RESEARCH ARTICLE OPEN ACCESS

A Machine Learning Based Classification And Prediction Technique For Ddos Attack Using Xg Boost Algorithm

Bhuwaneaswari K [1], Pavithra A [2]

M.sc., Department of Computer Science and Engineering, Dr. MGR Educational and Research Institute, Chennai, India

Faculty of Centre of Excellence in Digital Forensics, Dr. MGR Educational and Research Institute, Chennai, India

ABSTRACT

Distributed Denial of Service (DDOS) attacks are increasingly becoming a threat to the security of networked systems. However, this must not be discouraged by these challenges and instead work tirelessly to develop innovative strategies to overcome them. This study proposes a fresh and inspiring approach that leverages machine-learning techniques for prediction and categorization. Our solution uses a vast dataset of network traffic characteristics, including normal and malicious trends. This project follows a rigorous methodology, including pre-processing raw network data, extracting features, and training machine learning models. This article explores various methods that can identify patterns that indicate denial-of-service attacks, such as support vector machines, random forests, and neural networks. Here XG Boost algorithm is used majorly for its high accuracy rate. Our models can learn and generalize patterns with high accuracy because they are trained on labelled data sets representing normal and attack conditions. Moreover, our developed system incorporates an inspiring predictive component that allows it to identify potential attacks before they occur. In conclusion, our approach is a beacon of hope that can help protect networked systems from DDOS attacks.

Keywords: DDoS, XG Boost, DoS, Precision, Reconnaissance, Random Forest Algorithm.

I. INTRODUCTION

Cyber-assaults are developing at a humongous rate because of technological advancement. Anything that reason to create disturbance to the digital tool or the virtual assets, might be positioned below cyber-assaults. Based at the target, Cybercrime and their assaults are categorized accordingly. As cited above cyber assaults are performed via way of means of an outsider to get right of entry to an organization's asset or personal information. The intruder exploits the vulnerability with inside the framework of the device and assaults the vulnerable protection settings to have an entry.

Cybercrime is erected around the effective exploitation of vulnerabilities, and security brigades are always at a disadvantage because they must defend all possible entry points, while bushwhacker only needs to find and exploit one weakness or vulnerability. This asymmetry largely favours bushwhackers. The result is that indeed large enterprises struggle to help cybercriminals from monetizing access to their networks.

One of the primitive yet widely used types of Cyber-assaults is Dos attack. A denial-of-service (DoS) assault is a form of cyber assault wherein a malicious actor targets to render a pc or different tool unavailable to its meant customers via way of means of interrupting the device's normal functioning. DoS assaults commonly feature via way of means of overwhelming or flooding a centered system with requests till everyday site visitors is not able to be processed, ensuing in denial-of-carrier to addition users. A DoS assault is characterized with the aid of using the usage of a unmarried laptop to release the assault.

A Distributed Denial-Of-Service (DDoS) assault is a form of DoS assault that comes from many disbursed sources, which include a botnet DDoS assault. A Distributed Denial-Of-Service (DDoS) attack is a malicious try to disrupt the ordinary web website online site visitors of a centered server, service or network via overwhelming the aim or

DDoS assaults obtain effectiveness via way of means of using more than one compromised laptop structures as reassets of assault traffic. Exploited machines can consist of computer systems and different networked sources inclusive of IoT devices.

From an excessive level, a DDoS assault is like a surprising site visitor's jam clogging up the highway, stopping ordinary site visitors from arriving at its destination.

II. REVIEW OF LITERATURE

Amal M. Al-Eryani; Eman Hossny; Fatma A. Omara, et al., 2024 [1] proposed the Distributed Denial of Service (DDoS) attack is a widely spread attack that poses a major threat to organizations dependent on online services. DDoS attacks aim to disrupt services by overwhelming servers with fake traffic from multiple sources. Early and effective detection of DDoS attacks is important for mitigating their impact. Recently, the most widely used algorithms for detecting DDoS are based on Machine Learning (ML) and Deep learning (DL). The work in this paper focuses on providing a comparative study between recent ML algorithms that were tested using the CICDoS2019 dataset. The objective of this comparison is to determine the most effective ML algorithm for DDoS detection. Based on the comparative study results, it is found that the Gradient Boosting (GB) and the XGBoost algorithms are extraordinarily accurate and correctly predicted the type of network traffic with 99.99% and 99.98% accuracy respectively, in addition to, a low false alarm rate of approximately 0.004 for GB.

