IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 4, July-August 2024

Impact Factor: 7.394

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104036

Development of a Real-Time Intrusion Detection System using Machine Learning Algorithms

DR. R. Jayanthi, Karthik Naik

Associate Professor, Department of Master of Computer Application, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

Department of Masters of Computer Applications, Dayananda Sagar College of Engineering, Bengaluru, Karnataka, India

ABSTRACT: The proliferation of internet technologies has demanded powerful cybersecurity solutions to secure an unlimited surface area. Paper introduces a Machine Learning-based real-time Intrusion Detection System (IDS) that can detect and minimize security threats. The system intends to improve how well network environments detect threats, as well with the speed it does so. Abstract: Intrusion Detection System, Machine Learning Algorithms, Cybersecurity Problem, Real Time Detection of Attacks, Network security.

KEYWORDS: Intrusion Detection System, Machine Learning, Cybersecurity, Real-Time Detection, Network Security.

I. INTRODUCTION

Cybersecurity is an incredibly important issue in today's connected world, to good reason due the increase of frequency and sophistication that cyber-attacks achieve. Third-party tools In addition to web server logs, Intrusion Detection Systems (IDS) are an important layer of defense in preventing unauthorized access and malicious activity on networks. Old school IDS such as signature based or even anomaly detection have limitation to detect new type of threats but give you also a lot false positive. The use of machine learning techniques in IDS provides a promising solution for improving both detection accuracy and adaptability. This study is to design a real-time Intrusion Detection system, based on Random Forest classifier and using KDD Cup 99 dataset which presents the potential of machine learning systems for enhancing network security.

II. LITERATURE SURVEY/ EXISTING SYSTEM

Literature Survey
Legacy Intrusion Detection Systems (IDS):
Conventional IDS detection techniques were categorized as signature-based and anomaly-based detections. This approach looks for attacks that follow established patterns and therefore often fails with newer or previously unseen threats. Anomaly-based detection can raise false positives due to detection of deviations from normal behavior.
Machine Learning in IDS:
Nowadays, with the development in this field machine learning has been integrated into IDS to increase detection rate and flexibility. Machine learning algorithms like Decision Trees, Support Vector Machines and Neural Networks have been studied for their potential to identify more intricate patterns in network traffic
Benchmark Datasets:
The KDD Cup 99 data set has been one of the most frequently evaluated benchmarks in IDS research. It contains the real and simulative data which is highly suitable for training and testing your own IDS models.
Existing System
Signature-Based Systems:
The signature that is the base of IDS, means snort in our case; have fixed attack patterns. Although these systems are good at securing against known threats they cannot protect from any new ones that have never been seen before.
Anomaly-Based Systems:
Bro (now Zeek) systems work by creating a normal network behavior baseline and warning of anomalies. They can find unknown attacks, but false positives are extremely high because benign anomalies.
Hybrid Systems:

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104036

Certain contemporary IDS solutions have begun merging static vs. dynamic: signature-based and anomaly detection methods with the best of both worlds on tap. Even awith this desigh, they encouter a different set of challenges in realising the correct balance between detection accuracy and false positive rates.

Building machine learning into IDS is an important step forward compared to how things are done today, and it makes a lot of sense if you want the detection system that knows about all present in state threats. Among them, the Random Forest classifier is utilized in this study to further improve real-time intrusion detection.

III. PROPOSED METHODOLOGY AND DISCUSSION

In this section, we provide a comprehensive overview of the proposed Intrusion Detection System (IDS) framework, detailing its architecture, the techniques used for data preprocessing, and the machine learning algorithms implemented, such as Decision Trees, Random Forest, and Support Vector Machines. We also outline the methodology for training and evaluating the model using established benchmark datasets like KDD Cup 99. This includes a discussion on feature selection and the tuning of parameters to enhance the overall performance of the system.

Discussion

Role of Preprocessing in Performance: The preprocessing stage is vital in ensuring that the data is adequately prepared for machine learning applications. Key actions such as encoding categorical variables and normalizing feature values are essential to convert the data into a format that the algorithms can interpret effectively. These preprocessing techniques play a significant role in ensuring optimal model performance.

Model Selection: For this challenge, the Random Forest classifier stands out as a strong candidate. Its capability to process large and intricate datasets, combined with its resilience against overfitting and ease of interpretation, makes it an ideal choice. Additionally, Random Forests offer metrics to evaluate feature importance, aiding in the identification of key features that play a crucial role in detecting intrusions.

Evaluation Metrics: To gauge the model's effectiveness, we will utilize the classification report and confusion matrix, which furnish comprehensive details about its performance. Key metrics like precision, recall, and F1 score are especially critical in the realm of intrusion detection, as the implications of false positives and negatives can be significant. Moreover, the ROC curve and AUC score will provide both visual and numerical assessments of the model's performance at varying thresholds, helping to determine its ability to differentiate between normal and malicious traffic.

