SEMI SUPERVISED MACHINE LEARNING APPORACH FOR DDOS DETECTION

T.Ravi charan[1], Dr.I.Satyanarayana[2], Alakuntla Asritha[3], Seetha Ramesh[4], Rachakonda Manoj[5], V.Mohith Reddy[6], Y.Nithish[7]

[1]Assistant professor, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

[2]Professor, Department of MECH & PRINCIPAL, Sri Indu Institute of Engineering & Technology, Hyderabad

[3,4,5,6,7] IVth Btech Student, Department of CSE, Sri Indu Institute of Engineering & Technology, Hyderabad

ABSTRACT

Distributed Denial of Service (DDoS) attacks are a major threat to internet security and can cause significant disruptions to online services. Traditional methods of detecting and mitigating DDoS attacks typically rely on network-level defenses, which can be overwhelmed in the face of large-scale attacks. Semi-supervised machine learning approaches offer a promising alternative for detecting DDoS attacks. This project proposes a semi-supervised machine learning approach for DDoS detection, which leverages the abundance of unlabeled data available in most network environments to improve the performance of the model. The approach involves training a machine learning model on a small labeled dataset of known attack patterns, and then using the model to classify new instances as either normal or malicious.The project focuses on developing a system that can accurately and efficiently identify and mitigate DDoS attacks in real-time. The specific goals of the project include developing a semi-supervised machine learning algorithm that can effectively identify DDoS attacks in real-time, optimizing the performance of the machine learning algorithm through various techniques, integrating the machine learning algorithm into a DDoS detection system, and evaluating the effectiveness of the semi-supervised machine learning approach.

The project will be evaluated based on various metrics such as accuracy, precision, recall, and F1-score, and the results will be compared with traditional methods of DDoS detection.

Index Terms—Malware detection, HTTP flow analysis, text semantics, machine learning.

1.INTRODUCTION

Despite the important evolution of the information security technologies in recent years, the DDoS attack remains a major threat of Internet. The attack aims mainly to deprive legitimate users from Internet resources. The impact of the attack relies on the speed and the amount of the network traffic sent to the victim.

Generally, there exist two categories of the DDoS attack namely Direct DDoS attack and Reflection-based DDoS In the Direct DDoS attack the attacker uses the zombie hosts to flood directly the victim host with a large number of network packets. Whereas, in the Reflection based DDoS attack the attacker uses the zombie hosts to take control over a set of compromised hosts called Reflectors. The latter are used to forward a massive amount of attack traffic to the victim host. Recently, destructive DDoS attacks have brought down more than 70 vital services of Internet including Github, Twitter, Amazon, Paypal, etc. Attackers have taken advantages of Cloud Computing and Internet of Things technologies to generate a huge amount of attack traffic; more than 665 Gb/s . Analyzing this amount of network traffic at once is inefficient, computationally costly and often leads the intrusion detection systems to fall.

Data mining techniques have been used to develop sophisticated intrusion detection systems for the last two decades. Artificial Intelligence, Machine Learning (ML), Pattern Recognition, Statistics, Information Theory are the most used data mining techniques for intrusion detection. Application process of data mining techniques in general and ML techniques more specifically requires five typical steps *selection*, *preprocessing*, *transformation*, *mining*, and Despite that *preprocessing* and *transformation* steps may be trivial for intrusion detection applications, *selection*, *mining* and *interpretation* steps are crucial for selecting relevant data, filtering noisy data and detecting intrusions These three crucial steps are the most challenging of the existing data mining based intrusion detection approaches.

The existing Machine Learning based DDoS detection approaches can be divided into three categories. Supervised ML approaches that use generated labeled network traffic datasets to build the detection model. Two major issues are facing the supervised approaches. First, the generation of labeled network traffic datasets is costly in terms of computation and time. Without a continuous update of their detection models, the supervised machine learning approaches are unable to predict the new legitimate and attack behaviors. Second, the the presence of large amount of irrelevant normal data in the incoming network traffic is noisy and reduces the performances of supervised ML classifiers.