Anupama Mishra, Brij B. Gupta, Neena Gupta 2022 [2] proposed a Distributed denial of carrier assaults are not unusual places and excessive hazards to diverse computing eras like Cloud, IoT, and Block chain due to the disruption they cause to the offerings that are provided. Many unique varieties of DDoS assaults are there, every with a unique action, making it hard for community tracking and manipulating structures to pick out and save them. The goal of this study's paintings is to discover and pick a hard and fast of facts to symbolize DDoS assault occasions and assault site visitors' information. A pre-processing section is used to easily remodel the data, and afterward, the era of a version of device mastering for multi-class classification is done. This is finished to become aware of the diverse class of various styles of DDoS attacks. Here , The CIC dataset has been used for the test which incorporates all varieties of DDoS assault and a massive in range of records. Random Forest, Support Vector Machine, Naive Bayes, Decision Tree, XGBoost, and AdaBoost are six different types of machine learning algorithms employed in this research. From the results, AdaBoost achieves a high-quality accuracy of 99.87 in 27. Fours of computation time. Naive Bayes has the quickest computing time (3.2 s) with 94.15% accuracy, whereas Support Vector Machine has the slowest time, a lazy learner (229m26s for education and 0.2 S for prediction) and has a low accuracy of 95.73

Gürcan Çetin, Koray Çoşkun, et al., 2022 [3] The safety of facts sources is an exceedingly vital problem. The community infrastructure that allows net access, in particular, can be focused through attackers from quite a few countrywide and worldwide locations, resulting in losses for establishments that make use of it. Anomaly detection systems, from time to time referred to as Intrusion Detection Systems (IDSs), are designed to become aware of abnormalities in the networks. The functions of IDSs, however, are restrained through the algorithms and getting to know the capability used inside the background. Because of the complicated conduct of malicious entities, it's crucial to undertake powerful strategies that guarantee excessive overall performance even as being time efficient. The fulfillment price of the boosting algorithms in figuring out malicious community visitors was studied in this study. The boosting approach, one of the maximum used Ensemble Learning techniques, is general as a manner to deal with this challenge. In this work, Google Colab has been used to version famous boosting algorithms. The AdaBoost, CatBoost, GradientBoost, LightGBM, and XGBoost fashions were implemented in the CICID2017 dataset. The overall performance of the classifiers has been evaluated with accuracy, precision, recall, f1-score, kappa value, ROC curve, and AUC. As a result of the investigation, it was found that the XGBoost set of rules produced the finest consequences in phrases of f1-score, with 99.89 percent, and the AUC values were extremely near to 1, with 0.9989. LightGBM and GradientBoost models, on the other hand, are less effective in detecting attack types with little data

Ayaz Ali Khan, Hameed Hussain; Ismail, Muhammad Ismail Mohmand Muhammad 2022 [4] DDos are usually known as Denial of Services. In the present studies study, the writer laboured on a vintage KDD dataset. This article used a system studying technique for DDoS assault kind’s category and prediction. For this purpose, used Random Forest and XGBoost type algorithms. After making use of the gadget getting to know models, a confusion matrix is generated for the identity of the version performance. The common Accuracy (AC) of this recommended version is ~90%. By evaluating the paintings to the present studies' works, the accuracy of the illness willpower drastically progressed that's about 85% and 79%, respectively.

Anant Raj; Kolli Saivenu; Mumtaz Irteqa Ahmed; Sathvik, B,Sanjeetha et al., 2021[5] One of the primitive but rather powerful community assaults is the Distributed Denial-of-Service (DDoS). DDoS assaults are released with the aid of the intruder the use of compromised hosts referred to as botnets received with the aid of using the attacker host known as the botmaster, all being related to switches gift inside the equal environment. Numerous answers have been proposed to counter those assaults and save you career disruptions that have price many agencies a fortune. Strategies like XGBoost, Support Vector Machine (SVM), etc., have addressed the detection of DDoS attacks. Nonetheless confirmed the scope of development in detection speeds that may drastically lessen the provider unavailability time from a server ie, the sufferer of the DDoS attack. Thus, this article addresses those necessities to construct an optimal, reliable, and short DDoS detection and mitigation application. This application leverages the controller's functionalities, and continuously monitors the network traffic at a particular host interface (potential victim) to detect abnormal traffic. When the visitors are recognized as the ability DDoS attack, its mitigation is initiated. The utility makes use of the Cat Boost classifier, the boosting set of rules that has little or no prediction time and is relatively 8× quicker than XGBoost, due to its symmetric tree structure. It is tested to be proven reliable and efficient in detecting botnet-based DDoS attacks on the SDN environment with an accuracy of 98% and far less training time. Thus, proving that the proposed answer using the latest device getting-to-know version may be extra powerful in fast detecting and mitigating a DDoS attack.

