IoT Botnets anomalies detection using machine learning auto-encoders

Vijay Kumar Khatri
PG Scholar, Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India

Dr. Vandana Roy
Professor, Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India

Abstract

This study focuses on Machine Learning techniques for Internet of Things security threats detection. It seeks to investigate the feasible of using auto-encoders to detect IoT botnets. Botnets can develop Distributed Denial-of-Service (DDoS) attacks and present a major security concern in IoT networks, as there is no single method has demonstrated the potential to address this security threat. These methods often fail to meet Internet of Things (IoT) environments requirements, such as processing power and energy consumption. Auto-encoders offers one of the solutions to botnet detection. Future research needs to explore the opportunities that auto-encoders present inthe detection of IoT botnets.

Keywords: Machine Learning Detection, IoT, Botnets, DDoS

1. INTRODUCTION

The Internet of Things (IoT) is the interconnection of computer power with appliances, and it has the potential to displace the market for goods and applications that increase data creation, consumption, and device number [1]. IoT security vulnerabilities include non-user interfaces, insecure interaction protocols, sensitive data modification [2], middleware layer weaknesses, a plurality of assaults, and a lack of storage capacity [3]. Security is difficult to maintain against complex threats, which is why Intrusion Detection Systems (IDSs) based on machine learning algorithms detect rather than prevent. However, limited CPU and storage resources make applications challenging for use as embedded systems for smart cities [4], particularly botnets and IoT malware assaults via Distributed Denial of Service (DDoS). Having the capacity to conduct complex network-wide assaults like HTTP floods, the Mirai DDoS and botnets. For detection tactics, attack mutation and evolution are a worry, which may be addressed by ML and deep learning, which is also the goal of this work.

However, some security methods and features have a more limited scope for use due to the unique characteristics of IoT contexts, such as their constrained computation and storage capabilities. In the context of IoT, this study investigates the use of Machine Learning (ML)-based detection techniques and defences.

2. REVIEW OF LITERATURE

The focus of this part will be on describing IoT security flaws and botnet/DDoS assaults, followed by a description of machine learning detection of IoT botnets and DDoS, and finally a discussion of anomaly detection of IoT botnets using auto-encoders.

2.1 IoT SECURITY VULNERABILITIES AND BOTNET/DDOS ATTACKS

Numerous studies have examined the IoT security architecture, which is divided into four separate levels with various security implications: the application layer, the network layer, the device layer, and the service/application support layer [5]. These research all agree that there are different security flaws in each tier. Although a number of researchers have put out several IoT security architectural models [6], they all concur that no one IoT model can provide the best protection against all threats. Particularly vulnerable to security risks include assaults such as selective forwarding, Sybil attacks, Man-in-the-Middle (MiTM) attacks, and denial-of-service (DoS) attacks [7]. Botnets and DDoS assaults are the two types of attacks that draw the most attention, maybe because they have the potential to jeopardise the availability of information systems [8,9]. Table 1 lists probable attack methods and IoT security flaws.

Table 1. IoT security vulnerabilities and potential attack vectors [10]

IoT vulnerabilities	Attack vectors
Unreliable IoT interfaces	Weak credentials
Inadequate authorisation and authentication	Insecure login credentials
Insecure network services and software	IoT device attack and malware dissemination vector
Weak physical security	Unauthorized access to OS can be made possible through ports, SD cards, and storage media

