Malware Detection in Internet of Things using Machine Learning enabled Data Science Approach

Sunita Choudhary, Anand Sharma

Mody University of Science and Technology, Lakshmangarh, Sikar, Rajasthan, India

Abstract

Internet of Things (IoT) is measured as disseminated and unified arrangement of installed structures conferring by wired or mobile communication propels. With the extended use of IoT structure in every area, threats and attacks in these establishments are in like manner growing proportionately. In this way, wide considerations have been put to address the protection and security issues in IoT networks in a general sense through fundamental cryptographic techniques. Regardless, the created tools have various kinds of programming to be introduced and impression of the framework topology are not performed, so there is an issue that outwardly momentary irregularities can't be perceived.

Malware detection in the IoT networks is a rising issue in the space of IoT. In this paper, machine learning enabled data science approach for malware detection in IoT has been proposed.

Keywords

Internet of Things (IoT), Malware detection, Machine Learning, Data Science

1. Introduction

IoT is deliberated as widely interconnected and appropriated arrangements of device setup which are connected by wired or remote communication innovations [1]. It is additionally considered as the arrangement of actual things or items engaged with rules and protocols, communication capacities and storage as per the hardware devices, network topologies and computing capabilities that endows these things to collect, store and process the data.

The said things and devices in the IoT allude to the items by our daily life going from savvy house-hold gadgets, for example, smart meter, smart bulb, smoke alarm, temperature sensor, AC,IP camera, to more complex gadgets, for example, RFID (Radio Frequency Identification) gadgets, heartbeat indicators, sensors in garage, accelerometers, and a scope of different sensors in vehicles and so on [2].

ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25-27, 2021, Delhi, India

EMAIL: sunitadangi@gmail.com (S. Choudhary); anand_glee@yahoo.co.in (A. Sharma)

ORCID: 0000-0002-9995-6226 (A. Sharma)

The areas covered by the IoT incorporate, however not restricted to, energy, buildings, clinical, retail, supply chain, transportation, manufacturing, etc. That huge size of IoT networks fetches new difficulties, for example, the executives of these gadgets and things, sheer measure of information, communication, storage, computation, protection and security. There are broad explores casing these various parts of IoT (that are design, conventions, rules, applications, privacy and security) [3]. Be that as it may, the foundation of the commercialization of IoT framework is the privacy and security ensure just as purchaser fulfillment. The approach that IoT utilizes to empower the things, for example, SDN (Software Defined Networking), fog computing, and Cloud Computing (CC), likewise expands the scene of threats for the attackers.

2. Security Issues in the IoT Deployment

Privacy and Security are the principle factors in the business acknowledgment of IoT applications and installment. Presently Internet is the main target for cyber attacks from

hacking to access the secret information. It penetrates the security system that have unfavorably influenced various enterprises, for example, medical services and other businesses. The constraints of IoT gadgets and the complete framework they work in, represent extra difficulties for the devices and applications. Until this point, privacy and security issues have been broadly explored in the IoT area from alternate points of view, for example, communication security, information security, identity management, architectural security, malware examination, etc [4].

The inadequate safety efforts and absence of committed inconsistency location frameworks for these heterogeneous organizations make them defenseless against a scope of attacks, for example, spoofing, Denial of Service (DoS), data-leakage, and so forth. These can prompt terrible impacts; making harm equipment, disturbing the framework accessibility, causing framework power outages, and even truly harm people [5], [6]. Consequently, plainly the size of effect of the attacks executed on IoT organizations can change essentially. For instance, a moderately straightforward and apparently innocuous deauthentication attack can cause no huge harm, yet whenever performed on a gadget with basic importance, for example, a guiding wheel in a remote vehicle, it can represent a danger to human existence. Subsequently, clearly there is a significant gap in security necessities and protection abilities of presently accessible IoT gadgets. The primary concerns which make these gadgets smart are their computational power and heterogeneity regarding equipment, software, and protocols [7]. All the more explicitly, it is for the most part not practical for smart gadgets with limited computing capability, memory, data transfer capacity, and battery asset to execute computationally serious and dormancy touchy security undertakings that produce substantial calculation and transmission load [8]. Subsequently, it is absurd to expect to utilize intricate and hearty safety efforts. Also, given the variety of these gadgets, it is trying to create and send a security component that can suffer with the scale and scope of gadgets [9].

