

# Poster: A Semi-Supervised Framework to Detect Botnets in IoT Devices

Kashif Naveed<sup>\*</sup>, Hui Wu<sup>†</sup>

School of Computer Science and Engineering, UNSW Sydney, Australia

Email: <sup>\*</sup>mkashifn@gmail.com, <sup>†</sup>huiw@unsw.edu.au

**Abstract**—The number of IoT devices is growing at a rapid pace and the misuse of the shared communication channels has led to a new security challenge caused by botnets. Botnets are compromised IoT devices that are not only able to attack other devices but are also able to spread the infection in the network. In this work, we propose a novel Neural Networks based framework that can detect botnets in IoT devices. The key features of our work include (1) data labelling with minimal supervision with very high accuracy; (2) dynamic network updates to allow learning new attacks not yet discovered; (3) low detection latency to detect such attacks in real-time; and (4) detecting zero-day attacks. The evaluation was done on a dataset containing nine IoT commercial devices infected with BASHLITE Mirai. The experimental results demonstrate the usefulness of the framework providing highly accurate results with low-latency.

**Index Terms**—DDoS, NN, SOM, MSE, MLP, LOF, Deep Learning, Perceptron, Autoencoder

## I. INTRODUCTION

IoT devices at a large scale are being employed everywhere, especially in smart cities, to provide efficient services to the people including, but not limited to: (1) road and traffic management and safety; (2) public transport; (3) water and gas distribution; (4) electricity supply; (5) environment and building structure monitoring; (6) waste and recycling collection and administration; (7) street lighting; and (8) healthcare services. It is estimated that there will be more than 30 billion IoT device deployments by 2050 because more than 55% of 9 billion global population [1] would move to urban areas by then. An interconnection of such a large number of IoT devices is prone to attacks by intruders by making use of anomalous entities, such as botnets. The identification of such anomalous entities in an open research problem that is actively being investigated [2].

### A. Significance of the Problem

Hacked IoT devices can have significant impact on the daily life compared to a hacked email server. As an example, a hacked Smart Grid (SG) can impact lives of a large group of city by paralyzing a whole city [3]. Traditional approaches that work very well for big infrastructures like servers and data centers do not work for IoT devices because of various reasons including their limited processing and memory capabilities.

### B. The Advent of Botnets

Botnets are referred to as interconnected IoT devices that are capable of performing attacks on other devices and even

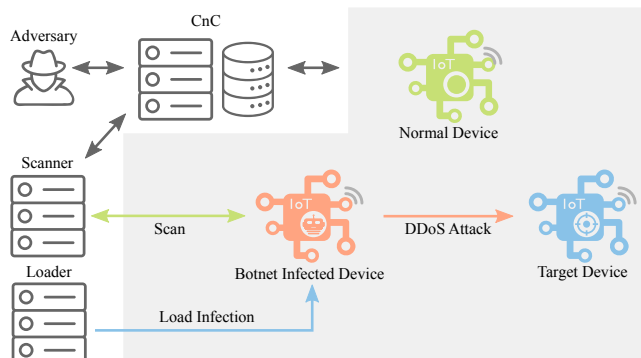


Fig. 1: Proliferation and Infection Caused by Botnets.

infect them to become botnets. An adversary can gain access to those devices by using a central machine, often referred to as a command and control (CnC) as shown in Figure 1. Two of the most common and open-source botnets are BASHLITE<sup>1</sup> and Mirai<sup>2</sup>.

### C. Key Requirements

Detecting anomalies in IoT devices at such a large scale is a challenging task and requires the inclusion of unique features in any system to work effectively. The first and foremost requirement is the capability to learn continuously so that new attacks can be detected. The second requirement is their ability to work without or with minimal expert supervision. Another requirement is the possibility for humans to guide the detection mechanism to limit the error rate. Aside from these functional requirements, the timing requirements mandate the system to perform such detection with the lowest possible latency to minimize the zero-day attacks.

### D. Our Contribution

Deep learning has proven its capabilities in almost every field including medical science, banking, computer vision, trading, real estate and home automation, to name a few. We have combined several deep learning techniques to provide an effective botnet detection mechanism. Our evaluation makes use of commercially available IoT devices infected by open-source botnets. The salient features of our work are: (1) the ability to detect botnets in real-time; (2) zero-day attack detection; (3) unsupervised classification; and (4) the possibility

<sup>1</sup><https://github.com/anthonygtellez/BASHLITE>

<sup>2</sup><https://github.com/jgamblin/Mirai-Source-Code>

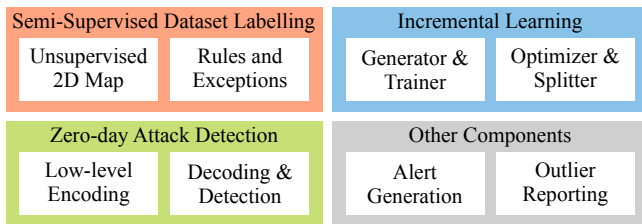


Fig. 2: Various Components of the Proposed Framework.

for experts to provide supervision to the learning process to minimize the false positive and false negative rates.

