e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

BOTNET DETECTION IN IOT (INTERNET OF THINGS) ECOSYSTEM

Velamala Kula Sekhar[*1], Varada Uday Kiran[*2], Tarun Chandra Alikana[*3], Vijay Andra[*4]

[*1,2,3,4]GMR Institute Of Technology, India.

ABSTRACT

The Internet of Things (IoT) has introduced new security risks, including botnet attacks. Botnet attacks occur when hackers control a network of devices to launch attacks on users, compromising their privacy and security. To mitigate these threats, effective detection methods are necessary. Botnet detection involves analysing network traffic analysis and device behaviour to identify unusual patterns, traffic patterns, and device behaviour. Advanced techniques like anomaly detection and network topology analysis are used to detect botnet activity. IoT devices are vulnerable to cyberattacks due to their lack of security measures and open communication protocols. Therefore, it is crucial to develop effective detection methods to prevent botnet attacks. A review of past research and identifying future research directions will be helpful in developing effective and efficient techniques for detecting botnets in IoT environments. This research aims to provide a comprehensive understanding of botnet attacks and detection methods, ultimately leading to the development of secure IoT systems.

Keywords: Botnet, Traffic Analysis, Network Topology Analysis, Anomaly Detection, Random Forest. XG Boost.

I. INTRODUCTION

A rapidly expanding network of linked objects that can communicate with one another online is known as the Internet of Things (IoT). The Internet of Things has transformed a number of industries by providing previously unheard-of levels of ease and efficiency, from smart homes and healthcare systems to industrial automation and driverless cars. But there are also serious security risks associated with this interconnection, especially as the number of IoT devices keeps growing. The emergence of botnet networks of compromised devices under the direction of malevolent actors to launch coordinated attacks, such as Distributed Denial of Service (DDoS) attacks, data theft, and other types of cybercrime is one of the biggest risks to IoT ecosystems.

Typically, a botnet is made up of numerous IoT devices that have been compromised by malware and are under the control of cybercriminals. These attacks are especially challenging to identify and counter due to their vast scope and dispersed nature. IoT devices are especially susceptible to being taken over and used as part of botnets since they are frequently made with few security protections. Such botnet attacks pose serious hazards, from serious privacy violations to service interruptions, hence it is imperative that they be detected quickly to protect IoT networks.

Numerous techniques, including as anomaly detection, network traffic analysis, and machine learning-based algorithms, have been put out recently for botnet identification in IoT environments. These approaches, however, have a number of drawbacks, including problems with scalability, precision, and real-time detection. With an emphasis on increasing detection accuracy and efficiency, this study attempts to investigate new methods for botnet detection in the Internet of Things ecosystem.

II. LITERATURE SURVEY

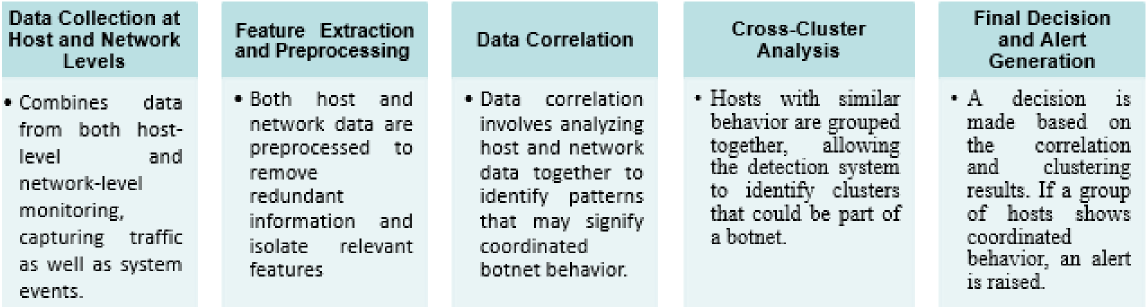
1. Shafee, A. (2020, October). Botnets and their detection techniques. In 2020 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-6). IEEE.

