



Intelligent Website Traffic Detection And Analysis Using Machine Learning

¹ Dr. M. Florence Dayana, ² S. Aswini

¹ Assistant Professor, ² Scholar

^{1,2} Department of Computer Science,

^{1,2} Bon Secours College for Women, Thanjavur

Abstract— One of the numerous uses for traffic analysis is to assess the effectiveness and security of network administration and operations. As a result, network traffic analysis is seen to be essential for enhancing network security and functionality. Various machine learning techniques for traffic analysis are covered in this study. The growth of artificial intelligence and increased network traffic need the development of new techniques for virus analysis, intrusion detection, Internet traffic classification, and other security-related tasks. Network issues may be effectively resolved by machine learning (ML). This study presents an overview of the methods utilised in the traffic analysis.

Keywords— Machine learning (ML), Internet traffic, performance and security.

I. INTRODUCTION

The development of smart devices has made networks more dynamic and diverse, which has forced network operators to look for fresh management ideas. As a result, one of the main

challenges is creating network architecture that can manage heterogeneity and optimise resource utilisation efficiency. A number of solutions, such as Network Slicing (NS) and Machine Learning (ML), have been put forth in this context. Combining these ideas can result in performance optimisation on a massive scale and autonomous and intelligent network resource management, both of which can meet 5G quality of service (QoS) criteria.

With the use of a training dataset or environment, machine learning (ML) is really used to solve complicated problems without the need for explicit programming. Its effectiveness has been confirmed, and it has shown promise in a number of fields, including network management. However, NS offers network as a service (NaaS) for many use cases, enabling infrastructure providers to grant multiple users access to their physical network.

It divides the network infrastructure into discrete network slices, each of which has distinct performance needs and resource requirements. For 5G networks and beyond, NS offers a number of

benefits. Initially, it can accommodate several tenants using the same physical network server. Second, NS may ensure service level agreements (SLA) and manage the diverse services with strict QoS. Third, it allows for a flexible and effective use of the limited resources by addressing the various service requirements rather than the "one-size-fits-all" principle.

II.LITERATURE SURVEY

With the incorporation of machine learning techniques, network traffic analysis—a crucial part of contemporary network management and security—has been going through a paradigm shift. This survey of the literature examines significant contributions and current developments in the field of machine learning-based network traffic analysis and prediction.

Techniques for Network Traffic Analysis: Lakhina et al. suggested a statistical measure-based approach for identifying network-wide traffic abnormalities. In order to identify departures from typical traffic patterns, they developed the idea of "flow level anomalies" and used a range of statistical measures [6]. An overview of IP flow-based intrusion detection methods was given by Sperotto et al. They talked on the benefits and difficulties in this setting and emphasised the need of flow-level data in detecting network risks.

Machine Learning for Network Security: A thorough analysis of network anomaly detection from a machine learning standpoint was provided by Ahmed et al. They divided anomaly detection approaches into five groups: SVM-based, information-theoretic, statistical, clustering, and closest neighbour [7]. He et al. (2017) introduced a network intrusion detection method based on deep learning. In order to identify intrusions with greater accuracy, they used a convolution neural network

(CNN) to extract intricate patterns from network traffic data [8].

Predictive modelling and optimisation: Zhang et al. suggested a hybrid model for network traffic prediction that combines support vector regression (SVR) and long short-term memory (LSTM) networks. Their methodology helped with resource allocation and congestion management by demonstrating accurate short-term traffic projections. A framework for dynamic network resource allocation based on reinforcement learning was presented by Zhang et al. By engaging with the network environment and adjusting to shifting traffic conditions, the model discovered the best resource allocation strategies [9].

Internet Traffic Categorisation: The problem of categorising encrypted network traffic was tackled by Sharma et al. To enable more efficient traffic analysis and management, they classified encrypted traffic into meaningful classifications using deep packet inspection and machine learning algorithms.

Network Traffic Plagiarism Detection: A framework for network traffic plagiarism detection was presented by Kwon et al. They developed a unique method for preserving data integrity in network communication by evaluating traffic patterns and spotting duplicate data packets.

