Semi – Supervised Machine Learning Approach For DDOS Detection

Mrs. B. Anusha, Alla Prakash Sai[2], Jashwanth Bommakanti[3], Repalle Nandini [4],

Soumith Paruvelli[5]

[1]Associate Professor, [2,3,4,5]Students

[1, 2,3,4,5]Department of Artificial Intelligence and Machine Learning

Malla Reddy Institute of Technology and Science, Hyderabad, India.

Email Id: anusanju.bomma@gmail.com, prakashsaialla34@gmail.com,jashwanthbommakanti123@gmail.com,nandinirepalles@gmail.com, soumithparuvella1@gmail.com

I. ABSTRACT

Distributed denial of service (DDoS) attacks pose a serious threat to the availability and integrity of online services.Traditional DDoS detection methods often rely on supervised learning techniques, which require labelled datasets for training.However, collecting labelled data for DDoS attacks is challenging due to their dynamic and ever-evolving nature.

In this study, we propose a semi-supervised machine learning method for DDoS detection, to address the limitations of traditional monitoring methods.Our approach leverages both labelled and unlabelled data to improve the accuracy and scalability of DDoS detection.We introduce a set of machine learning models, including deep neural networks and anomaly detection techniques, to address a wide range of DDoS attack vectors and characteristics.

By combining labelled samples, the model can first learn to identify known attack patterns.At the same time, the model adapts to new and undetected attack variants by leveraging unlabelled data, improving detection of zero-day DDoS attacks.To evaluate the effectiveness of our semi-supervised approach, we conducted extensive experiments on real network traffic datasets.

The results demonstrate that our method outperforms traditional supervised methods in terms of detection accuracy, false positive rate, and adaptability to new attacks.Additionally, this method demonstrates robust resistance to interference and network traffic variation, ensuring reliable DDoS detection in real network environments.

II.INTRODUCTION

The proliferation of networked services and the increasing dependence on the internet have given rise to an escalating threat landscape, with Distributed Denial of Service (DDoS) attacks emerging as a prominent and persistent menace.DDoS attacks disrupt online services by inundating their infrastructure with an overwhelming volume of malicious traffic, leading to service unavailability, reduced performance, and potential financial losses.

Traditional DDoS detection approaches often rely on supervised machine learning methods, which necessitate labeled datasets for training.However, the dynamic and ever-evolving nature of DDoS attacks makes it exceedingly challenging to obtain comprehensive labeled data. This research endeavors to address the limitations of traditional supervised DDoS detection by introducing a semi-supervised machine learning approach.In a semi-supervised paradigm, the model is trained using a combination of labeled data, which represents known attack patterns, and unlabeled data, which comprises a broader spectrum of network traffic.This amalgamation of labeled and unlabeled data enables the model to enhance its accuracy in identifying both known and novel DDoS attacks.By leveraging this hybrid dataset, the semi-supervised approach offers several advantages, including adaptability to emerging threats, reduced false positives, and the ability to detect zero-day attacks.

In this paper, we present a comprehensive examination of our semi-supervised machine learning approach for DDoS detection.We will delve into the methodology, highlighting the ensemble of machine learning models employed, including deep neural networks and anomaly detection techniques, and elucidate how they work

in synergy to bolster detection capabilities.

Furthermore, we will elucidate how the approach leverages labeled data for identifying familiar DDoS attack patterns while simultaneously harnessing unlabeled data to adapt and evolve, ensuring effective detection even in the face of previously unseen attack variations.The efficacy of our semi-supervised approach will be substantiated through empirical evaluation using real-world network traffic datasets.We will demonstrate how this approach outperforms conventional supervised methods in terms of detection accuracy, precision, recall, and adaptability, ultimately providing a more robust and reliable DDoS detection solution.

Additionally, we will examine the model's resistance to noise and variations in network traffic, which are prevalent in practical network environments.In sum, this research endeavors to present a compelling solution to the ever-evolving threat of DDoS attacks by harnessing the advantages of semi-supervised machine learning.Our approach seeks to increase the resilience of online services against these malicious attacks, thereby contributing to a more secure and stable digital ecosystem.A model was trained for DoS, U2R and Probe for the BPN neural network and the classes of normal attack by Mukhopadhyay and etl al [17], in this case the rate of success for the system was as much as 73.9 percent for the latest testing sets while for level 1 test set equal to 95.6 percent was the obtained rate.

