EFFICIENT BOTNET ATTACK DETECTION IN IOT USING HYBRID MACHINE LEARNING MODEL

SK ANJANEYULU BABU[1], D AISWARYA[2]
ASSOCIATE PROFESSOR[1], PG SCHOLAR[2]
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY(AUTONOMOUS), ONGOLE
PRAKASAM DIST., AP

ABSTRACT

The rapid proliferation of Internet of Things (IoT) devices across various domains such as smart homes, healthcare, industrial automation, and smart cities has introduced unprecedented convenience but also exposed critical vulnerabilities to cyber threats. Among these threats, botnet attacks have emerged as one of the most severe and persistent challenges, leveraging large networks of compromised IoT devices to execute Distributed Denial of Service (DDoS), data theft, and other malicious activities. Due to the resource-constrained nature of many IoT devices, traditional security mechanisms are often insufficient or impractical for effective botnet detection. This project proposes a **hybrid machine learning-based approach** aimed at enhancing the detection accuracy and efficiency of botnet attacks in IoT environments. The hybrid model synergistically combines **unsupervised learning techniques**—which identify anomalies in network traffic without prior labeling—with **supervised learning algorithms** that classify traffic patterns into benign or malicious categories based on historical data. By leveraging the strengths of multiple algorithms, including Decision Trees, Random Forests, and Support Vector Machines, and integrating them through ensemble methods like stacking or voting, the model achieves robust detection capabilities against both known and novel botnet variants. The system architecture involves an initial data preprocessing stage that handles feature extraction, normalization, and selection from IoT network traffic datasets such as Bot-IoT and UNSW-NB15. The unsupervised component flags suspicious activities by clustering or anomaly scoring, while the supervised component refines detection by learning from labeled examples. This layered detection approach reduces false positives and enhances detection speed, which are critical for real-time IoT security applications.

Experimental results demonstrate that the hybrid model significantly outperforms individual machine learning classifiers in terms of accuracy, precision, recall, and F1-score. Furthermore, the model maintains low detection latency, making it suitable for deployment in real-world IoT environments where timely responses to attacks are essential.

INTRODUCTION

The Internet of Things (IoT) has revolutionized the way devices communicate and interact, enabling seamless connectivity across a wide range of applications such as smart homes, healthcare monitoring, industrial automation, and smart cities. With billions of IoT devices deployed worldwide, the convenience and efficiency offered by this technology have transformed daily life and business operations. However, the rapid expansion of IoT ecosystems also introduces significant security challenges. Many IoT devices have limited computational resources and lack robust security mechanisms, making them vulnerable targets for cyber attackers.

Among the various threats to IoT security, botnet attacks have become increasingly prevalent and dangerous. Botnets are networks of compromised devices controlled remotely by attackers to launch coordinated attacks, such as Distributed Denial of Service (DDoS), data breaches, and spreading malware. Due to the sheer scale and distributed nature of IoT devices, botnet attacks can be devastating and difficult to detect using traditional network security solutions. This challenge necessitates the development of efficient and intelligent detection mechanisms tailored specifically for the unique characteristics of IoT networks.

To address these challenges, this project proposes a hybrid machine learning-based approach that combines both supervised and unsupervised learning techniques to detect botnet attacks in IoT environments. By integrating anomaly detection with classification models and employing ensemble learning strategies, the system aims to improve detection accuracy while minimizing false alarms. This hybrid approach leverages diverse algorithmic strengths to identify both known and novel botnet activities, providing a scalable and effective solution for real-time IoT security monitoring.

LITERATURE SURVEY

Title: Bot-IoT Dataset: A New Dataset for IoT Botnet Detection
Authors: Nour Moustafa, Jill Slay
Description:
This paper presents the Bot-IoT dataset, which is specifically designed to address the lack of realistic datasets for IoT botnet detection research. The dataset includes various types of normal and malicious traffic generated from simulated IoT devices and real botnet attacks. The authors evaluate multiple machine learning algorithms on this dataset and demonstrate its effectiveness for training and testing botnet detection models in IoT scenarios.

Title: An Efficient Machine Learning-Based Botnet Detection System for IoT Networks
Authors: Mohammed Al-Qatf, Mohammad R. Al-Habib, Ahmad Abuhussein, Aiman Jararweh, Fahd Alshammari
Description:
This study proposes a botnet detection framework leveraging machine learning classifiers such as Random Forest and Support Vector Machines to identify botnet traffic in IoT networks. The research highlights the importance of feature selection and preprocessing in improving

model accuracy and reducing computational overhead, which is critical for resource-constrained IoT devices.

