Machine-Learning Anomaly Detection for Early Identification of DDOS in Smart Home IoT Devices

Roland Lamptey
Department of Computer Science
City St George's, University of London
London, United Kingdom
Roland.lamptey@city.ac.uk

Mohammad Saedi
Department of Computer Science
City St George's, University of London
London, United Kingdom
Mohammad.Saedi@city.ac.uk

Vladimir Stankovic
Department of Computer Science
City St George's, University of London
London, United Kingdom
Vladimir.Stankovic.1@city.ac.uk

Abstract*—The rapid adoption of Internet of Things (IoT) devices in smart homes has introduced security vulnerabilities, with Distributed Denial of Service (DDoS) emerging as a critical threat. Exploiting the often-unsecured nature of these interconnected devices, such attacks overwhelm network resources, causing severe disruptions and privacy breaches. We present a novel anomaly detection system for early-stage DDoS attack identification in smart home IoT environments. Using NS-3 simulator, a realistic IoT network dataset was generated, capturing normal and malicious traffic. Key traffic features, e.g., packet size and inter-arrival times, were extracted to train two lightweight Machine Learning (ML) models: One-Class Support Vector Machine (OCSVM) and Isolation Forest (IF). OCSVM model achieved superior performance with accuracies from 96% to 99% for various attacks, while the IF model performed marginally worse. We offer a lightweight and scalable solution for real-time deployment in resource-constrained IoT environments, a significant step to enhance smart home security.

Keywords - Anomaly detection, DDoS attacks, smart home security, Internet of Things (IoT), machine learning, NS-3 simulator

I. INTRODUCTION

The rapid proliferation of Internet of Things (IoT) devices has revolutionized modern homes, seamlessly integrating smart technologies such as cameras, thermostats, and appliances into daily life. According to recent projections, there will be 41 billion connected IoT devices by 2030 [1]. While this enhances convenience, it also introduces significant cybersecurity challenges. Among these, Distributed Denial of Service (DDoS) attacks represent a critical threat. Exploiting the limited computational power and weak security of IoT devices, these attacks inundate networks with malicious traffic, rendering them inoperable and disrupting essential services [2] [3].

Despite their growing adoption, smart home environments are often designed with simplicity and ease of use, typically lacking adequate built-in security mechanisms. This makes them especially vulnerable to exploitation by attackers. DDoS attacks can disrupt communication, compromise device functionality, and breach household privacy. Existing security measures predominantly rely on traditional signature-based detection methods, which are ineffective against the dynamic and evolving nature of modern DDoS attacks [5]. This gap highlights the need for adaptive, scalable, and lightweight anomaly detection systems tailored specifically to the constraints of IoT devices in smart home networks [6]. Our research, as studies addressing security challenges in advanced networks, e.g., 5G Rogue Base Station (RBS) attacks [7], focuses on leveraging ML-based anomaly detection to mitigate DDoS attacks in IoT networks. As the simulation-based dataset generation proposed in [7], we develop realistic IoT traffic datasets to enhance model training for detecting anomalies.

The main contributions of this paper are as follows:

- *Simulated IoT Network Dataset*: A realistic smart home network is modelled using the Network Simulator-3 (NS-3), generating diverse traffic patterns to train and evaluate the anomaly detection models.
- *Lightweight Machine Learning (ML) Models*: 2 anomaly detection models, One-Class Support Vector Machine (OCSVM) and Isolation Forest (IF), are developed and optimized for resource-constrained IoT devices.
- *Comprehensive Performance Evaluation*: The proposed models are rigorously evaluated using key metrics: accuracy, precision, recall, and F1-score. They achieve high detection rates with minimal false positives.
- *Open Dataset Contribution*: The dataset generated in this study is made publicly available to support further IoT security research and innovation.

The rest of the paper is organized as follows: Sect. II reviews related work on DDoS detection in IoT networks and anomaly detection techniques. Sect. III describes the proposed approach, including the simulation setup, data generation, and model development. Sect. IV presents the performance evaluation of the proposed models. Sect. V discusses the findings, provides conclusions, and highlights potential future research.

II. RELATED WORK

The unique challenges of securing IoT environments, particularly within smart homes, require advanced solutions beyond traditional network defenses. This section examines existing techniques for DDoS and anomaly detection in IoT systems, highlighting the gaps that underscore the need for the specialized approach proposed in this paper.

A. DDoS Attacks and IoT Vulnerabilities

DDoS attacks are a critical concern in IoT environments due to the lack of standardized security practices and the heterogeneous nature of IoT devices. Many devices are resource-constrained making them vulnerable to attacks [8]. For instance, botnets such as Mirai have exploited these vulnerabilities, co-opting IoT devices into large-scale attacks without the users' knowledge [9]. This underscores the need for

robust detection mechanisms that can identify malicious activity early and prevent devices from being compromised. However, many existing systems rely on reactive measures, detecting attacks after they have already begun, which limits their effectiveness in preventing DDoS attacks [10].

