

Predicting Cyber Attacks Using Network Logs

¹ Nusrath Unnisa, ² Mrs. Supriya M

¹ Student, 4th Semester, Department of MCA, BIET, Davangere - 577004

nusrathunnisa849@gmail.com, Davanagere, Karanataka, India

² Assistant Professor, Department of MCA, BIET, Davangere-577004

supriyamallad@gmail.com, Davanagere, Karanataka, India

ABSTRACT: In modern times, where there is an abundance of connectivity in terms of digital technology, computers work through complex networks that increase their susceptibility to cyber-attacks. With time, businesses are becoming more dependent on the networking structure and, therefore, have become prone to such attacks, which are increasingly common and sophisticated, thus posing challenges to their control using traditional cybersecurity strategies. In most cases, traditional security frameworks are based on reaction methods since they only react to the incident after its occurrence. Apart from that, due to the time lag involved in detecting such threats, conventional security measures have proven to be ineffective in controlling the problem. This makes it necessary to implement more sophisticated means of preventing such attacks. Therefore, there is a growing demand for systems capable of predicting potential threats even before their occurrence.

The proposed project will employ the detection and identification of cyber attacks by conducting extensive analysis of network logs. Network logs are used for storing data concerning communication in the network. By conducting thorough analysis of these logs, any abnormality can be identified and classified as a potential threat to security.

In order to aid with the problem identification process, the authors of this paper make use of benchmark datasets, which include CICIDS2017 and NSL-KDD datasets. These are well-labeled data sets, which serve to provide valid network activities and different types of malicious network activities. Before applying any machine learning techniques on the datasets, extensive processing takes place in order to obtain important features of each case, which include source and destination IPs, network protocols, ports, sessions duration, and packet sizes.

There were multiple machine learning techniques applied for the purpose of classifying normal and malicious network activity; these include Random Forest, Support Vector Machine, and Logistic Regression classification algorithms. The performance of each algorithm was measured according to commonly accepted metrics, namely accuracy, precision, recall, and F1 measure.

From the results of the experiment, it becomes apparent that the proposed models were successful in the process of identifying malicious activities. This finding demonstrates great potential of data-driven intelligence-based solutions when it comes to network security.

Keywords: Cybersecurity, Cyber-attacks, Network security, Network logs, Anomaly detection, Machine learning, Random Forest, Support Vector Machine, Logistic Regression, CICIDS2017 dataset, NSL-KDD dataset, Feature extraction, Classification algorithms, Accuracy, Precision, Recall

I INTRODUCTION

Nowadays, there is hardly any place on earth where computer networks are not interconnected. In connection with this, cyberattacks have become increasingly frequent and complicated to control. Traditional security measures, which can only react once an attack has occurred, are already not enough. New ways are needed to identify threats before their occurrence and prevent possible damage.

In this research work, it is planned to predict potential cyber threats through the analysis of network traffic data.

Network traffic logs contain information about what occurs within a computer network. This way, one can find out about possible risk situations and, in this way, prevent damage.

In this research, network traffic logs for classification into normal behavior or malicious activities will be collected from reliable benchmarking datasets such as CICIDS2017 and NSL-KDD. Different types of machine learning classifiers will be used, including Random Forest, Support Vector Machine, and Logistic Regression for predicting the likelihood of certain network activity.

Classification results will be assessed by such metrics as precision, recall, and F1 score calculated based on the confusion matrix.

It all begins with the gathering of raw logs of events that originate from many different sources. This information contained in these logs is referred to as logs of information, and it includes such information as IP addresses of the source and destination, port numbers, time, type of protocols used, and many others.

It has the ability to work independently; the system is able to analyze network activities in real time for lessons learned incidents that would point out abnormal network traffic, repeated login attempts, abnormal access times, or packet behavior. Predictive maliciousness can be accomplished by detecting patterns, which may involve different types of attacks such as DDoS, repeated login attempts, port scanning, and malicious code distribution, among others.

No. Rather, the goal here is not detection of an attack after it happens, but more importantly, prediction of the attack before it happens.