Now the Malware is characterized as software intended to invade or harm a digital framework without the proprietor's educated assent. This is really a nonexclusive delineation for all sorts of cyber threats. A straightforward order of malware comprises of computer files or data infectors and independent malware. Another method of ordering malware depends on their specific activity: worms, rootkits, backdoors, spyware, trojans and so on as the ascent of malware on mobile phones has illustrated, if something is associated with the web, it's a likely road of cyber-attacks.

In this way, while the ascent of Internet of Things associated gadgets has carried various advantages to clients - in industry, the work environment and at home - it also has opened entryways for new digital criminal plans.

In contrast to mobile phones, IoT gadgets are frequently connected and disregarded, with the threat that the IoT camera you set up could turn out to be effectively open to outcasts - who might actually utilize it to keep an eye on your activities, be it in your working environment or in your home.

Such is the degree of the security stress with the IoT, police have cautioned about the threats presented by associated gadgets, while government bodies are running after methods of administering IoT gadgets in the near future, so we're not left with a harmful tradition of billions of gadgets that can undoubtedly be tainted with malware.

3. Machine Learning Approaches for malware Detection

Signature based strategies [10] is now getting more troublesome for detection of malware since all recent malware applications will in general have numerous polymorphic layers to dodge discovery or to utilize side components update themselves to a fresher variant at brief timeframes to evade detection by any specific antivirus programmer. For an illustration of dynamic malware analysis for detection of malware, by means of copying in a virtual platform, the intrigued reader can grasp [11]. Traditional strategies for the discovery of transformative infections are depicted in [12].

An outline on various ML techniques that were developed to detect malware are given in [13]. Here we are giving a couple of references to epitomize those strategies.

In [14], decision trees chipping away at n-grams are established to deliver preferable

outcomes over both the SVM (Support Vector Machines) and Naïve-Bayes classifier.

In [15] Hidden Markov Models are utilized to identify whether a specified program record is a variation of a past program document. To achieve a comparative objective, [16] utilizes Profile Hidden Markov Models that have been recently utilized with extraordinary accomplishment for grouping examination in bioinformatics.

In [17], Maps are utilized to recognize examples of conduct for infections in Windows executable records.

4. Machine Learning enabled Data Science Approach

In this section, machine learning enabled data science approach for detection of malware in IoT has been described. Figure 1 illustrates the basic blocks for the proposed detection approach.

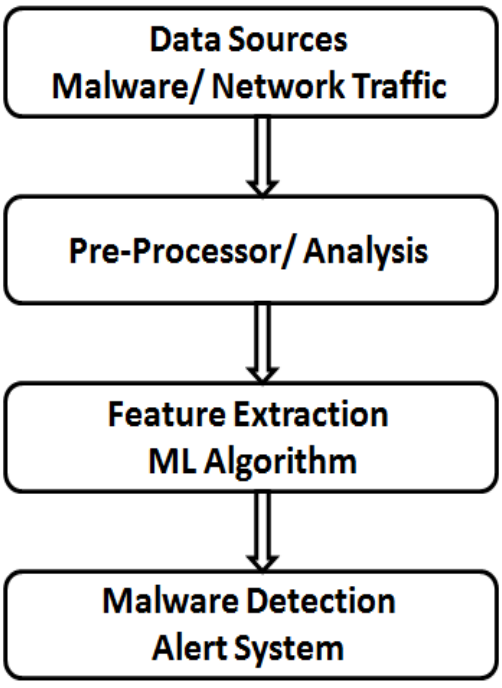

Figure 1: 1 Machine Learning enabled Data Science Approach for malware detection in IoT

The current machine learning approaches for malware detection roused us to propose and define a malware detection system which we emphatically accept will help in moderating the present testing issues. Figure 1 demonstrates various advances and interactions of the proposed malware detection framework. The concise conversation of its parts is as follows.

