International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com
doi : https://doi.org/10.32628/IJSRCSEIT

A Survey On Malware Detection for WebURLs and PE Files Using Machine Learning

Prof. J. N. Ekatpure, Archana S. Bichkule, Shital M. Kamble, Rutuja U. Lawand

Department of Computer Engineering, S. B. Patil College of Engineering, Indapur, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 Oct 2023
Published: 30 Oct 2023

Publication Issue
Volume 9, Issue 10
September-October -2023
Page Number
62-65

ABSTRACT

Malware such as Viruses, Worms, Trojans, Backdoors are some of the threats to computer system and internet.in recent years malware count is increased in millions. In the past few years millions of malwares were found in portable executable files which are downloaded from the internet. As the solution to this, it is highly desirable for users to detect such malware files, so that users can secure the devices as well as highly confidential data. Malware Detection System is an application which will detect the malwares from the portable executable files. The proposed system uses KNN algorithm to predict the malware files and legitimate files. so users can easily differentiate between them and secure their systems. The database will be generated by extracting maximum features of Portable Executable files which improves the accuracy of the model. The system implements pure machine learning algorithms to identify every malware file.

Keywords — Machine Learning, Malware, Portable Executable Files.

I. INTRODUCTION

○ PROJECT IDEA

Malware is one of the biggest security threat to the computer users regardless of the hardware and software they are using. Malware, or malicious software, is any program or file that is inserted intentionally by attacker to harm computer, network. Some of the malwares are computer viruses, worms, Trojan horses, ransomware and spyware So idea behind doing this project is to detect that whether particular file contain malware or not means it is legitimate file or malicious file by extracting the features of portable executable file.

○ MOTIVATION OF THE PROJECT

Detection of malware is important with dominance of malwares on the Internet because it provides early warning to computer system about malwares and cyber-attacks. so the idea behind this project helps to get protected and secured system.

II. LITERATURE SURVEY

1. Malware is a program or file which harms the computer, network or server intentionally. Malware is a recent problem, which affects the data, devices, etc. Prevention of malware attack is important to save highly confidential files and the devices. In this section, let's have a quick look of the existing Malware detection methodologies and related works.
2. "Malware Intrusion Detection for System Security" proposed by Mrs. Ashwini Katkar, Ms. Sakshi Shukla and Mr. Danish Shaikh in year 2021. Proposed system explains the importance of Malware Detection System. In this model Random forest and decision tree algorithms are used. Both the Algorithm gives maximum accuracy, but they used small size of dataset which may lead problem in future when it comes to large amount of datasets.
3. "A Survey and Experimental Evaluation of Practical Attacks on Machine Learning for Windows Malware Detection" proposed by Luca Demetrio, Scott E. Coull, Battista Biggio, Giovanni Lagorio, Alessandro Armando, Fabio Roli in year 2020. This paper provides the functionality of preserving manipulations to the Windows Portable Executable (PE) file format. This paper has the limitations that they didn't uses Random Forest and Decision tree so the accuracy might vary and also this is powerful in case of Denial-of-Service attacks (DOS) only.
4. "Malware Detection using Honeypot and Malware Prevention" proposed by Dhruvi Vadaviya, Mahesh Panchal, Dr.Abdul Jhummarwala and Dr. M. B. Potdar in year 2019. The main intension of this paper is to elaborate the seriousness of Malware problem and project the importance of online malware analysis. This paper explains only about the protection regarding the network attacks. In this paper authors has used Honeypot system to trace the details about the hacker or the unauthorized user who is accessing the details. Proposed paper only explains about the network safety includes recording and analysis of network activities and captures, and capture, to uncover evidence of the origin of device security attacks.
5. "A study to Understand Malware Behaviour through Malware Analysis" Om Prakash Samantray, Satya Narayan Tripathy and Susanta Kumar Das in the year 2019. In proposed system authors explains about the Malware Behaviour technique to 4 tect the malwares and mentioned the advantages and disadvantages of malware behavior technique. This paper focused on malware behavior, features and properties extracted by different techniques and decide whether to include them to create behavioral based malware signature.
6. After reviewing the above literatures we have proposed the malware detection system using KNN, random forest and decision tree algorithms in order to improve the accuracy and performance of the actual implemented system.

