International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 10 (2019) pp. 2417-2421

Survey for Anomaly Detection of IoT Botnets Using Machine Learning Auto-Encoders

Reem Alhajri
College of Computer Science & Information Technology
Imam Abdulrahman Bin Faisal University
P.O. Box 1982, Dammam 31441, Saudi Arabia

Rachid Zagrouba
College of Computer Science & Information Technology
Imam Abdulrahman Bin Faisal University
P.O. Box 1982, Dammam 31441, Saudi Arabia

Fahd Al-Haidari
College of Computer Science & Information Technology
Imam Abdulrahman Bin Faisal University
P.O. Box 1982, Dammam 31441, Saudi Arabia

Abstract

This study focus 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 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 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 in the detection of IoT botnets.

Keywords: Machine Learning Detection, IoT, Botnets, DDoS

1. INTRODUCTION

Interconnection of computing power with appliances comprises the Internet of Things (IoT), which has the capability to overtake the industry of applications and commodities that augments the data generation, consumption, and quantity of devices [1]. Areas of risk associated with IoT that can be compromised are non-user interfaces, non-secure interaction protocols, sensitive data modification [2], weakness in middleware layer, multiplicity of attacks and non-availability of storage capacity [3]. Complex attacks make security a challenge, which is catered by detection rather than prevention in the form of Intrusion Detection Systems (IDSs) based on Machine Learning algorithms. Nevertheless, limited computing power and storage capabilities make application difficult for usage as smart city embedded systems [4], especially botnets and IoT malware attacks via Distributed Denial of Service (DDoS). Mirai DDoS and botnets having capability to launch intricate network wide attacks such as HTTP flooding. Mutation and evolution of attacks is a concern for detection strategies, which can be catered through ML and deep learning that is also the aim of this study.

However, the specific characteristics of IoT environments, such as limited computing power and storage capacity limit the potential application of some of the security systems and features. This paper explores the application of ML-based detection methods and counterattacks in the context of IoT.

2. REVIEW OF LITERATURE

This section will focus on providing a details about IoT security vulnerabilities and botnet/DDoS attacks then it will provide a details about the machine learning detection of IoT Botnets and DDoS, lastly it will discuss the anomaly detection of IoT Botnets using Auto-Encoders.

2.1 IOT SECURITY VULNERABILITIES AND BOTNET/DDOS ATTACKS

The IoT security architecture is the subject of numerous studies, in which the entire structure is categorized into four distinction layers, each with its unique security implications: application layer, network layer, a device layer, and service/application support layer [5]. The consensus in these studies is that security vulnerabilities in each layer differ. While several researchers have proposed different models of IoT security architectures [6] , they all agree that no single IoT model can guarantee optimal security against different types of threats. In particular, the IoT network layer is exposed to many kinds of security threats, including selective forwarding, Sybil attacks, Man-in-The-Middle (MiTM) attacks, and DoS attacks [7]. Among these different kinds of attacks, botnets and DDoS attacks elicit the greatest focus, perhaps due to the potential impact of such attacks in terms of compromising the availability of information systems [8,9]. Table 1 shows IoT security vulnerabilities and potential attack vectors.

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

IoT vulnerabilities	Attack vectors
Insecure IoT interfaces	Weak credentials
Insufficient authentication and authorization	Insecure login credentials
Insecure software and network services	Vector for attacks of IoT devices and malware distribution
Weak physical security	Ports, SD cards and storage media, can allow unauthorized access to OS