Anomaly Detection and Ensemble Methods: Xu et al. presented an ensemble-based anomaly detection system that combines one-class support vector machines with autoencoders. Their method performed well in detecting abnormalities in the network.

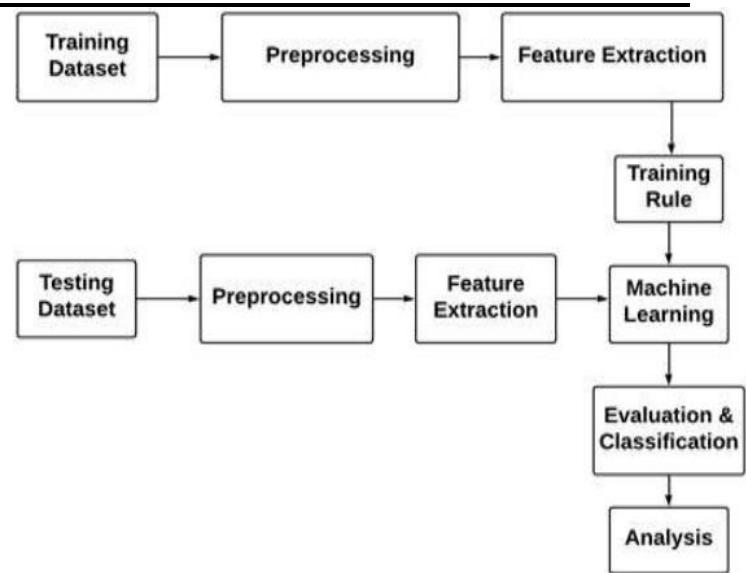
III. SYSTEM IMPLEMENTATION

Existing System

However, it is difficult to properly characterise the network state since the network behaviours vary based on a number of circumstances, including user mobility, location, and social events. Allocating resources to network slicing instances becomes more challenging in this situation. Furthermore, it took a lot of effort to manually create network slices due to the massive increase in new network traffic and apps. As a result, the procedure should be adaptable, respond to requests instantly, and be dependent on end-user traffic patterns.

Proposed System

ML provides these advantages by easily comprehending the traffic and automatically generating network slicing since it can examine a lot of data and quickly identify a valuable pattern from it. It makes it possible for the system to develop rules for task automation. The capacity of machine learning to handle complicated issues is its greatest benefit. Furthermore, the machine learning approach to the examination of these enormous datasets offers new findings. Consequently, by including these technologies into traffic analysis and NS, network operators will be able to provide self-optimizing, self-configuring, and self-healing networks, which ultimately offer zero-touch management. One might draw a broad contrast between supervised and unsupervised learning within the subject of machine learning.



Proposed Architecture

Methodology

1 Self-Similarity and TES: Analyse communication systems and identify threats by using Industrial Access Control & Security Systems.

Two Networks of Wireless Sensors WSN: This technology is applicable to big systems, including commercial applications, where security is essential to their operation. It is common practice to categorise assaults in wireless sensor networks in order to investigate trends and potential defences.

3 Analysis of Flow: Anonymity networks are found via flow analysis. It has a remarkable accuracy in identifying encrypted anonymity networks. Identifying anonymity networks, identifying network activity within encrypted apps, and profiling applications are the three primary uses of flow analysis.

4 User Intention-Based Traffic Dependence Analysis: This method makes use of pre-existing frameworks and algorithms to study network events and user behaviour on a host based on their login credentials. This can result in the detection of correlations, the identification of irregularities, and empirical evaluations of the precision, safety, and effectiveness of algorithms.

The five traffic anomaly detection algorithms fall into two major groups. The first category, which

includes pattern mining, similarity, and statistical techniques, finds flow outliers. Trajectory outliers, including offline and online processing, are derived in the second category.

6 Intrusion Detection System: These are hardware or software programs that watch a network or system for harmful activities.

Experimental Results



IV. CONCLUSIONS

The dataset with flow statistical characteristics is categorised using machine learning methods such as supervised and unsupervised with set of algorithms correspondingly. Determine and categorise the IP traffic using both machine learning approaches. Using a collection of algorithms, the supervised and unsupervised machine learning approaches categorise the dataset as either real-time or non-real-time traffic. When compared to the outcomes of supervised and unsupervised algorithms separately, the suggested approach produces superior results.