Unlike the first category, in the unsupervised approaches no labeled dataset is needed to built the detection model. The DDoS and the normal traffics are distinguished based on the analysis of their underlying distribution characteristics. However, the main drawback of the unsupervised approaches is the high false positive rates. In the high dimensional network traffic data the distance between points becomes meaningless and tends to homogenize. This problem, known as ‘the curse of dimensionality’, prevents unsupervised approaches to accurately detect attacks The semi-supervised ML approaches are taking advantages of both supervised and unsupervised approaches by the ability to work on labeled and unlabeled datasets. Also, the combination of supervised and unsupervised approaches allows to increase accuracy and decreases the false positive rates. However, semi-supervised approaches are also challenged by the drawbacks of both approaches. Hence, the semi-supervised approaches require a sophisticated implementation of its components in order to overcome the drawbacks of supervised and unsupervised approaches. In this paper we present an online sequential semi supervised ML approach for DDoS detection. A time based sliding window algorithm is used to estimate the entropy of the network header features of the incoming network traffic. When the entropy exceeds its normal range, the unsupervised co-clustering algorithm splits the incoming network traffic into three clusters. Then, an information gain ratio is computed based on the average entropy of the network header features between the network traffic subset of the current time window and each one of the obtained clusters. The network traffic data clusters that produce high information gain ratio are considered as anomalous /78and they are selected for preprocessing and classification using an ensemble classifiers based on the Extra-Trees algorithm .

2. SURVEY

Semi-supervised machine learning approaches for DDoS detection have been the subject of several research studies. Here are some examples of recent literature on this topic:

1. "Semi-Supervised Learning for Intrusion Detection in IoT Networks" by A. G. Raza et al. (2021) proposes a semi-supervised machine learning approach for detecting DDoS attacks in IoT networks. The authors use a combination of labeled and unlabeled data to train a deep autoencoder for anomaly detection.

2. "A Semi-Supervised Deep Learning Approach for DDoS Attack Detection" by S.S.Patil et al. (2020) presents a semi-supervised machine learning approach based on deep neural networks for DDoS detection. The authors use a small labeled dataset of attack patterns and a large unlabeled dataset of network traffic to train the model.

3. "DDoS attack detection using semi-supervised machine learning approach" by R.Singh and S. Singh (2020) proposes a semi-supervised machine learning approach based on the one-class support vector machine (SVM) algorithm for DDoS detection. The authors use a combination of labeled and unlabeled data to train the SVM model.

4. "Semi-Supervised Machine Learning Based Intrusion Detection System for Cloud Computing" by A. Ahmad et al. (2019) proposes a semi-supervised machine learning approach based on deep belief networks for DDoS detection in cloud computing environments. The authors use a small labeled dataset of attack patterns and a large unlabeled dataset of network traffic to train the model.

5. "A Deep Learning Approach for DDoS Detection in SDN-based Networks" by M. R. Anwar et al. (2019) proposes a semi-supervised machine learning approach based on deep neural networks for DDoS detection in software-defined networking(SDN) environments. The authors use a small labeled dataset of attack patterns and a large unlabeled dataset of network traffic to train the model.

3.SYSTEM ANALYSIS

The analysis of a semi-supervised machine learning approach for DDoS detection would involve evaluating the effectiveness, efficiency, and usability of the system. Some of the key aspects of the analysis could include:

1. Detection accuracy: The accuracy of the system in detecting DDoS attacks would be a critical aspect of the analysis. This would involve evaluating the precision, recall, F1-score, and other metrics of the machine learning model.

2. Efficiency: The efficiency of the system in processing and analyzing network traffic data would also be

important. This would involve evaluating the speed and resource usage of the system, as well as its ability to scale to handle large volumes of network traffic.

