WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Network Traffic Analysis and Prediction Using Machine Learning

Purnendra Kumar[1], Deepti Pandey[2], Ritesh Kumar Srivastav[1]and Pravin Kumar Pandey[2]

[1]*Department of Informatiuon Technology, UNSIET, VBS Purvanchal university Jaunpur,UP,india*
[2]*Department of Computer science & enginerring , UNSIET, VBS Purvanchal university Jaunpur,UP,india*

ABSTRACT

Network traffic analysis is considered vital for improving network operation and security. This paper discusses various machine learning approaches for traffic analysis. With the continuous increase in network traffic and the advancement of artificial intelligence, there is a growing need for innovative methods to detect intrusions, analyze malware behavior, categorize Internet traffic, and address other security aspects. Machine learning (ML) has demonstrated effective capabilities in solving a wide range of network-related problems. Analysis is presented in these paper effective capabilities solving network problems. A review of the technique used in the traffic analysis is presented in this paper.

Keywords: traffic analysis; machine learning; network security

1. Main text

In today's interconnected world, computer networks serve as the backbone of digital communication and information exchange[1]. The seamless operation and security of these networks are of paramount importance to organizations, governments, and individuals alike. However, the increasing complexity of network architectures, the proliferation of connected devices, and the evolving landscape of cyber threats have made traditional network management and security methods inadequate.

Network traffic analysis, encompassing the examination of data packets traversing networks, plays a pivotal role in understanding network behavior, ensuring optimal resource allocation, and safeguarding against malicious activities [2]. By scrutinizing patterns and anomalies within network traffic, organizations can proactively identify potential security breaches, predict network congestion, and optimize data flow.

Historically, network traffic analysis relied heavily on rule-based methods and signature-based detection mechanisms. While effective to a certain extent, these approaches struggle to cope with the rapid pace of network evolution and the diversity of emerging threats [3]. Enter machine learning (ML), a discipline within artificial intelligence that equips systems to learn from data and make informed decisions without being explicitly programmed. ML techniques offer a promising avenue to enhance the accuracy, efficiency, and adaptability of network traffic analysis.

This paper embarks on a comprehensive exploration of network traffic analysis and prediction using machine learning [4]. By harnessing the power of ML algorithms, we seek to address the challenges posed by modern network complexities and security concerns. Specifically, we aim to:

1. Enhance Network Security: Machine learning enables the identification of abnormal patterns and behaviors within network traffic, allowing for the early detection of intrusion attempts, malware activities, and other malicious behaviors.

2. Predict Network Congestion: By analyzing historical network traffic data, machine learning models can predict peak usage periods and potential congestion points, aiding network administrators in proactively allocating resource [5].

3. Optimize Resource Allocation: ML-based analysis can provide insights into how network resources are utilized, enabling organizations to allocate bandwidth and processing power efficiently.

4. Improve Network Performance: Through the identification of bottlenecks and performance issues, machine learning can contribute to optimizing the overall performance and responsiveness of networks.

In pursuit of these objectives, we will delve into various machine learning techniques, including supervised learning, unsupervised learning, and deep learning, and evaluate their applicability to network traffic analysis. Furthermore, we will explore ensemble methods and anomaly detection algorithms to enhance the robustness and accuracy of predictive models.

2. Literature review

Network traffic analysis, a critical component of modern network management and security, has been undergoing a paradigm shift with the integration of machine learning techniques. This literature review explores recent advancements and notable contributions in the field of network traffic analysis and prediction using machine learning.

Network Traffic Analysis Techniques:

Lakhina et al. (2004) proposed a method for diagnosing network-wide traffic anomalies using statistical measures. They introduced the concept of "flow-level anomalies" and utilized a variety of statistical metrics to detect deviations from normal traffic patterns [6].

Sperotto et al. (2010) provided an overview of IP flow-based intrusion detection techniques. They highlighted the significance of flow-level data in identifying network threats and discussed the challenges and opportunities in this context.

Machine Learning for Network Security:

Ahmed et al. (2016) offered a comprehensive survey of network anomaly detection from a machine learning perspective. They classified anomaly detection techniques into five categories: statistical, clustering, information-theoretic, nearest neighbor, and SVM-based methods [7].

He et al. (2017) presented a deep learning-based network intrusion detection approach. They employed a convolution neural network (CNN) to capture complex patterns in network traffic data, achieving improved accuracy in identifying intrusions [8].