B. Machine Learning for Anomaly Detection

Several studies have demonstrated the potential of ML for early detection of DDoS attacks in IoT environments. Kaur and Ayoade [11] highlight the importance of real-time anomaly detection mechanisms, especially given the limitations of traditional security tools that rely on pre-defined signatures and known attack patterns. Doshi et al. [12] propose using ML techniques on consumer routers to detect DDoS attacks launched by IoT devices by analysing network traffic flow. Similarly, Russell et al. [13] show the effectiveness of unsupervised learning techniques like the Local Outlier Factor (LOF) and OCSVM in detecting anomalies in IoT networks.

The challenge, however, is in the computational constraints of IoT devices in smart homes. Many detection systems, e.g., deep learning ones [14], require significant resources, making them unsuitable for real-time, on-device detection in environments with limited computational power. Jemal et al. [15] used convolutional neural networks to detect DDoS attacks, achieving high accuracy, but their method demands processing power beyond the ones of most smart home IoT devices.

The study in [20] investigates a progressive fuzzy C-means clustering (FCM) algorithm for anomaly detection in DDoS scenarios. Their clustering-based approach groups similar traffic patterns to differentiate between normal and attack behaviors. The FCM algorithm achieved low error rates and strong detection accuracy, making it an effective tool for distinguishing complex attack patterns. However, clustering-based methods, e.g., FCM, often require significant computational resources, making them less suitable for resource-constrained IoT environments. This highlights the importance of exploring lightweight models, e.g., OCSVM, that prioritize efficiency and scalability in real-time anomaly detection.

C. Real-Time Anomaly Detection for Smart Homes

While various ML techniques were proposed for DDoS detection, a key limitation remains: most systems are designed for enterprise or data-centre environments, where computational resources are abundant. These methods do not account for the resource limitations of smart home IoT devices, which require lightweight, real-time detection mechanisms. The study by Mishra and Pandya [17] shows that unsupervised learning offers a promising solution for detecting anomalies in IoT networks, as it does not require extensive labelled datasets, which are often difficult to obtain for all possible attack scenarios.

Hu and Tu [18] demonstrate the feasibility of using unsupervised clustering methods, specifically the fuzzy C-means clustering (FCM) algorithm, for detecting DDoS anomalies in network traffic. By grouping traffic patterns into clusters, FCM effectively distinguishes between normal and attack behaviors without relying on labeled data. However, the algorithm's computational overhead makes it less practical for real-time deployment in smart home environments with limited processing power. This study underscores the need for lightweight, scalable detection methods like the proposed OCSVM model, which is optimized for constrained environments while maintaining high accuracy.

Additionally, [10] propose using unsupervised learning methods, such as autoencoders combined with k-nearest neighbours (KNN), to identify DDoS attacks without the need for labelled training data. Their approach is particularly effective in environments where abnormal data points are sparse, as in early-stage DDoS attacks in IoT networks. This aligns with our focus on using lightweight unsupervised learning models that can detect deviations in traffic patterns while operating within the resource constraints of smart home IoT devices.

III. Proposed Approach

While much of the existing research has focused on ML techniques for DDoS detection in traditional networks, fewer studies have addressed the specific challenges of smart home IoT environments. The heterogeneity of devices, each with different communication protocols and computational capacities, complicates the implementation of uniform security measures. Our approach builds upon the existing work by developing an ML-based anomaly detection system optimised for the limited resources of smart home devices. We propose using lightweight algorithms, e.g., OCSVM, to detect early signs of DDoS based on deviations from normal traffic behaviour. Unlike more resource-intensive approaches, our system is designed to operate efficiently within the constrained environments of smart homes, providing early detection without overwhelming the device or network infrastructure.

A. Proposed Attack Model

We evaluate the anomaly detection system within a simulated smart home IoT environment by employing four common DDoS attack types: HTTP Flood, UDP Flood, SYN Flood, and Slowloris. Each attack targets distinct network vulnerabilities, enabling a comprehensive assessment of the model's ability to detect diverse anomalous traffic patterns.

An HTTP Flood is a volumetric attack that overwhelms a web server by sending excessive HTTP GET or POST requests, depleting its resources and causing service disruptions. It exploits real-world IoT vulnerabilities, particularly for devices that rely heavily on cloud-based services for functionality e.g., smart TVs. The simulation frequency was 500 requests per sec, average packet size 800 bytes, and durations ranged from short bursts (30 sec) to sustained traffic (5 min). The attack represents a high-bandwidth event, where malicious requests overshadow legitimate IoT traffic, causing severe service interruptions.