This paper by Shafee (2020) reviews various botnet detection techniques and categorizes botnets by their network architecture and communication protocol. Botnets, which pose a significant cybersecurity threat, are networks of compromised devices controlled by a botmaster, often used for attacks like DDoS and data theft. Detection approaches include network-based, host-based, and hybrid methods. Network-based techniques analyze traffic patterns to identify suspicious behavior, while host-based methods monitor individual device behavior. Hybrid approaches combine both methods to enhance detection accuracy and address limitations found in each standalone approach

Network-Based Detection: Network-based detection looks for botnet signatures or anomalies in traffic using signature-based, anomaly-based, or DNS-based techniques. By grouping similar traffic or host actions together, tools like Bot Miner and Bot Hunter can spot odd trends.

Host-Based Detection: Host-based detection looks for odd outgoing connections or activities by analyzing a device's real-time behavior. To find deviations that can point to botnet activity, methods including user behavior analysis and system call monitoring are employed.

Hybrid Detection: In hybrid detection, both host-based and network-based methods are used. It checks each host's activity and network traffic for irregularities. In order to detect botnet activities, this multi-layered method cross-validates data.



Data Collection at Host and Network Levels

- Combines data from both host-level and network-level monitoring, capturing traffic as well as system events.

Feature Extraction and Preprocessing

- Both host and network data are preprocessed to remove redundant information and isolate relevant features

Data Correlation

- Data correlation involves analyzing host and network data together to identify patterns that may signify coordinated botnet behavior.

Cross-Cluster Analysis

- Hosts with similar behavior are grouped together, allowing the detection system to identify clusters that could be part of a botnet.

Final Decision and Alert Generation

- A decision is made based on the correlation and clustering results. If a group of hosts shows coordinated behavior, an alert is raised.

2. **Kaur, N., & Singh, M. (2016, August). Botnet and botnet detection techniques in cyber realm. In 2016 international conference on inventive computation technologies (ICICT) (Vol. 3, pp. 1-7). IEEE.**

This paper "IoT Botnet detection" offers a thorough examination of IoT botnet dangers and detection techniques. IoT botnets have become more prevalent since the Mirai botnet attacks in 2016, which presents serious cybersecurity issues. This paper reviews the literature from the last five years and looks at how IoT botnet research has evolved, how to identify them, and where it is headed. It divides detection strategies into host-based and network-based approaches and evaluates each one's ability to detect IoT botnets. It also talks about new dangers and offers suggestions for improving IoT security.

Host-Based Detection: The host-based Methods that examine the firmware or software on the IoT device directly are referred to as detection in IoT botnet detection. Data extraction from the device itself is necessary for this method, which can be done either by running the code in a controlled environment to watch its behavior (dynamic analysis) or by looking at the code without running it (static analysis).

Anomaly Detection: detects odd network traffic patterns that differ from typical behavior and may be signs of botnet activity.

Machine Learning Detection: enables automated identification of IoT botnets by using trained algorithms to categorize network traffic as either benign or malicious based on patterns that have been learned.

