WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

ML-Based Network Anomaly Detection

Rudra Kareliya[1], Nikhil Mokal[2], Prem Shah[3], Harsh Solanki[4], Sneha Shingare[5]

[1,2,3,4] Second year Student, Cyber Security Engineering, Shah & Anchor Kutchhi Engineering College, Mumbai, India
[5]Senior Project Mentor, Cyber Security Engineering, Shah & Anchor Kutchhi Engineering College, Mumbai, India

ABSTRACT—

This paper introduces an anomaly detection system empowered by machine learning to address the shortcomings of traditional signature-based methods in combating modern cyber threats like zero-day exploits and encrypted traffic. By leveraging machine learning, this system offers dynamic and proactive protection against evolving threats in global communications impacted by the Internet. This research explores the transformative role of machine learning in enhancing cybersecurity, emphasizing the need for adaptive defenses in response to continually evolving cyber threats.

Keywords— Anomaly detection, machine learning, cybersecurity, zero-day exploits, encrypted traffic, dynamic protection, proactive defense

I. Introduction

The digital age has seen unprecedented progress in communication, connecting billions of devices around the world. However, with these connections come many cybersecurity risks, from malware to advanced cyber-attacks such as zero- day exploits. Once effective signature-based detection methods are struggling to keep pace with the rapidly evolving threat landscape. Encrypted traffic hides malicious activity from traditional detection methods, making it harder to detect. In response to these findings, this research paper proposes a paradigm shift for machine learning- based anomaly detection for network security. By analyzing patterns that indicate suspicious behavior, machine learning algorithms offer a proactive defense mechanism against emerging threats. This paper examines the limitations of conventional detection methods, presents a proposed machine learning frame- work, and demonstrates its usefulness in improving system defenses against modern cyber threats. Through this research, we aim to highlight the importance of adopting an adaptive and intelligent approach to protect digital infrastructure as cyber security risks evolve.

II. PROBLEM DEFINITION

In the digital age, traditional signature-based detection methods are inadequate against evolving cyber threats like zero-day exploits and encrypted traffic, posing significant challenges for network security. To address this critical issue, there is a pressing need for a paradigm shift towards proactive defense mechanisms, particularly in the realm of network anomaly detection. This research aims to bridge this gap by proposing a comprehensive framework comprising feature engineering, model training, and anomaly detection components tailored specifically for network security. Through this innovative approach, we seek to bolster system defenses against modern cyber threats, emphasizing adaptability and intelligence in safeguarding digital infrastructure.

III. METHODOLOGY

The methodology focuses on comprehensive data acquisition from diverse network sources, ensuring a holistic view of network activity. This involves gathering data from switches, routers, and controllers, along with external threat intelligence feeds. Rigorous preprocessing and feature engineering techniques are then applied to clean and normalize data while selecting the most relevant network traffic features. Advanced machine learning algorithms, including deep learning models, random forests, and support vector machines, are leveraged for anomaly detection, emphasizing interpretability. Evaluation metrics and hyperparameter tuning optimize model performance.

Continuous Improvement:

Evaluation metrics like precision and recall gauge model performance, while hyperparameter tuning enhances effectiveness through iterative experimentation. Actionable insights are prioritized, enabling the system to provide detailed alerts to network administrators. Automated responses, such as quarantining infected devices or blackholing malicious IP addresses, are triggered for high-confidence anomalies. Integration with Security Orchestration and Automation (SOAR) platforms streamlines incident response, ensuring swift threat mitigation and network reconfiguration as necessary.

Hardware/Software Requirements:

Hardware Requirements:

- Standard desktop or laptop computers with memory and disk space per user: 512MB RAM, 1GB of disk, 0.5 CPU core.

Software Requirements:

- Compatible with Windows, macOS, or Linux, with Python installed.
- Python installed (version compatibility may vary).
- Compatibility with Google Colab for cloud-based collaboration.
- Microsoft Word application.

Additional Requirements:

- Reliable network infrastructure.
- Development environment with an IDE and version control system.

IV. Industrial Survey

Survey: Network anomaly detection has become a crucial task due to the exponential growth of network traffic, leading to an

increase in anomalies such as cyber-attacks, network failures, and hardware malfunctions. In this context, researchers have explored various techniques to detect and mitigate these anomalies, ensuring the security and stability of computer networks. This literature survey aims to explore two specific algorithms, K-Means and Normalized Cut, for network anomaly detection, as outlined in the assignment provided.

-K-Means Algorithm:

K-Means is a widely used clustering algorithm with the objective of minimizing the Sum of Squared Error (SSE) within each cluster. Despite its popularity, K-Means has certain limitations. It may fail in handling non-spherical shaped data and is not applicable to non-numerical data. Additionally, the choice of the number of clusters (K) is critical, and the algorithm may get stuck at local minima, necessitating random restarts. However, K-Means offers advantages such as speed and simplicity.

-Normalized Cut Algorithm:

Normalized Cut is a clustering algorithm that operates by performing eigen-decomposition on the Laplacian matrix derived from the similarity matrix of the data. The algorithm aims to partition the data into clusters while minimizing the normalized cut criterion. Normalized Cut offers advantages such as the ability to handle non- spherical data shapes and arbitrary shape clusters.

