Hybrid Machine Learning Framework For Accurate Botnet Attack Detection In Iot Networks

[1]B. Prashant,[2]Manyam Pavan Sai,[3]Gurijala Venkata Durga Prasad,[4]Jetti Poojitha,
[5]Sanagala Dinesh Sri Raja Ratnam Naidu
[1] Associate Professor & Head of the Department, Department of CSE-DATA SCIENCE,
Eluru College of Engineering and Technology
[2,3,4,5]B. Tech Student, Department of CSE-DATA SCIENCE, Eluru College of Engineering and Technology

ABSTRACT

The widespread adoption of Internet of Things (IoT) devices has introduced significant security challenges, particularly the rise of botnet attacks that exploit vulnerable interconnected systems to launch large-scale distributed cyber threats. This study presents a hybrid machine learning–based framework for effective detection of botnet activities in IoT networks. The proposed approach integrates multiple classification algorithms to leverage their complementary strengths while incorporating feature selection, ensemble learning, and anomaly detection techniques to improve detection capability and reduce false positive rates. Experimental evaluation performed on benchmark IoT botnet datasets demonstrates that the proposed hybrid model achieves a detection accuracy of 98.7%, along with improved precision, recall, and overall robustness compared with individual machine learning classifiers. The framework provides a scalable, reliable, and efficient solution for enhancing security in modern IoT environments.

Keywords: Internet of Things (IoT), Botnet Attack Detection, Hybrid Machine Learning, Cybersecurity, Network Intrusion Detection, Anomaly Detection, Deep Learning, IoT Security, Network Traffic Analysis, Artificial Intelligence.

I. INTRODUCTION

The Internet of Things (IoT) ecosystem has expanded rapidly, connecting billions of devices across healthcare, smart homes, transportation, and industrial systems. However, the limited computational capacity, weak authentication mechanisms, and lack of security updates in IoT devices make them highly vulnerable to cyberattacks. Among these, botnet attacks have become particularly destructive, allowing attackers to hijack thousands of devices and orchestrate large-scale attacks.

Traditional intrusion detection systems often struggle with the high dimensionality of IoT data, evolving attack patterns, and imbalanced datasets. Machine learning (ML) provides a promising solution by learning patterns of malicious traffic and detecting anomalies. Yet, single ML models often suffer from overfitting, low generalization, or poor adaptability to new threats.

To address these challenges, this work proposes a hybrid machine learning model that integrates multiple classifiers with feature selection and anomaly detection. The proposed system aims to provide higher detection accuracy, reduce false positives, and ensure scalability in real-world IoT deployments.

II. LITERATURE SURVEY

1. Machine Learning Approaches for IoT Intrusion Detection

Abstract:
Moustafa & Slay (2015) introduced the UNSW-NB15 dataset for evaluating intrusion detection systems. Using classical ML models such as Decision Trees and Support Vector Machines, the study achieved promising accuracy in detecting

intrusions. However, the models faced challenges with imbalanced data, leading to increased false negatives when classifying rare botnet attacks.

2. Bot-IoT Dataset for Network Forensics in IoT

Abstract:
Koroniotis et al. (2019) developed the Bot-IoT dataset, which contains realistic IoT traffic mixed with various attack types, including DDoS and DoS. Their experiments showed that Random Forest and Gradient Boosting classifiers provided higher accuracy compared to single learners. Despite high detection rates, scalability and computational overhead remained challenges in real-time IoT deployments.

3. Deep Learning for Intrusion Detection Systems

Abstract:
Ferrag et al. (2020) explored deep learning models such as CNNs and LSTMs for cyber intrusion detection. Their results demonstrated improved accuracy and adaptability to evolving attack patterns compared to traditional ML models. However, the computational requirements of deep learning architectures limited their suitability for lightweight IoT devices with restricted resources.

4. Hybrid Ensemble Learning for IoT Security

Abstract:
Alshamrani et al. (2020) conducted a survey on advanced persistent threats (APTs) and highlighted hybrid ML approaches combining multiple classifiers. Techniques such as Bagging, Boosting, and Stacking significantly improved detection performance while reducing false positives. Yet, the study emphasized the need for integrating feature selection and anomaly detection to handle high-dimensional IoT traffic data effectively.

