WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DETECTION OF MALWARE IN DOWNLOADED FILES USING VARIOUS MACHINE LEARNING MODELS

JOTHEESHWARAN J

Master of Computer Applications,
M.G.R. EDUCATIONAL AND RESEARCH INSTITUTE, Chennai, Tamil
Email: jotheesh251001@gmail.com

ABSTRACT :

The rapid spread of computer networks has changed the way network security is approached. However, with a very easy access, computer networks are very vulnerable to numerous attacks from the intruders, which makes threats to the network huge and potentially catastrophic. On the other hand, researchers have come up with a number of Malware Detection Systems (MDS) that can trace attacks on different platforms corre- sponding to such obstacles. These setups use a wide range of possible means, like the misuse detection and the anomaly detection, where each of them has the advantage part of the strengths of the applicable working environment. The majority of these methods are complementary, as some of them are more suitable for cer- tain environments than others. A new malware detection infrastructure that is the fundament of the classes of threats leading to the correct taxonomy and the realization of the detection's founding principles together with the most important operational features in the field is hereby introduced by this project. To give the system dependence and improved precision of the detection, the new model employs computer learning- based algorithms such as Random Forest (RF) and Artificial Neural Network (ANN). The performance of the algorithms is validated on the basis of detection accuracy, targeting to provide a resilient and self-adapting solution to combating dynamic IT security threats.

Keywords: Malware Detection Systems, Anomaly Detection, Threat Classification, Data-driven Security

1. Introduction

The computer networks revolution has touched every corner of the globe and resulted in significant improve- ments in communication and information exchange. However, at the same time, it has created very serious security problems. Networks are more vulnerable to intruders as they become more accessible, resulting in malicious attacks, such as malware that can be a big loss for both the systems and the data. The existing security measures are not always good at finding and reacting to the new and innovative threats. To solve this, MDS, which stands for Malware Detection Systems, have been designed by scientists as a combination of the misuse detection and the anomaly detection methods to catch the cyber criminals running different tasks on a number of platforms. The present paper suggests a new method of malware identification which relies on machine learning models, especially Random Forest (RF) and the Artificial Neural Network (ANN), in order to improve the accuracy and consistency of the results. These computational methods can find very complicated models in the network behavior and then decide which category the attacks belong to. Besides that, the authors offer a classification scheme based on the detection concepts and the operating features of the proposed system. The classification scheme is designed to fully meet the demand for a flexible and suf- ficient antivirus program of the current cybersecurity problems.

2.. REVIEW OF LITERATURE

The evolving intricacy of cyber threats necessitates the development of smarter and more adaptable malware detection systems. Traditional signature-based methods perform well against known threats but frequently miss novel or evolving malware. Researchers now utilize machine learning methods including Random For- est (RF) and Artificial Neural Network (ANN) to enhance detection capabilities. The robustness of RF com- bined with its high accuracy and capacity to manage extensive datasets with numerous features makes it ideal for real-time analysis. ANN's design based on human brain function enables it to detect intricate non-linear data relationships that are crucial for identifying advanced malware patterns. Research indicates that merging these algorithms helps decrease false positive rates while boosting system dependability. Modern Malware Detection Systems (MDS) rely on both misuse detection to identify known patterns and anomaly detection to recognize unusual behavior as essential components. Research indicates that employing a combined method using both techniques results in enhanced adaptability for multiple environments. The combination of RF and ANN with MDS marks a major advancement toward developing more precise and expandable cybersecurity systems.

3. METHODOLOGY

a. Data Collection and Preprocessing:

During this step, various malware and benign network traffic datasets are collected from reputa- ble sources. Raw data is cleaned to remove extraneous or corrupted records. Feature extraction meth- ods are used to recognize meaningful features applicable to malware activity. Afterwards, data normaliza- tion is performed to provide uniform scaling over features, improving model training efficiency.