II RELATED WORK

Intrusion detection and cybersecurity analytics have been extensively studied using data mining, machine learning, and artificial intelligence techniques. Early foundational work by Charu C. Aggarwal in *Data Mining: The Textbook* highlights core techniques such as clustering, classification, and anomaly detection, which form the backbone of modern intrusion detection systems (IDS). Similarly, Stallings W. discusses essential principles of network security, including authentication, encryption, and threat modeling, providing a framework for understanding cyber threats.

Machine learning libraries such as Scikit-learn have enabled practical implementation of IDS models. These tools support supervised and unsupervised learning algorithms widely used in anomaly detection tasks. Research by Kim J. and Lee H. explores advanced detection techniques, emphasizing hybrid approaches that combine signature-based and anomaly-based methods to improve detection accuracy.

Anomaly-based intrusion detection has gained attention due to its ability to detect unknown attacks. Kumar S. and Singh R. propose using network log analysis combined with machine learning models to identify deviations from normal behavior. Their work demonstrates improved detection rates compared to traditional rule-based systems. Datasets play a critical

role in evaluating IDS performance. The KDD Cup 99 Dataset, analyzed by Mahbod Tavallae et al., remains one of the most widely used benchmarks, despite criticisms regarding redundancy and outdated attack patterns. It has nevertheless influenced the development and evaluation of numerous IDS models.

The role of big data in cybersecurity is emphasized by Nir Kshetri, who discusses how large-scale data analytics can enhance threat detection and response capabilities. Additionally, modern AI approaches outlined by Stuart Russell and Peter Norvig provide theoretical grounding for intelligent systems capable of adaptive learning in dynamic threat environments.

Industry insights, such as the Cisco Systems cybersecurity reports, highlight evolving threat landscapes, including ransomware, phishing, and advanced persistent threats (APTs). These reports reinforce the need for adaptive and intelligent IDS solutions.

Finally, implementation of such systems is often supported by programming environments like Python Software Foundation, whose Python language offers extensive libraries and frameworks for data analysis, machine learning, and cybersecurity applications.

III METHODOLOGY

The suggested approach highlights the way the system for prediction of the cyber-attack is created and operates. The initial step in this case is collecting data. Collection of logs is done through devices such as firewall systems, intruders, detectors, routers, application servers. These logs are used to log user activities, network activities, logon and activities, and possible threats in the network. Log data is inaccurate and contains many redundancies, among other unnecessary information.

With the cleaning of the logs complete, the next step involves feature extraction. This step involves identification and selection of significant features from the log data. These features include IP activity, number of requests, logon failure attempts, odd port usage, and _proto activity.

Since the logs were not visible features at the beginning, this procedure will help the model understand.

The next process would be model training. Model training includes learning algorithms such as decision tree random forest SVM, and neural networks by training the model on previously obtained data containing normal activity and attacks.

Once the model works effectively, it can be deployed for live monitoring. For this purpose, logs in the network are constantly being monitored. These logs will again undergo the above-mentioned processes before being fed to the model which then analyzes them and gives its verdict whether the activity is normal or suspicious. An alert is generated once any suspicious activity is identified.

Last but definitely not the least would be a feedback mechanism where after going through experts' analysis of any alert raised by the system regarding attacks whether the attack was genuine or not, feedback will be provided.

Conclusion

Thus, the suggested methodology adopts an intelligent and iterative process for translating raw network logs into meaningful predictions of cyber attacks. The method is adaptive, scalable, and can effectively detect threats in a constantly changing cyber security landscape.

IV RESULT AND DISCUSSION

Snapshots:

