International Journal of Computer Science and Information Technology Research (IJCSITR)
Vol. 7, No. 1, January-June, 2026, pp. 8-20
Journal ID: 9471-1297
website: www.ijcsitr.com
DOI: https://doi.org/10.63530/IJCSITR_2026_07_01_002

AI-Driven VPN Threat Detection and Preventive Mechanism

Amaka Eugenia Ngozi[1,*], Okpalla Chidimma Lilian[2], Ezea Jonathan Ikechukwu[3], Ibeneme-Sabinus Ifeoma Livina[4], Nworuh Godwinner Emeka[5], Atomatofa Emmanuel Oghenero[6], Gloria Ngozi Ezeh[7], Ugbor Ihechiluru Chinwe[8], and A.A. Galadima[9]

[(1,4,5,6,8,9)] Department of Cybersecurity, School of Information and Communication Technology, Federal University of Technology, Owerri, Imo State, Nigeria.

[2]Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Owerri, Imo State, Nigeria.

[3]Department of Information Technology, First Bank Nigeria Ltd, 35 Marina Lagos, Nigeria.

[7]Department of Information Technology, School of Information and Communication Technology, Federal University of Technology, Owerri, Imo State, Nigeria.

***Corresponding Author: Amaka Eugenia Ngozi**

Abstract

The widespread adoption of remote work has significantly established the Virtual Private Networks (VPNs) as a critical security tool since data remains vulnerable to Man-in-theMiddle (MITM) attacks at the local network layer. Hence, the data is being intercepted and manipulated by adversaries before it enters the secure tunnel. However, this research addresses this vulnerability by designing, implementing, and evaluating a real-time hybrid detection and prevention system for MITM attacks on VPN connections. Similarly, a controlled laboratory environment was used to simulate ARP spoofingbased MITM attacks against a commercial VPN client. Thus, the network traffic was analyzed to identify distinctive signatures of interception and integrated into a Python-based detection engine using Scapy. The signature-based and rule-based techniques were combined to monitor ARP traffic for inconsistencies such as IP-MAC binding violations and Unsolicited ARP replies. Interestingly, a 98.7% detection rate was achieved with an average response latency of 340 milliseconds. An automated mitigation module successfully neutralized 96.8% of attacks by broadcasting corrective ARP packets to restore legitimate network mappings.

Keywords

MITM Attack, VPN, ARP Spoofing, Signature-Based Detection, Network Security.

How to Cite: Amaka Eugenia Ngozi, Okpalla Chidimma Lilian, Ezea Jonathan Ikechukwu, Ibeneme-Sabinus Ifeoma Livina, Nworuh Godwinner Emeka, Atomatofa Emmanuel Oghenero, Gloria Ngozi Ezeh, Ugbor Ihechiluru Chinwe, and A.A. Galadima. (2026). AI-Driven VPN Threat Detection and Preventive Mechanism. *International Journal of Computer Science and Information Technology Research (IJCSITR)*, 7(1), 8-20.
DOI: https://doi.org/10.63530/IJCSITR_2026_07_01_002

Article ID: IJCSITR_2026_07_01_002

I. Introduction

The remote workers' adoption of Virtual Private Networks (VPNs) has significantly impacted operational services by enabling seamless job delivery (Coro, 2024; Qollakaj et al., 2025) However, most VPNs are designed primarily to protect data within the tunnel rather than detecting interception before data enters the tunnel (Hasan & Malik, 2024a; Scott & Giancarlo, 2024; Shingloo et al., 2024) Research shows that adopting advanced AI to detect attack threats could potentially reduce challenges, but implementing such systems on employee devices would be complex and could lead to numerous complications (Hasan & Malik, 2024a; Nair & Lakshmikanthan, 2020) Consequently, VPN attacks most frequently occur when the VPN is turned off, giving attackers a chance to block the connection or strike before it is reactivated (Gentile et al., 2024; Putra et al., 2024) Unfortunately, VPNs remain vulnerable to Manin-the-Middle (MITM) attacks, which have caused substantial damage to users and networks alike (Elmaghraby et al., 2024; Sun et al., 2021) Without timely measures to secure VPN data, potential damage could surpass previous incidents (Coro, 2024; Tuyisabe et al., 2025) These vulnerabilities often stem from weaknesses in authentication, assumptions of a secure network, a lack of detection for ARP spoofing, no user notifications for suspicious activity, the absence of traffic inspection tools, reliance on encryption alone, and no tracking of suspicious activity

related to MITM attacks(Nair & Lakshmikanthan, 2020; Scott & Giancarlo, 2024; Shingloo et al., 2024)