The said malware and considerate executable files are treated as data sources. The pre-processing and analysis are finished with data science. This cycle is a basic advance and incorporates rule age and knowledge data discovery (KDD) validation. Further the extracted features acquired through this phase are continually checked and approved utilizing cross-system validation and profound observing process. This is ended to conquer the difficulties presented by the adversary. Data science tools and machine learning execution make the feature extraction and overall process more productive and successful.

The preparation of testing and training dataset is further processed by the ML techniques or classifiers. Here, we applied hostile protection and algorithmic biasness defense to moderate the impacts on the decision making cycle. The end-product is further transferred to the detection and alert system which handles the important strides to retain the framework secured against any cyber-attacks.

5. Experiment Setup and Results

The test method was executed in two different operating systems to be specific, Linux 4.1. also, Windows 10 which introduced 8 center Core i5 processor with 8GB RAM. Moreover, two VMs Oracle VirtualBox 4.2.16 have been utilized in this work. These VM's are utilized to gather and analyze the malware tests. First VM is using CentOS Linux and the second VM is Windows 10. In addition, different tools are additionally used to set up the tests, for example, WEKA 3.9.4 (the data mining and ML tool) and MATLAB 2019b.

To assess the evaluation of the proposed method, firstly the said dataset is isolated into two different groups: Training group and Testing group. The said training dataset has been partitioned in some malware and some goodware to stay away from the awkwardness. The training dataset consisting of 2000 examples is divided in 1000 malware and 1000 goodware. The equivalent apportioned is acted

in the testing-dataset which likewise contains 2000 examples in total as 1000 malware and 1000 goodware.

Table 1 exhibits the correlation of our proposed technique with the current research and work. The precision evaluation appeared by Pajouh et al. [18], Darabian et al. [19] and Khammas B.M. [20] are contrasted and proposed strategy. Nonetheless, their procedures need extra time because of the dismantle cycle which isn't reasonable to encounter the clients necessities of IoT organization, while the proposed method kill this extra preparing in light of the fact that the highlights are extricated straightforwardly from crude parallel document. In addition, their outcomes are not mirroring the genuine precision because of little dataset that they utilized.

Table 1
Comparison of dataset and accuracy

	Method	**Dataset (M/G) Malware / Goodware**	**Results (Accuracy) (%)**
Pajouh et al. [18]	Recurrent Neural Network	281 M 270 G	98.18
Darabian et al. [19]	ML	247 M 269 G	99
Khammas B.M. [20]	ML	1000 M 1000 G	96.7
Proposed method	**ML + DS**	**1000 M 1000 G**	**98.6**

6. Conclusion

The majority of the security issues are perplexing and the arrangements can't be distinct. For example, in the event of privacy and security difficulties, like, intrusion or DoS, there is a likelihood of false-positives which will deliver the answers for be inadequate contrary to those attacks. Moreover, that will likewise diminish the customer trust and accordingly debasing the viability of IoT framework.

Subsequently, an all-encompassing privacy and security methodology for IoT has been developed from the current security arrangements as machine learning enabled data science malware detection approach that is evolutionary, robust, intelligent, and scalable mechanism to address malware detection in IoT.

These days, gadgets interfacing with the internet are broadly spread in everywhere on the world. In this paper, we inspected the capability of utilizing a blend of machine learning and data science to detect IoT malware. The best outcomes accomplished around 98.6% of accuracy utilizing machine learning enabled data science approach. Future exploration will extend the proposed way to deal with look at the other machine learning methods with data science tools for IoT malware detection.