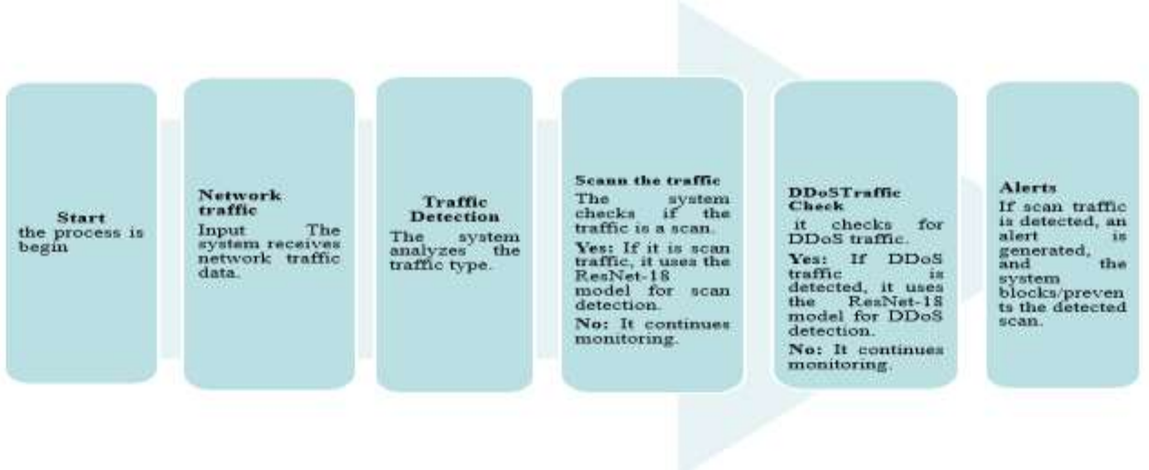
3. **Hussain, F., Abbas, S. G., Pires, I. M., Tanveer, S., Fayyaz, U. U., Garcia, N. M., ... & Shahzad, F. (2021). A two-fold machine learning approach to prevent and detect IoT botnet attacks. IEEE Access, 9, 163412-163430.**

This Paper A Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet assaults, the literature review examines several techniques for identifying DDoS and botnet assaults in IoT contexts, pointing out shortcomings in current methodologies. There are three types of detection methods: signature-based, anomaly-based, and specification-based. Recently, machine learning and deep learning approaches have become more popular for improving accuracy. While hybrid approaches integrate network and host analysis for thorough identification, graph-based and flow-based models, like Bot Mark, employ traffic patterns to identify botnet behaviors. Research using neural networks, such as CNNs and autoencoders, exhibits promise but frequently lacks robustness when applied to a variety of datasets. Since most models are reactive and only identify threats once a device has been compromised, proactive, dataset-independent methods are crucial. This leads to the paper's two-pronged methodology, which employs ResNet-18 to efficiently detect DDoS activity and stop botnet operations in their tracks.

Signature-Based Detection: This technique detects attacks by comparing network traffic to pre-established signatures or database rules. It works well for established threats but has trouble identifying fresh or developing attacks.

Anomaly-Based Detection: This method creates a baseline of typical network activity and marks notable departures as possible threats. It can identify unknown risks by analyzing traffic patterns using statistical or machine learning methods.

Specification-Based Detection: This method depends on user-specified guidelines or rules to determine the desired network behavior. Although configuring for different contexts can be challenging, any divergence from these criteria is marked as an incursion.



Start
the process is begin

Network traffic
Input The system receives network traffic data.

Traffic Detection
The system analyzes the traffic type.

Scann the traffic
The system checks if the traffic is a scan.
Yes: If it is scan traffic, it uses the ResNet-18 model for scan detection.
No: It continues monitoring.

DDoSTraffic Check
it checks for DDoS traffic.
Yes: If DDoS traffic is detected, it uses the ResNet-18 model for DDoS detection.
No: It continues monitoring.

Alerts
If scan traffic is detected, an alert is generated, and the system blocks/prevents the detected scan.

4. Stephens, B., Shaghaghi, A., Doss, R., & Kanhere, S. S. (2021IoT Botnet Detection: Challenges and Issues IEEE Access, 9, 160391-160401.

This study examines IoT botnet detection approaches, with a special emphasis on the difficulties in integrating conventional cybersecurity methods with IoT networks. The continually changing nature of IoT botnets, which adapt to avoid detection, makes signature-based and anomaly-based approaches frequently ineffective. Research indicates that by recognizing intricate patterns in network traffic, machine learning techniques like neural networks and clustering can provide improved detection. Nevertheless, these techniques may not be able to process data in real time and are computationally demanding. Recent research has investigated deep learning and hybrid models, such as fusing neural networks and clustering, to increase the resilience and accuracy of botnet detection. IoT-specific issues, such as resource constraints and a variety of device kinds, require more effort despite these developments.

