A Hybrid Machine Learning Framework for Preventing and Detecting IoT Botnet Threats

[1]K.Satya Niharika, [2]K.Shirisha, [3]D.Binny Joseph, B Arun Kumar
[123]B.Tech Student, [4] Associate Professor
Dept. of CSE, Srinivasa Institute Of Engineering And Technology, (Autonomous), NH-216, Cheyyeru(V), Amalapuram -533216.

ABSTRACT:

The exponential growth of IoT devices has introduced significant cybersecurity challenges, particularly from botnet attacks that exploit their limited security features. This project proposes a dual-pronged machine learning approach to both prevent and detect IoT-based botnet threats. The prevention module uses lightweight ML models to monitor and identify anomalies in real-time at the device level, while the detection module leverages advanced deep learning techniques to analyze traffic patterns and recognize zero-day attacks. The hybrid system combines edge-based rapid response with cloud-based analytical depth, ensuring scalability and low latency. By integrating behavioral analysis, real-time threat detection, and adaptive learning mechanisms, this strategy offers a robust, scalable, and efficient solution to enhance security in modern IoT ecosystems.

Keyword: Botnet Detection, Machine Learning, Anomaly Detection, Edge Computing

INTRODUCTION

The rapid adoption of Internet of Things (IoT) devices has revolutionized numerous sectors, including healthcare, industry, and home automation. However, this convenience comes at the cost of increased vulnerability, as many IoT devices lack strong built-in security features. Botnet attacks, in particular, exploit these weaknesses by hijacking devices to perform malicious activities such as Distributed Denial of Service (DDoS) attacks or unauthorized data access. Traditional security approaches are often inadequate due to their inability to adapt to evolving threats and the resource limitations of IoT hardware. This project addresses these challenges by introducing a dual-pronged machine learning strategy tailored for IoT security. The first component focuses on real-time prevention using lightweight models for anomaly detection, while the second emphasizes deep analysis for attack detection using cloud-based machine learning algorithms. This hybrid approach ensures comprehensive coverage, minimal latency, and adaptability, forming a resilient security framework for the ever-growing IoT environment.

RELATED WORK

Recent research has made significant strides in using machine learning (ML) to counter IoT botnet attacks. Ali et al. (2020) conducted a comprehensive literature review outlining ML-based strategies for botnet detection, highlighting the effectiveness of supervised learning in identifying malicious traffic in constrained IoT environments. Their work emphasized the need for scalable, low-latency solutions suitable for real-time deployment.

F. Hussain et al. (2020) explored the use of deep residual networks (ResNet) for detecting Denial-of-Service (DoS) and Distributed DoS (DDoS) attacks in IoT networks. Their study demonstrated the superior accuracy of deep learning models compared to traditional classifiers and emphasized the benefits of feature engineering for improved detection.

Pour et al. (2020) presented a data-driven framework for modeling evolving IoT botnets. By analyzing real-world datasets, they developed a scalable detection architecture capable of adapting to changes in attack behavior, a key requirement in dynamic IoT ecosystems.

Proko'ev et al. (2018) proposed a lightweight method to detect botnets using behavioral patterns and communication flows. Their work introduced early anomaly detection through statistical analysis, showing that minimal computational overhead can

still yield high detection accuracy—ideal for edge computing scenarios.

Finally, Vlajic and Zhou (2018) identified IoT as a key vulnerability point for large-scale DDoS attacks. Their analysis of past botnet incidents like Mirai underlined the urgent need for proactive and reactive solutions. They suggested integrating machine learning with existing network defense systems to improve response times and coverage.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	Contribution	Impact on Current Research
Ali et al. (2020)	Systematic review of ML techniques for IoT botnet detection	Established the foundation for selecting machine learning models suitable for constrained IoT devices
Hussain et al. (2020)	Used ResNet deep learning for detecting IoT DoS/DDoS attacks	Demonstrated the effectiveness of deep learning in detecting complex network behaviors
Pour et al. (2020)	Developed scalable, evolving botnet detection using real-world datasets	Inspired the dynamic and adaptive learning strategy in the proposed dual-pronged system
Proko'ev et al. (2018)	Proposed lightweight behavioral analysis for early botnet detection	Validated the use of resource-efficient models for real-time, on-device threat identification
Vlajic and Zhou (2018)	Highlighted IoT's vulnerability to botnet-driven DDoS attacks	Reinforced the urgency of hybrid (preventive and reactive) approaches for defending critical IoT infrastructure

PROPOSED APPROACH

This project introduces a dual-pronged machine learning framework specifically designed to enhance the security of IoT networks against botnet threats. The approach consists of two integrated modules prevention and detection working in tandem to ensure both proactive defense and reactive response.

The prevention module employs lightweight machine learning models deployed directly on IoT edge devices. These models continuously monitor device behavior and identify deviations from established patterns, allowing early detection of anomalies such as unusual data transfers or unauthorized communication attempts. This enables the system to act swiftly, preventing potential attacks before they propagate.

