Assessing DDoS Detection Accuracy through Semi-Supervised Techniques

[1]Dr.J. SIRISHA, [2]M. HARIKA, [3]M. PRABHU DASU, [4]N. SAI HARSHITHA, [5]M. PREM CHANDU

[1]*Assistant Professor, Information Technology, Prasad V. Potluri Siddhartha Institute of Technology*

[2,3,4,5] *IV Year B. Tech Information Technology, Prasad V. Potluri Siddhartha Institute of Technology*

Abstract— Despite the proliferation of advanced Machine Learning (ML) techniques in DDoS detection, this pervasive attack remains a significant menace to the Internet's integrity. Existing ML based DDoS detection methods fall into two categories: supervised and unsupervised approaches. This paper synthesizes insights from existing research endeavors, and enhance DDoS detection through machine learning methodologies, specifically focusing on semi-supervised techniques for analysis purposes. By harnessing the power of semi-supervised ML, we employ a succession of algorithms including Naive Bayes, Support Vector Machines (SVM), and Logistic Regression, focusing on factors crucial for detection accuracy. This paper mainly focuses on assessing various ML algorithms. In summary, this paper presents the potential of semi-supervised ML in augmenting detection accuracy.

Keywords— DDoS detection, Machine Learning, SVM, Naïve Bayes, Logistic regression

I. INTRODUCTION

A DDoS attack (Distributed Denial-of-Service) is a cyber-attack where attacker disrupts the normal internet traffic of a server, service or network by flooding the server with many requests. This slows down or complete shuts down of server as the resources provided by server will be exhausted, finally results in obstructing the legitimate users requests.

The distributed nature of the attack through a network of devices controlled by the attacker, makes it hard to block the attack. DDoS attacks pose risks such as downtime, financial losses, and reputational damage, necessitating robust defence strategies for organizations reliant on continuous online operations.

DDoS attacks disrupt online services by drenching target systems with spurious requests, rendering them inaccessible to legitimate users. Conventional detection methods struggle to keep pace with the evolving tactics of attackers, highlighting the need for more adaptive strategies. Semi-supervised machine learning offers a dynamic approach to discern subtle variations in traffic patterns indicative of DDoS activity.

Our study delves into assessing the accuracy of semi-supervised machine learning techniques for detecting DDoS attacks by combining labeled data showcasing known attack patterns with unlabeled data representing normal network behaviour. Through the utilization of Naive Bayes, SVM, and Logistic Regression algorithms, we try to analyse how accurately these algorithms would detect DDoS attack using semi-supervised techniques.

II. LITERATURE STUDIES

As there are various approaches proposed for detecting DDoS attack, in this section we summarize some of the recent works in DDoS detection.

- In [1], This paper conducted analysis of the ML methods for Botnet DDoS attack detection to

help developers in selecting right methods and for further research.

- In [2], Using ping of death technique and ML, they achieved 99.76% classification accuracy with the Random Forest algorithm on NSL-KDD dataset samples, demonstrating DDoS detection.

- In [3], they used semi-supervised k-means clustering and hybrid feature selection algorithm and concluded that these methods outperformed the benchmarks in detecting the attacks.

- In [4], They used prototype detector to identify and mitigate the DDoS attacks by providing filtering rules to reduce their impact.

- In [5], It was a study on detection of low-rate and high-rate doss attacks by using four entropy measures and also information divergence measures.
- In [6], Discussed about Internet-firewall approach that intercept attack packets before reaching the victim.

- In [7], This paper present an AI DDoS attack detection method based on neural networks. To detect the normal and abnormal attacks, they made use of server resources and network traffic.

- In [8], This paper talks about multivariate correlation analysis and the results show that it is highly accurate in detecting malicious network traffic.

- In [9], In this paper, they made used of SDN which is an emerging new network managing platform. They demonstrated that their methods can quickly locate DDoS victims and attackers by using a constrained number of flow monitoring rules.

- In [10], Author evaluated BayesNet, Naïve Bayes, J48 and Random Forest to detect attacks. Principal component analysis (PCA) is used for feature selection and also WEKA ML workbench is used to classify attack types.