-DBSCAN Algorithm:

While not directly explored in the provided context, the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is briefly dis- cussed as another clustering algorithm suitable for anomaly detection. DBSCAN excels in identifying clusters of arbitrary shapes and is robust to noise. However, it requires careful tuning of hyperparameters and may suffer from computational inefficiency for large datasets. The implementation of Normalized Cut involves calculating the Laplacian matrix, per- forming eigen-decomposition, and using the eigen- vectors as input to a K-Means clustering algorithm. Evaluation measures similar to those used for K- Means are employed to assess the performance of Normalized Cut. Empirical evaluations suggest that Normalized Cut outperforms K-Means in terms of anomaly detection, with fewer anomalies detected.

V. PROBLEM IDENTIFIED

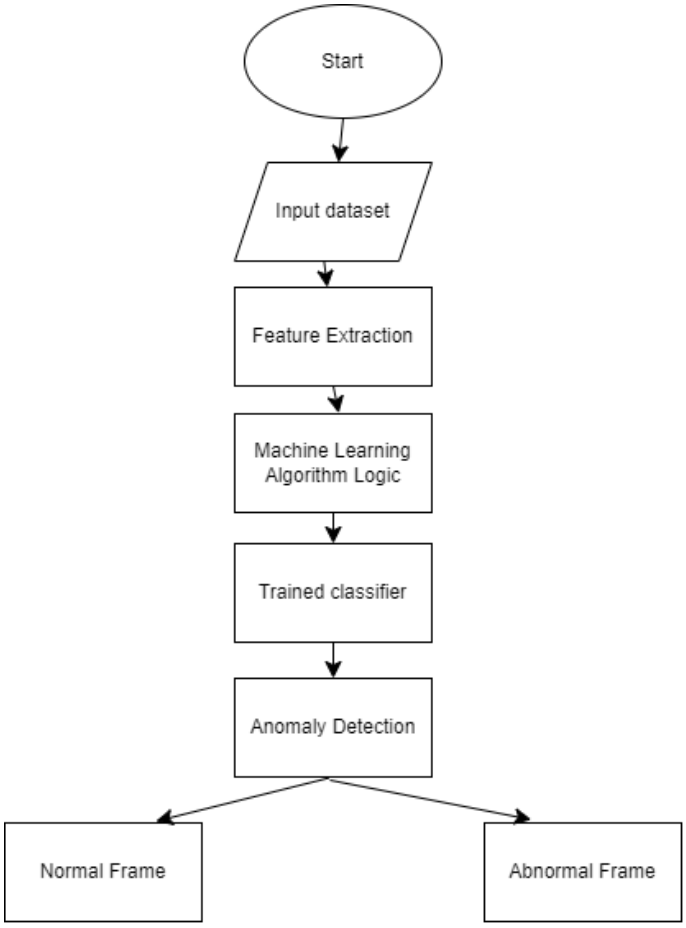
Developing a machine learning-based network anomaly detection system presents several challenges that developers must address to ensure the system's accuracy, reliability, and ethical deployment. Anomaly detection in network traffic is critical for identifying suspicious activities and potential threats while minimizing false positives and ensuring minimal impact on network performance.

One primary challenge lies in training and fine-tuning machine learning models to effectively distinguish between normal and anomalous network behavior. This requires robust feature selection, careful dataset curation, and continuous model optimization to adapt to evolving network patterns and emerging threats. Ensuring high detection rates while minimizing false alarms demands a sophisticated understanding of network protocols, traffic patterns, and attack vectors.

Moreover, scaling the anomaly detection system to handle large-scale networks with diverse traffic patterns poses significant technical hurdles. Developing algorithms that can efficiently process vast amounts of network data in real-time, while maintaining low latency and resource consumption, is essential for practical deployment in enterprise or service provider environments.

Ensuring the ethical deployment of an anomaly detection system is crucial. Developers must consider privacy implications, data protection measures, and the potential impact on user experience. Striking a balance between effective threat detection and respecting user privacy requires transparent communication, ethical data handling practices, and compliance with relevant regulations such as GDPR or CCPA.

VI. FLOWCHART



VII. CONCLUSION

In This study proposed a novel clustering-based approach for network anomaly detection by employing K-Means, Normalized Cut, and DBSCAN algorithms. These unsupervised techniques were applied to a dataset from Kaggle, enabling the identification of distinct traffic patterns and detection of anomalies. The results demonstrated the effectiveness of our approach, outperforming traditional methods. A key contribution was the adaptation of these clustering algorithms to the network anomaly detection domain.

Furthermore, we created graphs to visualize and analyze the identified clusters and anomalies in the dataset. While our approach showed promise, limitations include parameter sensitivity and computational complexity for large-scale data. Future work could explore ensemble methods, incremental learning, and scalable implementations.

The practical implications of our clustering-based anomaly detection lie in cybersecurity and network monitoring, enabling early detection of threats, intrusions, and performance issues. The unsupervised nature allows for identifying previously unseen attacks. In summary, this study presents a novel clustering-based approach for effective network anomaly detection, contributing to the advancement of this critical field and showcasing potential for real-world applications in cybersecurity.

VIII. REFERENCES

1. https://www.mdpi.com/1424-8220/23/11/5059
2. https://ieeexplore.ieee.org/document/9407
3. https://www.sciencedirect.com/science/article/pii/S1110016823006014

4. https://ieeexplore.ieee.org/document/8855360
5. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00669-1
6. https://ieeexplore.ieee.org/document/9634972
7. https://section.iaesonline.com/index.php/IJEEI/article/view/773/0