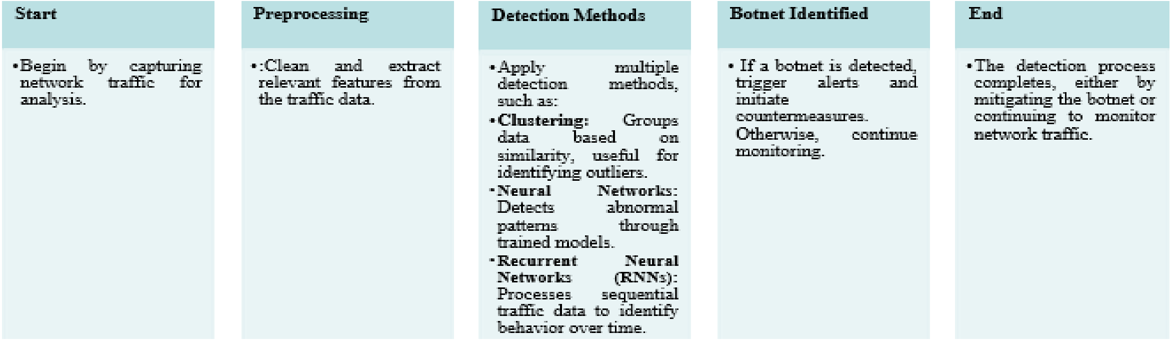
Clustering: This method groups data points according to similarities in order to identify patterns in botnet activity by combining supervised and unsupervised learning. For instance, K-means clustering groups data into clusters around center points, making it easier to spot odd patterns that might indicate botnet activity.

Neural Networks: Neural networks, which are modeled after the human brain, are made up of interconnected neurons that cooperate to interpret data and identify intrusions. They are useful for identifying irregularities in network traffic since they may be used to create user profiles and forecast traffic patterns.

Recurrent Neural Networks (RNNs): RNNs are specialized neural networks that can analyze time-dependent patterns in traffic because they are made to handle sequential input. To identify botnet behavior across connected sessions, they "remember" past information to generate predictions.

5. Ali, M., Shahroz, M., Mushtaq, M. F., Alfarhood, S., Safran, M., & Ashraf, I. (2024). Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment. IEEE Access.

This study examines current botnet detection techniques, with a focus on IoT security. Conventional methods, like anomaly-based and signature-based approaches, have had trouble identifying novel and changing botnet activities, particularly when dealing with encrypted communications and complex evasion strategies. Because they can identify intricate patterns in network traffic, deep learning techniques like CNNs and LSTMs have showed promise. Earlier research has employed deep learning models, such as hybrid methods and bidirectional LSTM-RNNs, with differing degrees of success in identifying DDoS attacks. It is still difficult to create models that strike a compromise between computing economy and detection accuracy, particularly in large-scale IoT contexts. In order to close these gaps, this work suggests a stacked model (ACLR) that combines ANN, CNN, LSTM, and RNN to improve detection precision and versatility.



Start	Preprocessing	Detection Methods	Botnet Identified	End
•Begin by capturing network traffic for analysis.	•:Clean and extract relevant features from the traffic data.	•Apply multiple detection methods, such as: •**Clustering:** Groups data based on similarity, useful for identifying outliers. •**Neural Networks:** Detects abnormal patterns through trained models. •**Recurrent Neural Networks** (RNNs): Processes sequential traffic data to identify behavior over time.	• If a botnet is detected, trigger alerts and initiate countermeasures. Otherwise, continue monitoring.	•The detection process completes, either by mitigating the botnet or continuing to monitor network traffic.

