Heterogeneity Tolerance in IoT Botnet Attack Classification

Samuel Kalenowski, David Arnold, Mikhail Gromov, and Jafar Saniie

Embedded Computing and Signal Processing (ECASP) Research Laboratory (http://ecasp.ece.iit.edu)
Department of Electrical and Computer Engineering
Illinois Institute of Technology, Chicago IL, U.S.A.

***Abstract* – Due to the rapid adoption of Internet of Things (IoT) technologies, many networks are composed of a patchwork of devices designed by different software and hardware developers. In addition to the heterogeneity of IoT networks, the general rush-to-market produced products with poor adherence to core cybersecurity principles. Coupled together, these weaknesses leave organizations vulnerable to attack by botnets, such as Mirai and Gafgyt. Infected devices pose a threat to both internal and external devices as they attempt to add new devices to the collective or to perpetrate targeted attacks within the network or against third parties. Artificial Intelligence (AI) tools for intrusion detection are popular platforms for detecting indicators of botnet infiltration. However, when training AI tools, the heterogeneity of the network hampers detection and classification accuracy due to the differences in device architecture and network layout. To investigate this challenge, we explored the application of a Neural Network (NN) to the N-BaIoT dataset. The NN achieved 94% classification accuracy when trained using data from all devices in the network. Further, we examined the model's transferability by training on a single device and applying it to data from all devices. This resulted in a noticeable decline in classification accuracy. However, when considering cyberattack detection the model retained a very high true positive rate of 99.6%.**

Keywords – Internet of Things (IoT), Artificial Intelligence, Botnets, Neural Networks

I. INTRODUCTION

Integration of the Internet of Things (IoT) has accelerated the growth in edge computing and the decentralization of modern computer networks. Utilization of edge computing elements permits faster responsiveness for target applications, such as consumer, commercial, and industrial use cases. IoT devices often trade computational resources for a lower price point in order to encourage adoption and increase network coverage. However, due to limited computing power and sparse encryption capabilities, IoT lacks defensive measures to protect themselves from attack [1]. Further, IoT devices need to be easily accessible via the local network and misconfiguration can easily expose them to the wider Internet. As a result, savvy attackers can use simple attack vectors to compromise a large number of devices. Compromised devices are then organized into a botnet to disrupt operations or facilitate further attacks [2-4]. Botnets are dangerous to both institutions with infected devices and without, as infected devices can be brought offline or directed to attack third-party organizations. Early identification of an attack is paramount to network defense to prevent propagation and prevent further attacks. Additionally, detection should occur at the edge of the network in order to increase responsiveness to attacks.

In order to address these challenges, our research focuses on the implementation of Artificial Intelligence-based Intrusion Detection Systems (IDS) for the edge of the network. Due to the presence of large numbers of unique IoT devices, modern networks have high heterogeneity. AI tools are popular for anomaly detection and are frequently applied in cyberattack detection roles. These tools are trained on a mixture of normal network activity and known attack indicators [5-12]. Training works well with known attacks and network architectures, however, novel attack vectors or changes in the network architecture can often lead to a low detection rate. Due to the presence of large numbers of unique IoT devices, modern networks often have high heterogeneity. High heterogeneity is a challenge for training AI-based IDS as normal network traffic varies greatly between networks. Ideally, these models should be able to maximize their performance on unknown IoT devices while minimizing the number of unique devices required to train the model. To achieve these goals, we evaluated the performance of a Neural Network (NN) under different training conditions.

Through the remainder of this paper, we will present our work regarding the application of Artificial Intelligence towards heterogeneity tolerance in IoT botnet attack classification. In Section II, we will discuss related work and our previous experience in IoT botnet attack detection. Section III presents the N-BaIoT dataset and challenges we encountered while collecting our results. Next, we will discuss our Neural Network model for botnet attack detection and classification in Section IV. Results will be provided in Section V. Finally, we'll wrap up our observations and discuss potential future work in Section VI.