The UDP Flood sends a high volume of UDP packets to random ports on a target device, forcing search for listening applications and exhausting processing power. This attack is particularly disruptive for devices like smart speakers and thermostats, which are optimized for lightweight operations. In the simulation, packet sizes ranged from 500 to 1200 bytes to mimic real-world, with a frequency of 800 packets per sec. The attack durations varied from transient bursts of 1 min to prolonged stress tests of up to 5 min. This attack stressed both bandwidth and computational resources, providing a challenging test for the anomaly detection model's ability to handle high-volume and irregular traffic.

A SYN Flood exploits the TCP handshake process by

initiating numerous incomplete connections, leaving them half-open. This prevents legitimate traffic from being processed, creating delays, or blocking access entirely. For IoT devices with limited networking capacity, such as smart thermostats, this attack can significantly disrupt operations. The simulation parameters included 300 connection attempts per sec, with durations of 1 min for transient tests and up to 3 min for sustained loads. The incomplete connections were designed to simulate resource exhaustion, replicating the real-world impact of such attacks on constrained IoT devices.

The Slowloris attack employs a "low-and-slow" approach, sending partial HTTP requests over extended periods to keep server connections open. Unlike high-bandwidth attacks, Slowloris gradually drains server resources without causing obvious traffic spikes, making it challenging to detect. The simulation settings included packet sizes of 200 to 400 bytes, with a frequency of 1 packet every 10 sec. The attack durations ranged up to 10 min, maintaining persistent connections. This scenario tested the anomaly detection system's sensitivity to subtle, low-volume anomalies that could otherwise evade traditional detection mechanisms.

Together, these 4 attack types represent a diverse set of challenges for IoT anomaly detection systems, capturing both high-bandwidth and subtle, low-rate attack scenarios. To evaluate the detection system's robustness, these attack scenarios were simulated in a smart home environment using the NS-3 simulator, featuring common IoT devices. The network supported internal and external communications, enabling realistic interaction scenarios. Metrics such as packet size, inter-arrival times, and flow duration were recorded to differentiate normal and attack traffic, with specific devices targeted to assess bandwidth and computational impacts. These parameters provide valuable insights into the system's ability to adapt to varied real-world attack patterns and improve IoT security.

B. Data Generation and Traffic Simulation

To simulate a realistic smart home IoT environment, the NS-3 network simulator was configured with several assumptions to reflect typical residential setups. The simulated devices – smart thermostats, security cameras, speakers etc. – were designed to accurately represent real-world IoT traffic patterns, including packet generation rates, communication protocols, and usage behaviors. These configurations were based on known usage patterns from established studies [19]. The network was assumed to have limited security mechanisms, consistent with the relatively low-security standards of consumer IoT devices, to realistically model their vulnerability to DDoS attacks. While NS-3 cannot fully replicate all real-world conditions, it effectively captures key behaviors such as traffic flow, and response delays, enabling the generation of realistic attack and normal traffic for this study.

Several additional assumptions were made to create a simulation environment that reflected realistic conditions commonly found in residential smart homes:

1) Internet Speed: The simulated network operated with a typical residential broadband speed for home internet in developed regions: 100 Mbps download and 20 Mbps upload. This bandwidth supports multiple devices simultaneously without overwhelming capacity, enabling realistic usage scenarios such as streaming and device updates. This ensured DDoS traffic patterns stood out against regular network usage.

2) Device Usage Patterns: Each IoT device's behavior was programmed to mimic standard usage patterns based on real-world studies, ensuring that the simulated data closely mirrored realistic traffic patterns observed in residential settings. For instance, the smart camera was configured to activate motion detection at regular intervals, generating high-bandwidth video streams during motion events; te smart thermostat sent periodic temperature updates every 30 sec to replicate routine functionality; the smart TV was set to stream video content during typical usage hours, emulating the high bandwidth demands of media consumption; etc. These realistic patterns provided a diverse dataset, accurately capturing the range of traffic behaviors expected in modern IoT environments.

3) Wi-Fi Signal Quality: For simplicity, Wi-Fi signal quality was assumed to remain stable throughout the simulation, with no significant drops or interference. This ideal condition allowed for cleaner analysis by ensuring that detected network anomalies were solely due to traffic patterns rather than external factors like signal degradation or packet loss. This approach also simplified result interpretation by isolating the effects of DDoS attacks from unrelated connectivity issues.