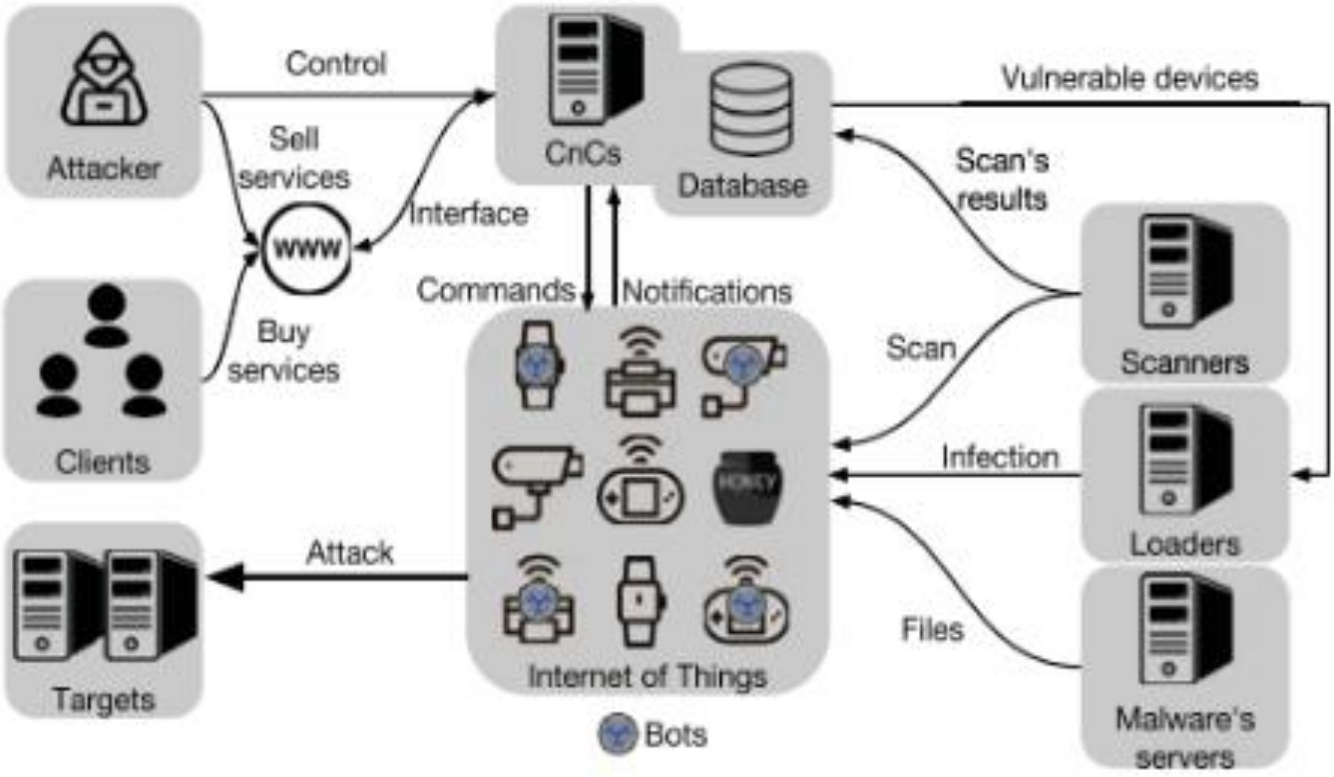

Figure 1. Typical IoT botnet ecosystem [11]

Mirai and Bashlite represent the most important botnet malware that has the potential to cause DDoS attacks via escalation in IoT networks [11,12]. Both Mirai and Bashlite share one common feature, they exploit vulnerabilities in known authentication credentials. A typical IoT botnet ecosystem comprises of a command and control server, which provides the attack interface, scanners for probing devices, bots or the infected devices, loaders for logging into vulnerable devices, malware servers, and databases (See Figure 1) [11].

2.2 MACHINE LEARNING DETECTION OF IoT BOTNETS AND DDOS

Botnet detection methods described in literature involve either specific operational steps or utilization of detection approaches. Several studies have addresses IoT related botnet detection operational steps, focusing on concepts such as Software Defined Networking (SDN) collaborative schemes [13], and employment of discrimination functions [14]. These methods focus on the early stages of attack propagation and execution within the C&C server. Since IoT network botnets evolve rapidly, mutated attack tools may evade existing detection systems. A viable solution to this problem is to focus on further steps in the IoT botnet operations. This requires network-based botnet detection methods.

In [15], the researchers proposed a model for classifying network-based botnet detection methods depending on detection techniques, detection sources, and detection algorithms. In terms of detection techniques, network-based detection could be either anomaly-based detection, which focuses on botnet behavior and protocol behavior, or fingerprint-based detection methods. Based on detection schemes, the major categories include detection approaches that rely on normal sources, like virtual internal networks and real networks, and detection approaches that rely on botnet sources, such as honeypots, virtual networks, and simulated solutions. The detection algorithms can involve six approaches: instance-based learning, supervised learning, semi-supervised learning, unsupervised learning, use of heuristic rules, and signal processing [15]. Supervised detection algorithms use methods such as Hidden Markov Model (HMM), Bayesian statistics, Artificial Neural Networks (ANNs), Decision Trees, and Support Vector Machine (SVM). Supervised detection algorithms focus mainly on clustering techniques. The assessment focused on various detection algorithms, albeit without mentioning auto-encoders.

2.3 ANOMALY DETECTION OF IOT BOTNETS USING AUTO-ENCODERS

Auto-encoders have received heightened scholarly interest in the recent past, as potential cybersecurity tools [16,17]. While auto-encoders have been studied extensively, there is paucity of research work on their application in IoT contexts. Even though, a number of researchers have shown the possibility of implementing auto-encoders [18]. Luo and Nagarajan [18] presented auto-encoder neural networks for anomaly detection in Wireless Sensor Networks (WSNs) embedded in IoT

environments. The detection algorithm comprised of two components; one located within sensors and the other placed in the IoT cloud. The evaluation demonstrated that the unsupervised learning features of the auto-encoder neural network allowed adaptation to unexpected changes in IoT networks. There have also been attempts to compare various deep learning models for network intrusion detection, including the "Vanilla Deep Neural Net (DNN)", "Self-Taught Learning (STL)", and "Recurrent Neural Network (RNN)" [19]. The findings revealed that the STL detection model was robust in environments with unclean data, suggesting suitability in IoT.