However, this poses severe threats to confidentiality, integrity, and trust required to have on sensitive data being transmitted between users and corporate networks(Elmaghraby et al., 2024; Guerra-Manzanares et al., 2025; Qollakaj et al., 2025) Alarmingly, the existing VPN technologies lack the potential to detect the occurrence of these threats and cannot provide a real-time prevention of the attack from occurring (Deepika & Abirami, 2024; Kumar, 2024; Michael Oladipo Akinsanya et al., 2024) Furthermore, most detection mechanisms relied on static rules, proving ineffective to detect advanced adversaries with dynamic spoofing and packet injection techniques (Razooqi & Pekar, 2025; Zscaler ThreatLabz team, 2024). This has also invariably increased the risk of identity theft, data breaches, and unauthorized access to most enterprise resources (Hasan & Malik, 2024b; Jain et al., 2025; Ponnuru, 2024; Sun et al., 2021; Zohaib et al., 2024).

Therefore, this research developed an improved system with the potential to detect and prevent MITM attacks within the domain of VPN connections, integrating intelligent monitoring, with traffic pattern analysis and encryption validation mechanisms. Nonetheless, the strategies for defense mechanisms span protocol hardening, runtime detection, and architecture changes. These protocols measure, however, robust certificate verification, mutual authentication, pinning the certificate, and authentication encryption usage. Additionally, the moving-target techniques will be adopted to make the VPN endpoint harder for an on-path attacker to exploit. Obviously, moving the VPN or spreading the trust makes it insufficient for a single path compromise. On the other hand, the runtime detection approach will use flow features, TLS metadata checks and machine learning models to flag suspicious sessions.