- In [11], They have used Bro and Corsaro open sourced, CART decision tree and Naïve Bayes ML classifiers. They achieved high performance on backscatter dark net traffic without using IP addresses and port numbers.

- In [12], They proposed AutoEncoder based detection framework which uses normal traffic to build detection model to auto detect the attacks as the time goes.

- In [13], They evaluated various ML like Naïve Bayes, C4.5, SVM, KNN, K-means and c-means clustering. And the result is Fuzzy c-means clustering provides better accuracy in detecting the attacks.

- In [14], Here Deep learning is used to automatically extract high-level features from low-level ones and they built a network to learn patterns from network traffic and to identify network attacks and their activities.

- In [15], They applied various ML algorithms like KNN, SVM and Random Forest and used clustering, voting methods.

III. METHODOLOGIES

In here, we used methodology that centered around semi-supervised ML techniques to analyze the accuracy of DDoS attack detection. Semi-supervised learning is particularly chosen as it can deal with both labeled and unlabeled data which helps in dealing with real time problems.

A. Data Collection and Preprocesssing

We have chosen network traffic data set. It has various attributes such as “RID, protocol, ip_src, ip_dst, pro_srcport, pro_dstport, flags_ack, ip_flags, ip_flags_mf, ip_flags_df, ip_flags_rb, pro_seq,

pro_ack, frame_time, Packets, Bytes1, Tx_Packets, Tx_Bytes, Rx_Packets, Rx_Bytes". In preparing our dataset, we followed preprocessing steps to ensure data quality and suitability for predictive tasks. We began by addressing missing values through strategies like imputation, removal. Subsequently, we performed feature engineering, extracting valuable information from datetime columns and converting categorical variables into numerical representations. Normalization was applied to numerical features to prevent dominance in modeling. Techniques such as oversampling and undersampling were employed to address potential class imbalance. Text data underwent preprocessing like tokenization and stopwords removal. We split the dataset into training and testing sets for evaluation. Additionally, outlier detection and removal procedures were implemented to enhance robustness, particularly in numerical features. Encoding of the target variable and removal of redundant features were conducted as needed.

B. Implimentation of DDoS Detection System

1) Remote User Perspective:

At remote user side authentication and registration functionalities are designed by making use of Django's authentication system. Here remote user searches whether it is a DDoS attack or not by entering details.

2) Service Provider Perspective:

The authentication mechanisms are developed to access the remote user data and also displays the attack type ratios using charts like line chart and pie chart.

C. ML Models

Model was trained and tested using multiple algorithms and various metrics are taken into consideration to evaluate the performance of the model. Algorithms used are SVM, Naïve Bayes and Logistic Regression whereas the metrics taken into consideration are accuracy, precision, recall, F1-score and confusion matrix.

1.Support Vector Machines (SVM):

SVM is employed as a classifier to predict the type of DDoS attacks based on various features extracted from network traffic data. It's trained on the dataset to learn patterns in the data and classify instances into different attack types such as "Normal," "Smurf," or "Fraggle."

2.Linear Regression:

Linear Regression is used in the project for regression analysis and predictive modeling on certain numerical features to understand their relationship with other variables in the dataset, such as predicting the number of packets or bytes transmitted.

3.Naive Bayes:

Naive Bayes is another classifier utilized in the project to predict DDoS attack types. It's particularly effective for text classification tasks, making it suitable for scenarios where textual features need to be analyzed, such as the prediction of attack types based on textual information extracted from network logs or packet headers.

IV. RESULTS AND DISCUSSIONS

In this study, we tried to evaluate various algorithms and find out which one has more accuracy for detecting 'Distributed Denial of Service' (DDoS) attacks. On application of SVM, Logistic Regression and Naive Bayes algorithms, all three algorithms achieved the same accuracy of 95.94% on the dataset. This implies that each algorithm performed equally in distinguishing between different classes of network traffic.

The results of our model are visually represented using various graphical formats like pie charts and line charts. These visualizations represent data in more intuitive form which helps everyone to easily understand the findings.