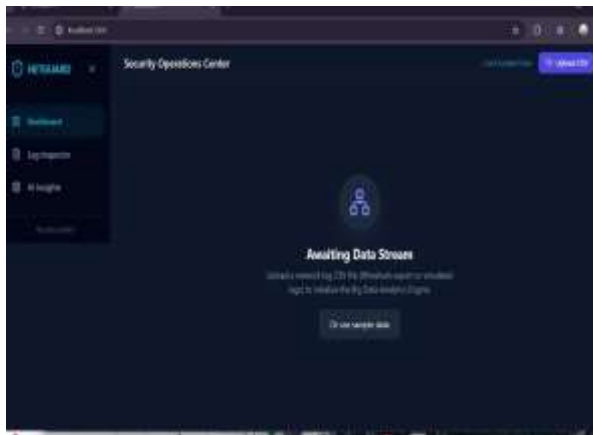


Fig: Dashboard Panel

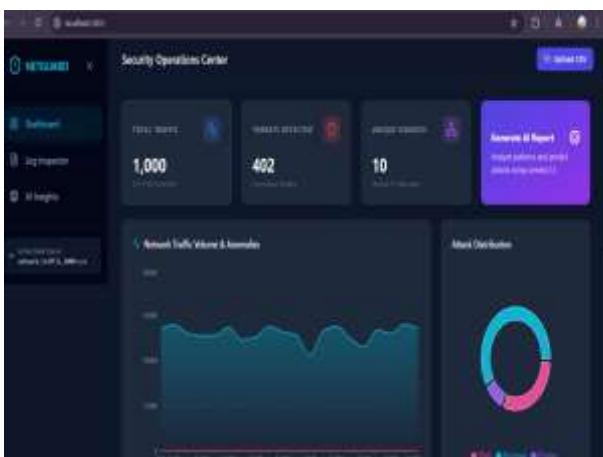


Fig: Log Inspector



Fig: Security Operations Center

The "Predicting Cyber Attacks Using Network Logs" project was successfully completed and implemented, where suspicious activities could be predicted using data obtained from network logs.

The main functionalities realized in the system are the following:

The system is capable of importing network logs in CSV format and processing the information contained.

Identified whether the traffic pattern observed was normal or suspicious according to some defined rules and anomaly detection algorithms.

The dashboard shows:

Network traffic total

Number of threats found

Sources of the data (IP address)

The Log Inspector tool enables further analysis of specific logs.

The Security Operation Center (SOC) section emphasizes suspicious IP addresses along with their threat level such as:

Low

Medium

High (critical threat)

Snapshots provided are:

Dashboard Section:

Shows the system ready for receiving data and initialization of the analysis engine.

Log Inspector section:

Displays analyzed data along with:

Trend of traffic

Number of threats detected

Attack distribution

SOC Section:

Provides list of IP addresses with associated threat probability and level.

The proposed solution illustrates how analyzing the network logs can assist in predicting cyber attacks.

The solution proves its capability in identifying anomalous traffic trends that may indicate cyber attacks, including:

Hacking

IP behavior
Traffic bursts
Using visualization tools (charts and graphs) enhances:
Understanding of network behavior
Making decisions by security analysts
Classifying attacks in order of importance facilitates:
Prioritizing the attacks
Reducing the response time in case of critical attacks
Scalability is also an advantage, as the solution can process many logs at once.

V CONCLUSION

The project titled “Predicting Cyber Attacks Using Network Logs” demonstrates the role of machine learning methods in enabling early detection of security threats across networked systems. By systematically collecting and analyzing large volumes of network log data the system is able to perceive uncharacteristic shapes that may specify malicious activity. Key interactive landscapes are taken out from the logs and used to classify events as either legitimate network behavior or possible cyberattacks. This proactive detection approach enables organizations to recognize threats at an early stage reducing the risk of serious security breaches. The implementation results show that automated prediction models can significantly minimize the need for manual log monitoring while improving detection accuracy. It is also capable of detecting typical attack activities, like

port scanning, repeated login attempts, and sudden surges in network traffic. This will be the key to unlocking intelligent systems capable of identifying and responding swiftly to cyber threats as they emerge.

VI REFERENCES

1. Aggarwal C. C. (2015). Data Mining: The Textbook. Springer.
2. Stallings W. (2018). Network Security Essentials: Applications and Standards. Pearson.
3. Scikit-Learn Developers. (2023). Scikit-learn Machine Learning Library Documentation.
4. Kim J. & Lee H. (2018). Techniques.” International Journal of Security and Networks.
5. Kumar S. & Singh R. (2020). “Anomaly-Based Intrusion Detection Using Network Logs.” IEEE Conference on Cyber security.
6. Kshetri N. (2017). The Emerging Role of Big Data in Cyber security. Elsevier.
7. Russell S. & Norvig P. (2021). Artificial Intelligence: A Modern Approach. Pearson.
8. Tavallaei M. et al. (2009). “A Detailed KDD Cup 99 Dataset.” Computational Intelligence for Security.
9. Cisco Systems. (2023). Cyber security Threat Trends Report.
10. Python Software Foundation. (2023). Python Programming Language Documentation.