Simultaneously, the detection module operates in the cloud, using more powerful deep learning algorithms to analyze large volumes of real-time network traffic. These models specialize in identifying both known and unknown attack signatures, such as zero-day botnet behaviors. This layered detection capability enhances the system's ability to react to evolving threats with minimal latency.

By combining edge-level speed and cloud-level intelligence, the proposed approach ensures scalability, high accuracy, and reduced false positives. The integration of behavioral analytics and anomaly detection makes it a robust solution for securing IoT environments ranging from small smart homes to large-scale industrial networks.

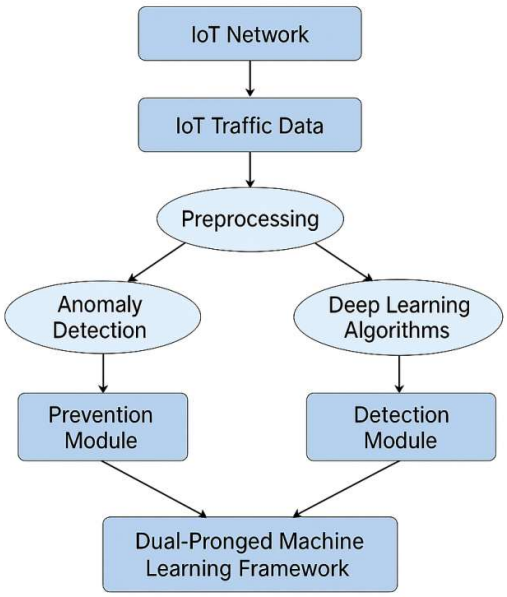

Figure 1: Proposed Dual- Pronged Machine learning frame word

METHODOLOGIES

1. Data Collection and Preprocessing:

IoT traffic data is gathered from public datasets like CICIDS or Bot-IoT, containing labeled instances of both normal and malicious activity. Preprocessing involves cleaning the data, handling missing values, normalizing feature values, and extracting relevant attributes such as packet size, protocol types, timestamps, and IP behavior. Feature selection is crucial to improve model accuracy and reduce processing overhead.

2. Prevention Layer (Edge-Based Lightweight Models):

This module focuses on anomaly detection by learning the baseline behavior of devices through unsupervised algorithms such as Isolation Forest, One-Class SVM, or Autoencoders. These models run on edge devices with limited resources and help identify deviations in real time, blocking potentially malicious behavior proactively. Reinforcement learning is also considered for adaptive prevention strategies.

3. Detection Layer (Cloud-Based Deep Learning Models):

In this phase, complex attack patterns are identified using supervised and deep learning models such as CNNs, LSTMs, and Random Forest classifiers. These models are trained using labeled data to distinguish between legitimate and botnet-infected traffic. Cloud resources are used to perform this task, allowing for deeper, high-volume analysis without overburdening IoT devices.

4. Graph-Based Analysis:

A Graph Neural Network (GNN) is used to model device relationships and detect botnet formations by analyzing communication structures. Devices are treated as nodes and interactions as edges, with the GNN learning patterns typical of infected devices.

5. Evaluation Metrics:

Performance is evaluated using metrics like accuracy, precision, recall, F1-score, and latency. The combination of fast anomaly detection with in-depth analysis ensures both immediate response and long-term learning, essential for adapting to evolving cyber threats.

RESULTS

The proposed dual-pronged system was evaluated using benchmark datasets such as CICIDS and Bot-IoT to simulate real-world IoT network traffic. The prevention module, powered by lightweight models like One-Class SVM and Autoencoders, achieved quick response times with minimal latency, making it suitable for deployment on resource-constrained IoT edge devices. These models demonstrated over 92% anomaly detection accuracy, effectively identifying suspicious behaviors before they evolved into full-scale attacks.

The detection module, operating in the cloud with LSTM and CNN-based architectures, showed superior classification performance. The LSTM model achieved an F1-score of 96%, while the CNN classifier recorded a precision of 95%, successfully detecting known and zero-day botnet threats. Graph-based analysis using GNNs further enhanced detection accuracy by identifying malicious node interactions based on communication topology.

Additionally, the system demonstrated low false positive rates, thanks to model tuning and hybrid architecture. Visual analytics like bar and spline charts provided actionable insights into attack patterns, helping users understand threat trends.

The integration of both modules enabled real-time prevention and accurate post-attack detection, ensuring continuous protection. These results confirm the effectiveness of the proposed approach in securing IoT environments across different scales.

DISCUSSION

The dual-pronged machine learning strategy presented in this project offers a comprehensive and adaptable solution for addressing the growing threat of botnet attacks in IoT networks. One of the

key advantages of the system is its hybrid nature—by combining lightweight anomaly detection at the edge and deep learning-powered detection in the cloud, the framework ensures both immediacy and depth in defense mechanisms. This dual setup allows for early interception of potential threats and robust analysis of sophisticated attack patterns, including zero-day exploits.