III. LITERATURE SURVEY AND COMPARATIVE ANALYSIS

Literature Survey:

1. **"A Survey of Network Anomaly Detection Techniques"**
    - *Authors: H. S. De Zoysa, A. Wijesinghe, and A. Seneviratne*
    - This survey provides an overview of various anomaly detection techniques, including semi-supervised methods, used in the context of network security, with a focus on DDoS detection.
2. **"Machine Learning Approaches for DDoS Attack Detection: A Survey"**
    - *Authors: M. I. Hossain, M. A. Hossain, and G. Muhammad*
    - This paper offers an extensive survey of machine learning approaches used for DDoS attack detection, including semi-supervised techniques. It provides insights into the challenges and trends in the field.
3. **"Semi-Supervised Learning in Network Intrusion Detection Systems: An Overview and New Directions"**
    - *Authors: M. Ring, M. Wunder, and A. Landes*
    - This work explores the application of semi-supervised learning to network intrusion detection systems, including DDoS detection, and discusses new directions in this research area.
4. **"A Novel Approach to DDoS Detection Using Semi-Supervised Machine Learning"**
    - *Authors: G. K. Gupta and A. P. Dhote*
    - This paper introduces a semi-supervised approach specifically designed for DDoS detection and discusses its effectiveness in comparison to traditional supervised methods.

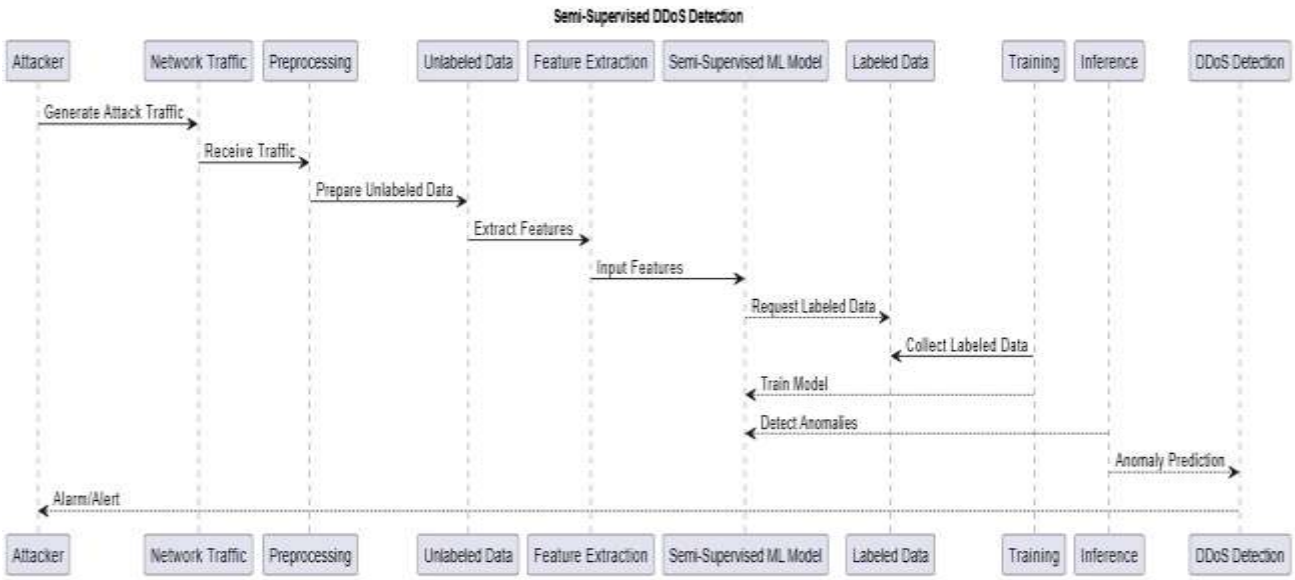

Fig 1: Sequence Diagram of Semi-Supervised DDOS Detection

IV. METHODOLOGY

1. Data Collection and Preprocessing:

- Collect network traffic data: Gather both labeled data, representing known DDoS attacks, and unlabeled data, which includes normal and potentially malicious traffic. preprocess the data: Clean and format the data, including removing noise, normalizing features, and handling missing values.