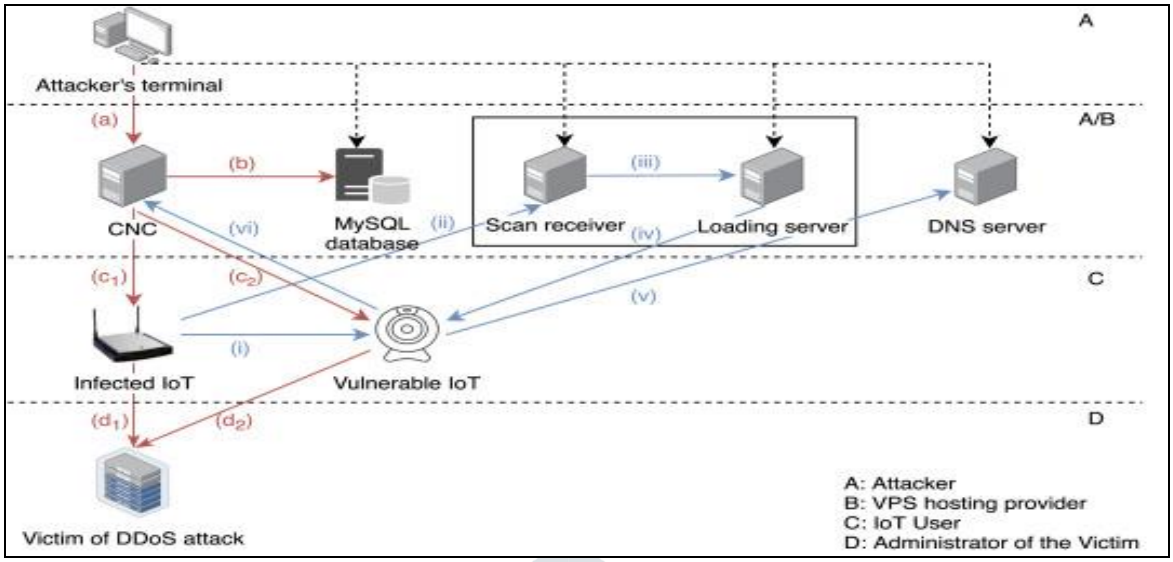

Figure 1. Typical IoT botnet ecosystem [11]

The two most significant botnet malware families that have the ability to escalate into DDoS assaults in IoT networks are Mirai and Bashlite [11,12]. Bashlite and Mirai both take use of known flaws in login credentials. The components of a typical IoT botnet ecosystem include a command and control server, which offers the attack interface, scanners for probing devices, bots or infected devices, loaders for logging into susceptible devices, malware servers, and databases as shown in Figure 1. As shown in Fig. 1, an attacker can start a DDoS assault by submitting a specific command over Telnet from a Remote Terminal to the Computerized Numerical Control (CNC) server (step a). The target of the attack is then delivered to the infected IoT devices (or bots) in step c_1 after the command has been simultaneously recorded historically on the MySQL database server (step b). The already active bots would then comply with the CNC's directive and deliver a deluge of network packets to the designated victim server (step d_1).

A compromised IoT device may also search the network from a variety of IP addresses for other compromised IoT devices (step i). The bot would notify the Scan Receiver of any discovered vulnerable devices (here, "vulnerable" refers to IoT/Linux systems with shoddy Secure Socket Shell (SSH) and Telnet user credentials) by reporting the IP address, user credential, kind of service, and other relevant information (step ii). The details about the susceptible device would be proactively gathered by the Loader as soon as a fresh report is received. As seen in Fig. 1, the Scan Receiver and the Loader were thought to be on the same computer because, by default, the Scan Receiver would add the information from the susceptible device to the Operating System's Standard Output stream (or stdout), which is always being watched by the Loader (step iii).

The Loader would then access the exposed device and upload the malicious software (step iv). The newly infected IoT device will then be configured as a new bot, which will then need to register with the CNC server (step vi). However, there is one phase that we emphasize here that was largely disregarded by the study that has already been done and is crucial for forensic investigators. In order to interact with the CNC server, the susceptible device must obtain its IP address from a hard-coded DNS server (step v). The same thing occurs when an infected device has to communicate with the Scan Receiver. This configuration allows the attacker to relocate all other servers to a new IP address while the DNS server is still operational.