II. RELATED WORKS

Regarding related work, research has been conducted in applying machine learning and artificial intelligence in detecting botnet attacks. The main comparison with our work will be the N-BaIoT paper by Meidan et al [13]. We have selected the dataset used by the authors for our analysis, so it allows us to offer the best comparison on accuracy and performance. During their research, Meidan et al trained a deep autoencoder on nine devices within an IoT network. The researchers observed a very high cyberattack detection rate with low false positive rates. Our research will contribute to the literature by examining a simpler network's ability to achieve similarly high cyberattack detection

rates. Additionally, we will examine the transferability of trained models by training on a single device and applying the model to the other devices.

In addition to the work by Meidan et al, other work has been done regarding cyberattack detection in networks with high heterogeneity. In [14], Zhou et al. considered the application of game theory for attack detection in an IoT network. Their work focused on balancing energy consumption with detection efficiency. A network-based IDS was considered in [15] with an extreme learning model that achieved 97.7% accuracy for cyberattack detection. Finally, an on-device detection system was considered in [16] with a variety of machine learning techniques applied to the network, including support vector machines, neural networks, naïve bayes classifier, and a decision tree.

During our previous work, we examined the application of computer vision as a pre-processing tool for detecting IoT botnet attacks. Computer vision was selected as it allowed us to group temporally related packets for analysis in a lightweight fashion compared to popular models such as Long Short-Term Memory (LSTM). During pre-processing we were able to set the number of packets and features that were included in our images to be processed. After pre-processing, we evaluated the accuracy of a Neural Network, Autoencoder, and Convolutional Neural Network. When examining the results, we considered the accuracy when training over the entire dataset and when training over a specific device. After applying our pre-processing, we observed that the models were transferable between devices within the network. Since this was applied to cyberattack detection only, our future goals were to expand our research to classification as well. This research fits into this overarching objective as we are examining attack classification with only the Neural Network stage. Future work will apply our computer vision pre-processing we used previously.

III. Dataset

For our research into detecting botnet activity, we selected the N-BaIoT dataset [13]. This dataset includes real traffic data from infected IoT devices during the propagation and attack phases of the Mirai and Gafgyt botnets. The recorded traffic was converted from .pcap to .csv using a feature extractor in order to vectorize the data. A total of 115 features are present within the dataset, representing statistical data on the network packets. Statistics were calculated based on streams of data, which represented recent traffic from the packet host. Nine different commercial IoT devices with varied functions were used to generate the dataset, so the effect of device heterogeneity on botnet detection can be studied quite effectively. These devices are listed below in Table I.

While the N-BaIoT paper applied an autoencoder to solely provide binary classification of traffic as either benign or malicious, the dataset was labeled to distinguish between five malicious behaviors for both botnets. These behaviors included a vulnerable device scan and four flooding attacks. The additional attack classification allows us to train and test multi-class classification models in order to distinguish between botnet behaviors. Identifying individual attack patterns from the botnet provides insight into the motivations of the attacker and may influence incident response and recovery. For instance, an attacker focused on expanding the botnet may pose a lower organizational risk compared to an attacker seeing out additional information regarding the network. Classification labels for the dataset are presented in Table II below.

TABLE I. N-BaIoT IoT Devices

ID	Device Name
1	Danmini_Doorbell
2	Ecobee_Thermostat
3	Ennio_Doorbell
4	Philips_B120N10_Baby_Monitor
5	Provision_PT_737E_Security_Camera
6	Provision_PT_838_Security_Camera
7	Samsung_SNH_1011_N_Webcam
8	SimpleHome_XCS7_1002_WHT_Security_Camera
9	SimpleHome_XCS7_1003_WHT_Security_Camera

TABLE II. N-BaIoT Classification Labels

ID	Label Name	Description
0	benign	Uninfected device traffic
1	gafgyt.combo	Combined TCP/UDP flooding
2	gafgyt.junk	Sending spam data
3	gafgyt.scan	Scanning for vulnerable devices
4	gafgyt.tcp	TCP flooding
5	gafgyt.udp	UDP flooding
6	mirai.ack	TCP ACK flooding
7	mirai.scan	Scanning for vulnerable devices
8	mirai.syn	TCP SYN flooding
9	mirai.udp	UDP flooding
10	mirai.udpplain	Optimized UDP flooding