Hemalatha E, Subhashini Peneti et al., 2021 [6] Denial of Service attacks have been one of the most frequent attacks. Correctional measures like Whitelisting/ Blacklisting Ip addresses etc are implemented. The major goal of any DoS attack is to bring down the reputation of the victim organization. A detection system to detect and prevent DoS attacks is made mandatory. The usage of Machine Learning techniques tops all the other techniques. There are lots of records to be had approximately DoS attacks, the system getting to know algorithms can locate styles of those DoS attacks and for that reason follow those styles to new requests and classify them as malicious or harmful requests. This article uses the CICIDS2017 dataset. The dataset has data related to requests of 7 days a week. Among those, Wednesday's dataset contains records related to types of DoS attacks. Even all even though strategies like XGBoost, and neural networks may be applied, random wooded area offers a outstanding performance.

Hassan A. Alamri; Vijey Thayananthan 2020 [7] Software-defined networking (SDN) is a rising community structure that addresses the quandary of the conventional community by supplying centralized control through a crucial controller that decouples the management and records planes. But, this improvement has made the controller an intense goal for malicious customers to execute assaults including Distributed Denial of Service (DDoS) assaults. Several schemes have been proposed to mitigate DDoS attacks in SDN, however the demanding situations nevertheless exist. This paper proposes a DDoS mitigation scheme for SDN to make certain correct attack detection and green community aid usage. The scheme employs stages: a bandwidth management mechanism and an Extreme Gradient Boosting (XGBoost) Algorithm. The bandwidth management mechanism makes use of an adaptive bandwidth profile-primarily based threshold and bandwidth management set of rules that cause the XGBoost set of rules in case of threshold violations. The use of a couple of bandwidth profiles in setting the edge guarantees the edge's adaptively to not forget the community site visitors' version and decrease the packet drop ratio, which suggests an incredible result. The XGBoost set of rules classifies community site visitors' float that violates a fixed threshold for ordinary or atypical site visitors. The evaluation of the overall performance of our scheme with the use of CICDDoS2019, NSL-KDD, and CAIDA datasets. Demonstrated our proposed answer in real time with the SDN environment. The consequences acquired display that our scheme protects SDN in opposition to DDoS assaults with excessive accuracy, low error, and green usage of community resources. The proposed gadget executed 99.9 accuracy in detecting DDoS assaults with a low false-nice charge of 0.0002% in SDN.

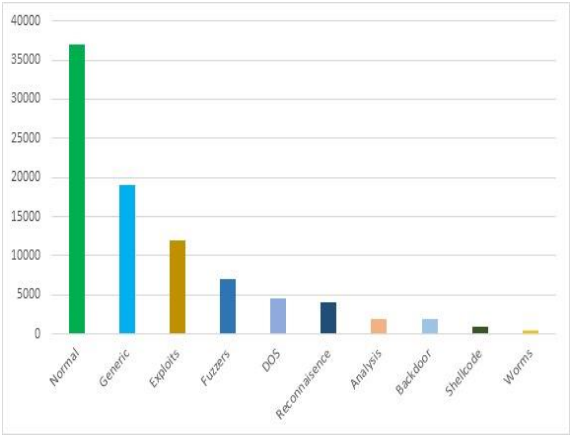
Bhoopesh Singh Bhati, Fadi Al-Turjman, Garvit Chugh, Nitesh Singh Bhati 2020 [8] Network attacks are increasing day by day. To mitigate them, a device has been created, that actively detects intrusions and assaults in an internal network. The device that detects those styles of assaults and intrusions is referred to as an intrusion detection device (IDS). The assaults are of kinds, acknowledged and unknown. The IDSs are capable of shielding regarded assaults as they're designed in particular. As the use of the Internet is developing each day, the attacks are growing tremendously and they all aren't acknowledged to an IDS without the right upgradation, which is dangerous because it will now no longer get detected through the IDS and depart the device open to threats. Therefore, an IDS needs to now no longer simply come across the acknowledged attacks but even offer safety from unknown assaults. In this article, an ensemble-primarily based totally IDS the use of

XGBoost is presented. There have been preceding studies on the subject and with the assistance of progressed technologies, it turns into viable to enhance the performance and accuracy of the ensemble primarily based totally IDS. This article points to a scheme that suggests usage of XGBoost with an ensemble primarily based totally IDS can offer higher consequences as XGBoost is primarily based totally on the tree boosting tool gaining knowledge of algorithms, which allows in handling a smoother "bias-variance" trade-off. The test is implemented at the KDDCup99 dataset and the recorded accuracy of the proposed technique through this test is 99.95