The purpose of this investigation is to locate the owner of the machines powering the servers in Fig. 1. The forensic examination of compromised IoT devices and DDoS attack victims is outside the purview of this study. Instead, we concentrate on the systems in the attacker's control (regions "A" and "A/B").

2.2 IoT BOTNET AND DDOS DETECTION USING MACHINE LEARNING

Botnet detection techniques mentioned in the literature either use particular operational stages or detection techniques. The operational processes for IoT-related botnet detection have been addressed in several research, with an emphasis on ideas like Software Defined Networking (SDN) collaboration schemes [13] and the use of discriminating functions [14]. These techniques concentrate on the initial phases of attack execution and dissemination within the C&C server. IoT network botnets change quickly, therefore mutated attack tools could get past current detection methods. Putting your attention on the next phases in the IoT botnet activities is a workable answer to this issue. Network-based botnet detection techniques are required for this.

In [15], the researchers suggested a classification scheme for network-based botnet detection systems based on the methods of detection, the sources of detection, and the algorithms for detection. Network-based detection approaches may employ fingerprint-based detection methods or anomaly-based detection, which concentrates on botnet activity and protocol behaviour. According to detection strategies, the two main types include detection approaches that rely on botnet sources like honeypots, virtual networks, and simulated solutions, as well as detection approaches that rely on regular sources like virtual internal networks and actual networks. Six techniques may be used in the detection algorithms: instance-based learning, signal processing, supervised learning, unsupervised learning, and semi-supervised learning [15]. Algorithms such as the Hidden Markov Model (HMM), Bayesian statistics,

Artificial Neural Networks (ANNs), Decision Trees, and Support Vector Machine are used in supervised detection (SVM). Clustering strategies are the major focus of supervised detection systems. The evaluation concentrated on several detection techniques, although auto-encoders were not mentioned.

2.3 IoT BOTNET ANOMALY DETECTION USING AUTO-ENCODERS

Recent years have seen a rise in scholarly interest in auto-encoders as prospective cybersecurity tools [16,17]. Although auto-encoders have been extensively explored, little study has been done on how to use them in IoT environments. The use of auto-encoders has been demonstrated by a number of researchers despite this [18]. In order to identify anomalies in Wireless Sensor Networks (WSNs) integrated in Internet of Things (IoT) settings, Luo and Nagarajan [18] introduced auto-encoder neural networks. The detection method was made up of two parts, one of which was housed inside sensors and the other on the IoT cloud. The assessment proved that the auto-encoder neural network's unsupervised learning capabilities allowed for adaptability to unanticipated changes in IoT networks. The "Vanilla Deep Neural Net (DNN)", "Self-Taught Learning (STL)", and "Recurrent Neural Network (RNN)" have all been used in attempts to evaluate various deep learning models for network intrusion detection [19]. The results showed that the STL identification model was reliable in settings with dirty data, pointing to its applicability in the IoT.

Table 2 compares several auto-encoder-related strategies that have been suggested, outlining the benefits and drawbacks of each method.

3. EXPERIMENTAL WORK/PROPOSED TECHNIQUES

Since auto-encoders have demonstrated significant potential for the anomaly detection of IoT botnets, the proposed study will analyse misconceptions and incorrect information regarding the viability of such an application through an exhaustive literature survey and transformational research design. Future research should concentrate on creating an ML-based auto-encoder model that is particularly made to identify IoT botnets like Mirai and evaluate the model using the best cybersecurity practises for IoT devices. A lab setting for testing the anomaly detector for IoT devices was suggested by Meidan et al. [20]. Replicating the simulation environment will be the goal.

4. DISCUSSION AND ANALYSIS

A preliminary examination of the literature reveals that while no one ML strategy is capable of detecting all forms of security vulnerabilities in IoT, ML techniques offer reliable tools for anomaly detection in IoT contexts. Further investigation reveals that due to both particular and broader weaknesses in IoT networks and devices, botnets and DDoS assaults pose a significant security issue in IoT environments. A comparison of the main themes in the research is shown in Table 3.