7. References

[1] O. Novo, N. Beijar, and M. Ocak, “Capillary Networks - Bridging the Cellular and loT Worlds ,” IEEE World Forum on Internet of Things (WF-IoT), vol. 1, pp. 571–578, December 2015.

[2] F. Hussain, Internet of Things; Building Blocks and Business Modles. Springer, 2017.

[3] K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, “On security challenges and open issues in internet of things,” Future Generation Computer Systems, vol. 83, pp. 326 – 337, 2018.

[4] J. Granjal, E. Monteiro, and J. S. Silva, “Security for the internet of things: A survey of existing protocols and open research issues,” IEEE Communications Surveys Tutorials, vol. 17, pp. 1294–1312, third quarter 2015.

[5] Vishu Madaan, Dimple Sethi, Prateek Agrawal, Leena Jain, Ranjit Kaur, “Public Network Security by Bluffing the Intruders Through Encryption Over Encryption Using Public Key Cryptography Method”, International Conference on Advanced Informatics for Computing Research (ICAICR’17), pp. 249 -257, Springer, Mar 2017.

[6] Cyber hackers can now harm human life through smart meters—smart grid awareness. https://smartgridawareness.org/2014/12/30/hackers-can-now-harm-human-life/. (Accessed on 02/06/2020).

[7] Securing the internet of things: A proposed framework - cisco. https://www.cisco.com/c/en/us/about/security-center/secure-iot-proposed-framework.html. (Accessed on 08/07/2020).

[8] Liang Xiao, Xiaoyue Wan, Xiaozhen Lu, Yanyong Zhang, and Di Wu. IoT security techniques based on machine learning. arXiv preprintarXiv:1801.06275, 2018.

[9] Eirini Anthi, Shazaib Ahmad, Omer Rana, George Theodorakopoulos, and Pete Burnap. Eclipseiot: A secure and adaptive hub for the internet of things. Computers & Security, 78:477–490, 2018.

[10] I. Santos, Y. K. Penya, J. Devesa, and P. G. Garcia, "N-grams-based file signatures for malware detection," 2009.

[11] K. Rieck, T. Holz, C. Willems, P. D¨ussel, and P. Laskov, "Learning and classification of malware behavior," in DIMVA '08: Proceedings of the 5th international conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer-Verlag, pp. 108–125, 2008.

[12] E. Konstantinou, "Metamorphic virus: Analysis and detection," Technical Report RHUL-MA-2008-2, M.Sc. thesis, 93 pages, 2008.

[13] P. K. Chan and R. Lippmann, "Machine learning for computer security," Journal of Machine Learning Research, vol. 6, pp. 2669–2672, 2006.

[14] J. Z. Kolter and M. A. Maloof, "Learning to detect and classify malicious executables in the wild," Journal of Machine Learning Research, vol. 7, pp. 2721–2744, December 2006.

[15] M. R. Chouchane, A. Walenstein, and A. Lakhotia, "Using Markov Chains to filter machine-morphed variants of malicious programs," in Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference, pp. 77–84, 2008.

[16] M. Stamp, S. Attaluri, and S. McGhee, "Profile hidden markov models and metamorphic virus detection," Journal in Computer Virology, 2008.

[17] I. Yoo, "Visualizing Windows executable viruses using self-organizing maps," in VizSEC/DMSEC '04: Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security. New York, NY, USA: ACM, pp. 82–89, 2004.

[18] Haddad Pajouh, H., et al., A deep Recurrent Neural Network based approach for Internet of Things malware threat hunting. Future Generation Computer Systems 85: p. 88-96, 2018.

[19] Darabian, H., et al., An opcode- based technique for polymorphic Internet of Things malware detection. Concurrency and Computation: Practice and Experience: p. e5173, 2019.

[20] Ban Mohammed Khammas, "The Performance of IoT Malware Detection Technique Using Feature Selection and Feature Reduction in Fog Layer" IOP Conf. Series: Materials Science and Engineering 928, 022047, 2020.