In Table 3. We can find a comparison shown between proposed methods related to auto-encoders describing advantages and weakness for each proposed method.

3. EXPERIMENTAL WORK/PROPOSED TECHNIQUES

Since auto-encoder s have shown great promise for anomaly detection of IoT botnets, the proposed study will examine the feasibility of such an application through extensive literature review and transformative research design to investigate myths and false knowledge on auto-encoder anomaly detection in IoT networks. Future studies should focus on the development of a model ML-based auto-encoder designed specifically to detect IoT botnets, such as Mirai as well as evaluation of the model based on best practices in cybersecurity for IoT devices. Meidan et al. [20] proposed a lab setup for evaluating the anomaly detector for IoT devices. The idea will be to replicate the simulation environment.

4. DISCUSSION AND ANALYSIS

Preliminary literature review shows that ML techniques provide robust tools for anomaly detection in IoT environments, yet no single ML method can detect all types of security threats in IoT. Further analysis shows that botnets and DDoS attacks present an important security threat in IoT contexts, due to specific and general vulnerabilities in IoT devices and networks. Table 2 provides a comparison of the major themes in the studies.

Table 2. Major themes found in the literature on ML-based detection in IoT

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

The analyzed literature contributes in the research discourse focusing on network-based detection of IoT security threats, particularly IoT-based botnets, which often lead to DoS attacks. The papers show consensus that a viable detection system should meet the unique constraints of IoT environments, including shortage of computing power and storage capacity. While ML techniques have demonstrated ability to detect botnets targeting various network devices, some of these methods are unfeasible in IoT networks. Future work should examine the viability of auto-encoder s in anomaly detection.

Based on Table3 the approaches proposed by [16] and [19] present the most promising results of the studies reviewed. These methods use deep auto-encoder s to automatically extract features and training data, which significantly increases the accuracy of the detection. In addition, auto-encoder s can be device-based and/or network-based implying that the features can be extracted from device or traffic data. Since most IoT devices have low memory capacity, it is possible to deploy the model's deep learning capabilities in a cloud environment and therefore only implement lightweight models on the devices and network for detecting anomalies. As a result, it is fairly easy to maintain continuous monitoring of the devices and network. In addition, it is relatively fast to detect devices and networks that have been compromised due to low burden on device memory.

5. CONCLUSION AND FUTURE WORK

While the IoT offers important benefits, it also poses security risks. This study has revealed that IoT botnets represent a major security threat because it can evolve into IoT botnet and DDoS attacks, which affects the availability of IoT devices and networks. highlighting the performance parameters for ML detection techniques against IoT Botnet attack. The common ML classifiers for IoT security using Random Forest classifiers which demonstrated feasibility for deployment in IoT environment.

The main contribution of this study was to examine the feasibility of using auto-encoder s to enable effective botnet detection and prevent DDoS attacks. Auto-encoder s could be useful for network-based detection of security threats.

The study findings indicate ML techniques have advanced to the extent of detecting specific threats to IoT networks. However, the IoT attacks continue to evolve to evade existing detection systems. A viable solution is to use network-based detection approaches, but currently such solutions do not exist or have not demonstrated effective performance. Future research should involve aggregating the desirable features of an auto-encoder and mapping the security requirements for a botnet detection system. Modeling these requirements can lead to a framework for developing effective detection systems that address the security threat posed by IoT botnets.

Table 3. 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 require clean training data • Testing set is not affected by outliers hence has consistent results • High recall	• Low precision • Time consuming in high rank matrix dimensions • Potential for high false negative rate for data with numerous sparse components	Robust Deep Auto-encoder
17	• Deployable to low memory network devices since it uses online processing • Lightweight and scalable across multiple IoT devices • Faster runtime since it operates as a single auto-encoder	• Dependent on external libraries for capturing and parsing raw packets • Trade-off between detection and packet processing • Anomaly detection is based purely on the RMSE hence it is prone to false positives during heavy but normal network activity	Auto-encoder neural networks
18	• Effective in detecting anomalies since it is both network-based and host-based • Low computational load on devices • Minimal communication overhead between devices and network • High accuracy in detection • Low false positive rate • Unsupervised learning facilitates adaptability to dynamic environments	• Uses proxies for intermediate objective functions which may lead to orthogonal results • Significantly affected by input errors • Sensitive to device and network changes in WCN • Results in transfer learning from auto-encoder to neural network	Auto-encoder neural networks
19	• Network-based which enhances the detection rate • Automatically extracts features from packet headers • High detection accuracy (99.82%) • Low false positive rate	• Training is time consuming • Data is pre-processing resulting in the loss of some features	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 Threats and 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 WSN for 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 Comput.*, vol. 17, no. September, pp. 12–22, 2018.