Table 3. The literature's main themes on ML-based detection in IoT

Major themes	Sources
IoT vulnerabilities to DDoS and botnets	[5,7,8,9,11,12]
IoT security architectures	[6,7]
ML techniques for detecting botnets (network-based detection)	[14,15]
IoT botnet detection with auto-encoders	[16,17,18,19]

One ML method cannot identify all types of security flaws in IoT, according to a preliminary review of the literature, but ML techniques provide trustworthy tools for anomaly detection in IoT contexts. Further research demonstrates that botnets and DDoS attacks offer a serious security risk in IoT environments because of both specific and general flaws in IoT networks and devices.

The methods suggested by [16] and [19] offer the most encouraging outcomes of the papers we analyzed, according to Table 3. These techniques dramatically improve detection accuracy by using deep auto-encoders to automatically extract features and training data. Additionally, auto-encoders may be network-based or device-based, indicating that the features may be retrieved from traffic or device-related data. It is feasible to deploy the model's deep learning capabilities in a cloud environment and just build lightweight models on the devices and network for anomaly detection because the majority of IoT devices have minimal memory capacities.

5. CONCLUSION AND FUTURE WORK

While the IoT has many advantages, there are security threats as well. This study has shown that IoT botnets provide a serious security risk since they can develop into DDoS assaults and IoT botnets, which reduce the availability of IoT networks and devices. showing the ML detection methods' performance metrics against IoT botnet attacks. The most widely used ML classifiers for IoT security make use of Random Forest classifiers, which have shown to be deployable in IoT environments.

This study's major contribution was to investigate if auto-encoders might be used to effectively detect botnets and stop DDoS attacks. For network-based security threat identification, auto-encoders may be helpful.

The results of the study show that ML approaches have developed to the point where they can recognize certain dangers to IoT networks. But IoT assaults are constantly changing to avoid being discovered by current security measures. Utilizing network-based detection methods is a workable alternative, but at the moment, none exist or have shown to function well. Future study should map the security criteria for a botnet detection system and aggregate the ideal characteristics of an auto-encoder. The foundation for creating efficient detection systems that counter the security threat posed by IoT botnets may be developed by modelling these needs.

Table 2. Comparison of proposed IoT Botnet detection methods

Ref #	Advantages of the proposed method(s)	Weakness of the proposed method(s)	Proposed Method
16	Does not need clean training data, the testing set is not impacted by outliers, and the findings are consistent.	Time-consuming in high rank matrix dimensions, Low precision and data with many sparse components may have a high false negative rate.	Robust Deep Auto-encoder
17	Since it employs online processing, it can be deployed to network devices with limited memory. It is lightweight and scalable across many IoT devices. Quicker runtime since it only uses one auto-encoder.	Anomaly detection is only reliant on the RMSE, which makes it susceptible to false positives during periods of high but typical network activity. Dependent on other libraries for capturing and parsing raw packets.	Auto-encoder neural networks
18	Due to its network- and host-based nature, it is effective at detecting anomalies. Low computational load on devices. Minimal communication overhead between devices and the network. High detection accuracy. Low false positive rate. Unsupervised learning that makes it easier to adapt to changing environments.	Uses intermediary goal functions as proxies, which might provide orthogonal outcomes; is significantly impacted by input mistakes; is sensitive to changes in the device and network in the WCN; and yields transfer learning from the auto-encoder to neural network.	Auto-encoder neural networks
19	High detection accuracy (99.82%), Low false positive rate, Automatic feature extraction from packet headers, Network-based, which increases detection rate, High detection accuracy (99.82%).	Training takes time, and pre-processing of the data removes some characteristics.	Deep learning stacked auto-encoder

REFERENCES

[1] W. Zhou, Y. Jia, A. Peng, Y. Zhang, and P. Liu, "The Effect of IoT New Features on Security and Privacy : New Threats, Existing Solutions , and Challenges Yet to Be Solved," IEEE Internet Things J., vol. PP, no. c, p. 1,2018.