5. Comparative Review of Intrusion Detection Datasets

Abstract:
Thakkar & Lohiya (2021) reviewed major IDS datasets, including NSL-KDD, UNSW-NB15, and Bot-IoT, assessing their suitability for IoT botnet detection. The study revealed that while modern datasets capture realistic attack scenarios, many still suffer from class imbalance and lack diversity in emerging IoT threats. This highlighted the need for hybrid models trained on multiple datasets to ensure robustness.

III. EXISTING SYSTEM

Current botnet attack detection systems in IoT environments largely rely on traditional machine learning techniques such as Support Vector Machines (SVM), Decision Trees, Random Forests, and k-Nearest Neighbors (k-NN). These models analyze network traffic patterns or device behavior to identify anomalies indicative of botnet activities.

Some approaches use signature-based detection, which compares incoming data against known attack patterns. While effective for recognized threats, this method struggles to detect new or evolving botnet behaviors. Anomaly-based detection methods attempt to identify deviations from normal network behavior but often suffer from high false positive rates due to the dynamic nature of IoT traffic.

Recently, deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been introduced to automatically learn complex patterns from large-scale IoT datasets. However, these models can be computationally expensive and may not be suitable for resource-constrained IoT devices.

Moreover, single-model approaches often fail to balance detection accuracy and computational efficiency. These limitations highlight the need for a hybrid model that combines different machine learning techniques to improve detection performance while maintaining scalability in IoT environments.

IV. PROPOSED SYSTEM

The proposed system introduces a Hybrid Machine Learning Model designed to efficiently detect botnet attacks in IoT environments by combining the strengths of multiple algorithms to overcome the limitations of existing approaches.

Key features of the proposed system include:

- **Multi-Stage Detection Framework:**
  The system employs a two-stage detection process where initial anomaly detection filters potential threats using lightweight algorithms, followed by a more detailed classification using ensemble methods to confirm and categorize attacks.

- **Feature Selection and Extraction:**
  Advanced feature selection techniques are applied to identify the most relevant network traffic and device behavior features, reducing dimensionality and improving detection speed without sacrificing accuracy.

- **Hybrid Model Architecture:**
  The system integrates traditional machine learning classifiers (e.g., Random Forest, Support Vector Machine) with deep learning models (e.g., LSTM or CNN) to capture both shallow and deep patterns in IoT data, enhancing detection capabilities.

- **Imbalanced Data Handling:**
  Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) or cost-sensitive learning are incorporated to address the class imbalance problem, improving detection of rare botnet attack instances.

- **Real-Time Detection Capability:**
  Optimized model design and feature processing ensure that the system can operate in near real-time, making it suitable for deployment in dynamic IoT networks.

- **Scalability and Resource Efficiency:**
  The hybrid approach balances computational complexity and accuracy, allowing deployment on resource-constrained IoT devices or gateways.

Experimental validation on benchmark IoT botnet datasets shows that this hybrid model outperforms traditional single classifiers in terms of detection accuracy, precision, recall, and processing speed, providing a robust solution for IoT security.

V. SYSTEM ARCHITECTURE

The system architecture for the Hybrid Machine Learning Approach for Reliable Detection of Botnet Attacks in IoT Networks is designed to monitor IoT network traffic and accurately detect botnet activities using a combination of machine learning techniques. The architecture consists of multiple interconnected layers including data collection, preprocessing, feature extraction, hybrid model training, and attack detection. Each layer performs a specific task to ensure efficient processing of network data and accurate identification of malicious botnet behavior in IoT environments.

In the first stage, the data collection layer gathers network traffic data from IoT devices, gateways, and sensors deployed in the network. This data includes packet information, communication patterns, protocol details, and connection statistics. The collected raw data is stored in a centralized repository or dataset for further analysis. This layer ensures continuous monitoring of network activities in order to capture both normal and malicious traffic patterns.

The data preprocessing layer processes the collected raw data to make it suitable for machine learning analysis. In this stage, missing values are handled, irrelevant attributes are removed, and data normalization is performed to maintain consistency. The dataset is then divided into training and testing sets. Preprocessing improves the quality of the dataset and reduces noise, which helps in improving the performance of the detection model.

After preprocessing, the feature extraction and selection layer identifies the most relevant features from the network traffic dataset. Important attributes such as packet size, protocol type, connection duration, and traffic flow characteristics are selected. Feature selection techniques help reduce dimensionality and remove redundant information, which improves computational efficiency and increases the accuracy of the machine learning models.