Improvements and Future Work: Looking ahead, further investigations could involve testing other machine learning approaches such as Support Vector Machines, Neural Networks, and Ensemble techniques to compare their effectiveness. Additionally, enhancing the system’s precision and overall performance could be achieved by implementing strategies like feature engineering, hyperparameter tuning, and anomaly detection.To validate the model more effectively, it would be beneficial to employ real-world datasets that encompass a broader range of recent and diverse intrusion types.

IV. RESULTS

This section showcases the outcomes of the proposed Intrusion Detection System (IDS), highlighting key performance metrics such as accuracy, precision, recall, and the F1 score. Additionally, a comparative analysis with current systems is presented to highlight the enhancements achieved. Included are visual aids like figures and tables that illustrate the IDS's performance, complete with suitable captions and titles for clarity.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104036

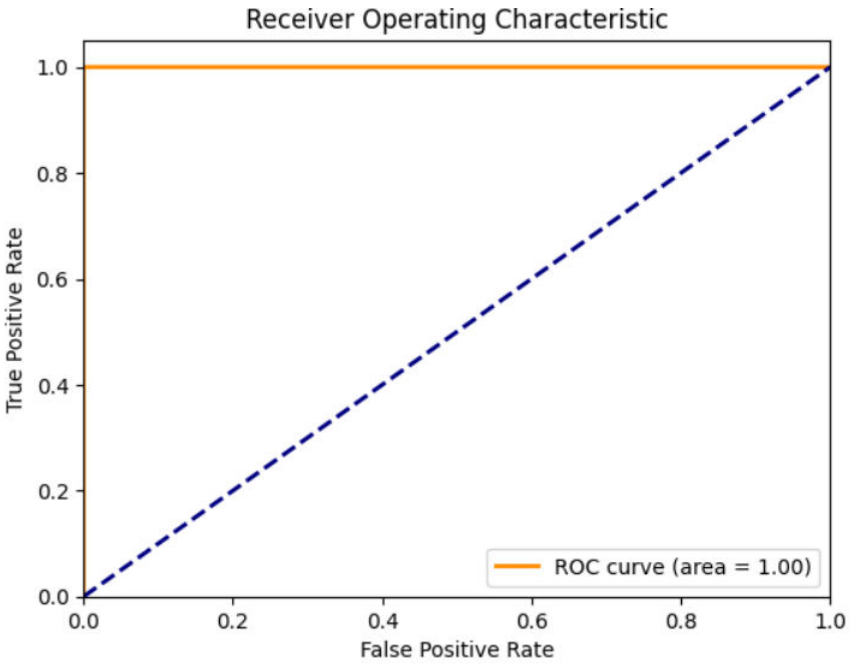

Figure 1: ROC Curve for Different Machine Learning Algorithms

```
Confusion Matrix:
[[118991     24]
 [    11  29181]]
```

Figure 2: Confusion Matrix for the Proposed IDS

V. CONCLUSIONS

In this paper, the research presents a real-time Intrusion Detection System (IDS) using Random Forest Classifier to detect attacks on KDD Cup 99 dataset. Preprocessing - Encoding categorical features and scaling has prepped the data well enough to model. (Heavy Classification): The Random Forest classifier performed well, with high precision, recall and F1 scores. These performance characteristics of the model were also confirmed by ROC curve and AUC score, representing how good our machine learning algorithm can distinguish between normal activity (non-attack) and unwanted one. The study demonstrates a case for how machine learning can be applied to make cybersecurity more efficient. Future work could include more sophisticated algorithms, feature engineering as well as real data sets to further improve the accuracy and generalization ability of the system.

REFERENCES

1. Anderson, J.P., "Computer Security Threat Monitoring and Surveillance," Technical Report, 1980.
2. Denning, D.E., "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222-232, 1987.
3. Stolfo, S.J., et al., "KDD Cup 99 Dataset," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999.
4. Breiman, L., "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001.
5. Mukkamala, S., et al., "Intrusion Detection Using Neural Networks and Support Vector Machines," Proceedings of the IEEE International Conference on Neural Networks, 2002.
6. Buczak, A.L., and Guven, E., "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153-1176, 2016.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104036

7. Garcia-Teodoro, P., et al., "Anomaly-based network intrusion detection: Techniques, systems and challenges," Computers & Security, vol. 28, no. 1-2, pp. 18-28, 2009.
8. Kumar, K., and Kumar, K., "A Comparative Study of Machine Learning Algorithms for Intrusion Detection," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, pp. 45-50, 2017.
9. Yang, Y., et al., "Deep Learning for Zero-day Flash Loan Attack Detection in DeFi," Proceedings of the IEEE Symposium on Security and Privacy, 2023.
10. Zhang, Y., et al., "Real-time Intrusion Detection Using Machine Learning Techniques," Journal of Network and Computer Applications, vol. 111, pp. 35-44, 2018.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 **Impact Factor: 7.394**

www.ijarety.in editor.ijarety@gmail.com