II. Research Methodology

Figure 1 depicts the methodology of this research

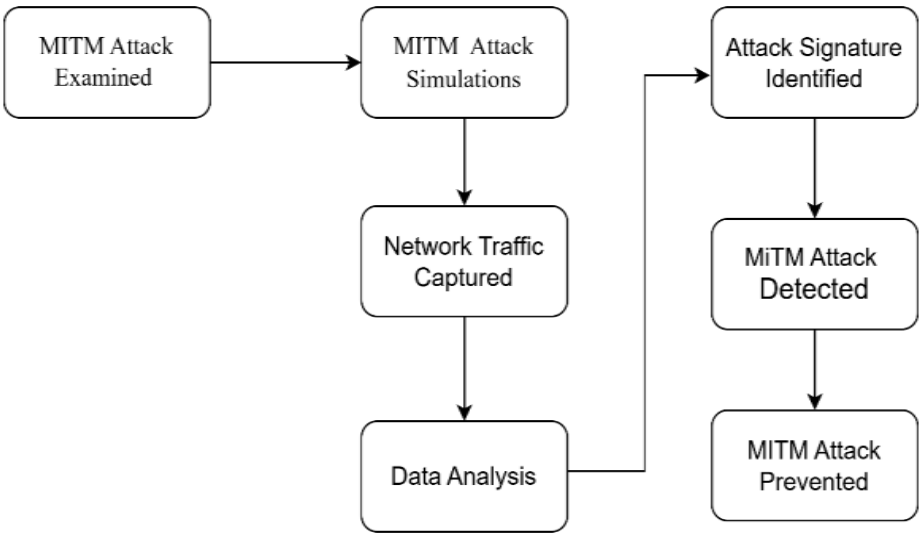

Figure 1: Research Methodology

Figure 1 shows a systematic process used to prevent MITM attacks on VPN connections for remote workers. Indeed, methodology followed a sequential process to ensure accurate data collection, realistic attack simulation, and practical validation of both detection and prevention mechanisms. Hence, this process began with a thorough understanding and examination of MITM behavior on VPN, with a comprehensive examination of how MITM attacks exploit the VPN infrastructure. Thus, this aimed at understanding the functional mechanisms, indicators, and vectors in a specified encrypted tunnel. Similarly, a simulated controlled environment was created to simulate realistic MITM attack scenarios against VPN connections. Ideally, a configuration of three systems for remote workers, an attacker, and a VPN server was done using a visualized tool. Consequently, the attacker's execution was focused on controlled interception attempts as observed in the laboratory environment and reproduced authentically with the behavior of MITM regarding the spoofed communication together with the traffic redirection.

The monitoring of the network traffic was captured during the simulation, obtaining realtime data for VPN sessions. However, the exchanged packets were recorded using Wireshark between victim, attacker and VPN server, noting the live network packets discovered on the victim's device. Similarly, captured data such as the sequences of encryption negotiation, handshake messages and metadata protocol. Notably, the essential traffic information was stored to maintain data privacy with total ethical standards compliance. This invariably has an

indication that the intrusion of MITM reveals multiple ARP unsolicited responses binding conflicting MAC address. Though the logs captured served primarily as dataset for subsequent analysis and identification of signatures.

Moreso, the network data captured was thoroughly examined to adequately identify both patterns and anomalies, while the legitimate VPNs traffics were distinguished from the recorded compromised sessions. Additionally, the analysis of protocol headers, packet timing and certificate exchanges were done to detect irregularities not limited to duplicate ARP responses, altered details of certificates, IP-MAC binding mismatched, and downgrades version of abrupt protocol. Then the baseline benign recorded was to establish a reference model. Interestingly, the MITM indicators were uniquely extracted during the comparison process for both legitimate and suspicious traces respectively, to form the foundational rule based and signature generation.

Hence, each signature described the observable pattern with its triggering condition and associated network parameters. Then, the rulebased detection parameters were defined, complementing the signatures, while setting aside the thresholds and their logical conditions to easily filter out the benign anomalies. Though this combination provided an enabling environment to capably recognize known attack traits, creating room for adaptive prevention. However, both the developed signatures and rulebased were integrated into a real-time detection and prevention framework, while the VPN traffic was closely monitored, and as well, matches against the stored attack signatures and rule conditions. Consequently, on detection of the suspicious patterns (spoofing attempt), the automated defensive action blocked the suspicious connection, triggering the alert notifications or enforcing re-authentication. The system, however, was evaluated, obtaining the metric performances, which ascertained the efficiency and reliability for securing remote workers' VPN communications.