4) External Server Reliability: External servers, e.g., streaming platforms, were assumed to be consistently available without outages or latency spikes. This idealization ensured that any detected anomalies could be attributed to malicious traffic or irregularities within the smart home network, rather than disruptions caused by external factors. This assumption provided a controlled environment, focusing solely on IoT device behavior under normal and attack conditions.

Given these assumptions, the simulation environment was made suitable for evaluating the anomaly detection system. It captured the complexities of modern IoT usage while maintaining controlled conditions to isolate the impact of DDoS attacks. These parameters ensured reproducibility and strengthened the practical relevance of the study, offering insights into the challenges and solutions for securing resource-constrained IoT networks.

C. Traffic Simulation Setup

The network traffic simulation was conducted in an Ubuntu 20.04.6 environment, chosen for its stability and compatibility with NS-3 [8]. This setup ensured efficient handling of simulated IoT traffic and provided a robust platform for accurate results. The smart home network topology included four common IoT devices: a smart speaker, smart TV, smart camera, and smart thermostat, reflecting real-world usage [9]. A Wi-Fi network was configured, mirroring typical smart home environments [10], and was set up to enable both internal and external communications to capture realistic data (Fig. 1).

NS-3's logging and packet tracing features were utilized to monitor all traffic flows. Key traffic metrics, e.g., packet size, inter-packet intervals, were recorded, as critical indicators of anomalous behavior, particularly in DDoS attacks [11].

1) Normal Traffic Collection The smart home environment was set up to simulate regular traffic patterns. Each device

performed standard tasks: e.g., the smart camera streamed video during motion events, the smart thermostat periodically sent sensor data via the MQTT protocol [16], etc. The diverse dataset is representative of normal smart home activities. The traffic was collected over a 10-min period, using PCAP files.

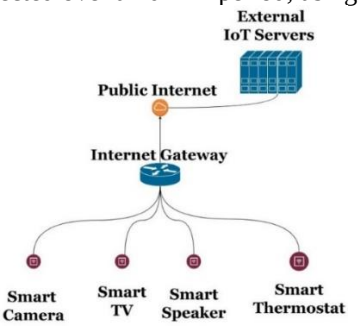

Fig. 1. Network Topology of Simulated Smart Home Environment

2) DDoS Attack Simulation. To generate attack traffic, 3 IoT devices (smart speaker, camera, and TV) were compromised to simulate various DDoS attacks, including HTTP Flood and SYN Flood attacks. These attacks aimed to overwhelm an Apache server simulated on an external node, reflecting real-world attack scenarios (Fig. 2) [6]. The traffic was captured in PCAP files for training and testing the models.

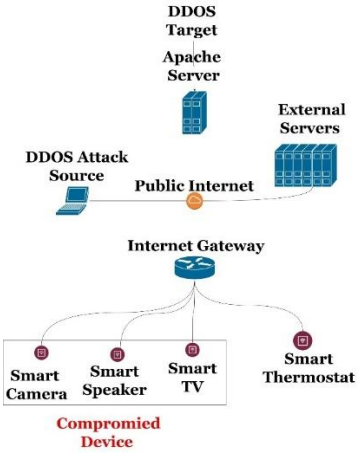

Fig. 2. DDoS attack scenario in a smart home IoT network

D. Detection Models

Two unsupervised ML models were used: One-Class Support Vector Machine (OCSVM) and Isolation Forest (IF). These models were chosen due to their ability to detect anomalies without requiring labeled attack data, which is often unavailable in IoT environments [9].

OCSVM: The ML model is suitable for scenarios where only normal data is available for training. The model employed Radial Basis Function (RBF) kernel, which is effective in capturing non-linear separations in network traffic anomalies. This configuration enabled the model to identify subtle deviations indicative of DDoS attacks. The key parameter ν, representing the proportion of data points treated as anomalies, was optimized at 0.05 after extensive testing to achieve a balance between minimizing false positives and detecting true anomalies. This lightweight design makes OCSVM particularly suitable for resource-constrained IoT environments.

IF: The ML model was chosen for its ability to isolate observations in a randomly partitioned feature space, effectively detecting anomalous behavior. Key parameters included *n* estimators, which defines the number of trees in the ensemble and was set to 100 to balance detection accuracy and computational efficiency, and the *contamination rate*, representing the expected proportion of anomalies in the dataset, which was set to 10%. These settings ensured reliable anomaly detection at a low computational overhead, making IF a feasible option for IoT devices with limited processing power.

E. Testing and Validation

The testing and validation process was designed to evaluate the performance of both models using a comprehensive dataset generated from simulated IoT devices under various DDoS attack scenarios (see Sec. III.A). Performance metrics *accuracy*, *precision*, *recall*, and *F1-score* were used to assess each model's effectiveness. AUC-ROC curves were generated to visually represent model performance across different threshold levels.