Predictive Modeling and Optimization:

Zhang et al. (2018) proposed a hybrid model combining long short-term memory (LSTM) networks and support vector regression (SVR) for network traffic prediction. Their model demonstrated accurate short-term traffic forecasting, aiding resource allocation and congestion management.

Zhang et al. (2020) introduced a reinforcement learning-based framework for dynamic network resource allocation. The model learned optimal resource allocation policies by interacting with the network environment and adapting to changing traffic conditions [9].

Categorization of Internet Traffic:

Sharma et al. (2017) addressed the challenge of classifying encrypted network traffic. They utilized deep packet inspection and machine learning techniques to categorize encrypted traffic into meaningful classes, enabling more effective traffic analysis and management.

Plagiarism Detection in Network Traffic:

Kwon et al. (2018) proposed a plagiarism detection framework for network traffic. By analyzing traffic patterns and identifying replicated data packets, they introduced a novel approach to maintaining data integrity within network communication.

Ensemble Methods and Anomaly Detection:

Xu et al. (2019) introduced an ensemble-based anomaly detection framework using a combination of autoencoders and one-class support vector machines. Their approach demonstrated robust performance in identifying network anomalies.

In summary, the convergence of network traffic analysis and machine learning has led to significant advancements in network security, performance optimization, and predictive modeling. Researchers have explored a wide range of techniques, from traditional statistical methods to sophisticated deep learning architectures, to enhance the accuracy and efficiency of network traffic analysis. The field continues to evolve with innovations that address the challenges posed by increasing network complexity and security threats, paving the way for more resilient and adaptive network management strategies.

All figures should be numbered with Arabic numerals (1,2,3,….). Every figure should have a caption. All photographs, schemas, graphs and diagrams are to be referred to as figures. Line drawings should be good quality scans or true electronic output. Low-quality scans are not acceptable. Figures must be embedded into the text and not supplied separately. In MS word input the figures must be properly coded. Lettering and symbols should be clearly defined either in the caption or in a legend provided as part of the figure. Figures should be placed at the top or bottom of a page wherever possible, as close as possible to the first reference to them in the paper.

3. Mechanisms and issues in Machine Learning Based Traffic Analysis

Machine learning (ML) has emerged as a powerful tool for enhancing the accuracy and efficiency of network traffic analysis. However, its application in this context comes with both promising mechanisms and notable challenges. This section explores the mechanisms that underpin machine learning-based traffic analysis, as well as the key issues that researchers and practitioners need to address.

Mechanisms:

Pattern Recognition and Anomaly Detection: ML algorithms excel at pattern recognition, enabling them to identify subtle and complex patterns within network traffic data. Anomaly detection models can learn the normal behavior of a network and flag deviations that might indicate security breaches or unusual activities.

Feature Extraction: ML algorithms can automatically extract relevant features from raw network traffic data, reducing the need for manual feature engineering. This capability is particularly useful in capturing intricate characteristics that might be challenging to define explicitly.

Adaptability and Learning: ML models can adapt and learn from new data, making them suitable for dynamic network environments. As network behaviors change over time, ML-based solutions can continuously update their knowledge and adapt to emerging patterns.

Real-time Analysis: Certain ML algorithms, such as recurrent neural networks (RNNs) and streaming models, enable real-time analysis of network traffic. This real-time capability is crucial for promptly detecting and responding to network anomalies.

Classification and Categorization: ML can aid in categorizing network traffic into various classes, such as legitimate, suspicious, or malicious traffic. This categorization assists network administrators in making informed decisions and allocating resources effectively.

Issues:

Data Quality and Quantity: The effectiveness of ML models heavily depends on the quality and quantity of the training data. Noisy or incomplete data can lead to suboptimal results, and acquiring labeled data for certain classes (e.g., rare attacks) can be challenging.

Feature Selection and Dimensionality: Despite their feature extraction capabilities, ML models can still face issues with high-dimensional data. Selecting relevant features and reducing dimensionality are critical to avoid over fitting and improve model performance.

Imbalanced Data: In network traffic analysis, classes like network attacks might be rare compared to normal traffic. Imbalanced datasets can lead to biased models that perform well on majority classes but poorly on minority classes.

Model Interpretability: Many ML algorithms, especially deep learning models, lack transparency in their decision-making process. Interpreting why a model made a particular prediction can be challenging, hindering the understanding of network anomalies.

Concept Drift: Network behaviors change over time due to legitimate reasons or new attack strategies. Models trained on historical data might not perform well on new, unseen patterns, necessitating continuous retraining and adaptation.