III. Model's Architecture

Figure 2 shows the architectural design of the model

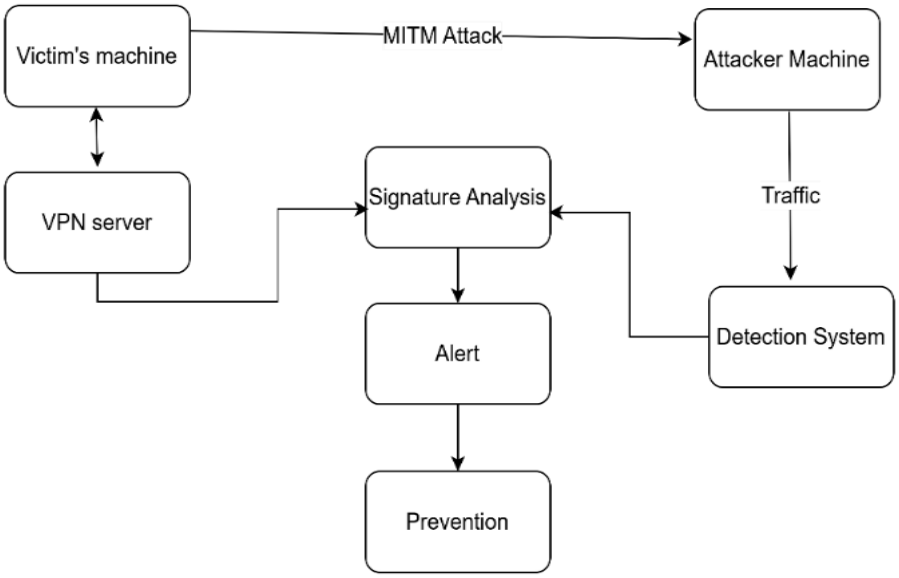

Figure 2: Model's Architecture

Figure 2 shows that the model's architectural design has three components: the victim machine (runs on Ubuntu Linux), the attacker machine (Kali Linux), and the detection system. Hence, these components worked together to simulate, observe, and detect MITM attacks on VPN connections. Thus, the victim's machine was securely connected to a legitimate VPN server, while a lightweight detection agent was installed for constant monitoring of the network. Similarly, this checks the ARP table, gateway MAC address, and DNS settings for any unexpected changes, such as a notification of a sudden change in the gateway's MAC address, to immediately flag a threat attempt.

On the other hand, the attacker machine, designed to perform a typical ARP spoofing attack, was maliciously used as a gateway between the victim and the VPN server. Hence, as the ARP spoofing attack begins, the attacker starts intercepting packets that have been transmitted and observes traffic patterns, session details, and DNS requests to further use them to execute more attacks. Contrarily, the defensive layer of the architecture (detection system), combined with real-time monitoring, signature-based detection, and automated alerting, analyzes traffic from the victim's network. Consequently, once the suspicious activity was confirmed, a notification of alert was raised on the local interface, and network traffic was blocked automatically.

Model's Defense Mechanism (MDM) against MITM attack.

Figure 3 shows a structured defense mechanism against MITM attack targeted at a VPN connection, using both signature and rule-based detection systems.

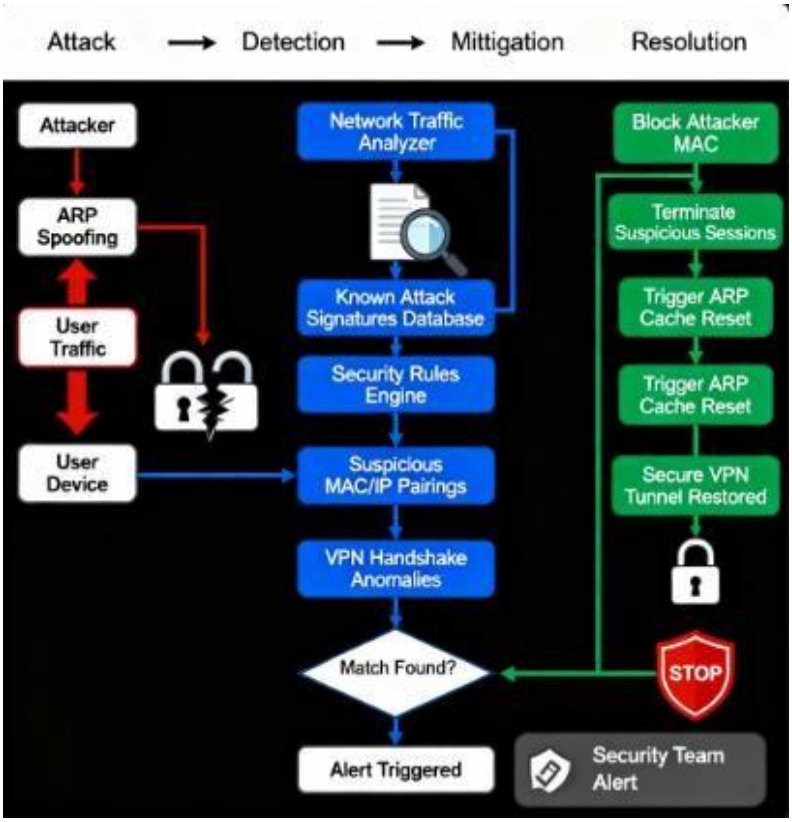

Figure 3: MDM against MITM attack

From the left side of this design, the attacker took a good position, attempting to perform ARP spoofing on a legitimate device, noting well that the VPN traffic has been compromised as indicated with the red arrow, together with the broken padlock icons. Consequently, as this malicious activity flows, it was detected and flagged with an alert notification. Fortunately, the attacker's MAC was blocked, and suspicious sessions were terminated, resulting in ARP cache reset to enable adequate restoration of the VPN connection with the secured network.

