WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Network Anomaly Detection Due To Intrusion.

Iliyaz Pasha M[1], Varun Gowda M[2], Sidharth T S[3], S Singaravelu[4].

[1]Assistanat Professor Department of Computer Science and Engineering, R L Jalappa Institute of Engineering Karnataka India
[2, 3, 4]StudentsDepartment of Computer Science and Engineering, R L Jalappa Institute of Engineering Karnataka India

Abstract-

Internet services are being increasingly exploited, making network security crucial in guarding against Distributed Denial of Service attacks. These attacks send in massive traffic from various sources, completely disrupting the internal operations and rendering systems unavailable to legitimate users for their use. Traditional defense measures such as firewalls and signature-based intrusion detection rarely detect new types or evolutions of attack methods. The system collects network traffic data and then performs feature extraction to build classification models that discern between normal and malicious behavior.

1. Introduction.

With the onset of such an evolutionary mechanism of internet connected systems, the cyber attacks have also increased considerably. One of the most disruptive and precarious types of attacks an adversary can perpetrate is a DDoS attack. In a DDoS assault, a large number of devices, which constitute a botnet at times, transmit copious amounts of traffic to the server or network. Being overwhelmed, the system slows or crashes and becomes unavailable to real users. The attacks can be launched on websites, online services, or even networks, leading to huge financial losses and damage to reputation. Networks have so far been protected by traditional security methods that include the firewall policy and signature-based intrusion detection systems. Nevertheless, these systems rely on attack patterns that are already known and are thus rarely effective against newer or varying types of attacks. With DDoS attacks being made more and more complex and larger in scale, it becomes imperative to employ sophisticated detection techniques that would notice aberrant behavior even if that specific attack has never been seen before. This sets the terrain for anomaly detection to play an important role. Anomaly detection intends to detect deviations or patterns that are unusual or unexpected in network traffic that may indicate an attack. It does not require any previous knowledge from attacks, as opposed to signature-based systems. Because of this, it is best-suited to first identify new zero-day attacks or maligned DDoS-strategies. Through this project, we propose an anomaly-based DDoS detection system using machine learning to examine network traffic data and extract key features.

2. Literature Review

Due to the growing implications DDoS attacks have on network security, a number of researchers have tried to devise a way to detect them over the years. Traditional methods, like signature-based intrusion detection systems (IDS) rely on a database of predefined patterns to identify attacks. These methods work well against existing threats; however, when attackers attempt to alter their strategies or use newer and previously unidentified approaches, these methods tend to fail. That is why Anomaly-based Systems which place emphasis on seeking unusual patterns within network traffic have been developed. Analyzing immense amounts of data and identifying deviations within said data is only possible with Machine Learning; it has emerged as a useful tool in this field. Other algorithms that have become popular when it comes to DDoS detection include Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Decision Trees and Random Forest. Many studies prove that these algorithms achieve higher accuracy rates provided the datasets are well prepared prior to being presented to the machine learning models. For instance, some studies point out that ensemble models like Random Forest tend to outperform individual classifiers because they are more resilient, handle imbalanced datasets better, and make more accurate predictions. Recently, deep learning methods such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNN) have also been used.

3. ProposedSystem.

In this project, we design a complete system for detecting and classifying DDoS attacks in distributed networks using cutting-edge machine learning models. Our approach applies the CIC-IDS 2017 dataset, which contains rich data of network traffic to be used during the training and testing phases. We are going to implement multiple algorithms, such as XGBoost, LightGBM, and CATBoost, which are known for their high performance in classification tasks. Additionally, we are going to perform Random Forest and Decision Tree algorithms to build an ensemble model. The feature

selection process will be done using Lasso regression to enhance the model by accentuating the most relevant features, improving both model accuracy and interpretability. In order to solve class imbalance, we will apply SMOTE (Synthetic Minority Over-sampling Technique) to enrich the data. Additionally, a Voting Classifier will be used to combine the individual merits of the models using bagging with Random Forest and Extra Trees. The overall goal is to develop a powerful detection system that can effectively defend against the risk of DDoS attacks.

Benefits of suggested system

[1] The proposed system uses the CIC-IDS 2017 dataset which is the reflection of the real-time modern network traffic making it a more applicable training platform for DDoS detection.

[2]Implementing high-perform

4. Methodology

This project sets out to identify the abnormal behavior of a network as a result of intrusion attempts using machine learning techniques. The work is done in a sequence which include data collection, preprocessing, feature selection, modeling, and evaluation. The first step is data collection, where a labeled dataset is used containing normal and intrusive network traffic. For this project, well known datasets like CICIDS2017 ,NSL-KDD or UNSW-NB15 can be used as they have diverse attack types such as DDoS, port scans, brute force attacks, among others. These datasets come with flow-based features from real or simulated network traffic.

