International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.061 ISSN: 2581-4621

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

Nila Roy[1], Sunandha Rajagopal[2], Bharath Vinod[3], Mohammed Ashique[4], Jefna Nazar[5]
[1,3,4,5]MCA, Department of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala,India
[2](Assistant Professor), Department of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery,Kerala,India

***Abstract* —** Now we are in an era of IoT. Lot of IoT hardware are ruling current market. Since we are connecting through Internet, security is a concern. Attacking to the system is big issue in this market. So lot of researches are going on this. To identify IoT botnet attack we are proposing a new method. First we generate individual behaviour baselines for different types of devices with Single conditional Variational Auto encoder model. Then detect with even minor deviations from baselines.

Key Points
- **Botnet**
- **Computational Modelling**
- **Machine Learning**
- **Computational Efficiency**
- **Computer Network Security**
- **Botnet Attack**
- **Heterogeneous IoT Devices**
- **IoT Security**

I. INTRODUCTION

Many Internet of Things (IoT) devices cannot afford robust on-host security techniques due to resource limitations in processing. As a result, they make numerous vulnerabilities vulnerable to being exploited and controlled as botnets that attack other crucial Internet infrastructure using Distributed Denial of Service (DDoS) attacks.

Machine learning models have been used in recent studies to detect assaults as anomalies by learning the data distribution of IoT devices' typical behaviour. The issues posed by the growing variety of IoT devices, which reduces the detection accuracy, have only been adequately addressed in a small number of studies. First, because of their unique purposes, IoT devices' typical behaviour patterns differ from one another (e.g., voice speakers and web cameras). Second, there is a good chance that a device's attack behaviour will be identical to another device's routine operation. Unfortunately, the previous works frequently assume that all devices have the same data distribution, which results in suboptimal model learning.

In this research, we present a multi-baseline modelling scheme (MBM-IoT) that uses a Conditional Variational Auto encoder (CVAE) to quickly create separate behaviour baselines for each type of IoT devices. Then, to detect attacks that deviate even slightly from the learned baselines, we develop a two-factor detection technique that combines the reconstruction error (RE) and Kullback-Leibler divergence (KLD) loss functions of CVAE.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.061 ISSN: 2581-4621

II. MBM-IOT BOTNET ATTACK DETECTION

MBM-IoT is comprised of two key designs: multi-baseline de- vice behaviour modelling and two-factor attack detection, as illustrated in Fig. 1.

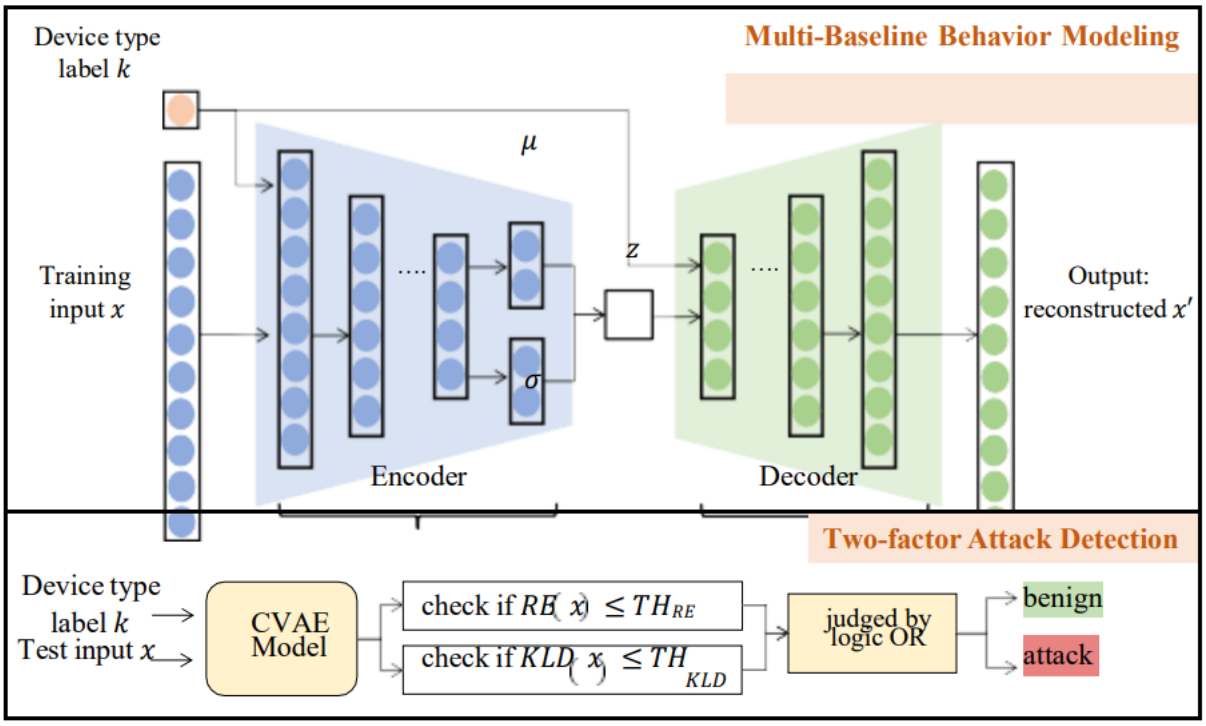

Fig. 1: Overview of MBM-IoT.