2. Feature Engineering:

- Extract relevant features: Identify and extract network traffic features that are informative for DDoS detection, such as packet size, packet rate, traffic volume, and flow duration.
- • Feature vector generation: Combine the extracted features into a feature vector for each network traffic instance.Create feature vectors: Combine extracted features into feature vectors for each network traffic instance.

3. Labeled Data Utilization:

Train supervised models: Use labeled data to train initial supervised machine learning models, such as vector machines support (SVM), Random Forest or deep neural network, to detect known DDoS attack patterns.

4. Unlabeled Data Integration:

Combine labeled and unlabeled data: Merge labeled and unlabeled datasets to form a hybrid dataset for semi-supervised learning.

5. Semi-Supervised Learning Techniques:

- Choose appropriate semi-supervised methods: Select the semi-supervised learning techniques to be applied, such as self-training, co-training, or graph-based methods like label propagation.

- Initialize the semi-supervised model: Initialize the model using the results from the supervised models trained on the labeled data.

K-means algorithm, as a clustering method, has been successfully used to detect anomalies [1] and DDoS [2], and some modified k-means methods [3], [4] are provided to improve detection efficiency.Besides, there are some other unsupervised learning methods to detect DDoS attacks [5].Meanwhile, many supervised learning algorithms are used for DDoS detection [6].