## II. BACKGROUND AND THREAT MODEL

Extremely powerful and sophisticated techniques exist that can attack IoT devices at a large scale. Such attacks can cause enormous amounts of damage in different ways including impacting the economy in a bad way. Popular embedded processors including ARC, PowerPC, x86, ARM and MIPS have been targeted by *Okiru*, an ELF malware [4]. These processors have been in use inside most common IoT devices as a system on chip (SoC).

A vulnerability in commercial internet routers has been found recently that can bypass the authentication mechanism and allow execution of an arbitrary piece of code remotely [5]. There have also been findings of widely used internet routers that can help gain access to millions of IoT devices by exploiting zero-day attacks even when their known control channels are blocked [6].

Our work assumes interconnected IoT devices in the presence of adversaries and anomalous sensor devices infected with botnets. Unlike other anomaly detection schemes, we do not assume the anomalous entities to be much smaller than normal devices. Our threat model includes the possibility of distributed denial-of-service attacks (DDoS) performed by infected devices that can flood the network to disrupt normal communication. Additionally, our work can work with the situations where such anomalies propagate within the network. This propagation is a key part of the botnets as shown in Figure 1.

## III. METHODOLOGY

We provide a framework that comprises of different sub-systems dedicated to performing various tasks to achieve high accuracy and low latency. Figure 2 presents the components of our framework and their brief descriptions are provided as follows:

- 1) **Semi-Supervised Dataset Labelling:** We combine unsupervised Self Organizing Maps (SOM) with expert-supplied rules to divide the dataset into normal and attack instances with very high accuracy.
- 2) **Incremental Learning:** Our work employs a continuous learning mechanism that is capable of keeping up with the emerging attacks that are not discovered yet.
- 3) **Zero-day Attack Detection:** We make use of Autoencoders to provide zero-day attack detection in real-time.

This enables administrators and experts to minimize the associated risks by making countermeasures promptly.

## IV. KEY ALGORITHMS

In this section, we present the key algorithms that are developed as part of our framework.

### A. Semi-Supervised Dataset Labelling

Kohonen map, or commonly known as Self-Organizing Map (SOM) [7], produces a low-dimensional, usually two, for a  $k$ -dimensional input. SOM is a type of unsupervised neural networks that is capable of providing separation of normal and anomalous entities without human intervention. SOM is inspired by *topographic map*, a principle used in neurobiology, to make artificial neurons *compete* with each other to *win* a certain position. Once all the multivariate dataset has been represented as a 2-D map of neurons, the binary classification can be done by applying a threshold to the *mean interneuron distance* (MID). This technique, without human supervision, does not produce a highly accurate output. Our semi-supervised dataset labelling algorithm makes use of a small subset of data labels to improve accuracy. Since this approach does not require manual labelling of the entire dataset, we can achieve a highly accurate classification output with minimal expert supervision.

### B. Incremental Learning

The incremental learning mechanism starts with building a neural network structure on the given data. The neural network selection algorithm evaluates several different structures defined by varying the *hyperparameters* (e.g., number of hidden units i.e., the layer count and the number of neurons within them) and evaluating their performance and selecting the best network. Once the best network has been chosen, it is iteratively updated to keep up with the new data. The dynamic network update algorithm keeps tuning the hyperparameters and even network split can occur when it finds that a single model is not enough to correctly classify all the devices in the network.

### C. Zero-day Attack Detection

The zero-day attack detection algorithm makes use of only the normal instances of data to extract a low-level representation by making use of a special type of neural networks called autoencoders. These networks generate a high reconstruction error even the input data containing both types of normal and anomalous instances are very close to each other. This mechanism guarantees a low-latency detection with very high accuracy.

## V. EVALUATION

In this section, we present the dataset used and the experimentation carried out to evaluate various aspects of our framework. Please note that we have made the evaluation data publicly available at Kaggle<sup>3</sup>.

<sup>3</sup><https://www.kaggle.com/mkashifn/iot-botnet-detection>







Doorbell	PnV Camera	Baby Monitor
 Danmini Ennio	 Provision PT737E PT838	 Philips B120N10
Thermostat	Security Camera	Webcam
 Ecobee	 Simple Home XCS71002 XCS71003	 Samsung SNH1011N

Fig. 3: Nine Commercial Devices Used in the Experiments.

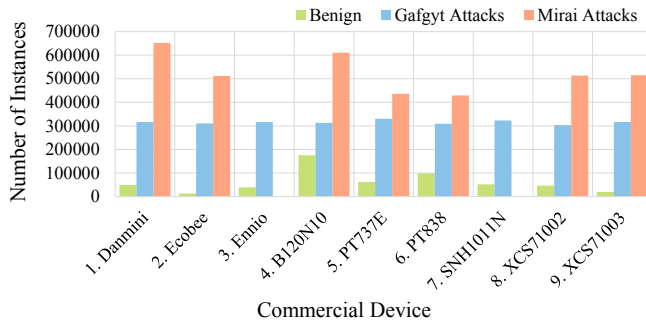


Fig. 4: Benign and Attack Data Attributes.

