IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114200

Detection of Malware for System Security

Preety Koli[1], Prof. D. M. Kanade[2], Priyanka Patil[3], Radhika Agrawal[4], Pratishtha Shelke[5]

Department of Computer Science Engineering, K. K. Wagh Institute of Engineering Education & Research, Nashik[1-5]

Abstract - The major goal of proposed system is to elaborate on the severity of the Malware problem and to project the importance of online malware analysis in Malware defense research. Malware is one of the most serious Internet security threats.The proposed machine learning architectures are capable of learning features from raw data.Malware detection requires advanced techniques to reduce malware threads that can disrupt computer operation [2].It may simply find the malware included in that file using these features. It is simple to detect using a classification Machine Learning algorithm and is equivalent to other approaches [3].The use of two detection methods increases the malware's security. It not only protects it from viruses transmitted over the internet, but it also protects it from malware installed on the device. As this malware system works on machine learning, it can be easy for it to be trained to detect new malware threats.We've demonstrated that the operations and behaviours typically associated with malware cannot be considered a critical component for malware detection because benign files can conduct them as well [3].

Keywords – Malware, Security, KNN Classifier, Threats

I. INTRODUCTION

Malware, which includes viruses, worms, Trojan horses, spyware, and other malicious software, is the most common computer attack [2].Malware detection requires advanced techniques to reduce malware threads that can disrupt computer operation. Malware is divided into seventeen types [2].Malware is a threat to the computer users regardless which operating systems and hardware platforms that they are using. Malware, often known as malicious software, is any programme or file designed to harm a computer, network, or server.Although each sort of malware has its own goal, the overall goal is to disrupt computer operations.

Malware assaults can crack weak passwords, penetrate deep into systems, propagate across networks, and disrupt an organization's or businesses regular operations. Malware can also lock up vital files, spam you with advertisements, slow down your computer, or reroute you to harmful webpages. The goal is to highlight that, while malware behaviour plays an important part in detecting malware samples, it cannot be totally relied upon [1].Malware detection is crucial with malware's prevalence on the Internet because it functions as an early warning system for the computer secure regarding malware and cyber attacks. It keeps hackers out of the computer and prevents the information from getting compromised.

As a result, the concept behind this project aids in the creation of a safe and secure system.We used numerous algorithms to test the correctness of each one in this project. This module detects the presence of malware in a certain file or dataset containing the feature. In the case of a dataset, it examines the features and determines whether or not malware is there.

II. BACKGROUND

Malware is any program or file that is intentionally harmful to a computer, network or server. 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 briefly go through the existing Malware detection methodologies and related works.

"Malware Intrusion Detection for System Security" proposed by Mrs. Ashwini Katkar, Ms. Sakshi Shukla and Mr. Danish Shaikh in year 2021. This paper explains the importance of Malware Detection System. Why there is need to detect malwares? Detecting the malware from portable executable files. In this model Random forest and decision tree

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114200

algorithms are used [5]. 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.

"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. In this paper authors have done different experiments to protect the windows operating system from the malware attacks. This paper also provides the functionality of preserving manipulations to the Windows Portable Executable (PE) file format [4]. It 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.

"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 [9]. 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. In this paper author has not used any algorithm to detect the malware from the portable executable files. Proposed paper only explains about the network safety includes recording and analysis of network activities and captures, to find out the proof about the source of the attacks to the device safety.

"A study to Understand Malware Behaviour through Malware Analysis" Om Prakash Samantray, Satya Narayan Tripathy and Susanta Kumar Das in the year 2019. In this paper authors explains about the Malware Behaviour technique to detect the malwares. In this paper authors also mentioned the advantages and disadvantages of malware behavior technique. Malware Behaviour technique is only good with the known attacks, but when it comes to unknown attacks the system fails [3]. This paper emphasizes malware behavior, characteristics and properties extracted by different analytic techniques and to decide whether to include them to create behavioral based malware signature.