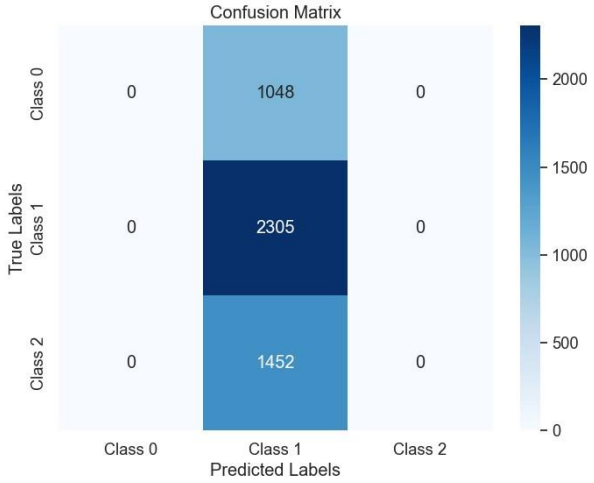

Fig 1: Confusion Matrix.

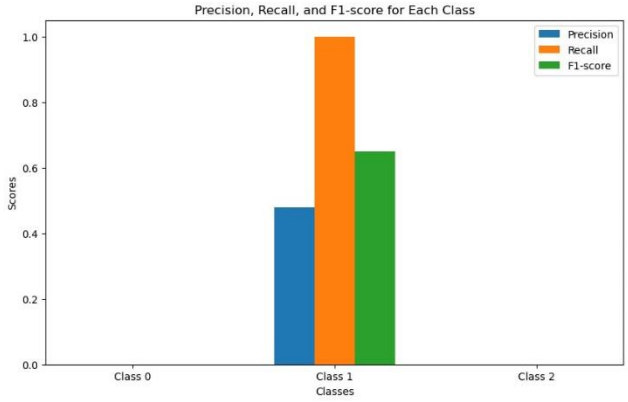

Fig 2: Bar graph representation of classification report.

In our dataset we mainly have two types of DDoS attacks which are, Smurf and Fraggle attacks. Our model detects whether it is a Smurf attack or Fraggle attack or normal i.e. not a DDoS attack.

1)Smurf:

In a Smurf attack, the attacker tricks many devices into sending data to a victim's network by pretending to be the victim. This flood of data overwhelms the victim's network, making it hard for real users to access its services.

2)Fraggle:

In a Fraggle attack, the attacker sends a flood of UDP packets to the broadcast address of a network, with the victim's IP spoofed as the source. This flood of packets, similar to a Smurf attack, causes the victim's network to be overwhelmed with traffic, disrupting its services for legitimate users.

The below pie chart represents the percentages of various DDoS types present in our network traffic dataset.

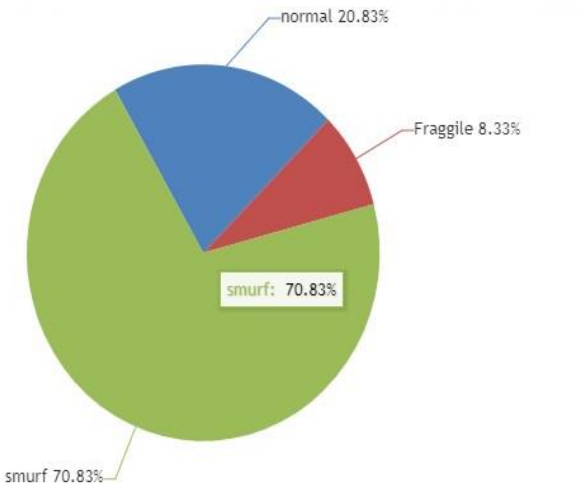

Fig 3: Pie chart representation of DDoS attack type.

V. CONCLUSION

On evaluating the efficacy of Naive Bayes, Support Vector Machine, and Logistic Regression algorithms for detecting Distributed Denial of Service (DDoS) attacks, all three algorithms achieved accuracy of 95.94% on the dataset. However, they classified different classes of network traffic and showcased their potential for reliable DDoS detection.