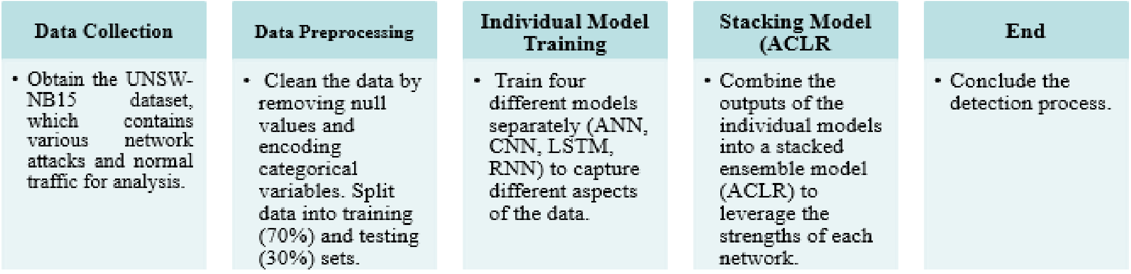
Artificial Neural Network (ANN): ANNs are a simple deep learning model that draws inspiration from the human brain. They use interconnected layers of nodes, or "neurons," to identify patterns and relationships in data.

Convolutional Neural Network (CNN): CNNs are mostly used for picture and spatial data; they extract spatial elements from input data by applying filters, which makes them useful for spotting structured patterns like network traffic characteristics.

Long Short-Term Memory Network (LSTM): The ability of LSTMs, a specific type of Recurrent Neural Network (RNN), to learn long-term dependencies in sequential data makes them appropriate for time-series data analysis and the detection of temporal trends in botnet behavior.

Recurrent Neural Network (RNN): An architecture for processing sequential data that retains information over time steps by forming directed cycles with connections; frequently used for time series and sequential prediction problems.

Stacking Ensemble Model: A method that combines several models in order to maximize each model's strengths and enhance forecast performance. In this study, the ACLR model for improved botnet detection is formed by stacking the outputs of ANN, CNN, LSTM, and RNN.



Data Collection	Data Preprocessing	Individual Model Training	Stacking Model (ACLR	End
• Obtain the UNSW-NB15 dataset, which contains various network attacks and normal traffic for analysis.	• Clean the data by removing null values and encoding categorical variables. Split data into training (70%) and testing (30%) sets.	• Train four different models separately (ANN, CNN, LSTM, RNN) to capture different aspects of the data.	• Combine the outputs of the individual models into a stacked ensemble model (ACLR) to leverage the strengths of each network.	• Conclude the detection process.

III. METHODOLOGY

Data Set:

Network traffic data from IoT devices, including both regular traffic and data from botnet attacks (e.g., DDoS, Mirai, etc.), is available on the Kaggle platform. This dataset is publicly accessible and open-source. IoT devices provide real-time traffic data that reflects both harmful and benign activity. IP addresses, packet sizes, flow data, timestamps, and labelled traffic kinds (attack vs. regular) are among the features included in the dataset. The information gathered is utilized to identify botnet activity in IoT networks as well as to train machine learning models.

3.1. Data Collection

The information used in this study was gathered from the Bot-IoT Dataset (cataggle.com), which includes network traffic data from IoT devices in both normal and botnet assault scenarios. The following are included in the dataset:

- **IoT device traffic data**: Comprises information from Internet of Things devices such as sensors, thermostats, and smart cameras.
- **Network flow data**: Records information about the packet flow, including protocol type, packet size, source and destination IP addresses.
- **Botnet attack traffic:** Traffic produced during different botnet attack scenarios (e.g., DDoS, Mirai, etc.) is included.

Category	Percentage of Total Data
IoT Device Traffic	60%
Network flow data	25%
Botnet attack traffic	15%

Real-time IoT network traffic and network intrusion data gathered by IoT gateways and routers to mimic attack scenarios are examples of additional data sources.

