Leveraging Random Forest for Distributed Denial Of Service Attack Defense.

Prof. R. Papalkar
Vishwakarma University, Pune, Maharashtra, India
P. Ingle, A. Joshi, S. Shah, A. Sayyed
Vishwakarma University, Pune, Maharashtra, India
H. Mohite, P. Ghatkar, M. Nasikwala
Vishwakarma University, Pune, Maharashtra, India

ABSTRACT: This paper presents a holistic approach to defending against Distributed Denial of Service (DDoS) attacks in today's dynamic cybersecurity landscape. By integrating innovative techniques and robust security measures, organizations can effectively mitigate these threats. Our methodology involves simulating DDoS traffic using Internet of Things (IoT) devices, facilitating rigorous testing and analysis. Upon detecting anomalous traffic, the system swiftly initiates blacklisting protocols, mitigating the impact by temporarily blocking malicious IP addresses. Model Input Filtering ensures accurate analysis by focusing solely on legitimate traffic. Leveraging machine learning algorithms like random forest, our DoS Detection Model identifies subtle attack signatures, promptly signaling ongoing threats. Additionally, a fortified Secured Server Network responds swiftly to detected threats, safeguarding network data confidentiality. Through meticulous integration, organizations can enhance their vigilance against DDoS attacks, fortifying the resilience and reliability of their network infrastructure.

1 INTRODUCTION:

In the ever-evolving landscape of cybersecurity, Distributed Denial of Service (DDoS) attacks continue to pose significant threats to the stability and integrity of network infrastructures. To combat these threats effectively, organizations must adopt a proactive defense strategy that encompasses various stages of detection and prevention. This paper proposes a comprehensive approach to mitigating DDoS attacks, leveraging a combination of innovative techniques and robust security measures. Beginning with the orchestrated creation of simulated traffic through Internet of Things (IoT) devices, our approach mirrors the behavior of real-world DDoS assaults, allowing for rigorous testing and analysis. Upon detection of anomalous traffic patterns, the system swiftly initiates blacklisting protocols, temporarily blocking identified malicious IP addresses to minimize the attack's impact. Subsequent Model Input Filtering ensures that subsequent analysis focuses exclusively on legitimate traffic, optimizing the accuracy of the detection model. Leveraging machine learning algorithms, such as random forest, our DoS Detection Model discerns subtle attack signatures, generating binary output to promptly signal the presence of ongoing threats. Finally, a fortified Secured Server Network acts as a central nerve center, equipped with robust security measures, to swiftly respond to detected threats and safeguard the confidentiality of network data. Through the meticulous integration of these components, organizations can attain a heightened level of vigilance against DDoS attacks, enhancing the resilience and reliability of their network infrastructure against potential compromises.

2 OBJECTIVE

This research paper aims to develop a comprehensive framework for detecting and mitigating Distributed Denial of Service (DDoS) attacks, focusing on the integration of IoT devices and machine learning algorithms. Firstly, the study will simulate DDoS attacks using IoT devices to generate controlled floods of traffic, replicating real-world scenarios for testing purposes. Then, a blacklisting mechanism will be implemented to identify and block malicious IP addresses responsible for the flood, mitigating further impact on the target network. Following this, a filtering mechanism will ensure that only legitimate traffic is processed by the detection model, enhancing its accuracy by reducing noise. The core of the framework involves designing and training a detection model, potentially utilizing machine learning algorithms like Random Forest, to identify patterns indicative of DDoS attacks within the incoming traffic. This model will generate alerts upon detecting suspicious activity, enabling timely response and mitigation efforts. Finally, a secure server network will be integrated to receive and process alerts, equipped with robust security measures to ensure the integrity and confidentiality of transmitted data. Through these objectives, the research seeks to contribute to more effective methods for enhancing the resilience and security of web services against DDoS attacks.