IV. Results

A. Simulation Results and Data Generation

1) Data Generated:

Over 4.1 GB of data was collected during a 1-hour simulation, with traffic split between normal operations and DDoS attack scenarios. Normal traffic, generated by regular IoT device operations, constituted the largest portion – 1.8 GB. Attack traffic, generated by various DDoS methods, accounted for the rest. HTTP flood traffic – 0.8 GB; SYN and UDP floods each – 0.6 GB. Slowloris attacks generated 0.3 GB, reflecting its low-volume nature. This dataset effectively captured both high-intensity and low-rate attack patterns, providing a robust foundation for training and evaluation of the anomaly detection models. The dataset ensures the models' ability to distinguish between typical IoT behavior and anomalous activity of DDoS.

2) Traffic Patterns:

Number of Packets Generated: A total of 6.1 million packets were captured during the simulation: 2.5 million packets correspond to normal traffic, and 3.6 million packets classified as DDoS attack traffic. The HTTP Flood attack generated the highest number of packets – 2.1 million, reflecting its high-intensity nature. In contrast, the Slowloris attack produced the fewest packets - 100,000, consistent with its low-rate, persistent strategy. Both UDP Flood and SYN Flood attack produced 700,000 packets. This distribution shows the significant variation in traffic volume between normal operations and different attacks, highlighting the diverse patterns utilized for model training and evaluation.

Packet Size Distribution: Normal traffic demonstrates a heterogeneous distribution of packet sizes and arrival times, reflecting typical IoT usage patterns such as sensor updates and media streaming. Conversely, DDoS attack traffic, particularly HTTP Flood and UDP Flood, exhibits sustained high-volume packet transmissions with uniform sizes, indicative of malicious intent. These variations in traffic patterns form the basis for anomaly detection in our model. The distribution of packet sizes revealed distinct patterns between normal and DDoS attack traffic. Normal traffic exhibited varied packet sizes, ranging from 400 to 800 bytes, reflecting the diverse nature of routine

IoT device communications such as video streaming, sensor updates, and command responses. In contrast, DDoS attack traffic clustered around larger, fixed packet sizes, primarily between 800 and 1600 bytes. This packet size is a hallmark of volumetric DDoS attacks, designed to overwhelm network resources with high-bandwidth traffic. The difference in packet size distributions between normal and attack traffic (Fig. 3) provides a critical feature for distinguishing anomalous behavior during model training and evaluation.

Inter-Arrival Times: Normal traffic exhibited dispersed inter-arrival times, with many packets arriving at intervals greater than 1 sec, reflecting the typical asynchronous nature of IoT device communications. In contrast, DDoS attack traffic displayed consistently shorter inter-arrival times, often approaching 0 sec, indicative of the high-intensity flooding behaviour characteristic of volumetric attacks. Fig 4. shows stark difference in inter-arrival times between normal and attack traffic: a critical feature used for distinguishing anomalous patterns during model training and evaluation.

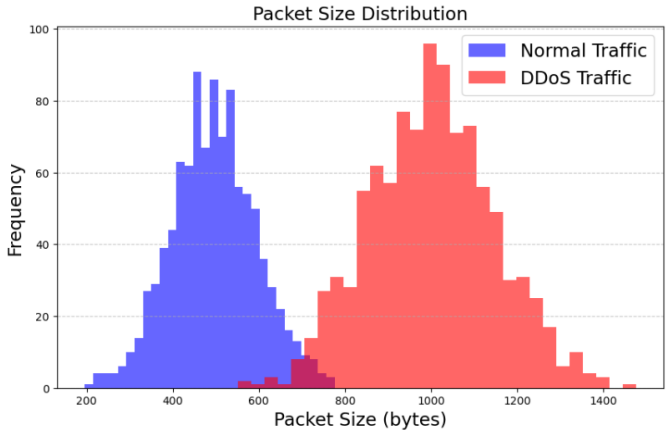

Fig. 3. Packet size distribution for normal versus DDoS traffic

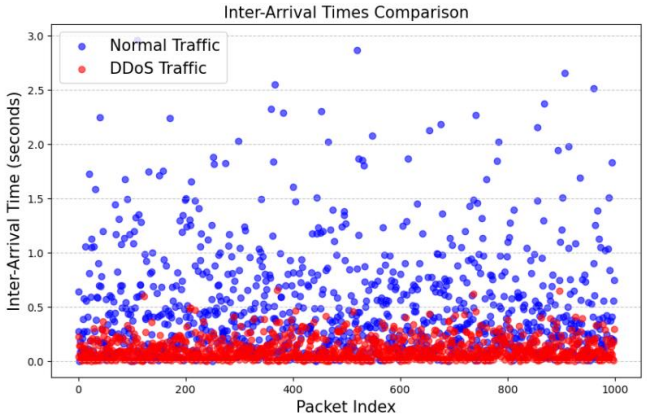

Fig. 4. Packet inter-arrival times for normal and DDoS traffic

B. Model Performance Evaluation

The OCSVM and IF models were trained on the dataset predominantly composed of normal traffic. Both models were evaluated on their ability to detect anomalies across various attack traffic types.