"A framework of Malware Detection Using Combination Technique and Signature" projected by Mohamad Fadli Zolkipli and Aman Jantan within the year 2018. These days' malware writers try and avoid detection by victimization many techniques like polymorphic, activity and additionally zero day of attack. However, industrial anti-virus or anti-spyware that used signature-based matching to detects malware cannot solve that sort of attack. So as to beat this issue, the authors projected a brand new framework for malware detection that mixes signature-based technique and genetic algorithmic rule technique [2]. Authors have combined signature-based detection, GA detection and signature generator. However additionally they need the issues with giant datasets and accuracy to observe malwares from the moveable possible files. And additionally this paper has genetic algorithmic rule and also the computations of GA area unit extremely pricey.

III. METHODOLOGY

- Data Collection

In this module a data sheet is imported to the system. This datasheet has features of various Portable executable files that is used to train the model and on basis of that model further process will be feature selection. This data contains various parameters of .exe files. This file is in .csv format.

- Data Pre-processing

In this module, basically cleaning is done. Cleaning it means removing redundant data and unnecessary data from the dataset. Filling the black gaps. Further dataset is passed in format of .csv file through the model to train it and can also perform the feature selection of independent variables that will be necessary for the execution of project.

- Training Model using KNN algorithm

In this module based on the feature selection model will be trained. Then algorithms will be used like knn for the accuracy and Decision Tree for comparison purpose. And then finally file will be passed to the trained model.

- Checking whether the file is Malware or Legitimate

After final execution of file result will be displayed and system can identify that the file is malware or legitimate.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114200

KNN Algorithm

K-Nearest Neighbor might be a Machine Learning rule that comes beneath supervised Learning. In KNN there is two fully totally different cases i.e. new case and on the market case. KNN just assumes these two case data and place the new data into the foremost similar category of available data. KNN does not build any assumption, it is a non-parametric rule. it's in addition referred to as a lazy learner rule as a results of it does not learn from the work set currently instead it stores the dataset and at the time of classification, it performs AN action on the dataset.

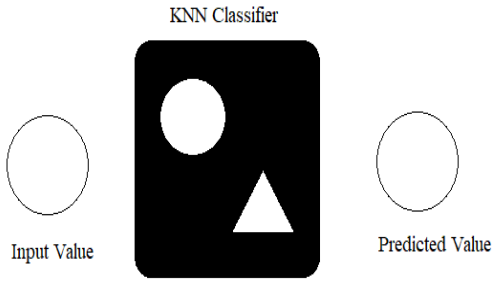

Figure 1. KNN Classifier

Algorithm of KNN Classifier:

- Step-1: choose the quantity K of the neighbors
- Step-2: Calculate the geometer distance of K variety of neighbors
- Step-3: Take the K nearest neighbors as per the calculated geometer distance.
- Step-4: Among these k neighbors, count the quantity of the info points in every class.
- Step-5: Assign the new information points to it class that the quantity of the neighbor is most.
- Step-6: Our model is prepared.

IV. RESULTS

Result Analysis& Discussion

- The Module is very helpful and productive for learning the detection of malware from the specified features.
- Module has flexible GUI which is understandable to any user.
- The module is divided into four sub modules as upload dataset, show dataset, clean dataset and prediction for test dataset. The visualization of each and every sub modules is clear and easy to understand for any user.
- Each sub module performs its specified functions as upload dataset upload the dataset from system by opening window, show dataset shows the uploaded dataset, clean dataset shows the cleaned dataset and after all the functioning it predicts the result according to the features as it is malware or not.

Classification of files

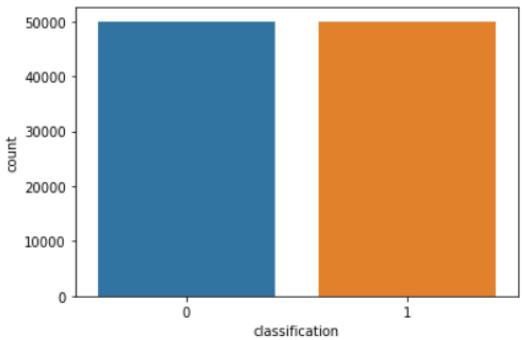

Figure 2. Classification

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114200

The distinction between malware and legitimate files is shown. There are 50 percent malware files and 50 percent legitimate files in the dataset.

