Final Year Design Project Proposal Document

VPN SpyGlass: VPN traffic analyzer

By

Hassan Abdullah	337275
Sameen Mubashar	346848

Supervisor:
Dr. Mehdi Hussain

Co-Supervisor:
Dr. Arsalan Ahmad

Bachelor of Science in Computer Science
(2020-2024)
Department of Computing
School of Electrical Engineering and Computer Science
National University of Sciences & Technology

ABSTRACT

The "VPN SpyGlass" project intends to develop a VPN traffic analyser tool that can improve network security by detecting and assessing VPN activities. The programme uses deep packet inspection techniques to distinguish between VPN and conventional network traffic and provides real-time monitoring via a customisable internet dashboard. The project solves VPN-related obstacles, such as security risks, privacy concerns, and network performance issues, by providing a comprehensive solution for effective network security management.

INTRODUCTION

The usage of VPNs is expanding as people become more concerned about their online privacy and security, as well as the desire to go around geo-restrictions and access region-specific material. **Controlling VPN use**, on the other hand, is critical for **preventing security concerns by monitoring and controlling encrypted traffic, protecting against possible misuse, optimizing network performance, and ensuring compliance with network policies and laws**. **Deep packet analysis is critical for comprehending network traffic** (Figure 1) since it inspects individual data packets, allowing for better security monitoring, threat detection, and network optimization. However, there has been little study towards identifying internal application activity via encrypted network traffic, particularly for VPNs. There is a dire need to research encryption solutions for today's cutting-edge network traffic for analysis and inspection.

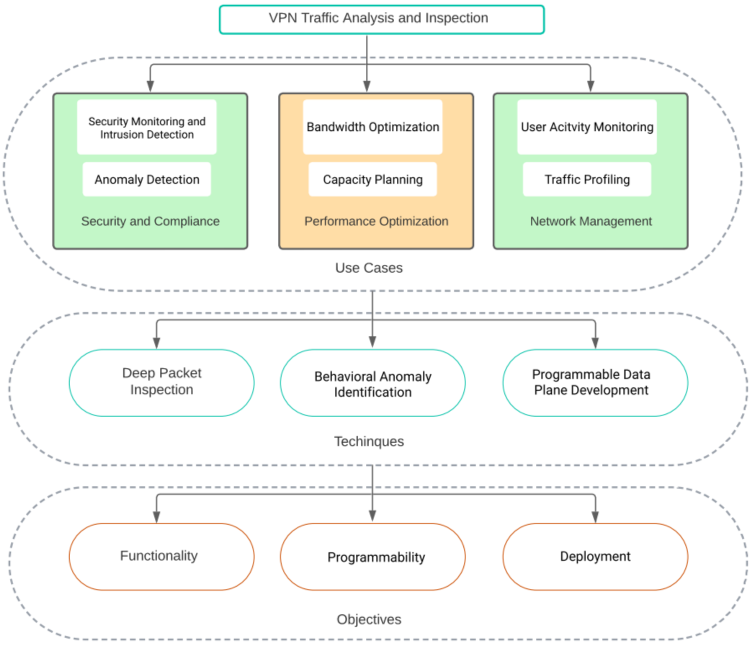
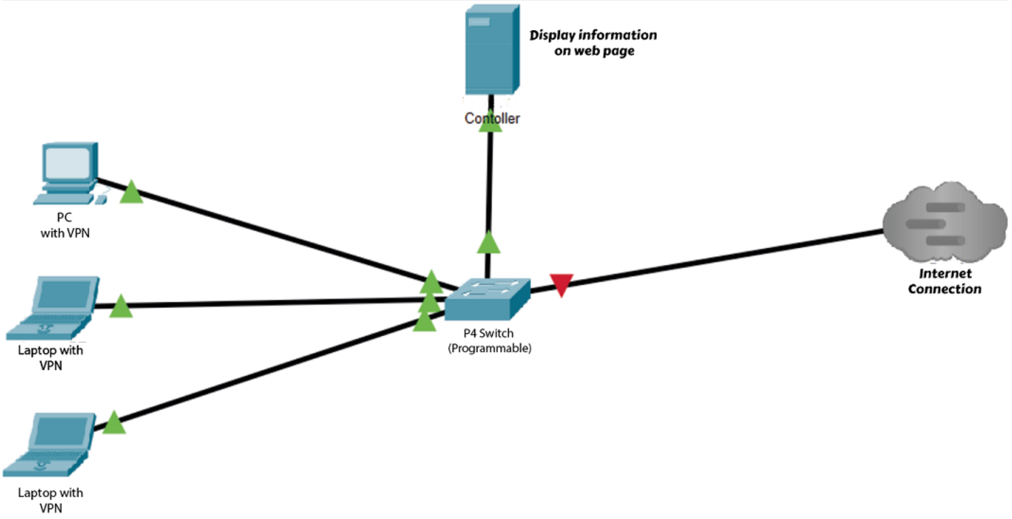