IV. Model's Implementation

Replicating the real-world network demanded for a suitable design of a controlled lab environment, to maintain isolation for effective security testing. Hence, the core infrastructure-based for this model consisted primarily of two different architectural systems for Kali Linx attacker's machine and victim's Windows 10 system, deploying both in a virtualized

environment together with a VMware workstation.

Moreover, the laboratory environment's configuration was done using Network Address Translation (NAT) network topology, enabling the controlled interaction between the victim's system and the attacker to maintain external connectivity in achieving realistic test scenarios. The IP address 192.168.224.128 was also assigned to the victim's Windows 10, together with the subnet mask of 255.255.255.0 and a configuration of a default gateway of 192.168.224.2. though the Kali Linux system's attacker operated from the IP address of 192.168.224.129 and created 24 standardized network segments for demonstrating ARP spoofing. Additionally, the comprehensive network validation was performed, ensuring adequate connectivity and baseline operations. Nonetheless, the displayed output of ipconfig in Figure 4 confirmed that the configuration of the Ethernet adapter with a link-local IPv6 address of fe80: e037:53c3:78e5: e342 and its critical part of IPv4 assignment established a strong foundation for subsequent attack scenarios.

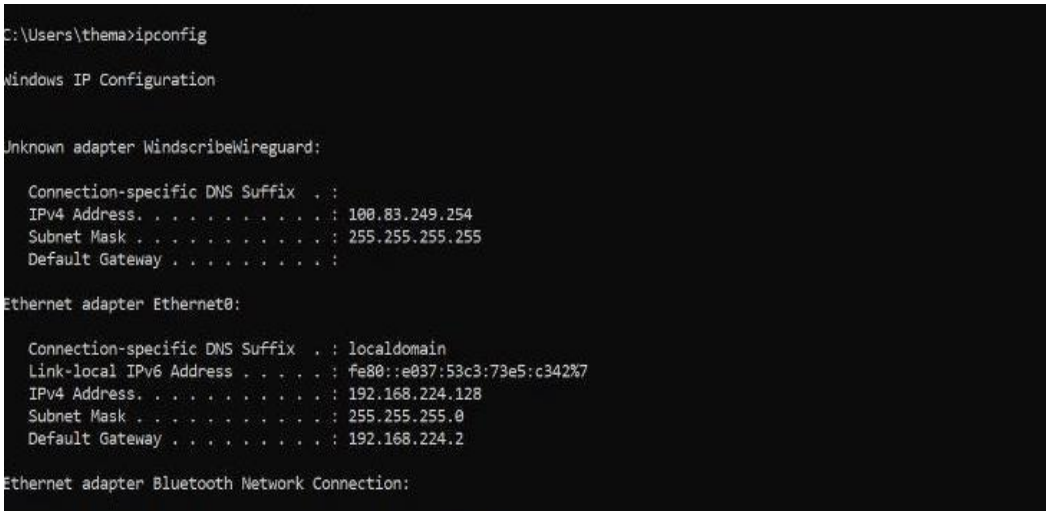

Figure 4: Windows IP configuration displaying

Ethernet adapter settings with IP 192.168.224.128 and gateway 192.168.224.2

However, enhancing the practical relevance of the testing environment called for the integration of the windscribe VPN, as shown in Figure 5. Nevertheless, the realistic scenario provided by the commercial VPN solutions for evaluating ARP spoofing attacks against encrypted communications showed a connection to the Atlanta Peachtree server and the external IP of 104.223.110.251, maintaining the internal configuration of the laboratory

network through the split-tunneling capabilities. Furthermore, the baseline data provided for both pre-attack and post-attack comparisons, with a provision of a gateway 192.162.224.2 and a MAC address of 00-50-56-f4-7c-ee for both pre-attack and postattack comparisons. Hence, the performance testing metrics for this baseline were established for comparison with post-attack conditions. Notably, the Ping tests using the Windows victim of 192.168.224.128 for the Kali attacker, excellently demonstrated smooth connectivity, averaging the round-trip times of 0ms and 0% packet loss as ascertained by multiple test iterations.

