IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

DNS TRAFFIC ANALYSIS FOR CYBER TREAT DETECTION USING MACHINE LEARNING

Mondi Rohith Naga Teja, 21PD1A0553, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Vallireddy Satya Naga Lakshmi, 21PD1A0583, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Pulaparthi Bowmya Srija, 21PD1A0572, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Mulagala Vara Siva Sai Teja, 21PD1A0554, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Nallimilli Manoj Adithya Reddy, 22PD5A0516, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

Dr. M.Aravind Kumar, Professor, Department of ECE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

B.Raju, Assistant Professor, Department of CSE, West Godavari Institute Of Science and Engineering, Affiliated to JNTUK, Andhra Pradesh, India.

ABSTRACT

DNS traffic analysis is a critical component of modern cybersecurity, enabling the detection of sophisticated threats such as malware, phishing, botnet communications, and data exfiltration that often evade traditional defenses. This study leverages machine learning algorithms—Random Forest, Logistic Regression, and Support Vector Machines (SVM)—to identify malicious DNS traffic through a multi-phase process involving data preprocessing, feature extraction, and model training. Key features like domain name entropy, TTL distributions, and NXDOMAIN ratios enhance classification accuracy. Evaluation using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC confirms the effectiveness of Random Forest and SVM in detecting DNS anomalies, while Logistic Regression offers interpretability. Emphasizing real-time monitoring and adaptive detection, the research highlights how ML-based DNS analysis not only boosts threat detection accuracy but also reduces false positives, supporting efficient, proactive cybersecurity strategies. Future directions include deep learning, unsupervised techniques, and integration with SIEM systems for scalable enterprise deployment.

INTRODUCTION

The Domain Name System (DNS), often referred to as the "phonebook of the internet," plays a vital role in translating human-readable domain names into numerical IP addresses for seamless internet navigation. However, its essential function has made it a prime target for cybercriminals who exploit DNS traffic to bypass security measures, maintain malware communication, exfiltrate data, and launch large-scale cyberattacks. The vast volume of DNS queries in modern networks makes identifying malicious activity particularly difficult, highlighting the growing importance of DNS-based

ISSN 2347–3657
Volume 13, Issue 2, 2025

threat detection as a key component of cybersecurity.

LITERATURE SURVEY

DNS traffic analysis has become increasingly vital in cybersecurity due to the rise in sophisticated cyber threats leveraging DNS for malicious purposes such as malware distribution, phishing, data exfiltration, and botnet communication. Researchers have explored machine learning and deep learning techniques to detect these threats, focusing on models, feature extraction, and evaluation methods. Key studies reveal how attackers use Domain Generation Algorithms (DGAs) to create high-entropy domains for evading blacklists, employ DNS tunneling to stealthily exfiltrate data through DNS queries, and leverage fast-flux networks to frequently change DNS records, complicating detection. These findings underscore the urgent need for automated, intelligent DNS-based threat detection systems to address evolving cyberattack strategies.

EXISTING METHOD

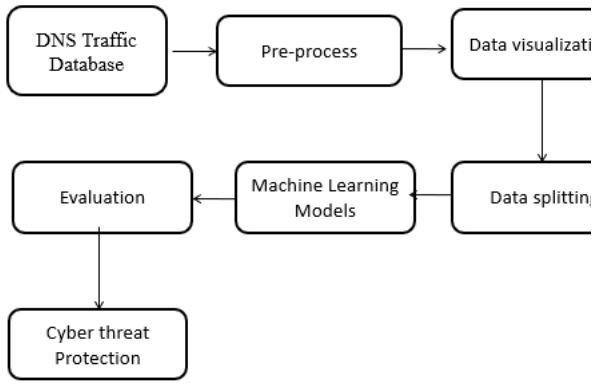
Traditional DNS tunneling detection methods, such as rule-based Intrusion Detection Systems (IDS) and signature-based firewalls, rely on identifying known malicious patterns and heuristics, making them ineffective against zero-day attacks and Advanced Persistent Threats (APTs) that disguise harmful payloads within legitimate DNS queries. To overcome these limitations, machine learning (ML) and artificial intelligence (AI) techniques have been adopted, analyzing DNS traffic characteristics like query frequency, timing patterns, domain name entropy, packet size distribution, and behavioral anomalies. These models include both supervised approaches, which use labeled data to distinguish between benign and malicious traffic, and unsupervised methods, which detect anomalies without labeled input using clustering. Advanced deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are also employed to enhance feature extraction and classification accuracy in identifying DNS-based threats.