Accuracy vs. k-value

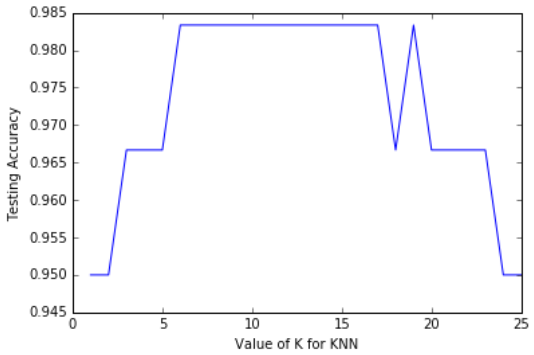

Figure 3. Accuracy vs. k-value

In this graph, predicted the accuracy for k values ranges from 0 to 25.. When a smaller number is utilised to train the model, the maximum feasible malware detection accuracy is achieved. When K=3 is utilised to prepare the model, the maximum malware detection accuracy is attained.

Actual Implemented Model vs. Proposed Model

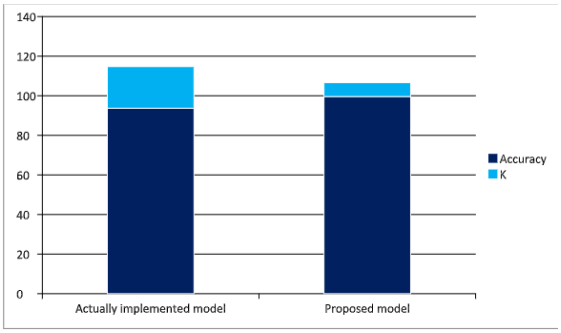

Figure 4. Actual Implemented Model vs. Proposed Model

The contrast of the actual implemented model and the proposed model is displayed in this graph. The actual implemented model has a 93.7% accuracy for the value of k is 21 while the proposed model has a 99.6% accuracy for the value of k is 7.

V. 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 aclassification Machine Learning algorithm based it's easy for the detection and is comparable with 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.

REFERENCES

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

DOI: 10.17148/IJARCCE.2022.114200

[2] Mohamad Fadli Zolkipli, Aman Jantan, "A Framework for Malware Detection Using Combination Technique and Signature Generation", 2021

[3]Om Prakash Samantray, Satya Narayan Tripathy, Susanta Kumar Das "A study to Understand Malware Behavior through Malware Analysis", 2019

[4] Luca Demetrio, Battista Biggio, Giovanni Lagorio, "A Survey and Experimental Evaluation of Practical Attacks on Machine Learning for Windows Malware Detection", 22 March, 2019

[5] Mrs. Ashwini Katkar, Ms. Sakshi Shukla and Mr. Danish Shaikh, "Malware Intrusion Detection for System Security", 2021

[6] 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.

[7] Sudarshan N P.Dass, "Malicious Traffic Detection System using Publicly Available Blacklist's" in IEEE Conference of International Journal of Engineering and Advanced Technology (IJEAT) - August 2019.

[8] 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.

[9] Dhruvi Vadaviya, Mahesh Panchal, Dr. Abdul Jhummarwala and Dr. M. B. Potdar, "Malware Detection using Honeypot and Malware Prevention", 2019

[10] 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.

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

[12] 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.

[13] 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.

[14] T. M. Mohammed, L. Nataraj, S. Chikkagoudar, S.Chandrasekaran, B. S. Man- junath "Malware Detection Using Frequency Domain-Based Image Visualization and Deep Learning" on 26 Jan 2021.

[15] M. D. Preda, M. Christodorescu, S. Jha and S. Debrey, G. Eason, B. Noble, and I. N. Sneddon, "A semantics-based approach to malware detection," ACM Trans. Program. Lang. Syst. 30, 5, Article 25, August 2008.

[16] L. Hanno, "Framework for Malware Resistance Metrics," QoP'06 in ACM, pp. 39-44, 2006.

[17] M. Christodorescu and S. Jha, "Testing Malware Detectors," ISSTA'04 in ACM, pages 34-44, 2004.

[18] H. Yin and D. Song, "Panorama: Capturing System-wide Information Flow for Malware Detection and Analysis," CCS'07 in ACM, pp. 116 – 127, 2007.