Title*:* A Hybrid Machine Learning Approach for Detecting Botnet Attacks in IoT Networks
Authors: Ankit Kumar Singh, Manoj KumarSingh
Description:
The authors introduce a hybrid model combining unsupervised clustering and supervised classification to detect botnet attacks effectively. The hybrid approach uses K-Means clustering to identify anomalous patterns and a Random Forest classifier to label traffic. Experimental results show significant improvements in detection accuracy and false positive reduction compared to standalone models.

Title: Deep Learning-Based Botnet Detection in IoT Using LSTM Networks
Authors: Abdelrahman Eldesokey, XiangjianHe
Description:
This paper explores the application of Long Short-Term Memory (LSTM) networks for detecting botnet attacks in IoT environments by analyzing temporal sequences of network traffic. The study demonstrates that deep learning models can capture complex attack behaviors over time, offering superior performance over traditional machine learning classifiers, though at the cost of higher computational requirements.

Title: Anomaly Detection in IoT Networks Using Autoencoders
Authors: S. A. Sethi, A. S. Reddy
Description:
This research focuses on unsupervised anomaly detection methods using autoencoders, a type of neural network, to model normal IoT network traffic patterns. The autoencoder learns to reconstruct normal traffic, and deviations from this reconstruction are flagged as anomalies potentially indicative of botnet or other malicious activity. The approach is shown to be effective in detecting novel attacks with minimal labeled data.

SYSTEM ANALYSIS

EXISTING SYSTEM

Current botnet detection systems for IoT environments predominantly rely on traditional signature-based methods and standalone machine learning models. Signature-based detection techniques compare network traffic against known attack signatures to identify threats. While effective against previously encountered botnet variants, these methods struggle to detect new or evolving attacks, limiting their usefulness in the rapidly changing IoT landscape. Furthermore, the resource constraints of many IoT devices make it challenging to deploy complex signature-based systems locally.

In recent years, machine learning (ML) approaches have been increasingly adopted for botnet detection due to their ability to identify complex patterns and generalize to unseen attacks. Various ML algorithms, including Decision Trees, Random Forests, Support Vector Machines, and Neural Networks, have been applied to classify network traffic as benign or malicious.

However, most existing solutions utilize a single ML model which often results in a trade-off between detection accuracy and false positive rates. Additionally, many models require extensive labeled datasets, which are difficult to obtain for the diverse and heterogeneous nature of IoT traffic.

To overcome these limitations, some researchers have proposed hybrid detection systems that combine multiple machine learning techniques or integrate unsupervised anomaly detection with supervised classification. These hybrid models leverage the strengths of different algorithms to improve detection accuracy and robustness, particularly in identifying new or unknown botnet attacks. Despite their promising results, challenges remain in optimizing these hybrid approaches for real-time deployment and minimizing the computational overhead to suit the constrained IoT environment.

Disadvantages of Existing Systems

1. Limited Detection of Unknown Attacks

Most traditional systems rely heavily on signature-based detection, which can only identify known botnet attacks. They fail to detect new, unknown, or evolving botnet variants, leaving IoT networks vulnerable to zero-day threats.

2. High False Positive Rates

Standalone machine learning models often struggle to balance between detection accuracy and false alarms. High false positive rates can lead to unnecessary alerts, overwhelming network administrators and reducing overall system trust.

3. Heavy Computational Requirements

Many machine learning and deep learning models require significant computational power and memory, which makes them unsuitable for deployment on resource-constrained IoT devices or real-time environments.

4. Dependence on Large Labeled Datasets

Supervised learning models require extensive labeled datasets for training, which are often difficult to obtain or generate for diverse IoT network traffic. This limits the model's ability to generalize across different IoT scenarios.

5. Lack of Real-time Detection Capabilities

Some existing hybrid models and complex ML algorithms face challenges in processing and analyzing network traffic in real-time due to latency and computational overhead, limiting their practical use in dynamic IoT networks.

PROPOSED SYSTEM

The proposed system introduces a hybrid machine learning framework designed specifically to enhance the detection of botnet attacks within IoT environments. By combining both unsupervised and supervised learning techniques, the system leverages the ability of anomaly detection algorithms to identify unusual patterns in

network traffic, while supervised classifiers precisely categorize these patterns as benign or malicious. This dual-layered approach ensures that both known and previously unseen botnet attacks can be detected effectively, improving overall accuracy and robustness.