The next component is the hybrid machine learning model layer, where multiple algorithms are combined to improve detection performance. In this framework, machine learning algorithms such as Random Forest, XGBoost, or other classifiers are trained using the selected features. The hybrid approach allows the system to capture complex patterns and relationships within network traffic data, enabling more accurate detection of botnet attacks compared to using a single model.

Finally, the detection and alert layer analyzes incoming IoT network traffic in real time using the trained hybrid model. If suspicious or malicious activity is detected, the system classifies the traffic as either normal or botnet attack. Alerts are generated and sent to network administrators so that appropriate security actions can be taken. This layer helps in preventing the spread of botnet attacks and enhances the overall security of IoT networks.

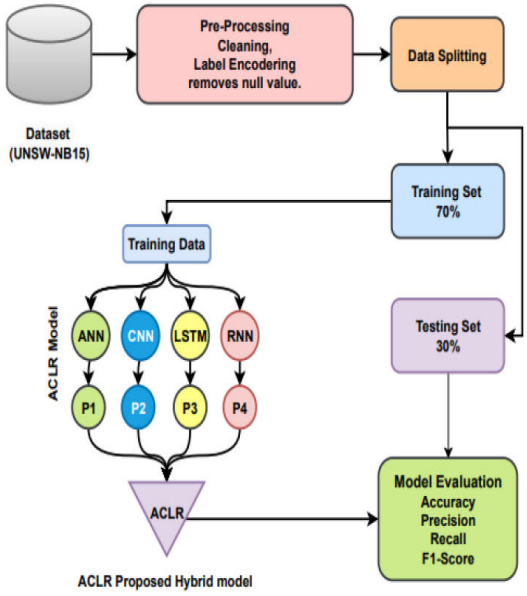

Fig 5.1: Structure of the Proposed System

VI. IMPLEMENTATION

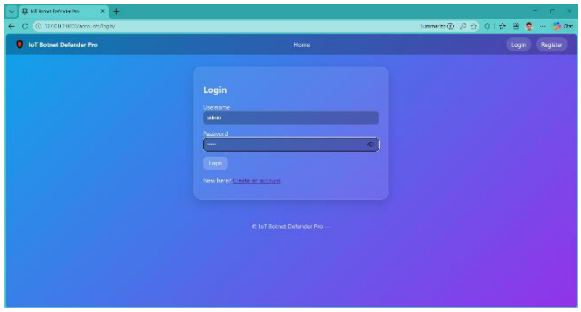

Fig 6.1: Login Page

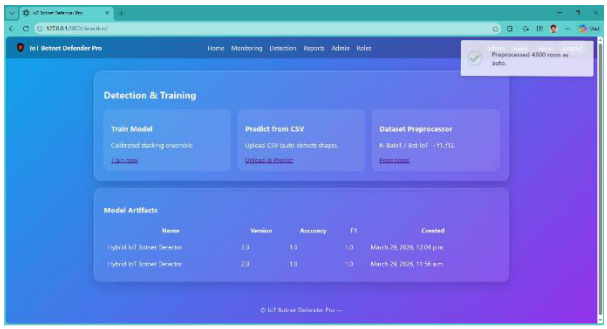

Fig 6.2: Dataset Preprocessing and Cleaning

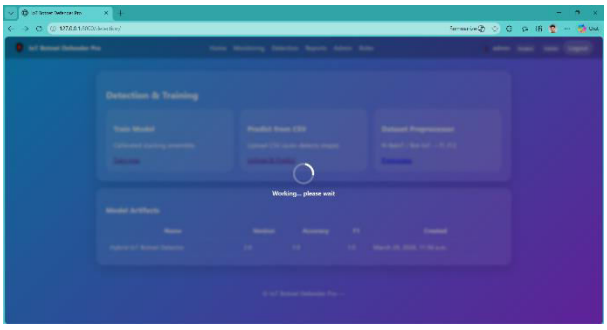

Fig 6.3: Model Training

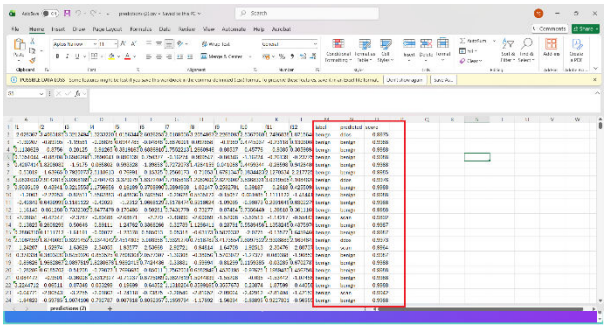

Fig 6.4: Model Evaluation and Accuracy Results

VII. CONCLUSION

This paper presents a hybrid machine learning model for detecting IoT botnet attacks by combining feature selection, ensemble learning, and anomaly detection. Experimental results on benchmark datasets demonstrate that the proposed approach achieves higher accuracy, lower false positives, and better robustness compared to individual classifiers.