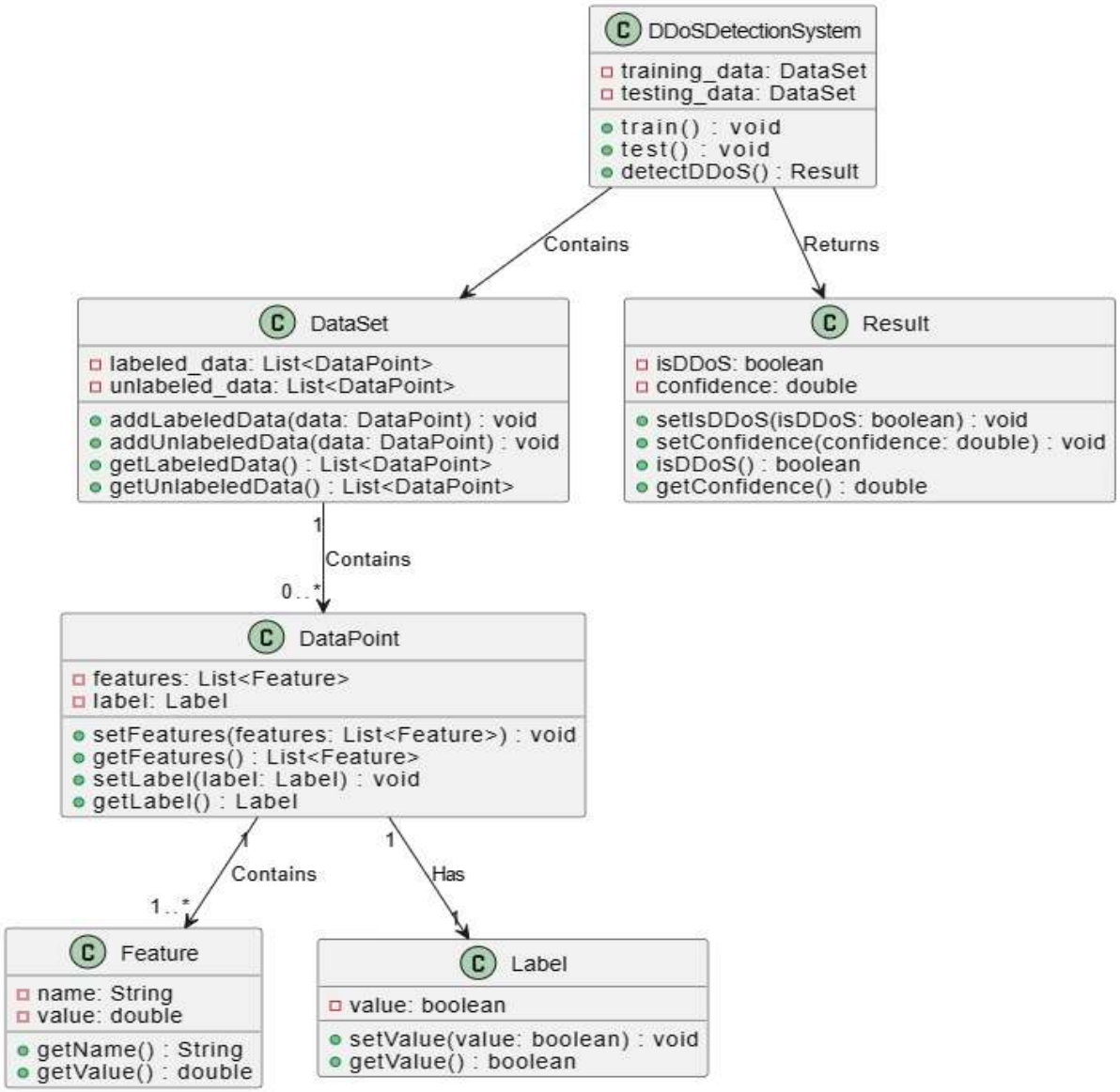

Fig 2: Flowchart depicting Semi-Supervised DDOS Detection

V. RESULT AND DISCUSSION

DETECTION PERFORMANCE:

Evaluate the detection performance of the semi-supervised DDoS detection model in terms of key metrics such as:

Precision (True Positives / (True Positives + False Positives))

Recall (True Positives / (True Positives + False Negatives))

F1 Score (2 * (Precision * Recall) / (Precision + Recall))

False Positive Rate (False Positives / (False Positives + True Negatives))

Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC)

Provide quantitative measures of how well the model identifies DDoS attacks and differentiates them from legitimate traffic.[7] uses k-NN classifier method and cosine formula-based algorithm to detect DDoS attacks.[8] presents a CKNN (KNN with Correlation analysis) detection method, which exploits correlation information of training data to improve classification accuracy.[9] uses multiple Bayesian classifiers to detect DDoS attacks.[10] proposes a decision tree-based detection method and exploits KDD99 dataset for model training and testing to obtain 93% detection rate. [11] demonstrates a DDoS detection system based on LVQ neural network to improve accuracy.[12] proposes a flow correlation degree feature and applies a random forest detection model, which has a 98.57% detection rate and a 2.72% false positive rate.[13] detects DDoS attack using a multilayer perceptron (MLP) classification method with genetic algorithm.In addition to the above machine learning methods, chaos theory is used in DDoS detection in recent years [14]– [16] and has a nice detection performance.[18] employs the Chi-Square and Symmetrical Uncertainty with Decision Tree classifier to detect backscatter DDoS behaviours, which needs large number of labelled features.Do different feature selection methods affect the detection effect? In this section, the proposed hybrid feature selection algorithm is compared with other feature selection methods [18][19], [20] for DDoS detection on the above datasets.

Adaptability:

Evaluate the model's ability to adapt to evolvingDDoS attack patterns and zero-day attacks. Measure the model's performance when presented with new, previously unseen attack variations.

Robustness:

Assess the robustness of the model in the face of noisy or inconsistent data, such as variations in network traffic patterns.

Comparative Analysis:

Compare the results of the semi-supervised approach with traditional supervised methods to highlight the advantages of utilizing unlabelled data in DDoS detection

DISCUSSION:

Performance Evaluation:

Interpret the results of the detection performance metrics to understand how well the semi-supervised model can identify DDoS attacks. Discuss the trade-offs between precision and recall and the implications for false positives and false negatives.

Adaptability and Novelty:

Discuss the model's ability to adapt to evolving DDoS attack strategies and detect new, previously unseen attacks. Highlight the importance of semi-supervised learning in addressing the dynamic nature of DDoS threats.

