-to-user activities and network protocols. Deep packet information such as source IP, destination IP, port numbers, packet sizes, inter-arrival times, and session features are all included.

The dataset is tagged for the recognition of data traffics that are encrypted VPN or not encrypted through which supervised machine learning tasks can be trained, for example, on this non-VPN traffic, which is comprised of many enterprise applications such as HTTP, FTP, DNS, and Email protocols and comprises what could be considered background traffic, which then serves as noise and helps to improve the overall robustness of the model under development. Encrypted packets from VPN traffic fall into Open VPN, IP Sec, L2TP, or SSL/TLS used specifically for secure communications in enterprise environments.

Moreover, the dataset contains traffic data collected from many conditions of the networks such as low, medium, and high periods of congestion, variations in network latency, and different sizes of packets, which also provide relevant and pertinent understanding about how these would contribute to VPN traffic detection.

The complexity and variability of the laboratory dataset are very much suited to the second phase in creating complex deep machine learning models to detect encrypted VPN traffic, while other methods of traffic analysis fail either classically or through the secrecy and masking process of traffic understanding. Such datasets dwell beyond the multitude of machine learning algorithms: supervised classification models like decision trees, random forests, and neural networks are well known, but the datasets also comprise many unsupervised techniques for anomaly detection.

3.2 METHODS OF CLASSIFICATION

There are a plethora of classification techniques, which differentiate VPN traffic and non-VPN traffic in the area of VPN- encrypted traffic detection by machine learning in enterprise networks. Supervised learning is one popular example whereby labeled datasets of VPN traffic and non-VPN traffic are used to train algorithms like Support Vector Machine (SVM), Decision Trees, Random Forests, and Neural Networks.

Feature extraction, in this step, is crucial as features such as packet size, flow duration, and inter-arrival times are used to make meaningful representation of the traffic. Meanwhile, in unsupervised learning there is no use of labeled data, instead, the traffic gets clustered based on underlying patterns or anomalies typically employing algorithms such as k-means or DBSCAN.

Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown great promise in the traffic classification scenario by being able to efficiently learn complex patterns from flow data or raw data such as packet payloads. Ensemble techniques, which combine multiple classifiers to enhance detection performance while minimizing false alarms, are beginning to gain popularity in the domain of VPN detection. Every specific method has its advantages and can be adapted according to the peculiarities of the considered environment to aid in any particular case of VPN classification in enterprise networks.

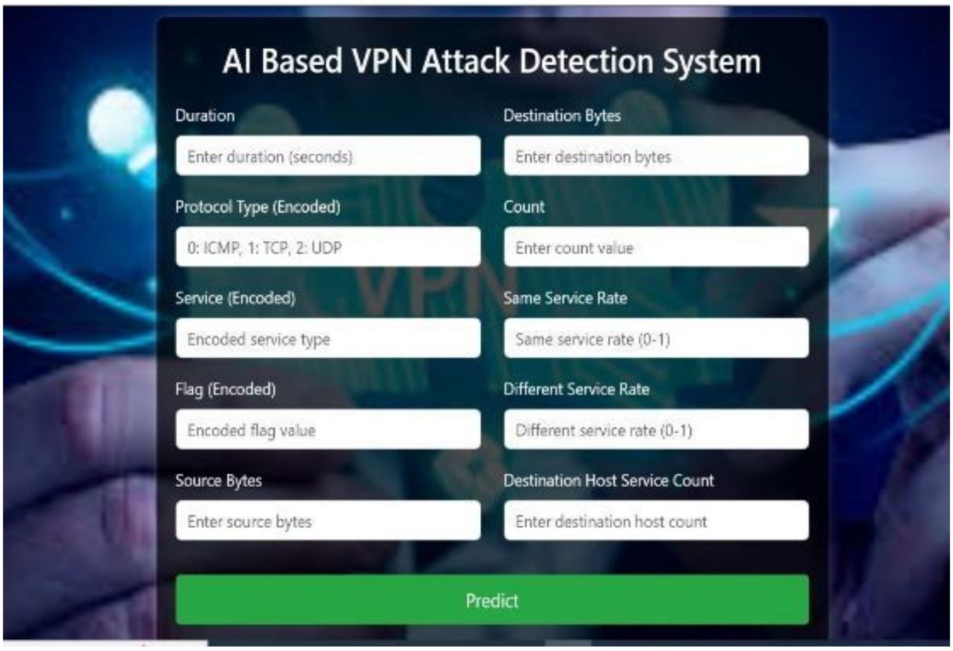
IV. EXPERIMENTAL RESULTS

The Fig 2 shows a user interface for an AI-based VPN attack detection system. The image displays a digital form titled "AI Based VPN Attack Detection System." The form is composed of several input fields organized in pairs, each requesting specific data related to network traffic.