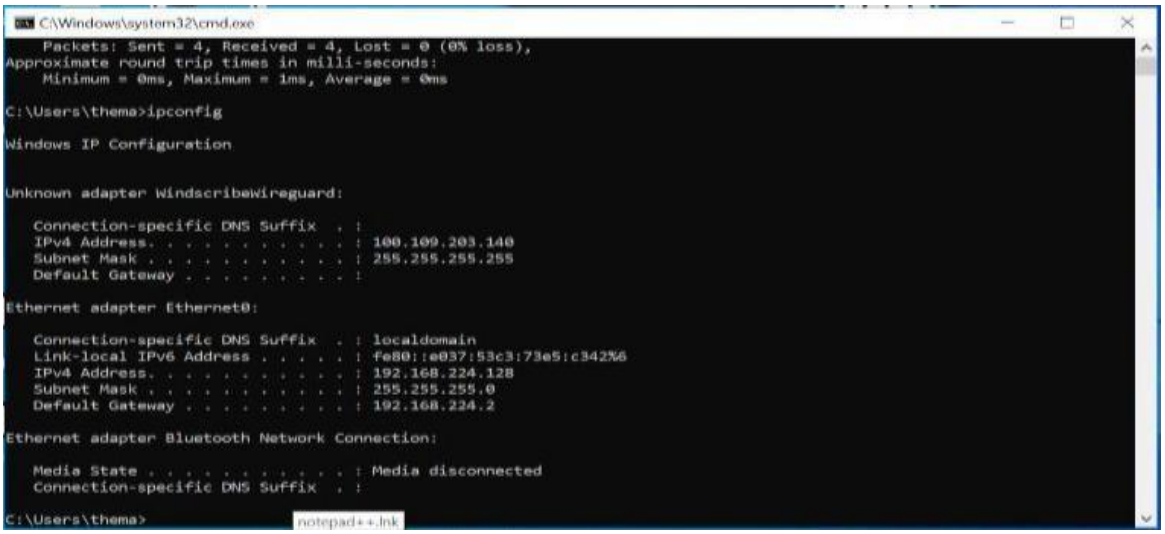

Figure 5: Detailed ipconfig output showing complete network adapter configurations

V. Results and Discussions

The outcome of the experimental setup using the metrics of detection accuracy, false-positive rate, response time, and mitigation success rate was discussed thus:

System Performance

Table 1 is experimental setup performances.

Metric	Observed Value
Detection Accuracy	98.7%
False negative rate	1.3%
False positive rate	2.3%
Average response time	340ms
Mitigation success	96.8%

The simulated detection system demonstrated a very strong capacity with easy identification of attempted spoofing at the VPN sessions. It was also obviously noted that all the tests conducted had an average record of 98.7% detection accuracy. However, this confirmed the effectiveness of the model, whereby the malicious traffic was distinguished from the legitimate VPN communications. Similarly, the false positive rate of 2.3% was recorded, maintaining insignificant misclassification when compared to a normal event of network attacks. On the other hand, a false negative rate of 1.3% recorded resulted in a detection of high sophistication and low-rate poisoning attacks. Thus, the average response time interval of 340 milliseconds was measured between the processes of detecting an attack and its mitigation for validating real-time responsiveness. Though, maintaining the real-time monitoring performance was achieved without any bias. Nevertheless, the spoofing attempts were actively identified and efficiently mitigated through restoration of legitimate ARP entries with a total blockage to all possible connections to the attackers. However, the detection engine, incorporating the IP and MAC addresses ensured a continuous verification of those addresses to easily flag off any potential attempt at spoofing. The alerts generated by the model contained a timestamp, a suspicious IP address, a legitimate MAC address and a spoofed MAC address from a legitimate database. While a detection of ARP spoofing initiated a multi-stage protocol response for restoring legitimate IP-MAC mapping using ARP packets within the network. This, however, ensured adequate provision of a countermeasure to effectively and diligently disrupt MITM attacks at a minimal detection time.