III. RESEARCH METHODOLOGY

To figure out the challenges and difficulties of the DDos attacks worldwide, prediction mechanism is being implanted by Machine Learning algorithms. Here a demonstration of the prediction of DDos attacks is been explained with Random Forest Algorithm and XG Boost Algorithm. Compared to Random Forest Algorithm, XG Boost Algorithm provides more accurate values. The UNSW-nb15 dataset is trained using Machine Learning techniques and the dataset is implemented using different algorithms for accurate value. Apart, from this the dataset is classified based on nine different types of DDoS attacks and each are given a specific, unique value. Once the given python code is executed, the respective value is displayed thereby the type of attack is predicted.
The given figure 1 depicts the graphical representation of all nine attacks and their entire count in the dataset.

Fig. 1: A graphical representation of the classification of the types of DDos Attacks



The dataset is obtained from the Australian Centre for Cyber Security. This UNSW-nb15 is a dataset which comprises of nine different types of DDos attacks and has a accuracy rate of 78.52 percent. The dataset consists of stings, a combination of alphabets and numerals. This given dataset is trained by different Machine Learning techniques. The dataset is to be cleaned to make it usable and executable to get the desired results. As mentioned above Random Forest Algorithm and XG Boost is implemented. In Random Forest Algorithm, the dataset is cleaned and trained in a manner that it gives a tree-like appearance as it continues to branch out. Basically, Random Forest is used to classify huge amount of data into smaller classes or categories. The accuracy rate for the prediction of DDoS attack by using Random Forest algorithm is 88.9 percent. To attain good accuracy rate and to prevent false positive and negatives, XG Boost algorithm is implied. Compared to Random Forest Algorithm, XG Boost algorithm classifies huge data in a more linear and narrow version. Also it responds and cleans the dataset at a higher speed than Random Forest algorithm. The accuracy rate of the XG Boost algorithm is 90 percent. Data Pre-processing is a phase where the raw data is gathered all together and it is made into several categories or classes upon the given algorithm. In this Pre-processing, the dataset is examined thoroughly. Using Python, the code for the implementation is written. The dataset is cleaned, strings are converted into numeric form with Python. Python has libraries which helps to divide the raw dataset into two i.e. Train data and Test data. Train data is making the obtained dataset to undergo training according to the commands and code.

The Train data is familiar with the commands and responds quickly.

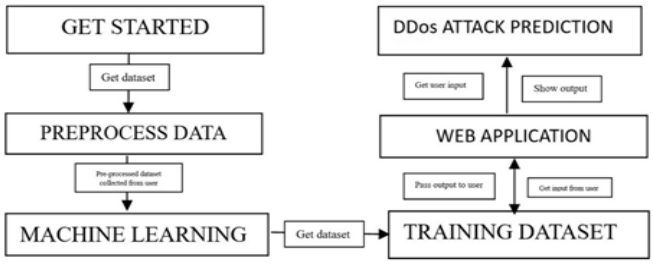


Fig. 2: Representation of the training of the dataset using Machine Learning Algorithms.

On the other hand, Test dataset is the one which is kept away from the working and the commands. The Test dataset answers to the

commands or request given to it. Test dataset is used for the implementation process. Once the dataset Pre-processing is completed, the data is recorded in the Jypyter notebook in rows and columns. New rules are added based upon the need of the problem and null sets are eliminated. Using get dummies and creating a confusion matrix, the dataset accuracy can be tracked.

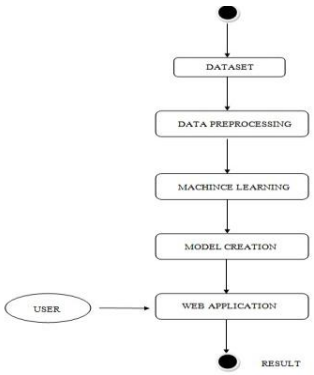


Fig. 3: Visual representation of the workflow.

The Pre-processed dataset, the model obtained from the training and test data set all are collaborated together to form the web application phase. Once the model is completed, an excel file is created by adding selected test dataset for implementation. Using Anaconda Prompt, a link is created using the given data. The link is executed on any browser. A text file contain the test data is uploaded. Once the file is uploaded, the type of attack is displayed on the screen.