Even though there are variations in their methodologies, all three algorithms presented consistent performance across various metrics, showcasing their robustness in differentiating between normal traffic and DDoS attacks.

REFERENCES

[1] Tuan, T. A., Long, H. V., Son, L. H., Kumar, R., Priyadarshini, I., & Son, N. T. K. (2020). Performance evaluation of Botnet DDoS attack detection using machine learning. Evolutionary Intelligence, 13(2), 283–294. doi:10.1007/s12065-019-00310-w

[2] Pande, S., Khamparia, A., Gupta, D., & Thanh, D. N. H. (2021). DDOS detection using machine learning technique. In Studies in Computational Intelligence. Recent Studies on Computational Intelligence (pp. 59–68). doi:10.1007/978-981-15-8469-5_5

[3] Gu, Y., Li, K., Guo, Z., & Wang, Y. (2019). Semi-supervised K-means DDoS detection method using hybrid feature selection algorithm. IEEE Access: Practical Innovations,OpenSolutions,7,64351–64365. doi:10.1109/access.2019.2917532

[4] Feinstein, L., Schnackenberg, D., Balupari, R., & Kindred, D. (2004). Statistical approaches to DDoS attack detection and response. Proceedings DARPA Information Survivability Conference and Exposition. Presented at the DARPA Information Survivability Conferenc&Exposition,Washington,DC,USA.doi:10.1109/discex.2003.1194894

[5] Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) An empirical evaluation of information metrics for low-rate and high-rate ddos attack detection. Pattern Recogn Lett 51:1–7

[6] Chang RKC (2002) Defending against flooding-based distributed denial-of-service attacks: a tutorial. IEEE Commun Mag 40(10):42–51

[7] Li, J., Liu, Y., & Gu, L. (2010, November). DDoS attack detection based on neural network. 2010 2nd International Symposium on Aware Computing. Presented at the 2010 2nd International Symposium on Aware Computing (ISAC), Tainan, Taiwan. doi:10.1109/isac.2010.5670479

[8] Jin, S., & Yeung, D. S. (2004). A covariance analysis model for DDoS attack detection. 2004 IEEE International Conference on Communications(IEEE Cat. No.04CH37577). Presented at the 2004 IEEE International Conference on Communications (IEEE Cat.No.04CH37577),Paris,France.doi:10.1109/icc.2004.1312847

[9] Xu, Y., & Liu, Y. (2016, April). DDoS attack detection under SDN context. IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, 1–9. Presented at the IEEE INFOCOM 2016 - IEEE Conference on Computer Communications, San Francisco, CA, USA. doi:10.1109/infocom.2016.7524500

[10] Jyoti, Navjot, and Sunny Behal. "A meta-evaluation of machine learning techniques for detection of DDoS attacks." 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2021.

[11] Balkanli, E., Alves, J., & Zincir-Heywood, A. N. (2014, December). Supervised learning to detect DDoS attacks. 2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS). Presented at the 2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS),Orlando,FL,USA. doi:10.1109/cicybs.2014.7013367

[12] Yang, K., Zhang, J., Xu, Y., & Chao, J. (2020, April). DDoS Attacks Detection with AutoEncoder. NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium. Presented at the NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary. doi:10.1109/noms47738.2020.9110372

[13] Suresh, M., & Anitha, R. (2011). Evaluating machine learning algorithms for detecting DDoS attacks. In Communications in Computer and Information Science. Advances in Network Security and Applications (pp. 441–452). doi:10.1007/978-3-642-22540-6_42

[14] Yuan, X., Li, C., & Li, X. (2017, May). DeepDefense: Identifying DDoS attack via deep learning. 2017 IEEE International Conference on Smart Computing (SMARTCOMP). Presented at the 2017 IEEE International Conference on Smart Computing (SMARTCOMP),HongKong,China.doi:10.1109/smartcomp.2017.7946998

[15] Aamir M, Mustafa S, Zaidi A (2019) Clustering based semi-supervised machine learning for DDoS attack classification. J King Saud Univ - Comput Inf Sci 33(4):436–446