XVII Simposio CEA de Control Inteligente

27-29 de junio de 2022, León

Convolutional Neural Network for Malware Detection in IoT Network

AlHamdi, M. J. M.[a,*], Lopez-Guede, J. M.[b], Rahebi, J.[c]

[a] *Systems and Automatic Control Department, University of the Basque Country (UPV/EHU), Spain*
[b] *Systems and Automatic Control Department, University of the Basque Country (UPV/EHU), Spain*
[c] *Department of Software engineering, Istanbul Topkapi University, Istanbul, Turkey.*

To cite this article: Alhamdi, M. J.M., Lopez-Guede, J. M., Rahebi, J. 2022. Convolutional Neural Network for Malware Detection in IoT Network. XVII Simposio CEA de Control Inteligente.

Abstract

The network has exploded in popularity in the twenty-first century in the last few years, becoming one of the most extensively utilized and prominent technologies. Nowadays, cyberattacks occurring and the variety, size, and intensity of cyberattacks are increasing. In this work, the machine learning method is used to predict Intrusion in the Internet of Things. Attacks on networks connected to smart cities or on intelligent transportation systems endanger the security of these networks. Studies show that IoT attacks can cost the network millions of dollars. DDoS attacks by malicious botnets and nodes on the IoT network are among the most malicious attacks on the network and can disable IoT application servers.

Keywords: Internet of Things, Malware Detection, Convolutional Neural Network.

1. Introduction

Machine learning models require training data regardless of the method utilized. There is no exception to this rule in machine learning models used in anomaly-based intrusion detection systems. Public data sets are distinguished from private data sets by their availability to the general public. In the lack of private sector inquiry, this study will only look at well-known public data sets. These are frequently employed by academics and researchers alike for benchmarking purposes. It contains attacks that are distinct from those of its parents, such as smurf and Neptune. There are also four different types of attacks classified as DoS, user to root, remote to local, and probing. It is, however, criticized for being very repetitive and for having damaged data (Cup, 1999). It was suggested that NSLKDD be used to solve these problems (Tavallaee et al., 2009). Although it mitigates statistical weaknesses in the data set, NSL-KDD still suffers from the lack of modern attacks and real-world samples. Additionally, the data collection includes packet capture (pcap) files of flowing traffic and characteristics derived from raw data. These characteristics are retrieved by the authors using the freely available argues tool and some custom scripts. The Canadian Institute for Cybersecurity has also compiled a collection of available data sets for IDS research (CIC). These are designated using the CICIDS format, for example, CICIDS2017 (Sharafaldin et al., 2018). Majority of prior work has been conducted on public or private data sets generated in lab settings using network traffic simulators. However, drawbacks associated with the use of artificially generated data sets may be described as follows: (1) a lack of real-world behavior profiling for a particular business network; (2) a model's lack of representation of real-world risks.

2. Literature review

The IoT has different applications and can be used in various fields such as agriculture. Nowadays, smart objects can be placed on farms to control the quality of crops and enhance their productivity by regular monitoring of the status of crops. It is also possible to control the sales network of a product in the market by monitoring the product distribution process. The IoT applications are not limited only to agriculture. According to figure 1, it can be also employed to enhance and upgrade security, productivity, and smartness to high levels:

IoT
Transportation
Healthcare
Industry
Agriculture
Market
Smart Cities
School
Vehicles

Fig. 1. The IoT applications (Pardini et al., 2019)

As seen in figure 1, the IoT can be used in smart homes and cities, agriculture, transportation, vehicles, healthcare, industries, education and business. A very important application of the IoT is in future smart homes by enhancing comfort greatly. In this technology, a set of nodes are placed in smart homes to collect data from the home environment and to control home appliances.

In research (Liu et al., 2020) and in 2020, a review of Android malware detection approaches based on machine learning is presented. Android apps are evolving rapidly in the mobile ecosystem, but Android malware is also emerging endlessly, and many researchers have explored the issue of Android malware detection and theories and methods from different perspectives. Have provided. Existing research shows that machine learning is an effective and promising way to detect Android malware. This article provides a comprehensive overview of machine learning approaches based on machine learning. They briefly introduced some areas related to Android applications, including Android system architecture, security mechanisms, and Android malware classification. This review will help academics get a complete picture of machine-based Android malware detection. They can then serve as a basis for future researchers to start new work and help guide research in this area.