Figure 1

The purpose of this research is to **look into the encrypted network traffic used by VPN providers to determine the traffic pattern, protocol identification, and participating servers**. The proposed methodology will capture traffic using a P4 Switch, as shown in Figure 2. Various VPNs will be installed on devices and linked to the internet through a wireless access point passing through switch. To filter out the corresponding apps, all internet traffic from the wireless access point will be routed through the P4 switch and information will be forwarded to the controller. Wireshark and other cutting-edge tools will be used to monitor encrypted data via trace file analysis for network traffic monitoring. The IP addresses, ports, packet lengths, and packet patterns will be utilized to identify the VPN and the exact VPN that is being used. It is worth mentioning that the IP addresses and ports are in plaintext, however the payloads are encrypted for the purpose of secrecy and privacy. The knowledge gathered from this study will aid in the **building of a Network Traffic Analysis for VPNs, as well as the ability to block VPN in a specific network.**



The project targets the issues that consumers have while navigating the VPN sector, with the goal of providing clear, unbiased information and tools to assist users in making educated VPN service selection decisions. By integrating robust network analysis with user-friendly interface design, the project allows users to control VPN usage, improve network security, and increase network visibility. The purpose is to enable network administrators to identify unauthorised VPN connections, categorise VPN traffic, and reduce security risks by leveraging modern technology and deep packet inspection algorithms.

PROBLEM STATEMENT

"The proliferation of VPNs poses a challenge to network admin, as these tools can bypass security measures and access restricted content pose security risks, violate agreements, and slow network performance. Moreover, Existing methods often struggle to identify VPN traffic due to encryption and masking techniques employed by VPN services."

LITERATURE REVIEW

In recent years, the proliferation of Virtual Private Networks (VPNs) has led to the development of various tools aimed at detecting and monitoring VPN network traffic. Machine learning approaches have been used in solutions such as those provided by Avnish Goel et al. and Gerard Draper-Gil et al. to analyses and categories VPN traffic patterns. But a lot of these current technologies are stand-alone programs that don't support real-time detection and need human input. On the other hand, VPN SpyGlass, our technology, is a breakthrough in VPN traffic analysis. In contrast to conventional programs that use database matching methods, VPN SpyGlass uses an advanced Python algorithm that has been trained on an ever-updating dataset of recognized VPN traffic patterns.

Compared to other similar tools in the market, our technology can offer zero-day protection against new VPN dangers thanks to this technique, which can detect such threats in real time even before they are categorized in databases that are already in place. Due to its dependence on static databases, other programs could find it difficult to identify more recent, unidentified VPN activity. In contrast, VPN SpyGlass is proactive and adaptable, always changing to counter new threats. VPN SpyGlass seeks to revolutionize VPN traffic monitoring by utilizing machine learning and real-time analysis to provide users with unmatched visibility and security in an ever-evolving digital environment.

2.1 Lightweight Anti DDoS Security Tool: Edge Level Filtering in SDN using P4.

The research paper titled "Lightweight Anti DDoS Security Tool: Edge Level Filtering in SDN using P4"[1] by Masumi Arafune, Bhargavi Goswami, Manasa Kulkarni, Nagarajan Venkatachalam, and Saleh Asadollahi focuses on addressing the challenge of packet injection attacks in Software Defined Networks (SDN). To provide edge-level filtering without relying on the control plane, the article presents a technique known as Lightweight Anti-DDoS Software (LADS). LADS seeks to provide a reliable and effective defense against malicious packet injection attacks by utilizing the P4 programming language to build filtering functions in edge switches.

The paper concludes that during packet injection assaults, LADS proved to be 100% successful in filtering malicious packets. The lightweight nature of the transition using LADS was demonstrated by the average 3% increase in CPU utilization. Measuring bandwidth revealed that LADS was able to keep genuine hosts' network performance intact while effectively isolating attackers. To improve efficiency, it is recommended that whitelisting be included for port blocking in future work. Provide a blocked port management feature to unblock hosts from being blocked indefinitely. Limitations include the quantity of hosts linked to a single switch and the kind of packet injection threats that LADS mitigates should be addressed.