3 METHODOLOGY

3.1 Dataset Used

Table 1. Comparative table for Bot-IoT dataset

Methods	Bot-IoT dataset					
	80 % of training data			K-fold 10		
SVM	60.760	61.538	60.494	70.868	71.728	71.286
SAE-LSTM	60.936	61.715	60.670	71.005	71.866	71.423
KNN	67.108	67.886	66.842	71.301	72.161	71.719
Light GBM	70.011	70.790	69.745	71.847	72.707	72.265
Deep NN	70.970	71.749	70.704	72.500	73.360	72.918
Cloud-based DL	71.290	72.069	71.024	83.588	84.448	84.006
Fog-based detection	71.610	72.388	71.343	84.793	85.653	85.211
TEHO-DBN	71.929	72.708	71.663	85.392	86.253	85.811
Multilayer perceptrons (MLP)	86.508	87.286	86.242	86.415	87.275	86.833

Deep CNN	87.112	87.891	86.846	87.017	87.877	87.435
CS-based Deep CNN	87.233	88.012	86.967	87.138	87.998	87.556

Table 2. Comparative table for UNSW-NB15 Dataset

Methods	UNSW-NB15 Dataset					
	80 % of training data			K-fold 10		
SVM	63.017	63.448	63.440	63.049	63.148	63.288
SAE-LSTM	63.407	63.838	63.830	63.431	63.530	63.669
KNN	63.537	63.968	63.960	63.558	63.657	63.797
Light GBM	63.667	64.098	64.090	63.778	63.877	64.017
Deep NN	63.797	64.228	64.220	63.910	64.009	64.149
Cloud-based DL	82.091	82.522	82.514	64.042	64.141	64.281
Fog-based detection	85.237	85.668	85.660	84.233	84.332	84.471
TEHO-DBN	86.434	86.865	86.857	85.869	85.968	86.108
Multilayer perceptrons (MLP)	86.673	87.104	87.096	86.106	86.205	86.345
Deep CNN	87.151	87.582	87.574	86.580	86.679	86.818
CS-based Deep CNN	87.630	88.061	88.053	87.053	87.152	87.292

In selecting a dataset for our research, we considered two options: the Bot IoT dataset and the Ton dataset. Upon examination of the Bot IoT dataset, we observed a significant class imbalance, with only 7 entries classified as normal and the remaining approximately 1.5 million entries categorized as DDoS attacks. Recognizing the importance of dataset balance to mitigate issues of overfitting, we opted not to utilize this dataset.

Turning our attention to the Ton dataset, we found a total of 7 million entries, with 7 lakh labeled as normal and the remainder as DDoS attacks. To ensure a balanced representation for training purposes, we randomly selected 7 lakh entries from both the normal and DDoS categories. Subsequently, we fine-tuned this balanced dataset to prepare it for our research objectives. This meticulous approach ensured that our dataset maintained a 50-50 distribution between normal and DDoS instances, thus enhancing the reliability and effectiveness of our model training process.

3.2 Models Used

Random Forest:

Methodology: Random Forest is an ensemble learning method that constructs a multitude of decision trees at training time. Each tree in the forest is trained on a random subset of the training data and makes predictions independently. The final prediction is determined by aggregating the predictions of all the trees (e.g., averaging for regression, voting for classification). Random Forests are known for their robustness against overfitting and their ability to handle large datasets with high dimensionality. However, they might not perform well on datasets with highly correlated features or noisy data.

Convolutional Neural Network (CNN):
Methodology: CNN is primarily used for image classification and recognition tasks. It consists of multiple layers of convolutional and pooling layers followed by fully connected layers. The convolutional layers apply filters to input data to extract features, while pooling layers reduce the spatial dimensions of the features. CNNs excel in tasks involving spatial data like images due to their ability to automatically learn hierarchical patterns.

Recurrent Neural Network (RNN):
Methodology: RNN is designed to handle sequential data by maintaining a state vector that captures information about previous inputs. This allows RNNs to model temporal dependencies in data. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are popular RNN variants that address the vanishing gradient problem. RNNs are suitable for sequential data like time series, natural language, and speech recognition tasks.