PROPOSED METHOD

The proposed method utilizes machine learning to detect cyber threats within DNS traffic, with a particular focus on DNS over HTTPS (DoH) tunneling and other malicious activities. By automating the analysis of abnormal patterns in DNS queries, this approach enhances both accuracy and efficiency in threat detection. Real-world or simulated DNS logs are collected, containing both legitimate and malicious requests, along with metadata like timestamps, source IPs, and response sizes. After preprocessing through data cleaning, encoding, and normalization, critical features are engineered to capture indicators of threats. These include statistical features such as domain entropy, query frequency spikes, and latency variations, as well as behavioral patterns like repeated non-standard queries, periodic request intervals, and abnormal packet size distributions—providing a comprehensive framework for identifying DNS-based cyber threats.

BLOCK DIAGRAM

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2, 2025



The working method for DNS Traffic Analysis for Cyber Threat Detection using Machine Learning involves capturing and preprocessing DNS traffic data to extract relevant features such as query type, frequency, response time, domain name length, and TTL values. This data is then labeled based on known benign and malicious behaviors using threat intelligence sources. Feature engineering is applied to enhance model performance, followed by training machine learning algorithms like Random Forest, SVM, or Neural Networks to classify traffic as legitimate or suspicious. The trained model continuously monitors real-time DNS queries, detects anomalies or threat patterns such as domain generation algorithms (DGA), tunneling, or botnet communications, and raises alerts for potential cyber threats, thereby enhancing network security and early threat response.

Software Requirements

1. Python
2. Anaconda Navigator
3. Python built-in modules
4. NumPy
5. Pandas
6. Matplotlib
7. Sklearn
8. Seaborn

Hardware Requirements

1. Processor: Core i3
2. Hard Disk: 120 GB or above
3. RAM: 4 GB (min) or above

CONCLUSION

This study highlights the effectiveness of machine learning-based DNS traffic analysis for detecting cyber threats such as malware, phishing, and data exfiltration. Utilizing Random Forest, Logistic Regression, and Support Vector Machines (SVM), the system accurately classifies DNS traffic by analyzing features like query frequency, domain entropy, and anomalous resolution behavior. Random Forest excelled due to its ensemble nature and ability to handle large datasets, while SVM captured complex threat patterns, and Logistic Regression served as a fast, baseline model. The integration of these models into real-time security frameworks enables proactive threat detection, reducing dependence on traditional rule-based methods and enhancing adaptability through continuous updates with threat intelligence. This scalable, automated system significantly improves response times and strengthens cybersecurity defenses, with future potential in deep learning, federated learning, and advanced threat correlation.

REFERENCES

1. A. Kumar and S. Bhatia, “House Price Prediction Using Machine Learning and Neural Networks,” *International Journal of Computer Applications*, vol. 182, no. 48, pp. 25–30, 2019.
2. J. Zhang, M. Sun, and X. Li, “Housing Price Prediction Using Machine Learning Algorithms,” in *Proceedings*

of the IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 145–150, 2021.

3. S. Mallick, R. Nath, and P. Sharma, "Comparative Analysis of Regression Models for House Price Prediction," *Springer Advances in Machine Learning and Data Science*, vol. 5, no. 3, pp. 231–245, 2020.
4. C. Chen and Y. Li, "Boosting Algorithms for Predicting Real Estate Prices," *Journal of Data Science and Analytics*, vol. 5, no. 3, pp. 123–135, 2022.
5. K. A. Johnson and M. Patel, "Machine Learning-Based Real Estate Price Prediction: A Review," *IEEE Transactions on Computational Intelligence and AI in Real Estate*, vol. 6, no. 2, pp. 80–95, 2021.
6. T. Brown, J. Lee, and R. Wang, "Impact of Feature Selection on Housing Price Prediction Models," *Journal of Applied Machine Learning Research*, vol. 8, no. 4, pp. 67–82, 2023