2.2 Detection of VPN Network Traffic

"Detection of VPN Network Traffic" [2] is the title of a research paper written by Avnish Goel, Rochak Kaushik, Apoorv Kashyap, S. Nagasundari, B. Devesha Reddy, and Prasad B. Honnavali from PES University in Bangalore, India's CSE Department in 2022. In the digital era, the article addresses the significance of Virtual Private Networks (VPNs) in guaranteeing data security and privacy. It emphasises the necessity of VPN detection to stop dangerous actions that might take use of VPNs for anonymity, such ransomware assaults.

In this study, we develop a virtual private network (VPN) based on a widely used network security protocol and employ multiple machine learning models to identify VPN traffic. To train and evaluate their algorithms, the scientists employed a standardised dataset from the Canadian Institute for Cybersecurity. They used techniques like Random Forest and Multilayer Perceptron (MLP) for classification, and they were successful in identifying VPN traffic with high accuracy rates.

As part of the technique, an AWS VPN server was set up, network traffic data was converted from PCAP to CSV format, and machine learning algorithms were used to analyse the data. To get insights, the authors plotted the data properties pertaining to both VPN and non-VPN traffic. Additionally, real-time data analysis was done to verify the models' predictions.

2.3 A Survey on Encrypted Network Traffic Analysis Applications, Techniques, and Countermeasure

The research paper, written by Sotiris Ioannidis and Eva Papadogiannaki and published on July 13, 2021[3], addresses the difficulties that arise from the growing use of network traffic encryption. It also explains the operation of encrypted traffic inspection, with a particular emphasis on functions related to middleboxes, network analytics, security, and user privacy. Classifying encrypted communication, identifying user behaviors, and examining quality of service/experience are all done with the use of methods like machine learning, deep learning, and neural networks. Scalability issues and the need to retrain models with new data present challenges. Machine learning is used for both malware and intrusion detection in encrypted networks and mobile devices. Evaluation criteria for machine learning include accuracy, precision, recall, and false-positive rates.

Further research into the scalability of machine learning models for classified encrypted traffic, the accuracy of these models, and the difficulties associated with malware detection on mobile devices and intrusion detection in encrypted networks are some possible directions for future work. Considering advancing encryption technologies and rising network traffic quantities, future research may concentrate on creating more effective and efficient methods for encrypted traffic analysis.

2.4 Characterization of Encrypted and VPN Traffic using Time-related Features

Written by [4] (Gerard Draper-Gil, Arash Habibi Lashkari, Mohammad Saiful Islam Mamun, and Ali A. Ghorbani, 2016), the research goes into the complexity of traffic categorization, concentrating specifically on encrypted and VPN data, and emphasises the importance of time-related factors in effectively categorising such information. The work uses machine learning algorithms such as C4.5 and KNN to provide a robust flow-based classification system that efficiently identifies VPN traffic and categorises encrypted traffic into various categories such as browsing, streaming, and VoIP. Using a dataset of encrypted communication with 14 distinct labels, the study illustrates the usefulness of time-related characteristics in traffic analysis, with accuracy rates over 80%. Notably, the study found that using lower flow timeout values improved classification accuracy, with C4.5 outperforming KNN.

The findings imply that time-related indicators can be used to reliably classify encrypted and VPN traffic. In the future, the researchers hope to broaden their study to include new applications and forms of encrypted communication, as well as investigate the use of time-based aspects in traffic categorization.