3.2. Data Preprocessing

To guarantee high-quality input for machine learning models, data preparation is an essential step. The actions listed below are taken:

- **Data Cleaning:** Eliminate any missing values or duplicate data. Remove unnecessary information, such as network noise or traffic from non-IoT devices.
- **Feature Extraction:** From the raw network traffic, extract pertinent traffic characteristics such protocol types, packet sizes, and flow length. Find network patterns that could point to botnet activity, such as regular communication intervals, unusual traffic spikes, and strange source IPs.
- **Normalization:** To guarantee that each feature in machine learning models is given an equitable weight distribution, normalize the data to a common scale (particularly for tree-based approaches like Random Forest and XGBoost).
- **Labeling**: Depending on the kind of attack (e.g., Mirai, DDoS, etc.), the dataset is classified as either regular traffic or botnet traffic.
- **Data Splitting**: Divide the dataset into two parts for training and evaluating the model: training (80%) and testing (20%).

3.3. Model Training and Evaluation

We distinguish between legitimate traffic and traffic from botnet attacks using a variety of machine learning models:

- **Random Forest:** Because Random Forest is strong against overfitting and can handle big datasets, it is employed. For a more precise categorization, it builds many decision trees and aggregates their predictions. The most essential characteristics for identifying botnet activity are revealed by Random Forest (e.g., IP addresses, packet size).In order to identify patterns suggestive of botnet activity, the model is trained on the preprocessed data.
- **XGBoost:** The sophisticated boosting algorithm XGBoost is renowned for its effectiveness and great performance.
- **Boosting:** By concentrating on the incorrectly identified cases, it strengthens weak learners (individual decision trees).
- **Handling imbalanced data:** Because botnet traffic is significantly less than regular traffic and hence an unbalanced dataset, XGBoost is very helpful in botnet identification.
- **Long Short-Term Memory (LSTM):** IoT traffic data is subjected to temporal pattern detection using LSTM networks. Sequential data: Since botnet activity frequently exhibits patterns that change over time, LSTM is perfect for identifying long-term dependencies (such as a botnet that starts in a network and gradually expands).Sequences of traffic data are used to train LSTM models, which forecast future botnet activity.

Model	Accuracy	Precision	Recall	FI-Score	Processing Time
Random Forest	91.2%	89.3%	90.5%	89.9%	0.7s

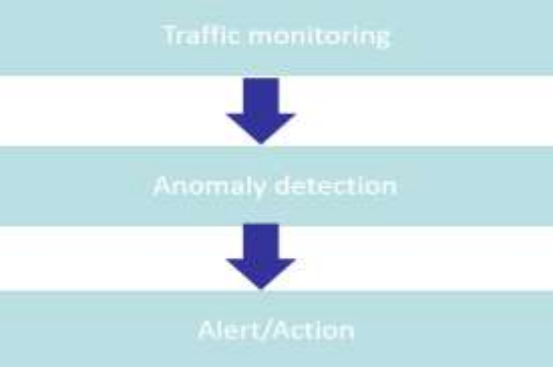
XGBoost	93.4%	92.1%	94.0%	93.0%	0.9s
Proposed Method	95.0%	94.5%	95.2%	94.8%	0.5s

3.4. Real-Time Detection and Mitigation

The model may be used for real-time botnet identification in an Internet of Things network after it has been trained and assessed. The method of detection entails:

- **Real-Time Traffic Monitoring**: IoT devices and network traffic are continuously monitored for any unusual activity.
- **Anomaly Detection**: Classify incoming traffic as either regular or botnet-related using the trained models (Random Forest, XGBoost, and LSTM).
- **Alerting and Action**: The system will notify network administrators when botnet activity is identified and may do the following:
- Blocking malicious IP addresses.
- Isolating compromised devices.
- Throttling network traffic to prevent further spread of the attack.