In the study (Chakravarty, 2020) and in 2020, presented feature selection and evaluation of licensed Android malware detection. Android malware is a pervasive threat to the security of mobile users' information. Android users often download apps from unauthorized and unreliable sources. Such programs may require multiple permissions from the user, and due to lack of awareness, the user can grant the necessary permissions. Android licenses are one of the most important sources of malware infection. By analyzing the permissions, it is possible to classify malware and malware in the database with the help of machine learning tools. There are a total of 330 licenses in Android apps. However, not all of them may contribute to the classification. In this paper, the proposed system examines the identification of the most effective permissions using feature reduction. In this research, decision tree, random committee, perceptron multilayer, sequential minimum optimization and random classifiers are used to evaluate the selected features. Experimental results show that five licenses can be almost complete, thus optimizing the malware detection system.

In the study (Haddadpajouh et al., 2020) and in 2020, they introduced a meta-heuristic and multi-kernel feature selection approach to identify the threat of IoT malware in the sensor layer and smart objects. IoT devices are increasingly targeted by malware due to their presence in a wide range of applications, including at home and in corporate environments. In this paper, they propose a method for detecting malware on IoT devices using the Gray Wolf optimization method and a multi-core support vector machine. This model is trained with IoT sample malware codes. The training data set consists of 271 benign samples and 281 malicious samples of Cortex A9 and is evaluated using cross-validation method. They validated the proposed model in terms of its ability to detect IoT malware that had not been seen before. Their proposed model, compared to the previous model of deep neural network, which requires more than 80 seconds to train the same data, requires only 20 seconds to train. In general, the proposed multi-core backup vector machine method performs better in terms of accuracy than deep network and IoT malware detection techniques based on fuzzy methods, while reducing computational cost and training time significantly reduced.

Host-based examines activity on endpoints, namely hosts, depending on the data collecting method. In other words, System resources including memory, CPU, and disk I/O are monitored by HIDS, as well as programs to look for unusual behavior. To that degree, their jurisdiction extends to the host on which they function. The majority of HIDS search for indications of a danger in locations such as operating system log files or the audit logs of another piece of software (Bhattacharyya & Kalita, 2019). Alarms are generated based on specified criteria. Due to the fact that they function according to rules, their performance in detecting known intrusion attempts is very good. They are, however, ineffective in detecting new assaults (Buczak & Guven, 2015). On the other hand, there are modern public data sets for IDS research created by universities and/or research groups. Recognizing the need for a modern IDS data set, Moustafa and Slay proposed UNSW-NB15 data set (Moustafa & Slay, 2015). As with earlier examples, it was produced in a lab setting using fake malicious and benign instances generated using a commercial network simulation device. To be more precise, such tools allow the simulation of current attacks by automatically gathering data from Common Vulnerabilities and Exposures (CVE) web sites and then developing and delivering network assaults in response.

3. Methodology

Anomaly-based intrusion detection machine learning approaches have no inherent advantages or disadvantages. Researchers may be influenced by a variety of factors when choose which technique to use. The amount of computer power available, the availability of field experts to manually intervene in the data, and the amount of data to be processed in a second are some of the issues a researcher should consider to choose the method of choice. To summarize, the effectiveness of a particular strategy should be assessed on a case-by-case basis. The flow chart of proposed method is shown in figure 2.