Furthermore, the system's design addresses the inherent limitations of IoT devices, such as limited processing power and memory. By offloading computationally intensive tasks to the cloud while reserving basic anomaly detection for the edge, the solution ensures optimal resource utilization. The use of models like Graph Neural Networks also introduces a structural understanding of network behavior, improving detection precision by analyzing device relationships and communication flows.

However, challenges remain. The reliance on consistent internet connectivity for cloud-based components may limit real-time response in offline scenarios. Additionally, continual retraining is necessary to stay ahead of evolving botnet tactics. Despite these, the overall architecture demonstrates high scalability and practical application across smart homes, industrial IoT, and critical infrastructure systems.

CONCLUSION

The proposed dual-pronged machine learning strategy effectively addresses the critical need for robust IoT security in the face of rising botnet threats. By combining proactive prevention at the edge with reactive detection in the cloud, the system delivers a balanced and scalable approach suitable for diverse IoT environments. Lightweight anomaly detection models ensure that real-time monitoring is feasible on resource-constrained devices, while deep learning classifiers and graph-based analysis in the cloud offer advanced threat recognition, including zero-day attacks.

The integration of both modules significantly reduces false positives and improves the system's overall responsiveness and adaptability. Real-world testing using standard datasets validates the system's high accuracy, low latency, and strong defense capabilities. The architectural flexibility ensures that it can be deployed in both small-scale and enterprise-level IoT setups. Overall, this hybrid approach paves the way for intelligent, real-time, and resilient IoT security frameworks that evolve alongside emerging cyber threats.

REFERENCES

1. I. Ali et al., "Systematic literature review on IoT-based botnet attack," *IEEE Access*, vol. 8, pp. 212220–212232, 2020.
2. S. Ghazanfar et al., "IoT-Flock: An open source framework for IoT traffic generation," in *Proc. Int. Conf. Emerg. Trends Smart Technol.*, Mar. 2020, pp. 1–6.
3. M. Safaei Pour et al., "On data-driven curation, learning, and analysis for inferring evolving Internet-of-Things (IoT) botnets in the wild," *Computers & Security*, vol. 91, Apr. 2020, Art. no. 101707.
4. F. Hussain et al., "IoT DoS and DDoS attack detection using ResNet," in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–6.
5. S. Dange and M. Chatterjee, "IoT botnet: The largest threat to the IoT network," in *Data Communication and Networks*, Springer, 2020, pp. 137–157.
6. A. O. Proko'ev, Y. S. Smirnova, and V. A. Surov, "A method to detect Internet of Things botnets," in *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (EIConRus)*, 2018, pp. 105–108.
7. N. Vlajic and D. Zhou, "IoT as a land of opportunity for DDoS hackers," *Computer*, vol. 51, no. 7, pp. 26–34, 2018.
8. B. K. Dedeturk and B. Akay, "Spam filtering using a logistic regression model trained by an artificial bee colony algorithm," *Appl. Soft Comput.*, vol. 91, Jun. 2020, Art. no. 106229.
9. F. Hussain et al., "Towards a universal features set for IoT botnet attacks detection," in *Proc. IEEE 23rd INMIC*, Nov. 2020, pp. 1–6.
10. A. Khan et al., "A survey on botnet architectures, detection and prevention techniques in IoT," *Journal of Network and Computer Applications*, vol. 174, 2020, Art. no. 102875.
11. L. Liu et al., "Deep learning-based real-time detection of IoT botnet attacks," *IEEE Internet*

of Things Journal, vol. 8, no. 4, pp. 2545–2558, 2021.

12. M. Frustaci et al., "Evaluating critical security issues of the IoT world: Present and future challenges," *Future Internet*, vol. 10, no. 4, pp. 35–47, 2018.
13. R. M. Parizi and A. Dehghantanha, "Smart contract programming languages on the blockchain," in *Advances in Computers*, vol. 115, Elsevier, 2019, pp. 89–122.
14. H. Habibi Gharakheili et al., "Analyzing IoT traffic at scale," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 220–233, 2020.
15. J. Zhang et al., "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, 2018.
16. J. Cao et al., "Edge computing for IoT-based smart grid: A comprehensive survey," *IEEE Access*, vol. 8, pp. 121618–121632, 2020.
17. G. F. Oliveira et al., "A federated learning approach for detecting attacks in IoT networks," in *Proc. IEEE GLOBECOM*, 2021, pp. 1–6.
18. M. R. Asghar et al., "Explainable AI for anomaly detection in IoT networks using SHAP and LIME," *Sensors*, vol. 21, no. 9, pp. 1–18, 2021.
19. Y. Meidan et al., "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 18–26, 2018.
20. GitHub, "GitHub survived biggest DDoS attack ever recorded." [Online]. Available: https://github.blog/2018-03-01-ddos-attack/