IV. Artificial Intelligence Models

Our botnet detection model was developed as a fully connected neural network. The input layer has 115 nodes, reflecting the number of features in each data point, and the output layer has 11 nodes to match the number of classification labels. Two hidden layers were placed between with widths of 64 and 32. The hyperbolic tangent function was used as the activation function for the first three layers, and the softmax

function was used at the output so that the final vector would represent the prediction of the model as relative probabilities for each class.

Early experiments revealed that the model was prone to overfitting due to the high width of the model necessitated by the large number of attributes in the data. To remedy this, L1 regularization was applied to the first layer. This was intended to drive some of the weights in the input layer to zero since the number of input features is very high and many are highly correlated. The dropout of many nodes and overall reduction of capacity in the first layer caused by this regularization resulted in greatly reduced overfitting. A visualization of our final network is presented in Figure 1.

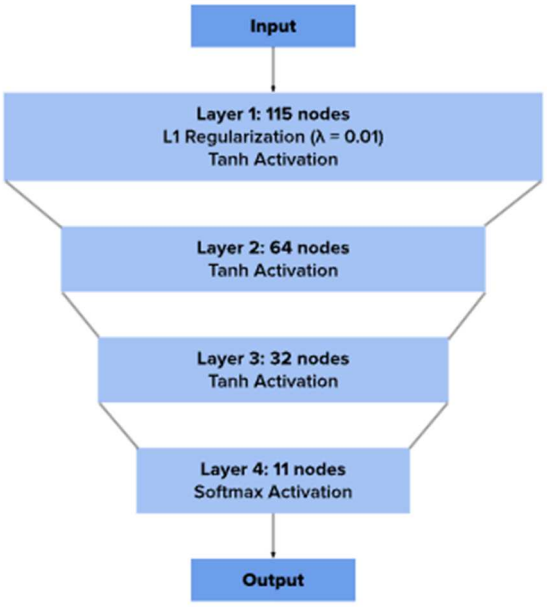

Fig. 1. Neural Network Structure

In order to study the effect of IoT device heterogeneity on the ability of a prediction model to identify botnet activity, five variants of the model were trained and evaluated. The first model was trained on data from the entire set of nine devices and tested on separate samples from these same nine devices. Since the number of samples per device per class in the original dataset varies greatly (from 20,000 to 200,000), it was critical that the training set be generated by choosing an equally sized subset from each label to prevent biasing the model toward any particular prediction. This model served as the baseline case where the prediction was performed on data from devices which were all seen during training. Each of the other four models were instead trained on data from only one of the nine devices and tested on data from the remaining eight. This meant that the prediction accuracy of the model would be evaluated using samples from totally unfamiliar devices. The four devices chosen to train these models were the Danmini Video Doorbell, the Ecobee Thermostat, the Philips B120 Baby Monitor, and the Provision PT-737E Security Camera. These were chosen from the original set of devices to maximize inter-device variability. All models were trained and run on the Nvidia Jetson Nano.

V. Results and Analysis

After completing the training process, we applied our baseline model to our test data and recorded the predictions over the 11 classes. The resulting classifications are displayed as a confusion matrix in Figure 2. The confusion matrix allows us to identify which classes were handled well, and which classes proved difficult to differentiate. Overall, the model was successful at differentiating between most of the classified activity, with near perfect classification between the predicted and observed gafgyt botnet activity. The model performed well with the mirai data, but struggled to differentiate between ACK, SYN, and UDPPLAIN activity from the botnet.