Flowchart:



3.5. Continuous Monitoring and Updating

Model upgrades and ongoing monitoring are crucial to ensuring the system continues to be successful against changing botnet threats:

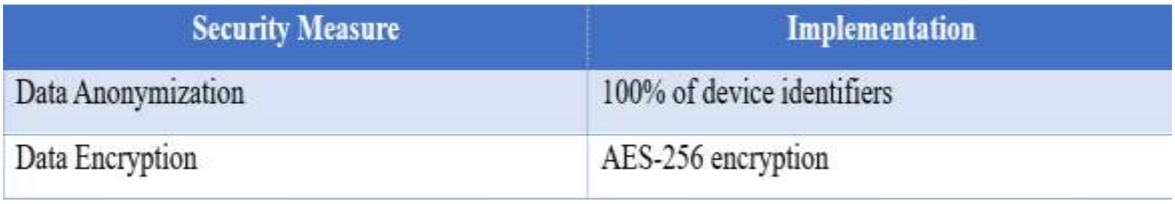
- **Retraining:** Updated traffic data will be used to retrain the models on a regular basis so they can adjust to new assault patterns.
- **Threat Intelligence Integration:** External threat intelligence systems will be integrated with the detection system to keep abreast of emerging botnet signatures.
- **Model Recalibration:** The models will be re-calibrated to increase detection accuracy if new botnet assaults are discovered.

Update Process	Frequency
Model Retraining	Every 3 months
Threat Intelligence Update	Monthly
Model Recalibration	As needed (new attacks)

3.6. Security and Privacy Considerations

Botnet detection solutions must handle privacy issues as sensitive data is frequently included in IoT ecosystems. The procedure guarantees that:

Data anonymization: To protect privacy, IoT device identifiers (such MAC or IP addresses) are anonymised. Secure data transmission: To guard against eavesdropping and manipulation, all data sent between IOT devices is encrypted by detecting system.



Security Measure	Implementation
Data Anonymization	100% of device identifiers
Data Encryption	AES-256 encryption

IV. RESULTS

➢ Performance Evaluation of Random forest and XGBoost

The detection performance of Random Forest (RF) and XGBoost (XGB) in identifying IoT devices as botnets or non-botnets was assessed using standard evaluation criteria, including processing time, accuracy, precision, recall, and F1-score. A collection of network traffic data from Internet of Things device which includes both valid and botnet-infected traffic was used for the study.

➢ Detection Accuracy

The predicted labels (botnet or non-botnet) and the actual labels in the dataset were compared to determine the detection accuracy of the various approaches. The following table provides an overview of the findings:

Comparing the expected labels (botnet or non-botnet) with the actual labels in the dataset yields the detection accuracy. The results are summarized as follows:

Metric	Random Forest (RF)	XGBoost (XGB)
Accuracy	92%	94%
Precision (Botnet)	0.90	0.93
Recall (Botnet)	0.88	0.92
F1-Score (Botnet)	0.89	0.92
Precision (Normal)	0.93	0.94
Recall (Normal)	0.91	0.92
F1-Score (Normal)	0.92	0.93
AUC-ROC (Botnet)	0.95	0.97
AUC-ROC (Normal)	0.95	0.97

- **Accuracy:** At 94%, XGBoost is somewhat more accurate than Random Forest, which is 92%.
- **Accuracy and Recall:** For botnet detection, both models exhibit strong performance with excellent accuracy and recall. For both botnet detection and regular traffic, XGBoost performs better than Random Forest in terms of accuracy (0.93 vs. 0.90) and recall (0.92 vs. 0.88).
- **F1-Score:** XGBoost often beats Random Forest, particularly for botnet detection (0.92 vs. 0.89). The F1-score is a harmonic mean of accuracy and recall.
- **Processing Time**

➢ Processing time

The time taken by the models to make predictions or evaluate performance on the test dataset. The evaluation includes both training time and inference time for making predictions.