Next, the raw data undergoes data cleaning to ensure it's consistent. This involves removing irrelevant or duplicate features while addressing missing values, turning categorical values into numerical ones (label encoding or one-hot encoding), and scaling the values of features through normalizing or standardization. Class imbalance is also addressed through oversampling in case the data is mostly made up of unobtrusive traffic making the attack data disproportionally smaller.

After the initial preprocessing stage, feature selection identifies attributes that have a great impact on anomaly detection. This step eliminates computation expense dan optimizes model accuracy. This can be accomplished through correlation analysis, mutual information, or importance scores via tree structures.

Once the dataset is prepared, machine learning models such as Random Forest, Decision Tree, Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN) are implemented to identify normal and intrusive network activities. Usually, these models use 70 to 80 percent of the data for training and the rest for testing. Models are also validated several times with different portions of the data to test reliability and overfitting.

The trained models undergo evaluation based on a set of performance measures, including accuracy, precision, recall, F1-score, and confusion matrix. These measurements indicate the efficiency of intrusion detection relative to false alerts. The selected model must maintain the capability to detect multiple types and methods of intrusion effectively.

4.1. Algorithms:

XGBoost: XGBoost (Extreme Gradient Boosting) is a patented speed and efficiency version of the gradient-boosting algorithm. In our DDoS Detection project, XGBoost is used to effectively class network data by complex feature management and imbalanced class handling. Its regularization techniques tackle overfitting, thus providing robust attack classification. Even more accuracy can be gained with hyperparameter tuning in tree depth, learning rates, and several other parameters, yielding an accurate model that identifies DDoS attack patterns precisely.

LightGBM: As a memory-efficient algorithm suitable for large datasets and low-latency environments, LightGBM (Light Gradient Boosting Machine) is fast and particularly well optimized for lower time-consuming environments. In this project, LightGBM is implemented for DDoS attack detection by dealing with large network data as well as achieving high accuracy in classification tasks. Through hyperparameter optimization, LightGBM improves model performance by altering leaf growth which boosts the speed as well as model performance. Its performance with sparse data andthe ease of managing categorical features improves DDoS detection accuracy and responsiveness.

CATBoost: CATBoost (Categorical Boosting) is a gradient-boosting algorithm that works very well with categorical data and therefore leads itself to be effective in DDoS attacks detection. The fact that CATBoost can work with data and only requires minimal cleaning and tempering preprocessing, coupled with its tendency to not overfit, bodes well for the project because it lowers the chances of model bias. Under hyperparameter tuning, CATBoost enhances its learning rate and depth further improving prediction accuracy. Its concentration on computational efficiency aids real-time threat classification and therefore serves the attack detection model ensemble well.

Random Forest + Decision Tree: Random Forest is an ensemble of Decision Trees which enhances the precision of prediction by averaging the output of the multiple trees. In our project, Random Forest and Decision Tree classifiers are employed to detect patterns in network traffic of a DDoS attack. Decision Trees create underlsimplified interpretable models while aggregation in Random Forest alleviates high variance and increases reliability. Performance can be improved by hyperparameter tuning around tree depth and the number of trees which makes this combination a sound primary strategy for detection of network anomalies.

Voting Classifier: Voting Classifier is a classifier which uses multiple models for one classification, combining their results to improve classification accuracy. In this project, Voting Classifier combines predictions from different classifiers where some classifiers use bagging such as Random Forest (RF) with stability from Extra Trees (ET).

5. Result & Discussion.

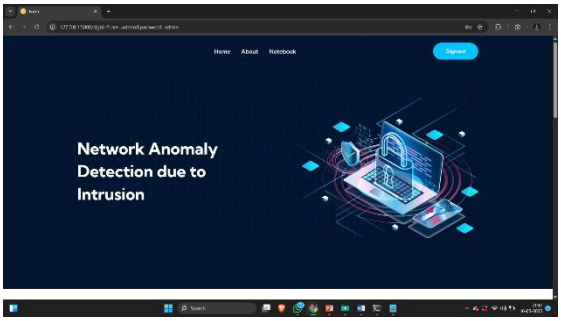

Fig.1 Home Screen

The main page of the project interface displays an orderly and professional design for the system called “Network Anomaly Detection due to Intrusion.” It shows a primary graphic depicting cybersecurity themes like secure data transfer and protection of systems. At the top, there is a navigation menu with links to Home, About, and Notebook pages, which provides access to the other parts of the application. There is a “Sign-out” button which is positioned at the upper right part of the screen which indicates an active session, probably an admin user. Such layout aids in the efficient management of network intrusion detection and monitoring, enhancing the overall user experience.

Fig.2 Attack Detection

This screen gives the output result interaction of the anomaly detection system. Upon completing the analysis, a definite alert message appears: "There is an Attack Detected! Type of Attack is DDoS" in bold red color. This means the model has effectively identified a Distributed Denial of Service (DDoS) attack. This clean layout with the option of Home, About, and Notebook ensures easy reading of results and further action by the user.