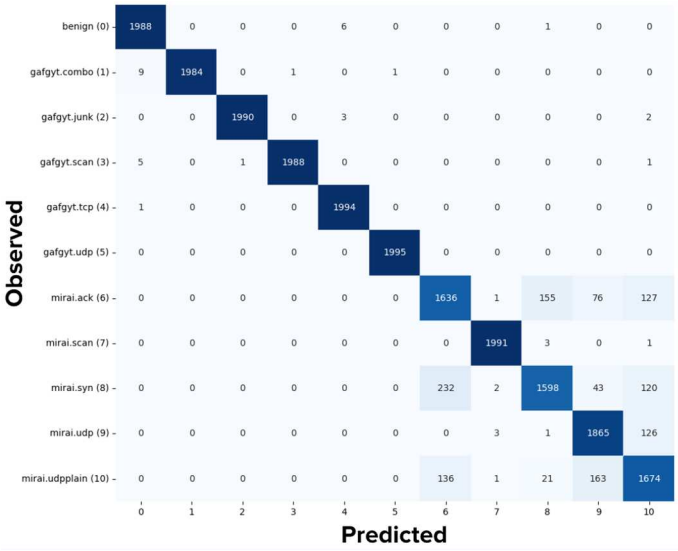

Fig. 2. Confusion matrix for the baseline model. Values in the matrix represent packets labelled by the baseline model.

In addition to examining the per-class predictions of our model, we also wanted to observe the separate benign vs malicious accuracy for general cyberattack detection purposes. For our analysis, we will introduce two new metrics, the True Malicious Rate (TMR) and the True Benign Rate (TBR). The True Malicious Rate represents the accuracy of the model in detecting any of our 10 botnet attack classes while our True Benign Rate represents the model's ability to correctly identify the benign data as benign. When compared to the overall accuracy, which indicated whether a predicted classification was correct, this gives us a more general impression of the pure botnet detection performance while overlooking the less serious mispredictions. In addition to our baseline, we also applied our models trained on a single device. Table III presents our results for our baseline (trained on all devices, training on the baby monitor, training on the thermostat, training on the doorbell, and training on the security camera. Similar to the observations from our confusion matrix, the baseline model was very successful at detecting the attack class and benign data with 94.34% accuracy overall and very strong TMR and TBR values at 99.93% and 99.65%, respectively.

TABLE III. BINARY CLASSIFICATION ACCURACY, TRUE MALICIOUS RATE (TMR), AND TRUE BENIGN RATE (TBR) FOR OUR MODEL

Training Set	Overall Acc.	TMR	TBR
All (Baseline)	94.34%	99.93%	99.65%
Baby Monitor	82.38%	99.95%	91.99%
Thermostat	81.76%	99.98%	80.53%
Doorbell	76.25%	99.60%	69.67%
Security Camera	64.36%	99.45%	64.01%

When we observe the overall accuracy of our trained models, we note that training on individual devices resulted in a significant decline in classification accuracy when compared to the baseline model. For instance, the security camera performed the worst at overall accuracy with a classification accuracy of 64.36%, meaning that it had low heterogeneity tolerance for classification. However, when considering general botnet activity and classifying whether an attack had occurred, the models fared much better. All models had an observed True Malicious Rate at 99%, indicating that our neural network had a high degree of heterogeneity tolerance on applications where catching malicious activity compared to the discrete classification. On the other hand, our True Benign Rate followed the overall accuracy, resulting in a higher rate of false positives among the datasets. Additionally, training with the baby monitor produced the greatest heterogeneity tolerance and shows that the device may be a good representation of generalized IoT activity. Nonetheless, these results indicate that it's possible to train our model using a small sample of different IoT devices and expect good performance when deployed to a diverse IoT environment.

VI. CONCLUSION

Overall, we were successful at developing a neural network model with high heterogeneity tolerance in identifying malicious IoT botnet behavior. When trained against all IoT devices within the N-BaIoT dataset, the model showed high classification accuracy for our 11 classes. Additionally, we showed that training on 4 individual devices yielded True Malicious Rate values of over 99%, with the baby monitor exhibiting the strongest overall classification accuracy as well. For our other devices, we observed a high false positive rate.