Support Vector Machine (SVM):
Methodology: SVM is a supervised learning algorithm used for classification and regression tasks. It works by finding the hyperplane that best separates different classes in the feature space. SVM can use various kernel functions to map input data into higher-dimensional spaces to find non-linear decision boundaries. SVMs are effective in high-dimensional spaces and are particularly well-suited for small to medium-sized datasets.

Simple Neural Network:
Methodology: A simple neural network consists of an input layer, one or more hidden layers, and an output layer. Each layer contains neurons that are interconnected and apply a nonlinear activation function to their inputs. These networks are trained using techniques like backpropagation and gradient descent. Simple neural networks are versatile and can be applied to a wide range of tasks.

Table 3. Comparative analysis of models used.

Sr	Models	Test Accuracy	Test Precision	Test Recall
1	SVM	96.32	96.32	96.32
2	SNN	99.84	99.84	99.84
3	RNN	99.97	99.97	99.97
4	CNN	99.84	99.84	99.84

5	RANDOM FOREST	99.99	99.99	99,99

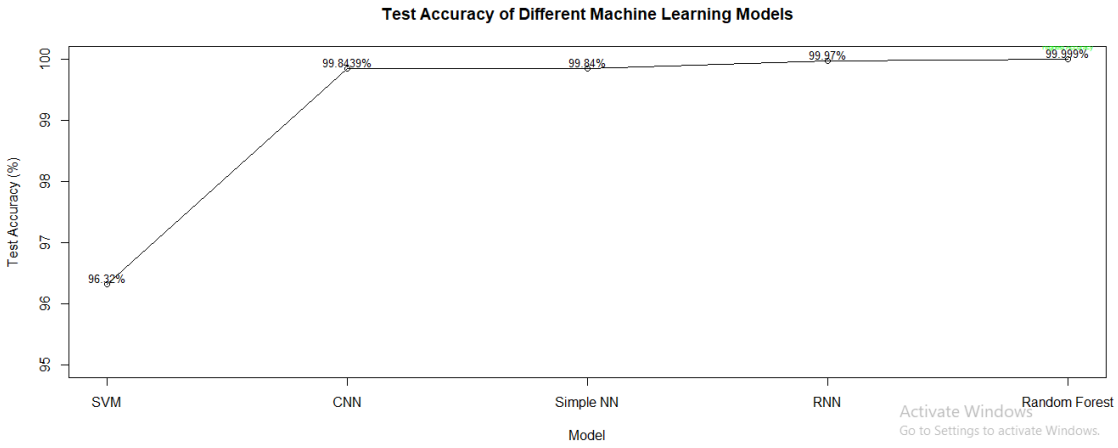

Fig 1.1 Accuracy

3.3 Proposed Model

Our proposed model for defending against Distributed Denial of Service (DDoS) attacks is structured to create a resilient network environment capable of effectively detecting and mitigating malicious traffic. To initiate this defense, IoT devices are employed to simulate traffic floods, replicating the behavior of real DDoS attacks and facilitating comprehensive testing scenarios. Upon detecting these simulated floods, an immediate blacklisting mechanism is triggered, swiftly identifying and blocking the originating IP addresses to proactively mitigate further disruption. Following blacklisting, a rigorous model input filtering process is implemented, meticulously sieving through incoming data to ensure subsequent analysis is focused solely on authentic traffic, thereby enhancing the precision of the detection model and reducing false positives.

Central to our defense strategy is the development of a specialized DoS detection model, potentially utilizing advanced machine learning algorithms like Random Forest. Trained on historical data, this model possesses the capability to discern subtle patterns indicative of DDoS attacks within incoming traffic. Once anomalies are detected, the model swiftly produces binary output, serving as an alert mechanism to promptly flag the presence of an ongoing attack. Complementing this, a fortified secure server network is established as the final layer of defense. Equipped with robust security measures including firewalls, intrusion detection systems, and encryption protocols, this network functions as a centralized hub for receiving output from the detection model and orchestrating rapid responses to detected threats. By seamlessly integrating these components, our proposed model aims to achieve a heightened level of accuracy in detecting DDoS attacks while maintaining resilience against potential breaches or compromises. Through proactive measures such as blacklisting, meticulous data filtering, and advanced machine learning-driven detection, our model endeavors to safeguard network integrity and ensure uninterrupted operation, even in the face of sophisticated cyber threats.