[2] C. Maple, "Security and privacy in the internet of things," J. Cyber Policy, 22, 155-184, DOI 10.1080/23738871.2017.1366536, 2017.

[3] M. Conti, A. Dehghantanha, K. Franke, and S. Watson,"Internet of Things security and forensics : Challenges and opportunities," Futur. Gener. Comput. Syst., vol. 78,pp. 544–546, 2018.

[4] H. Hindy et al., "A Taxonomy and Survey of Intrusion Detection System Design Techniques, Network Threatsand Datasets," vol. 1, no. 1, 2018.

[5] T. S. S. O. ITU, "Reference architecture for Internet of things network network capability exposure," 2017.

[6] F. Olivier, G. Carlos, and N. Florent, "New Security Architecture for IoT Network," Procedia - Procedia Comput. Sci., vol. 52, no. BigD2M, pp. 1028–1033, 2015.

[7] E. Leloglu, "A Review of Security Concerns in Internet of Things," Journal of Computer and Communications, pp. 121–136, DOI: 10.4236/jcc.2017.51010, 2017.

[8] A. Lohachab and B. Karambir, "Critical Analysis of DDoS — An Emerging Security Threat over IoT Networks," vol. 3, no. 3, 2018.

[9] R. Hallman, J. Bryan, G. Palavicini, J. Divita, and J. Romero-mariona, "IoDDoS — The Internet of Distributed Denial of Service Attacks : A Case Study of the Mirai Malware and IoT-Based Botnets IoDDoS — The Internet of Distributed Denial of Service Attacks A Case Study of the Mirai Malware and IoT-Based Botnets," no. April 2017.

[10] "OWASP IoT Top 10 2018," OWASP Internet of Things Project, 12-Apr-2019. [Online]. Available: https://www.owasp.org/index.php/OWASP_Internet_of_Things_Project#tab=OWASP_IoT_Top_10_2018_Mapping_Project.

[11] A. Marzano et al., "The Evolution of Bashlite and Mirai IoT Botnets," 2018 IEEE Symp. Comput. Commun., pp.813–818, 2018.

[12] M. Antonakakis et al., "Understanding the Mirai Botnet This paper is included in the Proceedings of the Understanding the Mirai Botnet," 2017.

[13] Hameed, Sufian & Ahmed Khan, Hassan. "SDN Based Collaborative Scheme for Mitigation of DDoS Attacks". Future Internet. 10. 23. 10.3390/fi10030023, 2018.

[14] K. M. Z. and Y. C. D. H. Summerville, "Automatic Feature Selection for Ultra-lightweight deep packet anomaly detection for Internet of Things devices," 2015 IEEE 34th Int. Perform. Comput. Commun. Conf., 2016.

[15] S. García, A. Zunino, and M. Campo, "Survey on network-based botnet detection methods," no. June 2013, pp. 878–903, 2014.

[16] C. Zhou & R. Paffenroth, "Anomaly Detection with Robust Deep Autoencoders," KDD'17, Halifax, NS, Canada, pp. 665–674, 2017.

[17] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune : An Ensemble of Autoencoders for Online Network Intrusion Detection," pp. 18–21, 2018.

[18] T. Luo and S. G. Nagarajan, "Distributed Anomaly Detection using Autoencoder Neural Networks in WSNfor IoT," no. May 2018.

[19] Quamar Niyaz, Weiqing Sun, Ahmad Y Javaid, "A deep learning based DDoS detection system in software- defined networking (SDN)," no. Dl, pp. 1–18, 2016.

[20] Y. Meidan and M. Bohadana, "N-BaIoT — Network- Based Detection of IoT Botnet Attacks Using Deep Autoencoders," IEEE Pervasive Computing, vol. 17, no. September, pp. 12–22, 2018.