During future work we are interested in establishing an IoT testbed to evaluate the model's cyberattack detection and classification capabilities on real-time data and different network configurations. We will also explore additional neural network models in order to determine whether it's feasible to increase the heterogeneity tolerance for the devices outside of the baby monitor.

REFERENCES

[1] M. Gromov, D. Arnold and J. Saniie, "Tackling Multiple Security Threats in an IoT Environment," in *2022 IEEE International Conference on Electro Information Technology (eIT)* , Mankato, 2022.

[2] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, "DDoS in the IoT: Mirai and Other Botnets," *Computer ,* vol. 50, no. 7, pp. 80-84, 2017.

[3] K. Angrishi, "Turning Internet of Things (IoT) into Internet of Vulnerabilities (IoV): IoT Botnets," https://arxiv.org/abs/1702.03681 [cs.NI], pp. 1-17, 2017.

[4] S. Dange, "IoT Botnet: The Largest Threat to the IoT Network," in *Springer International Conference on Computing, Power and Communication Technologies* , Greater Noida , 2019.

[5] M. Almseidin, M. Alzubi, S. Kovacs and M. Alkasassbeh, "Evaluation of Machine Learning Algorithms for Intrusion Detection System," in *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 2017.

[6] D. Arnold, J. Ford and J. Saniie, "Machine Learning Models for Cyberattack Detection in Industrial Control Systems," in *2022 IEEE International Conference on Electro Information Technology (EIT)* , 2022.

[7] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin and W.-Y. Lin, "Intrusion detection by machine learning: A review," *expert systems with applications,* vol. 36, no. 10, pp. 11994-12000, 2009.

[8] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, A. Khan and M. K. A., "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review," *Procedia Computer Science,* vol. 171, pp. 1251-1260, 2020.

[9] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," *Applied Sciences,* vol. 9, no. 20, pp. 4396-4423, 2019.

[10] N. F. Haq, A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah and D. M. Farid, "Application of Machine Learning Approaches in Intrusion Detection System: A Survey," *International Journal of Advanced Research in Artificial Intelligence,* vol. 4, no. 3, pp. 9-18, 2015.

[11] H. Karimipour, A. Dehghantanha, R. M. Parizi, K.-K. R. Choo and H. Leung, "A Deep and Scalable Unsupervised Machine Learning System for Cyber-Attack Detection in Large-Scale Smart Grids," *IEEE Access,* vol. 7, pp. 80778-80788, 2019.

[12] A. Delplace, S. Hermoso and K. Anandita, "Cyber Attack Detection thanks to Machine Learning Algorithms," *https://arxiv.org/abs/2001.06309 [cs.LG],* pp. 1-46, 2020.

[13] Y. Meidan, M. Bohadana, Y. Mathov, A. Shabtai, 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.

[14] M. Zhou, L. Han, H. Lu and C. Fu, "Intrusion Detection System for IoT Heterogeneous Perceptual Network," *Mobile Networks and Applications,* vol. 26, pp. 1461-1474, 2021.

[15] N. Hasan, Z. Chen, C. Zhao, Y. Zhu and C. Liu, "IoT Botnet Detection Framework from Network Behavior based on Extreme Learning Machine," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops*, 2022.

[16] A. Nitish, J. Hanumanthappa, S. P. Shiva Prakash and K. Kiril, "On-device context-aware misuse detection framework for heterogeneous IoT Edge," *Applied Intelligence,* pp. 1-27, 2022.

[17] M. Gromov, D. Arnold and J. Saniie, "Edge Computing for Real Time Botnet Propagation Detection," in *2022 International Conference and Expo on Real Time Communications at IIT (RTC)* , Chicago, 2022.

[18] A. Agniel, D. Arnold and J. Saniie, "Image Processing for Detecting Botnet Attacks: A Novel Approach for Flexibility and Scalability," in *2022 IEEE International Conference and Expo on Real Time Communications at IIT (RTC)*, Chicago, 2022.