For clarity, we define key evaluation metrics used in assessing model performance: *False Positive* (FP) - incorrectly identifying benign traffic as an attack; *False Negative* (FN) - failing to detect an actual attack; *Accuracy*: the proportion of correct classifications among the total number of cases: TP +TF / (TP+FP+TN+FN); Precision - the proportion of correctly identified attacks relative to total identified attacks: TP / (TP + FP); *Recall* - the proportion of actual attacks correctly identified: TP / (TP + FN); *F1-Score* - the harmonic means of precision and recall, offering a balanced measure of model performance.

The performance of the OCSVM (referred to as "O") versus IF model for detecting different attack types is given in Table I. The OCSVM model consistently outperformed the IF one across all attack types and key metrics. It performs better in detecting both high-volume DDoS attacks and subtle application-layer attacks like Slowloris.

The Receiver Operating Characteristic (ROC) curves show the performance of the OCSVM (Fig. 5) and IF (Fig. 6) models for detecting different types of attacks. The Area Under the Curve (AUC) values for each attack type highlight the classification effectiveness, and the evaluation metrics are provided for additional context.

TABLE I. OCSVM (O) VS IF MODEL PERFORMANCE

ATTACK TYPES	ACCURACY		PRECISION		RECALL		F1-SCORE	
	O	IF	O	IF	O	IF	O	IF
HTTP FLOOD	0.99	0.96	0.99	0.96	0.99	0.95	0.99	0.96
UDP FLOOD	0.98	0.95	0.98	0.94	0.97	0.95	0.98	0.95
SYN FLOOD	0.99	0.97	0.99	0.96	0.99	0.96	0.99	0.97
SLOWLORIS	0.96	0.94	0.95	0.93	0.96	0.94	0.96	0.94

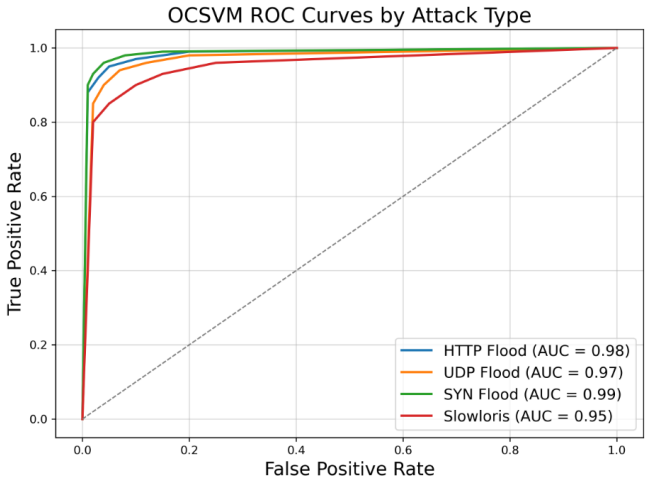

Fig. 5. ROC curves depicting the performance of the OCSVM model.

V. DISCUSSION AND CONCLUSIONS

We have simulated a smart home IoT network, generated diverse traffic data, and developed an effective anomaly detection system for early-stage DDoS attack detection. The NS-3 simulator replicated realistic traffic patterns from IoT devices, e.g., smart cameras, thermostats etc, producing a dataset of 4.1 GB. The anomaly detection models, particularly the One-Class

Support Vector Machine (OCSVM), demonstrated excellent performance across all tested DDoS attack types (HTTP Flood, SYN Flood, UDP Flood, and Slowloris).

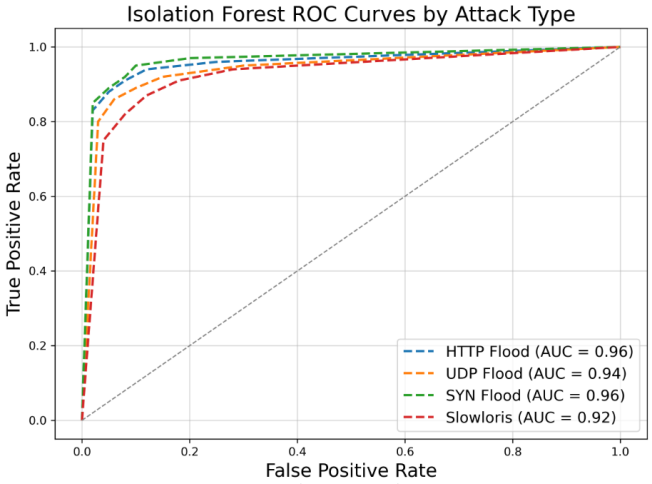

Fig. 6. ROC curves depicting the performance of the IF model.