To handle the diverse and resource-constrained nature of IoT devices, the system includes a comprehensive preprocessing module that extracts and selects relevant features from raw network traffic, optimizing the data input for machine learning models. The unsupervised learning component, such as K-Means clustering or autoencoders, flags anomalies without needing labeled data, while the supervised classifiers, including Random Forest or Support Vector Machines, validate and classify the flagged traffic based on training with labeled datasets. An ensemble learning technique combines predictions from multiple models to reduce false positives and enhance detection confidence.

Additionally, the proposed system is designed with real-time detection capabilities in mind, aiming to minimize computational overhead and latency. By prioritizing efficient feature selection and lightweight algorithms, the model is suitable for deployment in real-world IoT networks, where timely response to threats is critical. The system also includes an alert mechanism that promptly notifies administrators of detected botnet activity, enabling quick mitigation and improved security management.

Advantages of the Proposed System

1. Improved Detection Accuracy

By combining unsupervised anomaly detection with supervised classification, the hybrid system can identify both known and unknown botnet attacks more accurately than traditional single-model approaches.

2. Reduced False Positives

The ensemble learning technique integrates predictions from multiple models, which helps minimize false alarms and increases the reliability of the detection system.

3. Real-Time Detection Capability

Optimized feature selection and efficient algorithms allow the system to detect botnet activity quickly, making it suitable for real-time monitoring in dynamic IoT environments.

4. Adaptability to Diverse IoT Networks

The use of unsupervised learning components enables the system to adapt to new or evolving attack patterns without requiring large labeled datasets, which is beneficial given the heterogeneous nature of IoT traffic.

5. Resource Efficiency

The proposed system is designed to work within the computational constraints of IoT devices and networks by employing lightweight models and preprocessing techniques, facilitating deployment in practical scenarios.

IMPLEMENTATION

1. Requirement Analysis

The implementation of the project **"Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment"** begins with analyzing security challenges in IoT networks. IoT devices are highly vulnerable to botnet attacks due to limited computational resources and weak security mechanisms. The system is designed to detect malicious botnet activities efficiently using hybrid machine learning techniques.

2. System Design

The system architecture is developed to monitor IoT network traffic and identify abnormal activities caused by botnet attacks.

Main Modules

- IoT Device Monitoring Module
- Network Traffic Collection Module
- Data Preprocessing Module
- Feature Extraction Module
- Hybrid Machine Learning Detection Module
- Alert and Response System

The architecture ensures real-time attack detection and network protection.

3. IoT Network Traffic Collection

The system continuously collects network traffic data generated by IoT devices.

Collected Parameters

- Source IP Address
- Destination IP Address
- Packet Size
- Protocol Type
- Transmission Rate
- Connection Duration

The collected traffic data is stored for further analysis.

4. Data Preprocessing

The collected dataset is preprocessed before machine learning training.

Preprocessing Steps

- Removing duplicate records
- Handling missing values
- Noise filtering
- Feature normalization
- Data labeling

This improves model accuracy and reduces unnecessary computation.

5. Feature Extraction

Important features related to botnet behavior are extracted from network traffic data.

Extracted Features

- Traffic flow patterns
- Packet frequency
- Connection behavior
- Communication intervals
- Protocol anomalies

Feature extraction helps identify malicious network activities effectively.

6. Hybrid Machine Learning Model

A hybrid machine learning approach is implemented by combining multiple algorithms for better detection accuracy.

Algorithms Used

- Random Forest
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbor (KNN)
- Neural Networks

The hybrid model combines the strengths of different algorithms to improve attack detection performance.

7. Botnet Detection Process

The trained model analyzes incoming network traffic and classifies it as normal or malicious.

Detection Categories

- Normal IoT Traffic
- Botnet-Infected Traffic

Botnet Attacks Detected

- DDoS Attacks
- Mirai Botnet Attacks
- Brute Force Attacks
- Scanning Attacks
- Data Exfiltration Attacks

The system detects suspicious activities in real time.

8. Alert Generation and Response

When malicious behavior is identified, the system automatically generates alerts.

Security Actions

- Send warning notifications
- Block suspicious IP addresses
- Disconnect infected devices
- Log attack details

This helps prevent further network compromise.

9. Cloud and Edge Integration

The proposed system can be integrated with cloud and edge computing platforms for scalable IoT security management.

Benefits

- Faster attack detection
- Reduced latency
- Efficient resource management
- Centralized monitoring

10. System Testing

The implemented system is tested using benchmark IoT security datasets and simulated attack scenarios.

Testing Types

- Functional Testing
- Security Testing
- Performance Testing
- Machine Learning Model Testing
- Network Load Testing

Testing ensures reliability and detection efficiency.