Most importantly, the results obtained demonstrated the security challenges of MITM attacks posed to remote workers through the VPN-based unsecured connections. Indeed, the signature-based verification and automatic ARP threat detection with integration of real-time monitoring system, provided a very high accuracy and low-resource consumption for adequate deployment to user devices remotely. Perhaps, this simulation demonstrated that VPN encryption alone is insufficient to protect data against MITM interception across the local networks studied and requires a combination of encryption with real-time monitoring for adequate security measures.

Conclusions

This research was designed, implemented, and evaluated successfully using a hybrid rule-

based and signature-based detection system for the proper identification and prevention of MITM attacks on VPN network connections. Thus, a controlled simulation conducted in a virtual environment ascertained an accurate rate of 98.7% for malicious ARP spoofing activities. Also, 96.8% was confirmed to have a mitigation success rate, while an average response time of 340 milliseconds was maintained. However, these results demonstrated a severe gap beyond data encryption; rather, adequate security measures on the network access layer should be protected to avert traffic interception by MITM attack before or after encryption. Though this issue was addressed by this research, implementing real-time monitoring of ARP table integrity and restoring the legitimate gateway connections at any attempt of spoofing detection.

Recommendation

Future work is recommended for system capabilities extension, incorporating both anomaly-based and machine learning algorithms for the detection of unknown attack variations and improvement of adaptive responses. Moreso, large-scale testing is also recommended in real enterprise VPN environments to effectively conduct and evaluate interoperability, scalability, and performance in diverse traffic conditions.

References

[1] Coro, M. B. (2024). Information security in remote work: Strategies and challenges in a post-pandemic world. Revista Sistemática, 14(4), 995–999. https://doi.org/10.56238/rcsv14n4-020

[2] Deepika, C., & Abirami, Dr. K. (2024). Current Scenario About Virtual Private Network (Vpn) Cyber Security Threats. Futuristic Trends in Artificial Intelligence Volume 3 Book 11, 3, 101–112. https://doi.org/10.58532/v3bkai11p4ch3

[3] Elmaghraby, R. T., Abdel Aziem, N. M., Sobh, M. A., & Bahaa-Eldin, A. M. (2024). Encrypted network traffic classification based on machine learning. Ain Shams Engineering Journal, 15(2), 102361. https://doi.org/10.1016/j.asej.2023.102361

[4] Gentile, A. F., Macrì, D., Greco, E., & Fazio, P. (2024). Overlay and Virtual Private Networks Security Performances Analysis with Open Source Infrastructure Deployment. Future Internet, 16(8), 1–25. https://doi.org/10.3390/fi16080283

[5] Guerra-Manzanares, A., Caprolu, M., & Di Pietro, R. (2025). A comprehensive review on machine learning-based VPN detection: Scenarios, methods, and open challenges. Computer Science Review, 58(June), 100781. https://doi.org/10.1016/j.cosrev.2025.100781

[6] Hasan, M., & Malik, T. (2024a). AI-Enhanced VPN Security Framework: Integrating Open-Source Threat Intelligence and Machine Learning to Secure Digital Networks. European Conference on Information Warfare and Security, ECCWS, 764–772. https://doi.org/10.34190/eccws.23.1.2505

[7] Hasan, M., & Malik, T. (2024b). AI-Enhanced VPN Security Framework: Integrating Open-Source Threat Intelligence and Machine Learning to Secure Digital Networks. European Conference on Information Warfare and Security, ECCWS, 764–772. https://doi.org/10.34190/eccws.23.1.2505

[8] Jain, V. K., Aggrawal, J., Dangi, R., Prasad Shukla, S. S., Yadav, A. K., & Choudhary, G. (2025). Unmasking the True Identity: Unveiling the Secrets of Virtual Private Networks and Proxies. Information (Switzerland), 16(2), 1–18. https://doi.org/10.3390/info16020126

[9] Kumar, R. R. (2024). Security for Cloud D ata Protecting by VPN. 13(5).