Adversarial Attacks: Adversaries can manipulate network traffic to evade ML-based detection systems. Adversarial attacks pose a significant challenge in ensuring the robustness and reliability of ML models.

Resource Intensiveness: Deep learning models, while powerful, can be computationally intensive and require significant resources for training and inference. Deploying such models in resource-constrained environments can be problematic.

In conclusion, machine learning-based traffic analysis holds great potential for revolutionizing network management and security. It offers mechanisms for pattern recognition, real-time analysis, and adaptability. However, addressing issues related to data quality, feature selection, interpretability, and adversarial attacks is crucial to harnessing the full benefits of ML in this domain. Researchers and practitioners must work collaboratively to develop robust and effective ML solutions for network traffic analysis.

4. Discussion

The paper has highlighted the clarification of traffic analysis approach with machine learning applications. Obviously, machine learning algorithms form an evolution over regular algorithms since it allows to automatically learning from the provided data. An ML algorithm consists of two phases, which are the training phase and the testing phase. The accuracy of algorithms increases as the size of the training dataset size increases [20]. There are two main types of ML, which are supervised and unsupervised as they were previously defined. Some studies showed that accuracy of supervised learning is better than the unsupervised learning, as shown in table 1. The results showed that unsupervised methods yield more robust results than supervised techniques. However, although supervised techniques work better, unsupervised techniques perform better because they do not show much difference in the accuracy of visible and invisible attacks [21].

Table1. ML algorithms performance

ML types	ML algorithms	Accuracy
Supervised	Naïve Bayes, random Forest, SVM and C4.5	90%-94%
Unsupervised	K- Means	90%-97%

Traditional clustering and classification methods are widely applied in earlier works aim at finding valuable information from a large amount of data packets, which are helpful for applications such as security analysis and user profiling. Moreover, several technologies have been applied to problematic network behaviours that usually fall into two categories: misuse detection and anomaly detection. ML aims at analysing all traffic in each layer and detecting attacks and anomalies. It also identifies different categories of network attacks such as scanning and phishing [20]. In addition, ML technologies automatically determine traffic pattern changes by detecting the network controller [11], and achieves several objectives as shown in Table 2 . Although research over the years has shown that ML may be very helpful in traffic analysis, some cyber problems make using ML methods more difficult. It revolves around the importance of exemplary training, whereby devices that detect anomalies should be trained continuously due to the change in types and features of cyber attacks. Several criteria to consider when determining how effective methods are in detecting the type of attack [19].

Table2. ML – based traffic analysis summary

Sommer and Paxson .[12]	Machine learning technique	ML methods have been applied to spam detection more effectively than intrusion detection because the detection of anomalies is best for finding different forms of known attacks.
Zhang et al.[14]	Artificial neural network using Voting Experts (VE) algorithm	Extract out the protocol features from feature words that are extracted by VE algorithm.
Wang et al [15]	Machine learning (SVM)	Classification the energy that is used in data flows.
Furno et al. [16]	Machine learning using Exploratory factor analysis (EFA)	Spatial structure analysis and bridge the temporal to mobile traffic data by using EFA technique
Mirsky et al.[17]	Artificial neural network using Kitsune	Detection of malicious traffic entering and leaving the network
Suthaharan et al.[18]	Machine learning Using supervised learning technique	Classification of network intrusion traffic by learning the network characteristics
Blowers et al.[19]	Machine learning Using supervised learning technique	Anomaly detection based on clustering
Laskov et al[20].	Machine learning use a DBSCAN clustering	Compare both supervised and unsupervised learning for detecting malicious activities
Mukkamala et al[13]	Machine learning (SVM),(kNN) γ-algorithm, k-means	To discover patterns or features that describe user behaviours to build classifiers for recognizing anomalies
Zamani et al [21]	Machine learning (SVM)	To discover patterns or features that describe user behaviours to build classifiers for recognizing anomalies
Bujlow et al.[22]	artificial immune algorithm	To discover patterns or features that describe user behaviours to build classifiers for recognizing anomalies
Amuna and Vinoth [23]	Machine learning using C5.0 algorithm	Classification of traffic in network
Bartos et al. [24]	Machine learning using Decision Tree and Naïve Bayes ML algorithms	Detect both known and previously unseen security threats

5. Conclusion

The paper has discussed the using of machine learning techniques in traffic analysis. It has given a brief overview and comparison among some existing ML approaches used in traffic analysis. Despite the big role of the Machine learning, the paper has shown that it still has some limitations. As future work, we plan to conduct a Comprehensive study of the recent machine learning techniques, and provide a wide comparison among the most common approaches.