Accuracy, usability, and actionable insights may all be improved with future website traffic detection improvements. Here are some suggestions for improving a system that detects website traffic:

1. Real-Time Analytics Feature: Give real-time information on page views, user activity, and visitor numbers.

Improvement: Include session replays and live heatmaps to see how people move across the website in real time.

2. Advanced Segmentation Feature: Divide traffic according to user demographics, device type, browser, region, and referral sources.

Improvement: To find high-value user subgroups, apply AI-driven predictive analysis.

REFERENCES

- [1] J.S. Banerjee, A. Chattopadhyay, and A. Chakraborty. In cognitive radio networks, a non-uniform quantised data fusion rule reduces control channel overhead for cooperative spectrum sensing. The 7th International Advance Computing Conference (IACC) was held by IEEE in 2017. IEEE.
- [2] J.S. Banerjee, A. Chattopadhyay, and Chakraborty, A. Non-uniform quantised data fusion rule for cooperative spectrum sensing in cognitive radio networks to reduce control channel overhead and save data rate. p. 837-851 in Wireless Personal Communications, 2019 (104(2)).
- [3] A. Rueda. An examination of methods for characterising traffic in telecommunications networks. Canadian Conference on Electrical and Computer Engineering Proceedings, 1996. IEEE.
- [4] Zincir-Heywood, A.N., and Shahbar, K. How far can we push flow analysis to find encrypted anonymity network traffic? in NOMS 2018–2018 Management and Operations of IEEE/IFIP Networks
- [5] S. Axelsson. Intrusion detection systems: A taxonomy and survey. Technical report, 2000.
- [6] Wang, P., Y. Li, and C.K. Reddy, A overview of machine learning for survival analysis. ACM Computing Surveys (CSUR), 2019. 51(6): p. 110.
- [7] S. Agrawal, N. Namdev, and S. Silkari, recent developments in the categorisation of internet traffic using machine learning. 60: pp. 784-791, Procedia Computer Science, 2015.
- [8] Cheng, Y., et al., Bridging computer network research and machine learning: a survey. 1(1–4): pp. 1–15 in CCF Transactions on Networking, 2019. [9] S. Mukkamala, A. Sung, and G. Janoski. Intrusion detection: support vector machines and neural networks. in proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO. 2002.
- [10] Robust identification of smartphone apps by encrypted network traffic analysis, Taylor, V.F., et al. IEEE Transactions on Information Security and Forensics, 13(1), pp. 63–78, 2017.
- [11] Kim, J., et al., Clustered patterns for multivariate network traffic analysis. Computing, 2019. 101(4): p. 339-361.
- [12] M. Shafiq et al. methods for classifying network traffic and machine learning algorithms for comparison analysis. The IEEE held the 2nd International Conference on Computer and Communications (ICCC) in 2016.
- [13] Paxson, V. and Sommer, R. Outside the closed world: On detecting network intrusions with machine learning. 2010 IEEE Symposium on Privacy and Security. IEEE.
- [14] Zhang, H., et al. Traffic dependency analysis for anomaly identification based on user intention. in 2012. IEEE Symposium on Security and Privacy Workshops. 2012. IEEE.
- [15] Djenouri, Y., et al., A survey on urban traffic anomaly detection methods. IEEE Access, 7: pp. 12192–12205, 2019.
- [16] Zamani et al., M. a mobile agent and risk theory-based DDoS-aware intrusion detection system. Computational Intelligence and Security International Conference, 2009. IEEE.
- [17] K.C. Claffy, A. Dainotti, and A. Pescape, Traffic classification issues and future directions. pp. 35–40 in IEEE Network, 2012. 26(1).
- [18] Zhou, W., et al. Use feed-forward neural network to classify Internet traffic. 2011. IEEE, International Conference on Computational Problem-Solving (ICCP).
- [19] E. Guven and A.L. Buczak, A study of data mining and machine learning technologies for cyber security intrusion detection. p. 1153–1176 in IEEE Communications Surveys & Tutorials, 2015, 18(2).