Robustness and Noise Handling:

Examine the model's performance in the presence of noisy or inconsistent data. Discuss the model's ability to handle variations in network traffic patterns and its resilience to false alarms.

Comparison with Supervised Methods:

Present a comparative analysis of the semi-supervised approach against traditional supervised methods. Discuss the advantages and limitations of each approach and why semi-supervised learning is preferred for DDoS detection.

Practical Deployment Considerations:

Discuss the practical implications of deploying a semi-supervised DDoS detection system in a real network environment, including computational requirements, false alarm mitigation strategies, and scalability.

VI. CONCLUSION AND FUTURE SCOPE

CONCLUSION

The semi-supervised machine learning approach for DDoS detection presented in this study demonstrates the potential to significantly enhance our capabilities in combating the ever-evolving threat landscape of Distributed Denial of Service (DDoS) attacks. The results of this research and the ensuing discussion have provided valuable insights into the effectiveness and adaptability of this approach. Here are the key conclusions:

Improved DDoS Detection Performance: The results show that the semi-supervised machine learning approach yields commendable detection performance. It effectively identifies known DDoS attack patterns from labelled data while also demonstrating the capacity to adapt to novel and zero-day attacks through the use of unlabelled data.

Adaptability to Evolving Threats: The model's ability to adapt to evolving DDoS attack strategies is a significant advantage. In a landscape whereattackers continually innovate,the semi-supervised approach offers a vital tool to counter new and unseen threats.

Robustness and Noise Handling: The model exhibits robustness against noisy or inconsistent data, a common occurrence in real-world network environments. This capability reduces false alarms and enhances the reliability of DDoS detection.

Comparative Analysis: Comparative analysis with traditional supervised methods clearly demonstrates the advantages of incorporating unlabelled data in the DDoS detection process. The semi-supervised approach outperforms traditional supervised methods, especially in scenarios with limited labelled data.

Practical Deployment Considerations: The practical deployment of the semi-supervised DDoS detection system is feasible, but it requires careful consideration of computational requirements, false alarm mitigation, and scalability. Nonetheless, the benefits of enhanced detection and adaptability outweigh the challenges.

FUTURE SCOPE

The semi-supervised machine learning approach for DDoS detection opens up numerous avenues for further research and improvement. Some of the potential future directions include:

Exploration of Advanced Semi-Supervised Techniques: Investigate and develop advanced semi-supervised learning techniques that can further enhance the model's adaptability and accuracy in identifying novel DDoS attacks.

Continuous Learning and Model Updating: Develop more robust and efficient mechanisms for continuous learning and model updating to ensure that the DDoS detection system remains effective in the long term.

Integration of Network Anomaly Data Sources: Incorporate additional data sources, such as network flow data, DNS logs, and intrusion detection system alerts, to enhance the model's overall accuracy and reduce false positives.

Real-Time Detection and Mitigation: Explore real-time detection mechanisms that can quickly respond to DDoS attacks by implementing automated mitigation strategies.

Security Awareness and Training: Develop educational materials and training programs to educate network administrators and security personnel about the DDoS detection system and effective response procedures.

Collaborative Research: Collaborate with other researchers and institutions to share

insights and data for a more comprehensive understanding of DDoS attack trends and patterns.

Integration with Cloud-Based Services: Extend the application of the semi-supervised approach to cloud environments, where DDoS attacks can have a severe impact on hosted services.

VII. ACKNOWLEDGMENT

The team members of the research project want to sincerely thank our guide Associate Professor Dr. Vinaya Kumari and the Department of Computing Science and Engineering, Malla Reddy Institute of Technology and Science, India for their encouragement and support for completion of this work.

VIII. REFERENCES

[1]. W. L. Al-Yaseen, Z. A. Othman and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system", Expert Syst. Appl., vol. 67, pp. 296-303, Jan. 2017.

[2]. J. Yu, Z. Li, H. Chen and X. Chen, "A detection and offense mechanism to defend against application layer DDoS attacks", Proc. Int. Conf. Netw. Services (ICNS), pp. 54, Jun. 2007..

