# MBM-IoT: Intelligent Multi-Baseline Modeling of Heterogeneous Device Behaviors against IoT Botnet

Jianyu Wang
*University of Missouri-St. Louis*
St. Louis, Missouri, USA
jwgxc@mail.umsl.edu

Jianli Pan
*University of Missouri-St. Louis*
St. Louis, Missouri, USA
pan@umsl.edu

***Abstract*—Recent researches have applied various machine learning models to detect IoT botnet attacks. However, the heterogeneity of IoT devices' normal and attack behaviors was not well addressed, which resulted in high false positive/negative detection rates. To solve this issue, we propose a method that builds individual behavior baselines for different types of devices with a single Conditional Variational Autoencoder model, and then detects attacks with even minor deviations from the baselines. The evaluation results on the public N-BaIoT dataset show that our method outperforms the others with accuracy higher than 99.9% while introducing limited extra computational cost.**

***Index Terms*—IoT security, botnet attack detection, heterogeneous IoT devices, conditional variational autoencoder**

## I. Introduction

Due to computation resource constraints, many IoT devices cannot afford strong on-host security mechanisms. Thus, they expose various vulnerabilities to be compromised and controlled as botnets that launch Distributed Denial of Service (DDoS) attacks against other critical Internet facilities [1].

Recent researches have applied machine learning models to learn the data distribution of IoT devices' normal behaviors and detect attacks as anomalies [2]–[5]. However, few works have well addressed the challenges of increasing heterogeneity of IoT devices, which degrade the detection accuracy. First, the normal behavior patterns of IoT devices are distinct between each other due to their specialized functions (e.g., voice speakers and web cameras). Second, it is highly possible that the attack behavior of one device is similar to the normal activity of another device. Unfortunately, the existing works usually assume a single data distribution for all devices, which lead to suboptimal model learning.

In this paper, we propose a multi-baseline modeling scheme (MBM-IoT) that employs a Conditional Variational Autoencoder (CVAE) [6] to build distinct behavior baselines for each type of IoT devices efficiently in a single learning process. Subsequently, we design a two-factor detection algorithm that jointly utilizes reconstruction error (RE) and Kullback-Leibler divergence (KLD) loss functions of CVAE to identify attacks with even minor deviations from the learned baselines.

## II. MBM-IoT Botnet Attack Detection

MBM-IoT consists of two key designs: multi-baseline device behavior modeling and two-factor attack detection, as shown in Fig. 1.

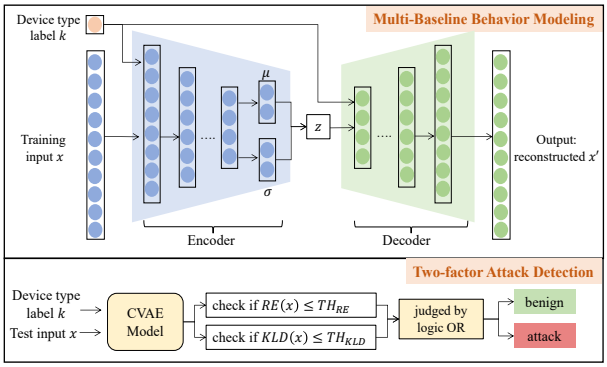


Fig. 1: Overview of MBM-IoT.

### A. Multi-Baseline Behavior Modeling

In order to build behavior baselines separately, we first conduct device type labeling on the data collected from different IoT devices. The type of a device is determined by its functionality, such as smoke detectors, voice assistants, web cameras, and etc. In this way, a behavior sample is profiled as $(x|k) = ([x_1, x_2, ..., x_n]|k)$, where $x_i$ is feature values and $k$ is device type.

The modeling process is formulated as a multimodal distribution learning problem, where each mode is the behavior data distribution of one device type. Our goal is using one CVAE model to learn behaviors of multiple types of devices in a joint training process. The model inputs are $x$ and $k$, and the output is the reconstructed feature values $x'$ conditioned to $k$. The learning objective of CVAE is to maximize the ensemble loss function of reconstruction error (RE) and Kullback-Leibler divergence (KLD) as below:

$$L_{CVAE}(x,k) = E[\log P(x|z,k)] - D_{KL}(Q(z|x,k) \parallel P(z|k))$$

The first item $E[\log P(x|z,k)]$ is the expectation of log-likelihood between $x$ and $x'$, which encourages the decoder of CVAE to reconstruct $x$ from its latent space variable $z$ generated by the encoder. Maximizing $E[\log P(x|z,k)]$ is equal to minimize the $RE(x,x')$, which is calculated by mean square error. The second item $D_{KL}(\cdot)$ represents KLD, which estimates the divergence between encoder's learned distribution $Q(z|x,k)$ and expected distribution $P(z|k)$.