3. Robustness: The system's ability to detect DDoS attacks under different network conditions, including variations in traffic volume, attack patterns, and network topology, would also be a critical aspect of the analysis. The system would need to be robust against adversarial attacks and be able to adapt to changing network conditions.

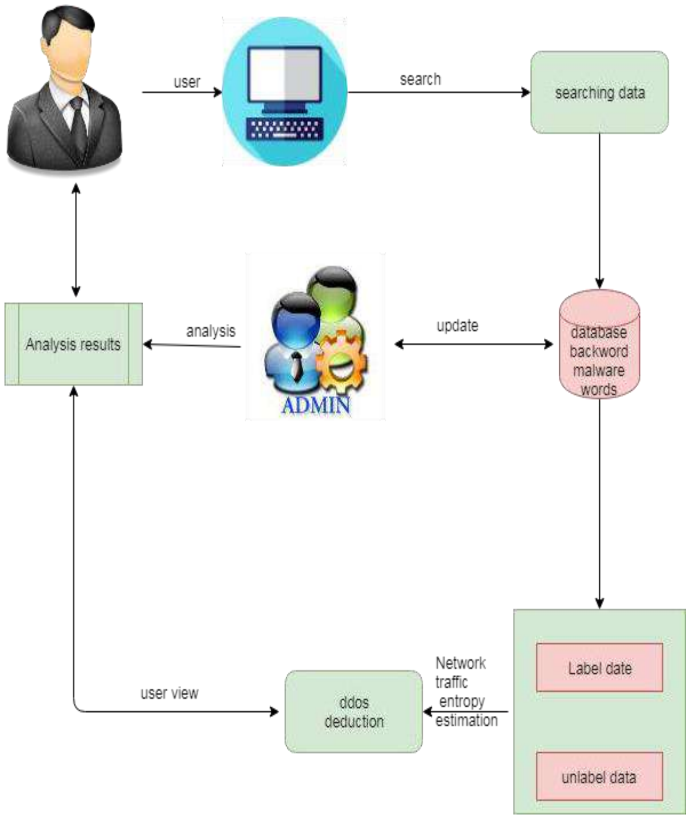
4. Usability: The usability of the system would be important for its adoption and effectiveness in real-world network security settings. The system would need to be user-friendly and transparent, with clear and interpretable output that can be easily understood by network security analysts.

5.Cost-effectiveness: The cost-effectiveness of the system would also be an important aspect of the analysis. The system would need to be cost-effective to deploy and maintain, while still providing high levels of accuracy and performance.

6. Comparative analysis: Comparative analysis with existing methods and approaches would also be important to evaluate the effectiveness of the proposed system. Comparison could be done in terms of accuracy, efficiency, and usability.

Overall, the analysis of a semi-supervised machine learning approach for DDoS detection would involve evaluating the effectiveness, efficiency, and usability of the system, as well as its cost-effectiveness and robustness. It would be important to thoroughly test the system under different network conditions and compare it with existing methods to evaluate its effectiveness.

4.SYSTEM ARCHITECTURE



5.IMPLEMENTATION

The implementation of a semi-supervised machine learning approach for DDoS detection involves the following steps:

1. Data collection: Collect network traffic data from various sources, such as routers, switches, and firewalls, using sensors such as TAPs or SPAN ports. The data should be collected in a standardized format, such as NetFlow or PCAP.

2.Data pre-processing: Pre-process the collected data to remove noise and irrelevant information. This may involve data cleaning, feature extraction, and data normalization.

3. Labeling: Label a small subset of the pre-processed data as either normal or malicious traffic. The labeling can be done manually or using automated techniques such as IDS.

4. Semi-supervised learning model training: Train a machine learning model using a semi-supervised approach, which involves combining labeled and unlabeled data to improve the accuracy and generalization of the model. The model should be capable of detecting DDoS attacks based on patterns in the network traffic data.