PROJECT OVERIEW/GOAL

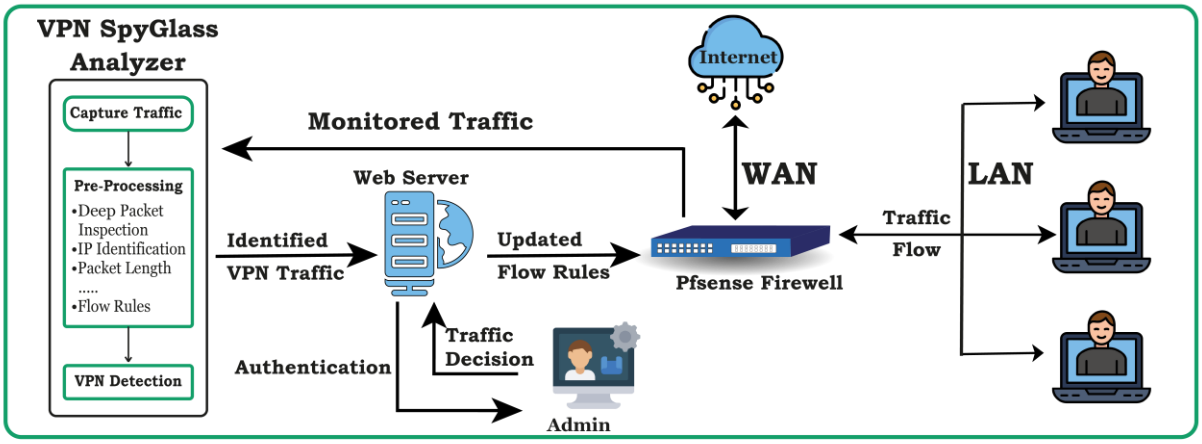
- Acquire a thorough comprehension of the analysis of encrypted network traffic.
- The goal of this study is to look into the difficulties in examining network traffic that has been tunneled or encrypted.
- Recognize and investigate the many applications and use cases for encrypted traffic analysis.

Outcomes of the FYP:

- 1x high-quality publication
- A straightforward encrypted network traffic analysis module capable of distinguishing between VPN and conventional traffic as well as particular VPN types.
- A web interface that provides real-time traffic analysis and allows us to disable VPN connections in a specified network.

PROJECT DEVELOPMENT METHODLOGY / ARCHITECTURE

The project's aims are divided into smaller objectives/modules, such as online dashboard creation, cloud database integration, hosting setup, and classification algorithm implementation. The system design consists of ReactJS/NextJS for the online interface, NodeJS for backend services, and Python for classification algorithms. The project process is agile, enabling flexibility and response to change requirements.



PROJECT MILESTONES AND DELIVERABLES

Phase 1:

Detection of VPN traffic using Deep Packet Inspection based on their IP, Port Number, and Packet Length. Understand P4, BMV2, and OpenFlow.

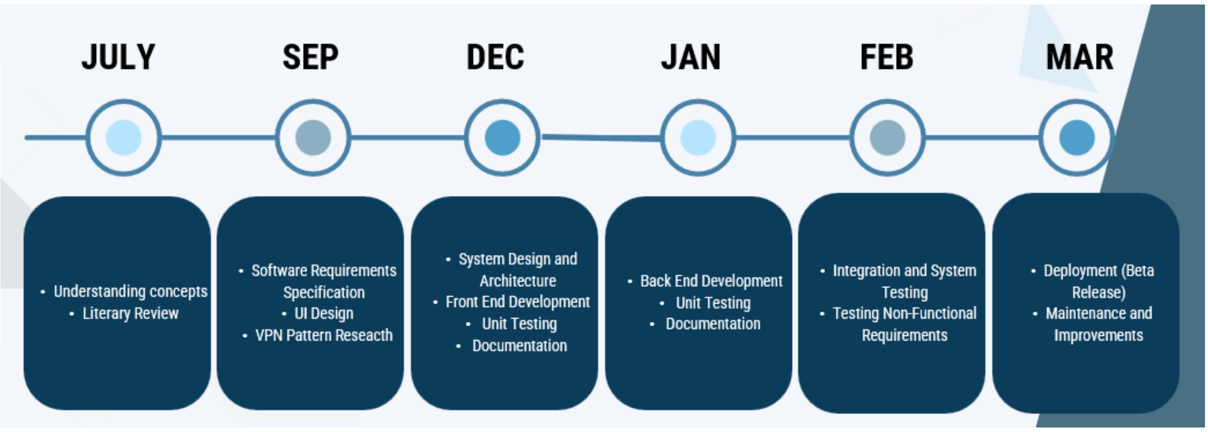
Phase 2:

Development of Web Page. Setting up of implementation environment.

Phase 3:

Implementation of code on BMV2 switch and integrating with the webpage.