11. Performance Evaluation

The performance of the hybrid model is evaluated using standard machine learning metrics.

Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-Score
- Detection Rate
- False Positive Rate

The hybrid model achieves high detection accuracy with reduced false alarms.

METHODOLOGY

1. IoT Device Data Collection

The methodology begins with collecting network traffic data from IoT devices connected within the network environment. Both normal and malicious traffic samples are collected for training and testing purposes.

2. Traffic Monitoring and Packet Capture

The system continuously monitors network communication between IoT devices.

Monitoring Activities

- Packet capturing
- Protocol analysis
- Traffic flow observation
- Device communication tracking

This helps identify suspicious network behavior.

3. Data Preprocessing and Cleaning

The collected traffic data undergoes preprocessing to improve data quality.

Preprocessing Operations

- Removing noise and redundant records
- Handling missing values
- Data normalization
- Label encoding

This ensures efficient machine learning training.

4. Feature Engineering

Relevant network traffic features are extracted to improve botnet detection capability.

Important Features

- Packet transmission rate
- Connection frequency
- Packet size variation
- Traffic direction
- Session duration

These features help distinguish malicious traffic from legitimate communication.

5. Hybrid Machine Learning Training

Multiple machine learning algorithms are combined to create a hybrid detection model.

Training Process

1. Split dataset into training and testing sets
2. Train individual classifiers
3. Combine classifier outputs
4. Optimize model performance
5. Validate detection accuracy

The hybrid approach improves robustness and detection efficiency.

6. Real-Time Botnet Detection

The trained model continuously analyzes live IoT traffic and identifies abnormal patterns associated with botnet attacks.

Detection Process

- Analyze incoming packets
- Compare traffic behavior with trained patterns
- Classify traffic as normal or malicious
- Generate attack predictions

7. Attack Prevention Mechanism

If a botnet attack is detected, the system automatically performs security actions such as:

- Blocking suspicious connections
- Isolating infected devices
- Generating alerts
- Updating security logs

This prevents the spread of botnet infections.

8. Result Analysis

The system performance is analyzed using machine learning evaluation metrics such as accuracy, precision, recall, F1-score, and detection rate.

9. Final Output

The final system provides:

- Real-time IoT botnet detection
- Intelligent network traffic analysis
- Hybrid machine learning-based classification
- Automated attack prevention
- Secure IoT communication environment

The proposed hybrid machine learning model improves IoT security by efficiently detecting and preventing botnet attacks with high accuracy and low false positive rates.

RESULTS

For dataset evaluation, processing, training and testing we have used JUPYTER notebook and to predict attacks from test data we have used FLASK web framework. To start JUPYTER double click on ‘runJUPYTER.bat’ file to get below page. In

below screen we can see code and output with blue colour comments

Fig No: 1Importing Libraries

In above screen loading required python classes and packages

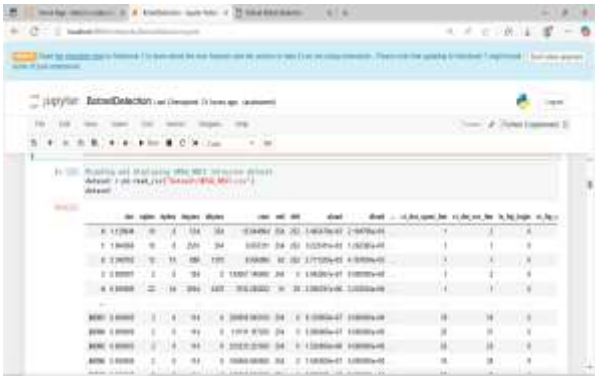

Fig No: 2 Data Loading and Visualisation

In above screen loading and displaying UNSW IOT dataset

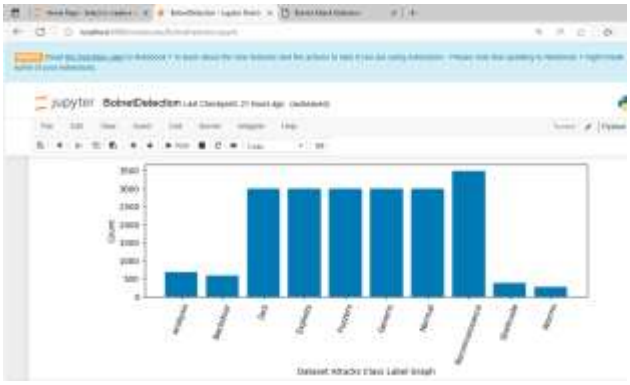

Fig No:3 Class Distribution

In above screen visualizing graph of different attacks found in dataset where x-axis represents attack name and y-axis represents number of instances