[19] Microsoft Corporation, "The Antivirus Defense-in-depth Guide," 2004, in press. [6] M. Apel, C. Bockermann and M. Meier, "Measuring Similarity of Malware Behavior," SICK 2009 in IEEE Explorer, pp. 891-898, October 2009.

[20] G. McGraw and G. Morrisett, "Attacking malicious code: report to the Infosec research council," IEEE Software, 17(5):33 - 41, Sept./Oct. 2000.

[21] S. Noreen, S. Murtaza, M. Z. Shafiq and M. Farooq, "Evolvable Malware," GECCO'09 in ACM, pp. 1569–1576, July 2009.

[22] F. Hsu, H. Chen, T. Ristenparty, J. Liz and Z. Su, "Back to the Future: A Framework for Automatic Malware Removal and System Repair," Proc. IEEE Annual Computer Security Applications Conference (ACSAC'06), IEEE Press, July 2006: 0-7695-2716-7/06

[23] Y. Zhou and M. Inge, "Malware Detection Using Adaptive Data Compression," AISec'08 in ACM, pp. 53 – 59, October 2008.

[24] Y. Ye, Q. Jiang and W. Zhuang, "Associative Classification and Post-processing Techniques used for Malware Detection," IEEE Explorer, 2009.

[25] S. Mehdi, A. K. Tanwani and M. Farooq, "IMAD: In-Execution Malware Analysis and Detection," GECCO'09 in ACM, pp. 1553- 1560, July 2009.

[26] S. M. Tabish, M. Z. Shafiq and M. Farooq, "Malware Detection using Statistical Analysis of Byte-Level File Content," CSI-KDD'09, pp. 23-31, June 2009.

[27] Y. Ye, D. Wang, T. Li and D. Ye, "IMDS: Intelligent Malware Detection System," KDD'07 in ACM, pp. 1043 – 1047, August 2007

[28] P.V.Shijo, A.Salim, Integrated Static and Dynamic Analysis for Malware Detection, the third ICoICT Conference, 2015

[29] Liu, Wu Ren, Ping Liu, KeDuan, Haixin., "Behavior-Based Malware Analysis and Detection", Proceedings of Int. Workshop on Complexity data mining, 10.1109/IWCDM.2011.17, 2011

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114200

[30] Michael Bailey, Jon Oberheide, Jon Andersen Z, Morley Mao, Farnam Jahanian and Jose Nazario, Automated classification analysis of internet malware, In: Kruegel C., Lippmann R., Clark A. (eds) Recent Advances in Intrusion Detection. RAID 2007, Lecture Notes in Computer Science, vol 4637. Springer, Berlin, Heidelberg, 2007

[31] Konra dRieck, Thorsten Holz, Carsten Willems, Patrick Dusse and Pavel Laskov, Learning and classification of malware behavior, in DIMVA2008, Springer Berlin Heidelberg, vol. 5137, 2008

[32] G. Wagener, R. State, and A. Dulaunoy, Malware behaviour analysis, Journal in Computer Virology, vol. 4, no. 4 2008

[33] K. Asmitha and P. Vinod, A machine learning approach for linux malware detection, in International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014

[34] M. Zolotukhin and T. Hamalainen, Detection of zero-day malware based on the analysis of opcode sequences, in IEEE 11th Consumer Communications and Networking Conference, 2014

[35] S. Josse, Secure and advanced unpacking using computer emulation, Journal in Computer Virology, vol. 3, no. 3 2007

[36] A.Singh and A. Lakhotia, Game-theoretic design of an information exchange model for detecting packed malware, in 6th International Conference on Malicious and Unwanted Software (MALWARE), 2011

[37] S. Association., IEEE-SA - Industry connections security group (ICGS), 2015

[38] T.-Y. Wang and C.-H. Wu, Detection of packed executables using support vector machines, in International Conference on Machine Learning and Cybernetics (ICMLC), 2011, vol. 2, 2011

[39] C. Linn and S. Debray, Obfuscation of executable code to improve resistance to static disassembly, in Proceedings of the 10th ACM Conference on Computer and Communications Security, ser. CCS 03. New York, NY, USA: ACM, 2003

[40] Victor Marak, "Windows Malware Analysis Essentials", PACKT Publishing.