III. LIMITATIONS OF EXISTING WORK

- Even though several new methods have been proposed by using these different malware detection approaches, no method could detect all new generation and sophisticated malware.
- For the known malware signature- and heuristic based detection approaches perform well. And, for an unknown and complicated malware behaviour-, model checking-, and cloud-based approaches perform better. However, some portion of malware could not be detected by using these approaches.

- Even though the trends in malware creation and detection approaches are changing rapidly, this study still can be considered as a key reference for the computer scientist and developers who work in this field.
- As a future work,new approach and method need to be proposed. To do that combining malware detection approaches can be one of the solutions among many.

IV. CONCLUSION

The proposed system will be able to predict the Malware in portable executable files. The proposed ML architectures, capable of learning features out of the raw inputs. Using these features it can easily find the malware present in that file. Using a classification Machine Learning algorithm based it's easy for the detection and is comparable with other methodologies. Using two methods for detecting, gives more security towards the malware. As this Malware detection system works on Machine Learning, it can be easy for it to be trained to detect new malware threats. The accuracy of the proposed malware detection model using KNN is tested, and the result shows that the model has 99.6accuracy on malware detection. Overall, the model is effective when the value of K is 7 and model is trained on the malware dataset using KNN algorithm.

- The purpose of malware analysis and detection is to collect and provide the information needed to revise system or network intrusions. Goals of malware detection system includes detecting intrusions by monitoring files in the system and classify them as legitimate and malware. Objectives of malware detection is as follows —

1. To study existing malware detection technique
2. To analyze existing malware detection techniques for greater security
3. To design and develop malware detection system
4. To compare and analyze the performance of proposed system with previously existing system
5. To validate, research, publishing in reputed general.

V. REFERENCES

[1]. G.D. Penna, L.D. Vita and M.T. Grifa, "MTA-KDD'19: A Dataset for Malware Traffic Detection" in ITASEC – 2020.

[2]. M. Gao, Li Ma, H. Liu, Z. Zhang, Z. Ning and J. Xu, "Malicious Network Traffic Detection Based on Deep Neural Networks and Association Analysis" in Sensors (Basel) - 6 March 2020.

[3]. Sudarshan N P.Dass, "Malicious Traffic Detection System using Publicly Available Blacklist's" in IEEE.

[4]. Paul Prasse, Lukas Machlica, Tomas Pevny, Jiri Havelka and Tobias Scheffer, "Malware Detection by Analysing Network Traffic with Neural Networks" in IEEE Conference of Symposium on Security and Privacy Workshops - May 2017.

[5]. Nancy Agarwal and Syed Zeeshan Hussain, "A Closer Look at Intrusion Detection System for Web Applications" in IEEE Conference of Security and Communication Networks Volume - 2018.

[6]. Gonzalo Marin, Pedro Casas, German Capdehourat, " DeepMal - Deep Learning Models for Malware Traffic Detection and Classification" on 10 March 2020.

[7]. Felipe N. Ducau, Ethan M. Rudd, Tad M. Heppner, Alex Long, Konstantin Berlin " Automatic Malware Description via Attribute Tagging and Similarity Embedding " on 15 May 2019.

[8]. Luca Demetrio, Scott E. Coull, B. Biggio, G. Lagorio, A. Armando, Fabio Roli “A Survey and Experimental Evaluation of Practical Attacks on Machine Learning for Windows Malware Detection” on 17 Aug 2020.

[9]. T. M. Mohammed, L. Nataraj, S. Chikkagoudar, S.Chandrasekaran, B. S. Manjunath “Malware Detection Using Frequency Domain-Based Image Visualization and Deep Learning” on 26 Jan 2021.

[10]. E. Gandotra, D. Bansal, and S. Sofat, “Malware analysis and classification: a survey,” Journal of Information Security, vol. 05, no. 02, pp. 56–64, 2014.

[11]. Shirkande, S., and M. J. Lengare. "A survey on various underwater image enhancement techniques." Int J Innov Res Comput Commun Eng 5.7 (2017): 13701-13710