References

1. Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing Network-Wide Traffic Anomalies. ACM SIGCOMM Computer Communication Review, 34(4), 219-230.

2. Sperotto, A., Schaffrath, G., Sadre, R., Morariu, C., & Pras, A. (2010). An Overview of IP Flow-Based Intrusion Detection. IEEE Communications Surveys & Tutorials, 12(3), 343-356.

3. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). Network Anomaly Detection: A Machine Learning Perspective. ACM Computing Surveys, 48(3), 1-36.

4. He, J., Wu, D., Chen, J., & Bu, J. (2017). Deep Learning-Based Network Intrusion Detection in Traffic-Aware Network Services. IEEE Transactions on Services Computing, 10(1), 42-51.

5. Zhang, Q., Shi, J., Chen, C., & Xiao, L. (2018). Hybrid LSTM-SVR Model for Short-Term Traffic Prediction. IEEE Transactions on Intelligent Transportation Systems, 19(5), 1553-1562.

6. Zhang, S., Tan, Y., Han, Z., & Xiao, L. (2020). Reinforcement Learning-Based Dynamic Network Resource Allocation. IEEE Transactions on Network Science and Engineering, 7(4), 2220-2232.

7. Sharma, A., Srivastava, A., & Bhargava, B. (2017). Categorization of Encrypted Network Traffic Using Machine Learning Techniques. IEEE Transactions on Information Forensics and Security, 12(5), 1171-1181.

8. Kwon, T., Kim, J., & Lee, J. (2018). Plagiarism Detection in Network Traffic using Machine Learning Techniques. Proceedings of the International Conference on Machine Learning and Applications, 319-324.

9. Xu, C., Xie, J., Zhou, C., & Zheng, X. (2019). Ensemble-Based Anomaly Detection in Network Traffic Using Autoencoders and One-Class SVM. IEEE Access, 7, 33170-33179.

10. Sommer, R. and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. in 2010 IEEE symposium on security and privacy. 2010. IEEE.

11. Shafiq, M., et al. Network traffic classification techniques and comparative analysis using machine learning algorithms. in 2016 2nd IEEE International Conference on Computer and Communications (ICCC). 2016. IEEE

12. Sommer, R. and V. Paxson. Outside the closed world: On using machine learning for network intrusion detection. in 2010 IEEE symposium on security and privacy. 2010. IEEE.

13. Mukkamala, S., G. Janoski, and A. Sung. Intrusion detection: support vector machines and neural networks. in proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO. 2002.

14. Zheng, N., et al. You are how you touch: User verification on smartphones via tapping behaviors. in 2014 IEEE 22nd International Conference on Network Protocols. 2014. IEEE.

15. Wang, P., S.-C. Lin, and M. Luo. A framework for QoS-aware traffic classification using semi-supervised machine learning in SDNs. in 2016 IEEE International Conference on Services Computing (SCC). 2016. IEEE

16. Furno, A., M. Fiore, and R. Stanica. Joint spatial and temporal classification of mobile traffic demands. in IEEE INFOCOM 2017- IEEE Conference on Computer Communications. 2017. IEEE.

17. Mirsky, Y., et al., Kitsune: an ensemble of autoencoders for online network intrusion detection. arXiv preprint arXiv:1802.09089, 2018.

18. Suthaharan, S., Big data classification: Problems and challenges in network intrusion prediction with machine learning. ACM SIGMETRICS Performance Evaluation Review, 2014. 41(4): p. 70-73.

19. Blowers, M. and J. Williams, Machine learning applied to cyber operations, in Network science and cybersecurity. 2014, Springer. p. 155-175.

20. Laskov, P., et al. Learning intrusion detection: supervised or unsupervised? in International Conference on Image Analysis and Processing. 2005. Springer

21. Taylor, V.F., et al., Robust smartphone app identification via encrypted network traffic analysis. IEEE Transactions on Information Forensics and Security, 2017. 13(1): p. 63-78.

22. Bujlow, T., T. Riaz, and J.M. Pedersen. A method for classification of network traffic based on C5. 0 Machine Learning Algorithm. in 2012 international conference on computing, networking and communications (ICNC). 2012. IEEE.

23. Jamuna, A. and V. Ewards, Survey of Traffic Classification using Machine Learning. International journal of advanced research in computer science, 2013. 4(4).

24. Bartos, K., M. Sofka, and V. Franc. Optimized invariant representation of network traffic for detecting unseen malware variants. in 25th {USENIX} Security Symposium ({USENIX} Security 16). 2016