The superior performance of the OCSVM model underscores its ability to detect anomalies effectively while maintaining computational efficiency. Its ability to distinguish between normal and anomalous traffic with minimal FP makes it a strong candidate for real-time anomaly detection in IoT networks. This makes it well-suited for deployment in resource-constrained IoT environments, such as smart homes.

A. *Model Strengths and Limitations*

The OCSVM model demonstrated exceptional accuracy and recall, proving effective in detecting early-stage DDoS attacks. Early detection is critical in real-world scenarios to mitigate the impact of attacks before they cause widespread disruption. By leveraging both packet-level features (e.g., packet size) and flow-level features (e.g., flow duration), the model achieved high efficiency while remaining computationally lightweight. This makes it particularly suitable for deployment on resource-constrained platforms, such as smart home gateways, where processing power and memory are limited. However, some legitimate traffic was still flagged as anomalous, particularly during periods of high network activity. In real-world deployments, such FPs could contribute to alert fatigue, potentially diverting attention from genuine threats.

The simulation evaluated the model's performance against a subset of DDoS attacks. While the results are promising, the model's effectiveness against other attack vectors, such as ICMP Flood or application-layer attacks, remains untested. This limitation impacts its generalizability to a broader range of DDoS scenarios. However, the unsupervised nature of OCSVM and its ability to identify deviations from normal traffic patterns may allow it to detect previously unseen or novel attack types.

B. *Recommendations for Future Work*

We plan to enhance the anomaly detection system to address zero-day attacks, which pose a significant challenge due to their evolving attack vectors. Incorporating *adaptive learning techniques* offers a promising solution to this issue: *Federated Learning* – the model can leverage distributed training across multiple IoT devices without sharing raw data, ensuring privacy and scalability. This approach enables continuous model updates as new traffic patterns emerge across different environments; and *Transfer Learning* – these techniques adapt pre-trained models to new, unseen attack patterns, reducing the time and resources required to train models on diverse datasets.

Additionally, expanding the scope of attack scenarios to include *application-layer and perception-layer DDoS attacks* would help evaluate the model's robustness against complex threats. For instance, application-layer attacks, such as HTTP Slow Post, exploit server resources in subtle ways, while perception-layer attacks target the physical devices themselves, disrupting sensor functionality. Testing the model's performance in real-world smart home environments, where traffic patterns are more dynamic and unpredictable, would also provide valuable insights into its practical deployment.

To further improve the dataset's diversity and model generalizability, collecting *real-world traffic* over an extended period, or extending the simulation, could capture a broader range of IoT behaviours and anomalies. This would enhance the system's ability to generalize across varying conditions.

Exploring *hybrid detection models* presents another promising direction. They can combine the strengths of OCSVM and other techniques, e.g., Deep Learning Approaches - Autoencoders or Graph Neural Networks (GNNs), and can detect complex relationships in traffic, identifying novel attack patterns, while Graph-Based Anomaly Detection can effectively model network interactions to detect relational anomalies.

Also, real-time deployment and integration with *automated mitigation strategies* would be critical steps toward operationalizing this system in real-world IoT networks. Such strategies might include *Traffic Throttling* to limit traffic from suspicious sources to minimize service disruption, or *IP Blacklisting* to block IP ranges that consistently demonstrate malicious behaviour.

Finally, incorporating *threat intelligence feeds* and *adaptive thresholds* would improve resilience against previously unseen attack strategies, enabling a more proactive IoT security.

C. *Dataset Justification and Availability*

While the dataset was synthetically generated using the NS-3 simulator, it closely mimics real-world IoT traffic patterns through carefully designed simulation parameters. Each device's behavior was modeled based on empirical data from real deployments, ensuring realistic traffic flows, latencies, and attack characteristics. Synthetic data generation is particularly advantageous in cybersecurity research, as it allows controlled experimentation and reproducibility, which are often challenging with real-world datasets due to privacy concerns and data availability. Previous studies have demonstrated the reliability of NS-3 for simulating IoT environments, further supporting its applicability in this study.

To support reproducibility and promote further research, a dataset subset, along with the code for the anomaly detection models, has been made available here. These resources enable researchers to replicate, validate, or adapt the models for similar IoT anomaly detection scenarios. The data generator, dataset and code repository can be accessed on request.