DELIVERABLES



JULY
- Understanding concepts
- Literary Review

SEP
- Software Requirements Specification
- UI Design
- VPN Pattern Reseacth

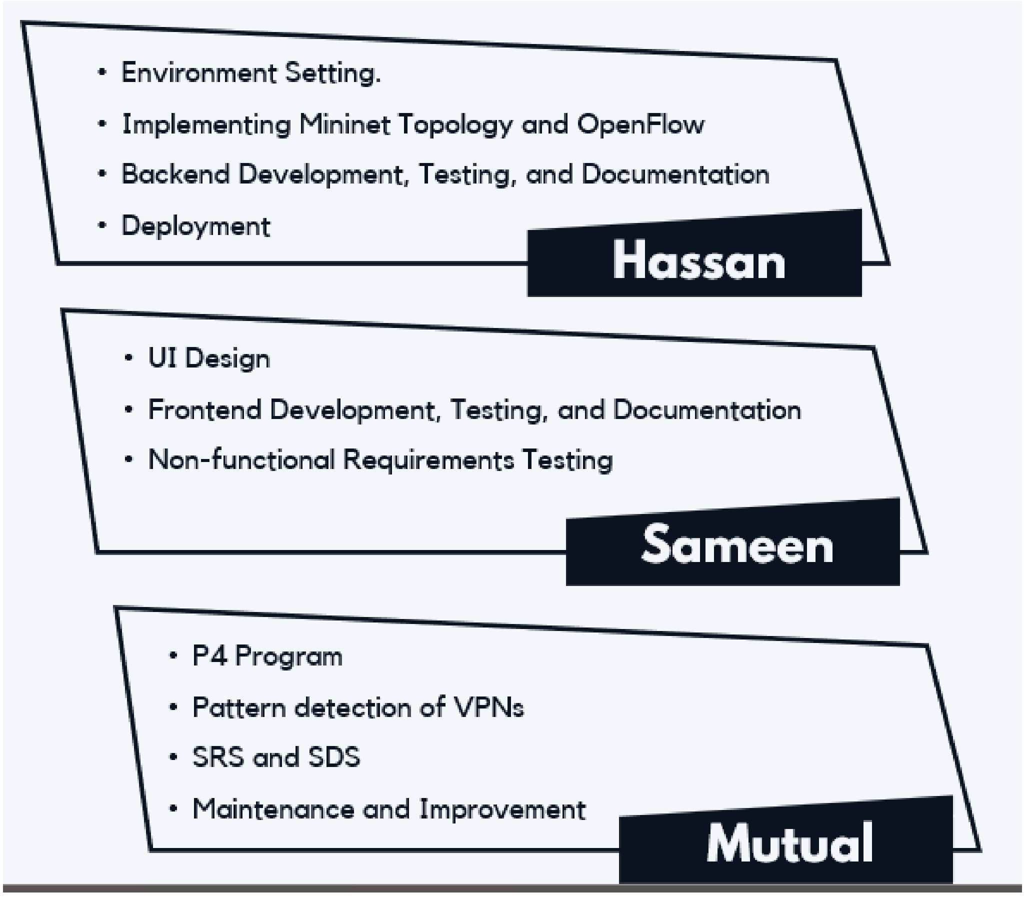
DEC
- System Design and Architecture
- Front End Development
- Unit Testing
- Documentation

JAN
- Back End Development
- Unit Testing
- Documentation

FEB
- Integration and System Testing
- Testing Non-Functional Requirements

MAR
- Deployment (Beta Release)
- Maintenance and Improvements

WORK DIVISION



- Environment Setting.
- Implementing Mininet Topology and OpenFlow
- Backend Development, Testing, and Documentation
- Deployment

Hassan

- UI Design
- Frontend Development, Testing, and Documentation
- Non-functional Requirements Testing

Sameen

- P4 Program
- Pattern detection of VPNs
- SRS and SDS
- Maintenance and Improvement

Mutual

COSTING

ITEM	COST
Domain for 1 year	PKR 1300
System to set up virtual Switch	PKR 60,000
(Optional) Edgecore DCS802-12.8T Programmable Data Center Switch	$ 32,000

REFERENCES

[1] Arafune, M., Goswami, B., Kulkarni, M., Venkatachalam, N., & Asadollahi, S. (2023). Lightweight anti ddos security tool: Edge level filtering in SDN using P4. 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT). https://doi.org/10.1109/icecct56650.2023.10179747

[2] Goel, A., Kashyap, A., Reddy, B. D., Kaushik, R., Nagasundari, S., & Honnavali, P. B. (2022). Detection of VPN network traffic. *2022 IEEE Delhi Section Conference (DELCON)*. https://doi.org/10.1109/delcon54057.2022.9753621

[3] Papadogiannaki, E., & Ioannidis, S. (2021). A survey on encrypted network traffic analysis applications, techniques, and countermeasures. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457904

[4] Draper-Gil, G., Lashkari, A. H., Mamun, M. S., & A. Ghorbani, A. (2016). Characterization of encrypted and VPN traffic using time-related features. *Proceedings of the 2nd International Conference on Information Systems Security and Privacy*. https://doi.org/10.5220/0005740704070414