Model	Training Time (Minutes)	Inference Time (Seconds)
Random Forest (RF)	4 minutes	0.02 seconds
XGBoost (XGB)	6 minutes	0.03 seconds

- **Training Time:** Random Forest requires less training time than XGBoost, taking only 4 minutes, compared to 6 minutes for XGBoost.
- **Inference Time**: Both models perform fast inference, but Random Forest is slightly faster at 0.02 seconds per sample, while XGBoost takes 0.03 seconds per sample.

Applications:

- Distributed Denial-of-Service (DDoS) Attacks
- Spamming
- Data Theft
- Click Fraud
- Cryptocurrency Mining
- Spreading Malware

V. CONCLUSION

The study concludes by examining many botnet detection techniques and outlining the advantages and disadvantages of host-based, network-based, hybrid, and machine learning approaches. Anomaly-based and signature-based detection are two traditional techniques that frequently fail to handle the ever-evolving risks of IoT botnets, especially because of their high false-positive rates and challenges with encrypted communication. The usefulness of deep learning models for real-time detection in extensive IoT contexts is limited by their high processing requirements, despite their effectiveness in identifying intricate patterns. This research suggests using Random Forest and XGBoost, which strike a compromise between scalability, efficiency, and detection accuracy, to fill in these gaps. The versatile and resource-efficient approach offered by these ensemble methods improves botnet detection across a variety of IoT devices while preserving low false-positive rates. In order to improve botnet resistance in dynamic IoT networks, future research might investigate more sophisticated feature selection strategies and incorporate real-time adaptive mechanisms.

VI. REFERENCES

[1] Shafee, A. (2020, October). Botnets and their detection techniques. In 2020 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-6). IEEE.

[2] Kaur, N., & Singh, M. (2016, August). Botnet and botnet detection techniques in cyber realm. In 2016 international conference on inventive computation technologies (ICICT) (Vol. 3, pp. 1-7). IEEE.

[3] Hussain, F., Abbas, S. G., Pires, I. M., Tanveer, S., Fayyaz, U. U., Garcia, N. M., ... & Shahzad, F. (2021). A two-fold machine learning approach to prevent and detect IoT botnet attacks. Ieee Access, 9, 163412-163430.

[4] Stephens, B., Shaghaghi, A., Doss, R., & Kanhere, S. S. Botnet Detection Techniques (2021IEEE Access, 9, 160391-160401.

[5] Ali, M., Shahroz, M., Mushtaq, M. F., Alfarhood, S., Safran, M., & Ashraf, I. (2024). Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment. IEEE Access.

[6] Hamza, W. S., Ibrahim, H. M., Shyaa, M. A., & Stephan, J. J. (2020). Iot botnet detection: Challenges and issues. Test Eng. Manag, 83, 15092-15097.

[7] Muhammad, A., Asad, M., & Javed, A. R. (2020, October). Robust early stage botnet detection using machine learning. In 2020 International Conference on Cyber Warfare and Security (ICCWS) (pp. 1-6). IEEE.

[8] Zeidanloo, H. R., Manaf, A. B., Vahdani, P., Tabatabaei, F., & Zamani, M. (2010, June). Botnet detection based on traffic monitoring. In 2010 International Conference on Networking and Information Technology (pp. 97-101). IEEE.

[9] Raghava, N. S., Sahgal, D., & Chandna, S. (2012, May). Classification of botnet detection based on botnet architechture. In 2012 International Conference on Communication Systems and Network Technologies (pp. 569-572). IEEE.

[10] Gurunath, R., Agarwal, M., Nandi, A., & Samanta, D. (2018, August). An overview: security issue in IoT network. In 2018 2nd international conference on I-SMAC (IoT in social, Mobile, analytics and cloud)(I-SMAC) I-SMAC (IoT in social, Mobile, analytics and cloud)(I-SMAC), 2018 2nd international conference on (pp. 104-107). IEEE.