REFERENCES

[1] Sinha, S. (2023) 'State of IoT 2021: Number of Connected IoT Devices Growing 9% to 12.3 Billion globally, Cellular IoT Now Surpassing 2 Billion', IoT Analytics. Available at: https://iot-analytics.com/number-connected-iot-devices/ [Accessed 25 Sep. 2024].

[2] Cloudflare (2024) 'What is a DDoS attack?' Cloudflare. Available at: https://www.cloudflare.com/learning/ddos/what-is-a-ddos-attack/ [Accessed 25 Sep. 2024].

[3] A. Munshi, N. A. Alqarni and N. Abdullah Almalki, "DDOS Attack on IOT Devices," *2020 3rd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, 2020, pp. 1-5, doi: 10.1109/ICCAIS48893.2020.9096818.

[4] R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," *2018 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 2018, pp. 29-35, doi: 10.1109/SPW.2018.00013.

[5] S. Vattikuti, M. R. Hegde, M. Manish, V. Bodduvaram and V. Sarasvathi, "DDoS Attack Detection and Mitigation using Anomaly Detection and Machine Learning Models," *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India, 2021, pp. 1-6, doi: 10.1109/CSITSS54238.2021.9683214.

[6] S. Vattikuti, M. R. Hegde, M. Manish, V. Bodduvaram and V. Sarasvathi, "DDoS Attack Detection and Mitigation using Anomaly Detection and Machine Learning Models," *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India, 2021, pp. 1-6, doi: 10.1109/CSITSS54238.2021.9683214.

[7] M. Saedi et al., "Generation of Realistic Signal Strength Measurements for a 5G Rogue Base Station Attack Scenario," in *Proc. 2020 IEEE Conference on Communications and Network Security (CNS)*, 2020, pp. 1-7, doi: 10.1109/CNS48642.2020.9162275.F`

[8] A. Zielonka, M. Wozniak, S. Garg, G. Kaddoum, M.J. Piran, and G. Muhammad, "Smart Homes: How Much Will They Support Us? A Research on Recent Trends and Advances," *IEEE Access*, vol. 9, pp. 26388–26419, 2021. doi: https://doi.org/10.1109/access.2021.3054575.

[9] A. Kumar, S. V. Rao, and D. Goswami, "NS3 Simulator for a Study of Data Center Networks," *IEEE Int. Symp. Parallel Distributed Computing*, pp. 224-231, 2013.

[10] J. Guo, G. Liu, Y. Zuo and J. Wu, "An Anomaly Detection Framework Based on Autoencoder and Nearest Neighbor," *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, Hangzhou, China, 2018, pp. 1-6, doi: 10.1109/ICSSSM.2018.8464983.

[11] J.-H. Han, Y. Jeon, and J. Kim, "Security considerations for secure and trustworthy smart home system in the IoT environment," *Int. Conf. Information Communication Technology Convergence*, pp. 1116–1118, 2015.

[12] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning DDoS detection for consumer IoT devices," *IEEE Secur. Privacy Workshops*, pp. 29–35, 2018.

[13] L. Özgür, V. K. Akram, M. Challenger, and O. Dağdeviren, "An IoT-based smart thermostat," *Int. Conf. Electrical Electronic Eng.*, pp. 252–256, 2018.

[14] A. Garg, A. Singh, K. Sharma, and V. Sharma, "A taxonomy for IoT in security DDoS attacks," *Int. Conf. Adv. Comput., Communication Control Networking*, pp. 1274-1281, 2022.

[15] A. Aguru and S. Erukala, "A Lightweight DDoS Detection Mechanism in IoT Networks using Entropy and Expectation of Packet Size," *IEEE Int. Symp. Smart Electronic Systems*, pp. 101–106, 2022.

[16] J. Guo, G. Liu, Y. Zuo, and J. Wu, "An Anomaly Detection Framework Based on Autoencoder and Nearest Neighbor," *Int. Conf. Service Systems Service Management*, pp. 1–6, 2018.

[17] N. Mishra and S. Pandya, "Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review," *IEEE Access*, vol. 9, 2021.

[18] Y. Hu and B. Tu, "Security Situation Assessment Model of DDoS Attack Based on Progressive Fuzzy C Clustering Algorithm," *2024 International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2024, pp. 1-4, doi: 10.1109/ICDSNS62112.2024.10691183.

[19] M. A. Salimee, M. A. Pasha, and S. Masud, "NS-3 Based Open-Source Implementation of MQTT Protocol for Smart Building IoT Applications," 2023 International Conference on Communication, Computing and Digital Systems (C-CODE), 2023, pp. 1-6, DOI: 10.1109/C-CODE58145.2023.10139859.