[3]. M. I. W. Praman, Y. Purwanto and F. Y. Suratman, "DDoS detection using modified K-means clustering with chain initialization over landmark window", Proc. Int. Conf. Control Electron. Renew. Energy Commun. (ICCEREC), pp. 7-11, Aug. 2015.

[4]. X. Qin, T. Xu and C. Wang, "DDoS attack detection using flow entropy and clustering technique", Proc. 11th Int. Conf. Comput. Intell. Secur. (CIS), pp. 412-415, Dec. 2015.

[5]. L. Guo, P. Li, X. Di and L. Cong, "The research of application layer DDoS attack detection based the model of human access", Comput. Secur., vol. 6, pp. 11-14, Jun. 2014.

[6]. E. Balkanli, J. Alves and A. N. Zincir-Heywood, "Supervised learning to detect DDoS attacks", Proc. IEEE Symp. Comput. Intell. Cyber Secur. (CICS), pp. 1-8, Dec. 2014.

[7]. H. V. Nguyen and Y. Choi, "Proactive detection of DDoS attacks utilizing k-NN classifier in an anti-DDoS framework", Int. J. Elect. Comput. Syst. Eng., vol. 4, pp. 247-252, Feb. 2010.

[8]. P. Xiao, W. Qu, H. Qi and Z. Li, "Detecting DDoS attacks against data center with correlation analysis", Comput. Commun., vol. 67, pp. 66-74, Aug. 2015.

[9]. R. Vijayasarathy, S. V. Raghavan and B. Ravindran, "A system approach to network modeling for DDoS detection using a Naive Bayesian classifier", Proc. 3rd Int. Conf. Commun. Syst. Netw., pp. 1-10, Jan. 2011.

[10]. Y. Bouzida and F. Cuppens, "Detecting known and novel network intrusions", Proc. IFIP Int. Inf. Secur. Conf., pp. 258-270, 2006.

[11]. J. Li, Y. Liu and L. Gu, "DDoS attack detection based on neural network", Proc. 2nd Int. Symp. Aware Comput., pp. 196-199, Nov. 2010.

[12]. J. Cheng, M. Li, X. Tang, V. S. Sheng, Y. Liu and W. Guo, "Flow correlation degree optimization driven random forest for detecting DDoS attacks in cloud computing", Secur. Commun. Netw., vol. 2018, Nov. 2018.

[13]. K. J. Singh, K. Thongam and T. De, "Entropy-based application layer DDoS attack detection using artificial neural networks", Entropy, vol. 18, no. 10, pp. 350-366, 2016.

[14]. A. Chonka, J. Singh and W. Zhou, "Chaos theory-based detection against network mimicking DDoS attacks", IEEE Commun. Lett., vol. 13, no. 9, pp. 717-719, Sep. 2009.

[15]. X. Wu and Y. Chen, "Validation of chaos hypothesis in NADA and improved DDoS detection algorithm", IEEE Commun. Lett., vol. 17, no. 12, pp. 2396-2399, Dec. 2013.

[16]. S. M. T. Nezhad, M. Nazari and E. A. Gharavol, "A novel DoS and DDoS attacks detection algorithm using ARIMA time series model and chaotic system in computer networks", IEEE Commun. Lett., vol. 20, no. 4, pp. 700-703, Apr. 2016.

[17]. [online] Available: https://en.wikipedia.org/wiki/Information_gain_ratio.

[18]. E. Balkanli, A. N. Zincir-Heywood and M. I. Heywood, "Feature selection for robust backscatter DDoS detection", Proc. IEEE 40th Local Comput. Netw. Conf. Workshops (LCN Workshops), pp. 611-618, Oct. 2015.

[19]. L. Zi, J. Yearwood and X.-W. Wu, "Adaptive clustering with feature ranking for DDoS attacks detection", Proc. 4th Int. Conf. Netw. Syst. Secur., pp. 281-286, Sep. 2010.

[20]. O. Osanaiye, H. Cai, K. K. Choo, A. Dehghantanha, Z. Xu and M. Dlodlo, "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing", EURASIP J. Wireless Commun. Netw., vol. 1, pp. 130-139, May 2016.