A. Multi-Baseline Behaviour Modelling

We initially device type classify the data gathered from various IoT devices in order to create behaviour baselines individually. The functionality of a gadget, such as that of a smoke detector, voice assistant, web camera, etc., determines its type. A behaviour sample is profiled in this way: (x k) = ([x1, x2,..., xn] k), where xi is a set of feature values and k is the type of device.
Each mode represents the behaviour data distribution of a particular device type, and the modelling procedure is conceptualized as a multimodal distribution learning problem. Our objective is to jointly train a single CVAE model to learn the behaviors of many device kinds. The reconstructed feature values x′ conditioned to k are the model output. The model's inputs are x and k. The CVAE learning goal is to optimize the ensemble loss function of Kullback-Leibler divergence (KLD) and reconstruction error (RE) as shown below:

LCV AE(x, k) = E[log P (x|z, k)]−DKL(Q(z|x, k) ‖ P (z|k))

The expectation of log-likelihood between x and x′ is the first item in E[log P (x z, k)], which motivates the CVAE decoder to reconstruct x from its encoder-generated latent space variable z. In order to reduce the RE(x, x′), which is determined by mean square error, one must maximize E[log P (x z, k)]. The second component, DKL(), stands for KLD, which calculates the difference between the predicted distribution P (z|k) and the learnt distribution Q(z|x, k) of the encoder.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.061 ISSN: 2581-4621

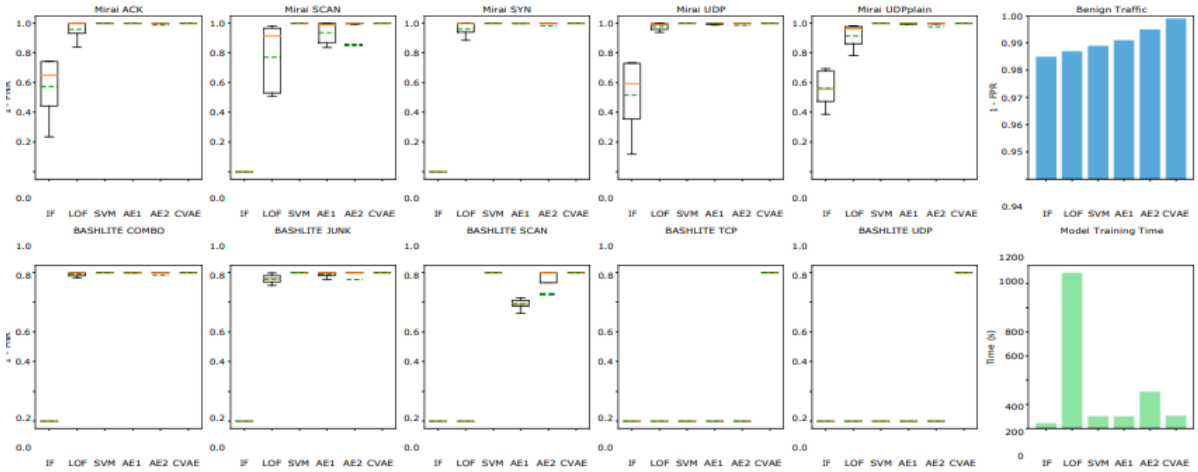

Fig. 2: The studies' findings for each detection model include training time, (1 FPR) on benign traffic (top-right), and (1 FNR) against ten botnet attacks (bottom-right).

B. Two Factor Attack Detection

Following modelling , we use CVAE to assess the abnormality of fresh behaviour samples. RE and KLD are used together as measurements based on how much they deviate from safe baselines. In comparison, KLD can catch small but systematic variations while RE is more sensitive to macro deviations in feature values. As a result, we suggest the following two-factor detection algorithm: (1) using the training data to define the anomaly thresholds of RE and KLD for each device type: THRE = mean(RE) + 3 std(RE) and THKLD = mean(KLD) + 3 std (KLD). (2) calculating the RE(x) and KLD of the new samples x. (x). (3) Determine if RE(x) or KLD(x) exceeds their respective thresholds. If so, x is recognized as an attack. Otherwise, x is acceptable.

III. EVALUATION OF DETECTION ACCURACY AND COST

For evaluation, we use the ten kinds of botnet assaults gathered from nine IoT devices in the public N-BaIoT dataset. As accuracy measures, false positive rate (FPR) and false negative rate (FNR) are used, where FPR represents the percentage of benign samples that are incorrectly identified as attacks and FNR represents the percentage of incorrectly identified attacks. Five other methods are used to compare the performance of MBM-IoT, including three traditional machine learning models that train a single IF, LOF, or one-class SVM model for all devices (without differentiating between device types) and two Auto encoder (AE)-based models that train a single model for all devices (denoted as AE1) or train one model per device (denoted as AE2). On a desktop computer with a 3.6 GHz 4-core CPU and 16 GB of RAM, all trials are conducted.

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 lack the ability to detect some specific attacks like BASHLITE TCP flooding with only little variations from device usual behavior, or have large variances in accuracy to detect various attacks (e.g., IF, LOF, AE1) (e.g., SVM, AE1, AE2). In addition, behaviour variability affects the first four models' rates of benign traffic identification, and AE2 is less capable of learning than CVAE. Additionally, CVAE training is finished in 107.62 seconds, showing that our model has a lower computational cost

DOI:10.46647/ijetms.2022.v06i05.061 ISSN: 2581-4621

than existing approaches.

CONCLUSION

In this paper, we presented MBM-IoT, an innovative IoT botnet attack detection method that addresses the challenges of heterogeneous IoT device behaviours. First, we used CVAE to create individual baselines for various types of devices. Then, we used the RE and KLD loss functions together to detect attacks with large or minor deviations from the baselines. MBM-IoT outperformed five wellknown machine learning models in terms of detection accuracy and cost over the public N-BaIoT dataset. Future work will include incorporating more behaviour features and IoT device types 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 Vulnerabilitiesin 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, “D¨IoT: 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.