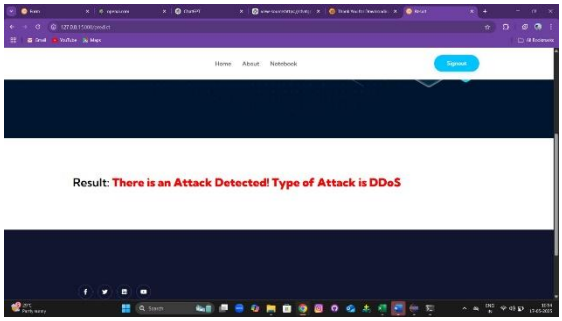

Fig.3 No Attack Found

This is a screen that presents the results of an anomaly detection system that processes network data as input. The system displays the message in bold red font: "There is an No Attack Detected, it is Normal!", which implies that no attack was found in the network activity and that the activity is safe. The layout is consistent as it would carry navigation tabs and a "Signout" button that would be featured prominently to make all things easy for the user.

6. CONCLUSION

The present times require the protection of networks more than ever, especially against threats to financial institutions. Thus, the ever-rising digital transformations in public life fast increase the variety of risks and attack strategies. Cyberspace offers many benefits and opportunities, although it also comes with challenges. To combat this complex structure, implementation of a highly modern security system that can fully distinguish among several stages of cyber intrusions is required. We have created an effective DDoS attack detection system by utilising hierarchical methods for machine learning, placing particular emphasis on Voting Classifier performance, which yields a promising accuracy rate of 100%. Using the CIC-IDS 2017 dataset, we present a multi-engine approach to malware classification based on advanced feature selection and data augmentation techniques to improve detection capabilities. This kind of comprehensive combine-and-conquer multi-brand approach shows that resilient and effective countermeasures against DDoS attacks can be worked out to secure the integrity and confidentiality of client information. Henceforth, our findings will contribute to increasing the efforts to strengthen network security parameters in the financial industry and beyond as there lies an urgent demand for trustworthy and efficient solutions in cyber-attack detection. We intend to improve our DDoS attack detection system by adding further ensemble machine learning models and deep learning architectures.

REFERENCES

[1] K. Alissa, T. Alyas, K. Zafar, Q. Abbas, N. Tabassum, and S. Sakib, ‘‘Botnet attack detection in IoT using machine learning,’’ Comput. Intell. Neurosci., vol. 2022, pp. 1–14, Oct. 2022.RESEARCH PAPER 2025.pdf

[2] C. Iwendi, Z. Jalil, A. R. Javed, T. G. Reddy, R. Kaluri, G. Srivastava, and O. Jo, ‘‘KeySplitWatermark: Zero watermarking algorithm for software protection against cyber-attacks,’’ IEEE Access, vol. 8, pp. 72650–72660, 2020.KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks | IEEE Journals & Magazine | IEEE Xplore

[3] N. Bakhareva, A. Shukhman, A. Matveev, P. Polezhaev, Y. Ushakov, and L. Legashev, ‘‘Attack detection in enterprise networks by machine learning methods,’’ in Proc. Int. Russian Autom. Conf. (RusAutoCon), Sep. 2019, pp. 1–6.https://www.researchgate.net/publication/336557310_Attack_Detection_in_Enterprise_Networks_by_Machine_Learning_Methods

[4] S. Siddiqui, S. Hameed, S. A. Shah, I. Ahmad, A. Aneiba, D. Draheim, and S. Dustdar, ‘‘Toward software-defined networking-based IoT frameworks: A systematic literature review, taxonomy, open challenges and prospects,’’ IEEE Access, vol. 10, pp. 70850–70901, 2022. https://www.researchgate.net/publication/353154816_Toward_Software-Defined_Networking-Based_IoT_Frameworks_A_Systematic_Literature_Review_Taxonomy_Open_Challenges_and_Prospects

[5] S. Agrawal, S. Sarkar, M. Alazab, P. K. R. Maddikunta, T. R. Gadekallu, and Q.-V. Pham, ‘‘Genetic CFL: Hyperparameter optimization in clustered federated learning,’’ Comput. Intell. Neurosci.,vol. 2021, pp. 1–10, Nov. 2021.https://onlinelibrary.wiley.com/doi/10.1155/2021/7156420

[6] M. S. Khan, N. M. Khan, A. Khan, F. Aadil, M. Tahir, and M. Sardaraz, ‘‘A low-complexity, energy-efficient data securing model for wireless sensor network based on linearly complex voice encryption mechanism of GSM technology,’’ Int. J. Distrib. Sensor Netw., vol. 17, no. 5, May 2021, Art. no. 155014772110186.https://journals.sagepub.com/doi/full/10.1177/15501477211018623