[10] Michael Oladipo Akinsanya, Cynthia Chizoba Ekechi, & Chukwuekem David Okeke. (2024). Virtual Private Networks (Vpn): a Conceptual Review of Security Protocols and Their Application in Modern Networks. Engineering Science & Technology Journal, 5(4), 1452–1472. https://doi.org/10.51594/estj.v5i4.1076

[11] Nair, S. S., & Lakshmikanthan, G. (2020). Beyond VPNs: Advanced Security Strategies for the Remote Work Revolution. International Journal of Multidisciplinary Research in Science, Engineering and Technology, 03(05), 1283–1294. https://doi.org/10.15680/ijmrset.2020.0305009

[12] Ponnuru, S. P. (2024). Advanced Network Security: Evaluating Firewall and Vpn Strategies for Robust Protection in Modern Telecommunications. International Journal of Research In Computer Applications and Information Technology (IJRCAIT), 7(2), 1359.

[13] Putra, S., Iqbal, M., Putera, A., & Siahaan, U. (2024). Network Security Design Using Virtual Private Network (VPN) Method By Utilizing Point To Point Tunneling Protocol (PPTP) Technology On Local Area Network (LAN). International Journal of Computer Sciences and Mathematics Engineering Journal Homepage: Www.Ijecom.Org International Journal of Computer Sciences and Mathematics Engineering Journal Homepage. www.ijecom.org

[14] Qollakaj, K., Larsson, L. E., & Memeti, S. (2025). Cybersecurity of remote work migration: A study on the VPN security landscape post Covid-19 outbreak. Array, 27(September 2024), 100437. https://doi.org/10.1016/j.array.2025.100437

[15] Razooqi, Y. S., & Pekar, A. (2025). VPN Traffic Analysis: A Survey on Detection and Application Identification. IEEE Access, 13(June), 132830–132848. https://doi.org/10.1109/ACCESS.2025.3592152

[16] Scott, V. A. B., & Giancarlo, S. A. (2024). Designing the VPN with Top-Down to Improve Information Security. International Journal of Advanced Computer Science and Applications, 15(6), 182–190. https://doi.org/10.14569/IJACSA.2024.0150620

[17] Shingloo, H., Mishra, S., Zambare, H., Jaiswal, A., & Kumar, M. (2024). Designing A VPN Using Open Network Infrastructure with Enhanced Security and Performance. 11(7), 40–46.

[18] Sun, Y., Wang, B., Wang, C., & Wei, Y. (2021). On Man-in-the-Middle Attack Risks of the VPN Gate Relay System. Security and Communication Networks, 2021. https://doi.org/10.1155/2021/9091675

[19] Tuyisabe, R., Ngugi, Dr. J., & Sumbiri, Dr. D. (2025). Leveraging Site-to-Site VPN and BGP Protocols to Enhance Digital Resource Sharing Among TVET Institutions in Rwanda. Journal of Information and Technology, 5(8), 36–46. https://doi.org/10.70619/vol5iss8pp36-46

[20] Zohaib, S. M., Sajjad, S. M., Iqbal, Z., Yousaf, M., Haseeb, M., & Muhammad, Z. (2024). Zero Trust VPN (ZT-VPN): A Systematic Literature Review and Cybersecurity Framework for Hybrid and Remote Work. Information (Switzerland), 15(11), 1–25. https://doi.org/10.3390/info15110734

[21] Zscaler ThreatLabz team. (2024). Zscaler ThreatLabz 2024 Phishing Report. https://www.zscaler.com/resources/industry-reports/threatlabz-phishing-report-2024.pdf