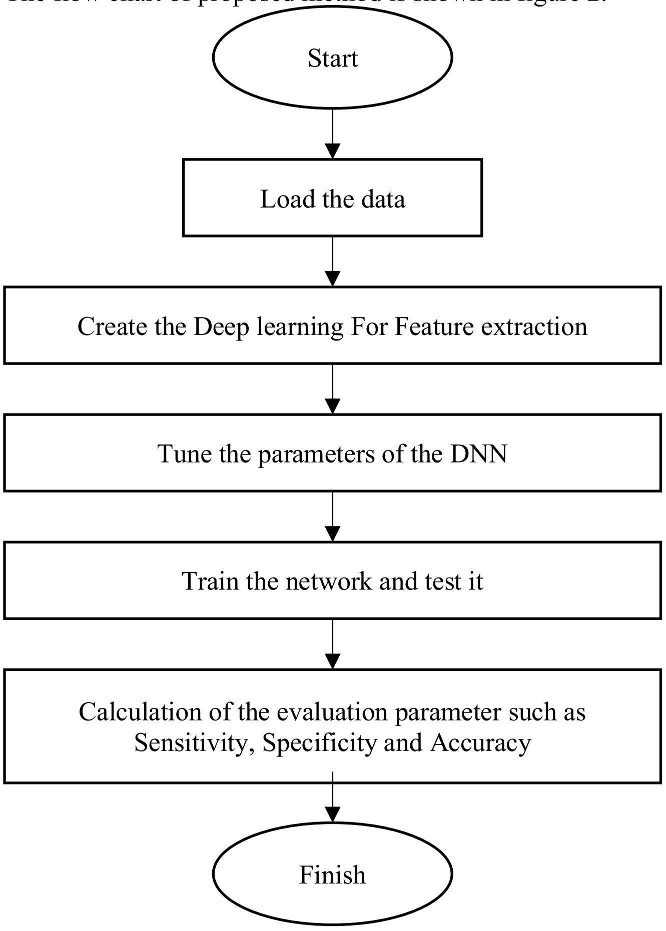
Start
Load the data
Create the Deep learning For Feature extraction
Tune the parameters of the DNN
Train the network and test it
Calculation of the evaluation parameter such as Sensitivity, Specificity and Accuracy
Finish

Fig. 2. Flowchart of proposed method

There are challenges in IoT network security because most IoT devices have limited power, in other words, their limited computing power makes it impossible to support strong security systems. Due to the weakness of the objects connected to the IoT network, it is difficult to encrypt and authenticate to prevent malicious cyber-attacks on this type of network. An intrusion detection system logically as a leading security solution can be of great help in solving this challenge. Network intrusion detection can be based on the non-intelligent mechanism of blacklisting or it can be based on anomalous learning methods. Machine learning and anomaly detection mechanisms play an essential role in protecting networks against various malicious activities. In this study, they applied different machine learning algorithms to effectively identify anomalies in the Internet penetration data network, and their results show that machine learning methods are highly accurate in penetration detection.

4. Conclusion

This study is investigated especially true in the DDoS detection, where previous hacks had cascading effects. Securing the network and maximizing its security While IoT devices may cause havoc on networks, they also provide a level playing field for users to adopt security measures that protect all connected devices.

References

Bhattacharyya, D. K., & Kalita, J. K. (2019). *Network anomaly detection: A machine learning perspective*. Chapman and Hall/CRC.

Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153–1176.

Chakravarty, S. (2020). Feature selection and evaluation of permission-based Android malware detection. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 795–799.

Cup, K. D. D. (1999). http://kdd. ics. uci. edu/databases/kddcup99/kddcup99. html. *The UCI KDD Archive*.

Haddadpajouh, H., Mohtadi, A., Dehghantanaha, A., Karimipour, H., Lin, X., & Choo, K.-K. R. (2020). A multikernel and metaheuristic feature selection approach for IoT malware threat hunting in the edge layer. *IEEE Internet of Things Journal*, *8*(6), 4540–4547.

Liu, K., Xu, S., Xu, G., Zhang, M., Sun, D., & Liu, H. (2020). A review of android malware detection approaches based on machine learning. *IEEE Access*, *8*, 124579–124607.

Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1–6.

Pardini, K., Rodrigues, J. J. P. C., Kozlov, S. A., Kumar, N., & Furtado, V. (2019). IoT-based solid waste management solutions: a survey. *Journal of Sensor and Actuator Networks*, *8*(1), 5.

Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, *1*, 108–116.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 1–6.

DOI: 10.18002/simceaci