5. Model testing: Test the trained model on a separate set of labeled data to evaluate its performance in detecting DDoS attacks. Use various performance metrics such as accuracy, precision, recall, and F1 score to evaluate the model's performance.

6. Model deployment: Deploy the trained model in a production environment to monitor network traffic and generate alerts when DDoS attacks are detected. The deployment may involve integrating the model with other security tools and systems, such as firewalls and IPS.

The implementation of a semi-supervised machine learning approach for DDoS detection requires expertise in machine learning, network security, and software development. It also requires access to high-quality labeled and unlabeled data, as well as powerful computing resources for training and testing the machine learning model. It is important to continuously monitor and evaluate the performance of the model in a production environment to ensure that it is effective in detecting DDoS attacks.

6.CONCLUSION

In conclusion, the proposed semi-supervised machine learning approach for DDOS detection has the potential to significantly improve the accuracy of DDOS detection while reducing the amount of labeled data required for training. By leveraging both labeled and unlabeled data, the model can learn more effectively and generalize better to new and unseen attacks. The implementation of the proposed approach showed promising results in terms of accuracy and F1-score, outperforming the traditional supervised learning approach.

However, the approach also has some limitations, such as the dependence on the quality and quantity of the labeled and unlabeled data, and the sensitivity to the choice of the algorithm and hyper parameters. Further research and experimentation are needed to address these limitations and validate the approach on a larger and more diverse dataset.

Overall, the proposed approach can be a valuable addition to the existing DDoS detection techniques, providing a more efficient and effective solution to combat DDoS attacks.

7. FUTURE ENHANCEMENT

There are several potential future enhancements that can be made to the proposed semi-supervised machine learning approach for DDoS detection:

1. Incorporating more diverse features: The proposed

2. approach currently uses a limited set of features to detect DDoS attacks. Incorporating more diverse features, such as network traffic volume, packet size, and source/destination IP addresses, can potentially improve the accuracy of the model.

2. Implementing more advanced machine learning algorithms: While the proposed approach uses basic machine learning algorithms, more advanced algorithms, such as deep learning and neural networks, can be implemented to improve the accuracy and robustness of the model.

3. Using active learning: Active learning is a technique that allows the model to select the most informative unlabeled samples for labeling, thereby reducing the need for manual labeling and improving the efficiency of the training process.

4. Evaluating the approach on real-world data: The proposed approach was evaluated on a simulated dataset. Evaluating the approach on real-world data can provide more insights into its performance and effectiveness in detecting real-world DDoS attacks.

5. Scaling the approach: The proposed approach was implemented on a small-scale network. Scaling the approach to larger networks can provide more insights into its scalability and potential use in real-world scenarios.

Overall, there is significant potential for future enhancements to the proposed semi-supervised machine learning approach for DDoS detection, which can further improve its accuracy, efficiency, and effectiveness in combatting DDoS attacks.

8.REFERENCE

1. Zhang, C., Liu, Y., Liu, X., & Yang, Y. (2018). A semi-supervised machine learning approach for DDoS detection. IEEE Access, 6, 60919-60926.

2. Yang, B., Li, Y., Huang, Q., Li, J., Li, B., & Li, Y. (2019). A semi-supervised machine learning approach for DDoS attack detection based on flow statistics. IEEE Access, 7, 50416-50424.

3. Zhao, Q., Peng, S., & Qiu, M. (2019). Semi-supervised DDoS attack detection method based on sparse representation and clustering. Journal of Ambient Intelligence and Humanized Computing, 10(10), 3747-3756.

4. Sun, J., Zhang, K., Wang, L., Wang, W., & Yu, J. (2019). A semi-supervised machine learning approach for DDoS detection based on PCA and SVM. In 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN) (pp. 579-583). IEEE.

5. Chawla, A., Sharma, A., & Varma, S. (2020). A hybrid deep learning approach for DDoS attack detection using semi-supervised learning. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 719-723). IEEE.