IV. *RESULT AND DISCUSSION*

Upon analyzing both Random Forest algorithm and XG Boost Algorithm, XG Boost is found to provide more accurate result. One of the major reason is that XG Boost follows a linear way of dataset analyzation and classification. On the other hand, Random Forest algorithm branches out the classification process and it appears to a tree-like structure. The precision reading of XG Boost is also higher than Random Forest algorithm for the given dataset, UNSW-nb15.

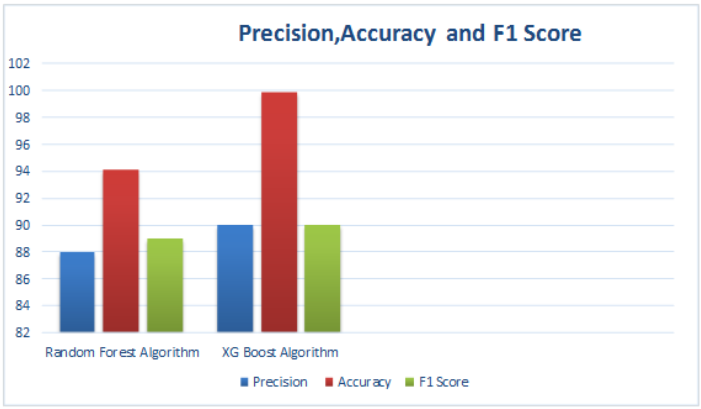


Fig 4: A Graphical representation of Random Forest Algorithm and XG Boost Algorithm along with their respective values.

CONCLUSION

The proposed DDos prediction system ensures a system app for the detection of DDos attack. Here, the UNSW-nb15 dataset from the Github repository that comprises information about the DDos attacks. Implementing Machine Learning approach, the given dataset is normalized and trained. This proposed DDoS prediction system ensures that the dataset which is given can predict the user upon the entry of some malicious DDoS attacks. In the near future, this model will be implemented on live systems and active environment and prediction and prevention of DDoS attacks could be done.

REFERENCES

[1] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ``Adversarial machine learning applied to intrusion and malware scenarios: A systematic review," IEEE Access, vol. 8, pp. 35403_35419, 2020.

[2] G. Karatas, O. Demir, and O. K. Sahingoz, ``Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset," IEEE Access, vol. 8, pp. 32150_32162, 2020.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ``BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset," IEEE Access, vol. 8, pp. 29575_29585, 2020.

[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ``Network intrusion detection based on PSO-xgboost model," IEEE Access, vol. 8, pp. 58392_58401, 2020

5] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ``Similarity based feature transformation for network anomaly detection," IEEE Access, vol. 8, pp. 39184_39196, 2020.

[6] L. D'hooge, T. Wauters, B. Volckaert, and F. De Turck, ``Classi_cation hardness for supervised

learners on 20 years of intrusion detection data," IEEE Access, vol. 7, pp. 167455_167469, 2019.

[7] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ``An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7, pp. 82512_82521, 2019.

[8] Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, ``Network intrusion detection based on supervised adversarial variational auto-encoder with regularization," IEEE Access, vol. 8, pp. 42169_42184, 2020.

[9] C. Liu, Y. Liu, Y. Yan, and J. Wang, ``An intrusion detection model with hierarchical attention mechanism," IEEE Access, vol. 8, pp. 67542_67554, 2020.

[10] S. U. Jan, S. Ahmed, V. Shakhov, and I. Koo, ``Toward a lightweight intrusion detection system for the Internet of Things," IEEE Access, vol. 7, pp. 42450_42471, 2019.

[11] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, ``Machine learning-based Technol., vol. 40, no. 1, pp. 215_229, Jan. 2021.

network vulnerability analysis of industrial Internet of Things," IEEE Internet Things J., vol. 6, no. 4, pp. 6822_6834, Aug. 2019.

[12] Y. Chen, B. Pang, G. Shao, G. Wen, and X. Chen, ``DGA-based botnet detection toward imbalanced multiclass learning," Tsinghua Sci. Technol., vol. 26, no. 4, pp. 387_402, Aug. 2021.

[13] X. Larriva-Novo, V. A. Villagrá, M. Vega-Barbas, D. Rivera, and M. S. Rodrigo, ``An IoT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets," Sensors, vol. 21, no. 2, p. 656, Jan. 2021.

[14] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, ``Network intrusion detection system: A systematic study of machine learning and deep learning approaches," Trans. Emerg. Telecommun. Technol., vol. 32, no. 1, p. e4150, Jan. 2021.

[15] M. Aamir, S. S. H. Rizvi, M. A. Hashmani, M. Zubair, and J. A. Usman, ``Machine learning classi_cation of port scanning and DDoS attacks: A comparative analysis," Mehran Univ. Res. J. Eng.