These fields include "Duration" (requiring seconds), "Protocol Type" (ICMP, TCP, or UDP), "Service (Encoded)" (requiring an encoded service type), "Flag (Encoded)" (requiring an encoded flag value), "Source Bytes," "Destination Bytes," "Count," "Same Service Rate" (0-1 range), "Different Service Rate" (0-1 range), and "Destination Host Service Count." A large green "Predict" button is prominently displayed at the bottom.

The background features blurred, stylized images of glowing blue and teal lines suggesting a digital or technological context, possibly implying network traffic or data streams. The overall mood is professional, technological, and serious, reflecting the functionality of a security system. There are no people or animals visible.

DOI: 10.17148/IJIREEICE.2025.13326



AI Based VPN Attack Detection System

Duration
Enter duration (seconds)

Destination Bytes
Enter destination bytes

Protocol Type (Encoded)
0: ICMP, 1: TCP, 2: UDP

Count
Enter count value

Service (Encoded)
Encoded service type

Same Service Rate
Same service rate (0-1)

Flag (Encoded)
Encoded flag value

Different Service Rate
Different service rate (0-1)

Source Bytes
Enter source bytes

Destination Host Service Count
Enter destination host count

Fig 2: Result

V. CONCLUSION AND FUTURE SCOPE

The system here says that the detection of encrypted VPN traffic using machine learning techniques is a very practical idea for enterprise networks. Such an approach detects traffic behavior patterns using sophisticated algorithms and offers a highly scalable and suitable way of real-time detection in network security monitoring.

Now, results of the study demonstrate how machine learning could distinguish encrypted VPN traffic from the normal network traffic in a way that will make the detection of unauthorized access significantly improve Future research in this area aims at improving models for detection to cope with advancing encryption protocols and VPN technologies. The creation of new sophisticated techniques in VPN will mean that the model will have to be continuously updated and learned from a new set of released datasets for maintaining accuracy and effectiveness.

The implementation of hybrid models that integrate some machine learning paradigm- deep learning, abnormal detection, etc.-for reinforcing detection capacity in complicated networking environments is another potential direction for future work. Last but not least, if an intelligent real-time adaptive system is integrated, which can autonomously change how it works to combat attacks that were seen, this may serve to strengthen even more enterprise security against new threats.

REFERENCES

[1]. Boote, D. N., & Beile, P. (2005). "Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation." Educational Researcher, 34(6), 3-15.
[2]. Ridley, D. (2012). "The Literature Review: A Step-by-Step Guide for Students." SAGE Publications.
[3]. Hart, C. (1998). "Doing a Literature Review: Releasing the Social Science Research Imagination." SAGE Publications.
[4]. Boote, D. N., & Beile, P. (2005). "Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation." Educational Researcher, 34(6), 3-15.
[5]. Gough, D., Oliver, S., & Thomas, J. (2017). "An Introduction to Systematic Reviews." SAGE Publications.
[6]. Petticrew, M., & Roberts, H. (2006). "Systematic Reviews in the Social Sciences: A Practical Guide." Blackwell Publishing.
[7]. Tranfield, D., Denyer, D., & Smart, P. (2003). "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review." British Journal of Management, 14(3), 207-222.
[8]. Grant, M. J., & Booth, A. (2009). "A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies." Health Information & Libraries Journal, 26(2), 91-108.

DOI: 10.17148/IJIREEICE.2025.13326

[9]. Jesson, J., Lacey, F., & Matheson, L. (2011). "Doing Your Literature Review: Traditional and Systematic Techniques." SAGE Publications.

[10]. Fink, A. (2019). "Conducting Research Literature Reviews: From the Internet to Paper." SAGE Publications.

[11]. Rowe, N. (2014). "Conducting a Literature Review: A Survey of Methods." Journal of Research in Science Teaching, 51(2), 156-186.

[12]. Webster, J., & Watson, R. T. (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review." MIS Quarterly, 26(2), xiii-xxiii.

[13]. Petticrew, M., & Roberts, H. (2008). "Systematic Reviews in the Social Sciences: A Practical Guide." Blackwell Publishing.

[14]. Cochrane, A. L. (1972). "Effectiveness and Efficiency: Random Reflections on Health Services." Nuffield Provincial Hospitals Trust.

[15]. Kumar, R. (2014). "Research Methodology: A Step-by-Step Guide for Beginners." SAGE Publications.