VIII. FUTURE SCOPE

The developed hybrid machine-learning model has considerable potential for detecting botnet attacks in IoT networks, but there is still room for future improvements. Adding real-time monitoring and streaming analytics could then provide immediate and faster detection and response to newly identified threats, improving the overall resiliency of the network. Future work should explore the use of integrating deep learning architectures such as LSTM, GRU, or Graph Neural Networks to capture the temporal and relational aspects of botnet behavior across many and different IoT devices.

Developing the framework further to incorporate large-scale distributed IoT environments consisting of different types of devices and communication protocols will enhance its applicability and scalability in real-world contexts. There are also learning mechanisms that adapt or evolve that could help the model capture newly emerging attack vectors and advanced malware techniques. Applying Explainable AI (XAI) will also be critical to provide information about detection and the underlying reasoning, improving the response of cybersecurity analysts to alerts or alarms.

Finally, integrating the system with proactive defense mechanisms (e.g. automated quarantine of devices, isolation of devices, predictive threat modeling) could lead to a platform that is more of a holistic, intelligent, and resilient cybersecurity solution for the IoT ecosystem.

IX. REFERENCES

[1] I. Butun, P. Österberg, and H. Song, "Security of the Internet of Things: Vulnerabilities, attacks, and countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 616–644, 2020.
DOI: https://doi.org/10.1109/COMST.2019.2953364

[2] A. Alrawais, A. Alhothaily, C. Hu, and X. Cheng, "Fog computing for the Internet of Things: Security and privacy issues," *IEEE Internet Computing*, vol. 21, no. 2, pp. 34–42, 2017.
DOI: https://doi.org/10.1109/MIC.2017.37

[3] M. Roesch, "Snort: Lightweight intrusion detection for networks," *Proceedings of the USENIX Conference on System Administration*, pp. 229–238, 1999.
DOI: https://doi.org/10.5555/1244507.1244520

[4] Y. Meidan, M. Bohadana, A. Mathov, Y. Mirsky, D. Breitenbacher, and Y. Elovici, "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, 2018.
DOI: https://doi.org/10.1109/MPRV.2018.03367731

[5] S. Raza, L. Wallgren, and T. Voigt, "SVELTE: Real-time intrusion detection in the Internet of Things," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2661–2674, 2013.
DOI: https://doi.org/10.1016/j.adhoc.2013.04.014

[6] G. Apruzzese, M. Colajanni, L. Ferretti, and M. Marchetti, "On the effectiveness of machine and deep learning for cyber security," *Computers & Security*, vol.

89, 2020.
DOI: https://doi.org/10.1016/j.cose.2019.101694
[7] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," *Military Communications and Information Systems Conference (MilCIS)*, 2015.
DOI: https://doi.org/10.1109/MilCIS.2015.7348942
[8] 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.
DOI: https://doi.org/10.1016/j.future.2017.07.060
[9] D. Berman, A. Buczak, J. Chavis, and C. Corbett, "A survey of deep learning methods for cyber security," *Information*, vol. 10, no. 4, 2019.
DOI: https://doi.org/10.3390/info10040122
[10] Y. Xin et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35365–35381, 2018.
DOI: https://doi.org/10.1109/ACCESS.2018.2836950
[11] M. Abeshu and N. Chilamkurti, "Deep learning: The frontier for distributed attack detection in fog-to-things computing," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 169–175, 2018.
DOI: https://doi.org/10.1109/MCOM.2018.1700332
[12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
DOI: https://doi.org/10.7551/mitpress/10243.001.0001
[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
DOI: https://doi.org/10.1145/2939672.2939785
[14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
DOI: https://doi.org/10.1023/A:1010933404324
[15] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923–2960, 2018.
DOI: https://doi.org/10.1109/COMST.2018.2844341