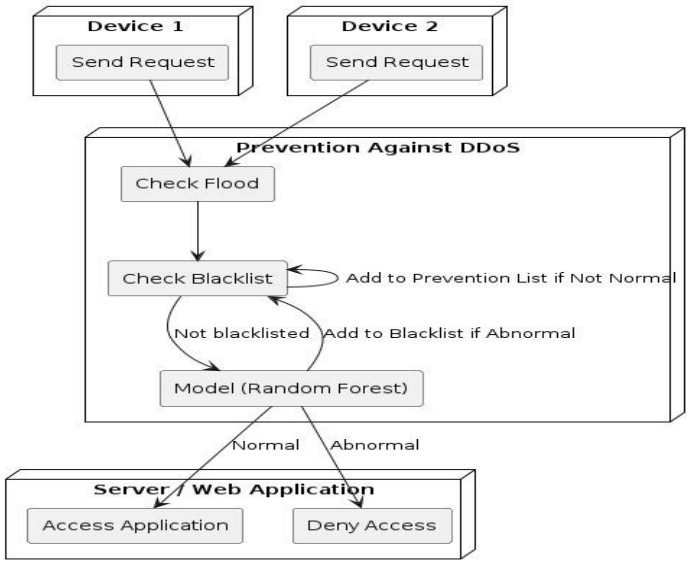

Figure 1.2 Model's Flow

flow:

1. Device 1/Device 2 Send Request:
The data flow begins with either Device 1 or Device 2 sending a request to the model.

2. Prevention Against DDoS:
The system initiates prevention measures against Distributed Denial of Service (DDoS) attacks by performing two checks:
Check Flood: The system examines if there's an abnormally high number of requests, indicating a potential flood attempt.
Check Blacklist: It verifies if the source IP address of the request is in a blacklist containing known malicious IP addresses.

3. Add to Prevention List if Not Normal:
If the request is abnormal (e.g., part of a potential flood or the source IP is in the blacklist), the system adds the source IP to a prevention list to block further requests from that IP.

4. Add to Blacklist if Abnormal:
If the source IP is not already in the blacklist and the request is abnormal, the system adds the source IP to the blacklist.

5. Model (Random Forest):
The system utilizes a Random Forest model to classify the request as either 'Normal' or 'Abnormal'. This model is trained on various features of past requests to discern the difference between normal and abnormal behavior.

6. Server Web Application Access Application:
Based on the model's classification, the system determines access to the server web application:
If the request is classified as 'Normal', access to the application is granted.
If the request is classified as 'Abnormal', access to the application is denied.

4 CONCLUSION

In conclusion, this paper presents a comprehensive framework for detecting and mitigating Distributed Denial of Service (DDoS) attacks, incorporating innovative techniques and robust security measures. By leveraging a combination of IoT devices, machine learning algorithms such as Random Forest, and secure server networks, the proposed model aims to enhance the resilience and reliability of network infrastructures against potential compromises. Through proactive measures such as blacklisting, rigorous data filtering, and advanced detection methods, the model endeavors to safeguard network integrity and ensure uninterrupted operation, even in the face of sophisticated cyber threats. The proposed model's effectiveness is demonstrated through comparative analysis, showcasing high accuracy and precision in detecting abnormal traffic patterns indicative of DDoS attacks. With its structured defense strategy and seamless integration of components, the proposed model offers organizations a heightened level of vigilance against DDoS attacks, contributing to a more secure cyber landscape.

5 REFRENCES

Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

Kumar, S., & Panigrahi, S. (2016). "Machine learning based detection and mitigation techniques for DDoS attacks: A survey." Journal of Network and Computer Applications, 66, 214-237.

https://research.unsw.edu.au/projects/toniot-datasets