#### A. Dataset Description

Meidan et al. [8] contributed a dataset comprising of nine commercially available devices infected with two botnets. The dataset is publicly available at the University of California Irvine Machine Learning Repository [9]. To make the data readily usable for the researchers, we have re-organized the files with a consistent naming structure and uploaded to Kaggle<sup>4</sup> as a public dataset.

Figure 3 presents the details about the nine commercial devices used in the experiment. The dataset contains instances of benign traffic as well as BASHLITE and Mirai attack data as shown in the chart presented in Figure 4. Each instance contains 115 traffic characteristics used as features by the neural network. The attacks contain: (1) SCAN commands to discover the vulnerable devices; (2) ACK, SYN, UDP and TCP flooding; and (3) combo attacks opening connections and sending spam data.

#### B. Experimental Results

We compared the results of our framework with: (1) N-BaIoT, an autoencoder based anomaly detection framework [8]; and (2) three commonly used anomaly detection algorithms named LOF, SVM and Isolation Forest [10].

As you can see from Figure 4, the dataset is imbalanced, i.e., the normal and anomalous classes are represented unequally. This means that a single measure of accuracy does not correctly reflect the quality. To overcome this challenge, we have included two other metrics: (1) false-positive rate (FPR); and (2) false-negative rate (FNR). Another aspect is to measure the system’s performance in terms of detection latency. Figure 5

<sup>4</sup><https://www.kaggle.com/mkashifn/baiot-dataset>

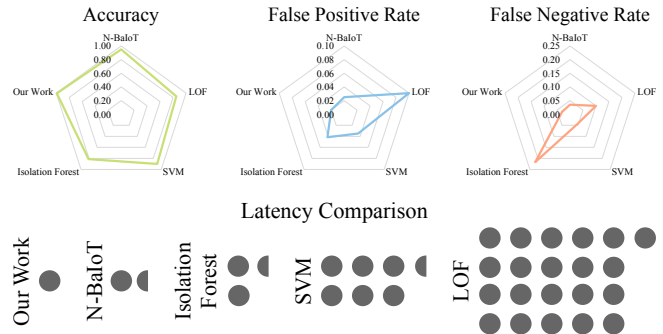


Fig. 5: Comparison of Accuracy, FPR, FNR and Latency.

compares these metrics and as you can see that our work achieves the highest accuracy and lowest false positive and false negative rates. Additionally, a small detection latency ( $150 \pm 80$  msec) is offered by our system compared to the other techniques.

## VI. CONCLUSION

We aimed to develop a novel botnet detection framework for IoT devices. Unlike other anomaly detection approaches, this framework does not assume that anomalies constitute a small portion of the entire dataset. This framework keeps learning over time and provides effective detection for attacks to be known in future. The experimental results prove the effectiveness in terms of accuracy and speed. It is an autonomous system that does not require manual labelling for the entire dataset and provides flexibility for the experts to supervise the learning progress for robust and reliable operation.

## REFERENCES

- [1] D. Evans, “The internet of things: How the next evolution of the internet is changing everything,” *CISCO white paper*, vol. 1, no. 2011, pp. 1–11, 2011.
- [2] M. Naveed and H. Wu, “Begonia: An efficient and secure content dissemination scheme for smart cities,” in *2019 IEEE Wireless Communications and Networking Conference (WCNC 2019)*. Marrakech, Morocco: IEEE, Apr. 2019.
- [3] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, “Cyber-physical security of a smart grid infrastructure,” *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.
- [4] R. Joven and D. Maciejak. (2018) Iot botnet: More targets in okiru’s cross-hairs. [Online]. Available: <https://www.fortinet.com/blog/threat-research/iot-botnet-more-targets-in-okirus-cross-hairs.html>
- [5] R. Millman. (2018) Satori creator linked with new mirai variant masuta. [Online]. Available: <https://www.scmagazineuk.com/satori-creator-linked-new-mirai-variant-masuta/article/1473395>
- [6] D. Goodin. (2017) 100,000-strong botnet built on router 0-day could strike at any time. [Online]. Available: <https://arstechnica.com/information-technology/2017/12/100000-strong-botnet-built-on-router-0-day-could-strike-at-any-time/>
- [7] T. Kohonen, *Self-Organizing Maps*. Springer-Verlag Berlin Heidelberg, 2001, vol. 30, no. 3.
- [8] Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, 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.
- [9] D. Dua and C. Graff, “UCI machine learning repository,” 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [10] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, “Deep learning for unsupervised insider threat detection in structured cybersecurity data streams,” in *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*, 2017.