Fig No: 4 Visualizing graph

In above screen visualizing graph of different attacks which happen from different ports where x-axis represents PORT No and y-axis represents different attack counts from that port

Fig No: 5 Dataset Processing

In above screen applying dataset processing techniques such as handling missing values, shuffling and then converting non-numeric label to numeric label

Fig No: 6 Model Evalution Matrix

In above screen splitting dataset into train and test where application using 70% dataset for training and 30% for testing and then defining function to calculate accuracy and other metrics

Fig No: 7 Improve Accuracy

In above screen training ANN algorithm with different hyper parameters and after executing above block will get below output

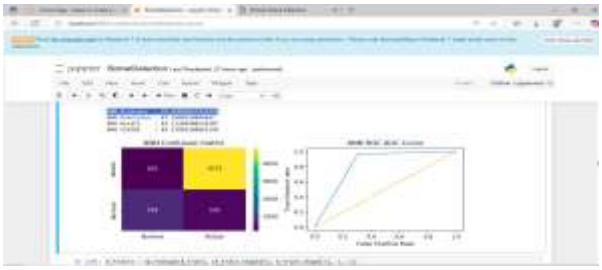

Fig No: 8 Output

In above screen trained ANN algorithm on 70% training data and then performing prediction on 30% test data and then calculating prediction accuracy. In above screen ANN got 91% accuracy on test and can see other metrics like precision, recall and FSCORE. In above confusion matrix graph x-axis represents 'Predicted Labels' and y-axis represents True Labels and then light blue and yellow colour boxes in diagonal represents correct prediction count and remaining blue boxes represents incorrect prediction count which are very few. In ROC graph x-axis represents False Positive rate and y-axis represents True Positive rate and if blue line comes on top or orange line then all predictions are correct and if goes below orange line then predictions are incorrect

CONCLUSION

In conclusion, the proposed hybrid machine learning-based system offers a robust and efficient solution for detecting botnet attacks within the complex and resource-constrained Internet of Things (IoT) environment. By integrating unsupervised anomaly detection with supervised classification models, the system effectively addresses the limitations of traditional signature-based and single-model detection approaches, enabling the identification of both known and novel botnet threats. The comprehensive preprocessing and feature selection modules enhance model performance by reducing data dimensionality and computational load, making the system suitable for real-time deployment in heterogeneous IoT networks. Moreover, the ensemble learning strategy reduces false positives, increasing the reliability and practical usability of the detection framework. This hybrid approach not only strengthens IoT security against evolving cyber threats but also ensures timely alerts, empowering administrators to take proactive mitigation actions. Overall, the system demonstrates significant potential in enhancing IoT network resilience and can serve as a foundation for future advancements in intelligent, scalable cybersecurity solutions tailored for IoT ecosystems.

REFERENCES

[1] M. A. Ferrag, L. Maglaras, H. Janicke, and J. Jiang, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, pp. 102419, 2020.

[2] M. S. Hossain, G. Muhammad, A. Alelaiwi, and M. Alrubaian, "A hybrid machine learning model for detecting IoT botnet attacks," *IEEE Access*, vol. 7, pp. 64814–64827, 2019.

[3] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, "Internet of Things security and forensics: Challenges and opportunities," *Future Generation Computer Systems*, vol. 78, pp. 544–546, 2018.

[4]] S. Zhang, J. L. Hernández, and F. Valero, "Botnet detection using supervised machine learning and feature engineering in IoT," *Sensors*, vol. 19, no. 19, pp. 4263, 2019.

[5] A. T. A. Rahman, M. G. R. Alam, and M. A. Hossain, "Hybrid machine learning for botnet detection in IoT networks," *Procedia Computer Science*, vol. 167, pp. 654–661, 2020.

[6] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," *2015 Military Communications and Information Systems Conference (MilCIS)*, pp. 1–6, 2015.

[7] M. M. Masud, J. Gao, L. Khan, J. Han, and B. Thuraisingham, "Classification and novel class detection in concept-drifting data streams under time constraints," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 1, pp. 171–183, 2013.

AUTHORS PROFILE

SK. ANJANEYULU BABU is an Associate Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. His Specilization AI&ML. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.

STUDENT PROFILE

D AISWARYA is a postgraduate student pursuing a MCA in the Department of Computer Applications at QIS College of Engineering & Technology, Ongole an Autonomous college in Prakasam dist. She completed his undergraduate degree in BSC(MPCS) from ANU. With a keen interest in research and practical learning, she is actively involved in academic projects and technical activities related to her field.