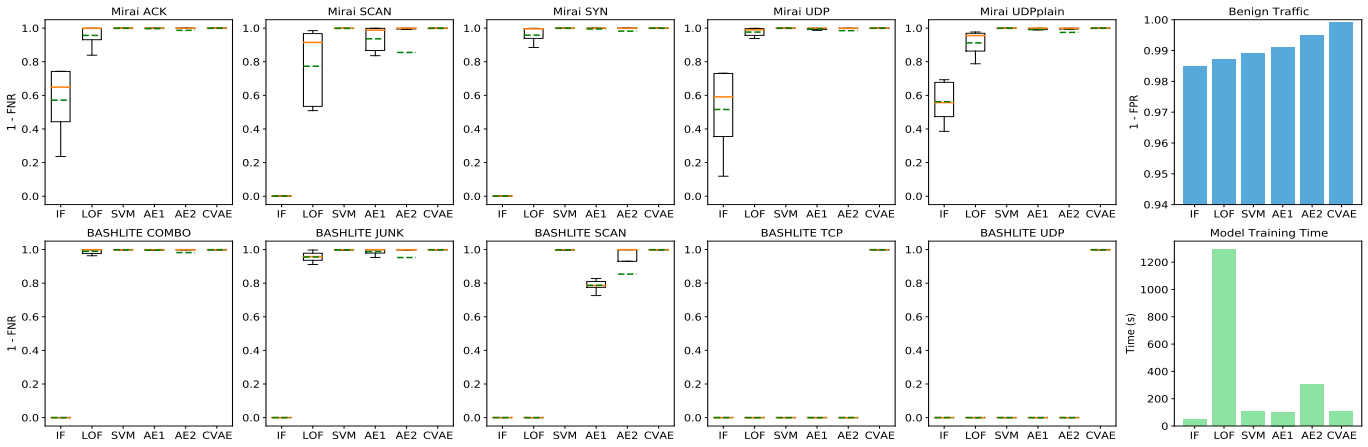


Fig. 2: The experiments results of all the detection models: $(1 - FNR)$ against ten botnet attacks (left five charts in the 1st and 2nd rows), $(1 - FPR)$ on benign traffic (top-right), and training time (bottom-right).

### B. Two-factor Attack Detection

Once the modeling is completed, we employ CVAE to evaluate the abnormality of new behavior samples. Based on their magnitude of deviations from benign baselines, RE and KLD are jointly utilized as metrics. Relatively, RE is more sensitive to macro deviations on features values, and KLD can capture minor but systematic deviations. Thus, we propose a two-factor detection algorithm as follows: (1) defining the anomaly thresholds of RE and KLD for each device type based on the training data: $TH_{RE} = mean(RE) + 3 * std(RE)$ and $TH_{KLD} = mean(KLD) + 3 * std(KLD)$. (2) measuring the new samples $x$'s $RE(x)$ and $KLD(x)$. (3) check if either $RE(x)$ or $KLD(x)$ is larger than their thresholds. If yes, $x$ is detected as attack. Otherwise, $x$ is benign.

## III. Evaluation of Detection Accuracy and Cost

We use the public N-BaIoT [4] dataset for evaluation, which contains ten classes of botnet attacks collected from nine IoT devices. False Positive Rate (FPR) and False Negative Rate (FNR) are adopted as accuracy metrics, where FPR indicates the proportion of benign samples predicted as attacks and FNR is the proportion of attacks predicted as benign. The performance of MBM-IoT is compared with five other methods, including three conventional machine learning models that train a single IF, LOF, or one-class SVM model for all devices (without differentiate device types) [2], and two Autoencoder (AE)-based models that train a single model for all devices (denoted as AE1) [3] or train one model per device (denoted as AE2) [4]. All the experiments are done on a desktop with a 3.6 GHz 4-core CPU and 16 GB RAM.

Fig. 2 shows the results, where the green (dash) and orange (solid) lines indicate mean and median values, respectively. We observe that CVAE performs the best against both benign samples and the ten classes of attacks and (lower than 0.01 FPR and FNR in both cases). The other models either have big variances of accuracy to detect different attacks (e.g., IF, LOF, AE1), or are not capable of detecting some specific attacks like BASHLITE TCP flooding with minor deviations from device normal behaviors (e.g., SVM, AE1, AE2). Besides, the benign traffic detection rates of the first four models are influenced by behavior heterogeneity, while AE2 has weaker learning ability than CVAE. Furthermore, the training of CVAE completes within 107.62s, which indicates that our model does not introduce much computation cost compared to other methods.

## IV. Conclusion

In this paper, we proposed a novel IoT botnet attack detection method named MBM-IoT that addressed the challenges of heterogeneous IoT device behaviors. First, we employed CVAE to build individual baselines for different types of devices. Then, we jointly utilized RE and KLD loss functions to detect attacks with macro or minor deviations from the baselines. Evaluation results demonstrated MBM-IoT's superior detection accuracy and cost compared to five well-known machine learning models over the public N-BaIoT dataset. Future works include utilizing more behavior features and more types of IoT devices to further validate the performance of our method.

## References


[1] F. Meneghello, M. Calore, D. Zucchetto, M. Polese and A. Zanella, "IoT: Internet of Threats? A Survey of Practical Security Vulnerabilities in Real IoT Devices," in IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8182-8201, Oct. 2019

[2] M. Eskandari, Z. H. Janjua, M. Vecchio and F. Antonelli, "Passban IDS: An Intelligent Anomaly-Based Intrusion Detection System for IoT Edge Devices," in IEEE Internet of Things Journal, vol. 7, no. 8, pp. 6882-6897, Aug. 2020.

[3] Y. Mirsky, T. Doitshman, Y. Elovici and A. Shabtai, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection", in 2018 Network and Distributed System Security Symposium, San Diego, CA, USA.

[4] Y. Meidan et al., "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders," in IEEE Pervasive Computing, vol. 17, no. 3, pp. 12-22, Jul.-Sep. 2018.

[5] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan and A. Sadeghi, "DÏoT: A Federated Self-learning Anomaly Detection System for IoT," 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, pp. 756-767.

[6] K. Sohn, H. Lee and X. Yan, "Learning structured output representation using deep conditional generative models," Advances in neural information processing systems 28 (2015): 3483-3491.