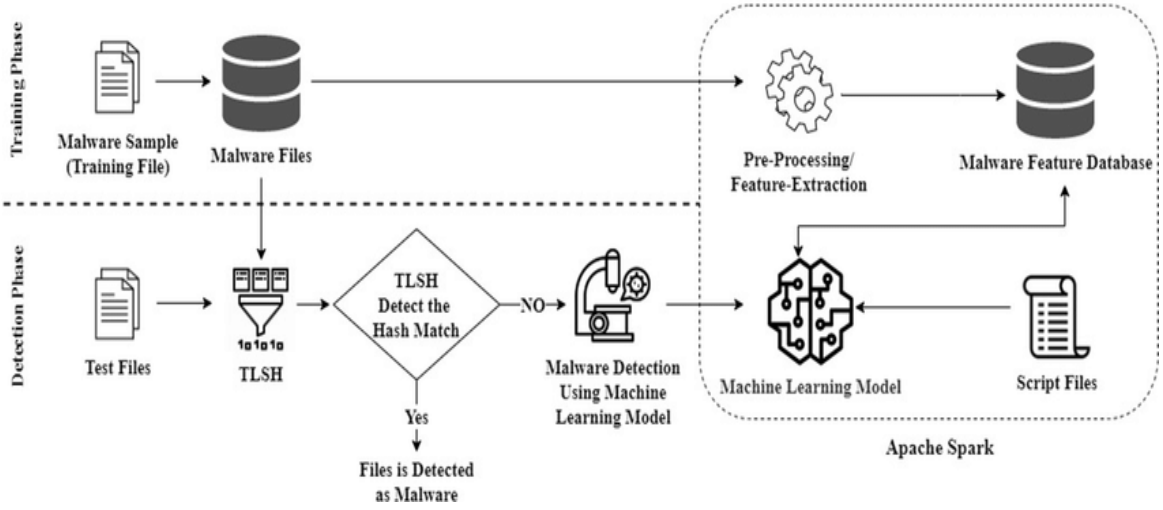
b. Algorithm:

The Random Forest (RF) algorithm is used as an ensemble learning algorithm that trains several decision trees to enhance classification efficiency and avoid overfitting. At the same time, the Artificial Neural Net- work (ANN) is built with multiple layers for extracting intricate, non-linear relations from the data. Hy- perparameter tuning is done for both models in order to maximize performance depending on the character- istics of the dataset at hand.

c. Model training and evaluation:

Both RF and ANN models are trained on a labeled dataset split into training and testing subsets to avoid bi- ased assessment. Performance metrics such as accuracy, precision, recall, and F1-score are com- puted to evaluate the detection ability of each model. The outcomes are compared to identify the strengths and limitations of each algorithm to detect malware effectively.

4. MODULE



5. SYSTEM IMPLEMENTATION

The implementation of the system is initiated by the collection and preprocessing of network traffic data, both normal and malicious samples. Feature extraction methods are utilized to determine attributes rel- evant in distinguishing malware from normal behavior. The processed data is then used to train two machine learning models: Random Forest (RF) and Artificial Neural Network (ANN). Both models are trained on la- beled datasets to acquire patterns linked with malware activities. Hyperparameter tuning is conducted to re- fine model performance. The models are deployed by the system after they have been trained to clas- sify real-time network data. Classification results are logged and examined for accuracy and false positives. A user interface is built to provide display of detection results and system status. The system is tested in a multitude of simulated network environments to ensure robustness and adaptability. Lastly, perfor- mance metrics are logged to assess the overall impact of the malware detection system.

6. Result

The suggested malware detection system based on Random Forest (RF) and Artificial Neural Network (ANN) was tested on benchmark datasets to quantify its performance.Both models had high accuracy in rec- ognizing malicious behavior, where RF delivered marginally higher precision and lesser false positive rates, while ANN was superior to detect sophisticated, non-linear patterns. The comparison revealed that merg- ing both methods would benefit from their individual strengths in terms of enhanced detection perfor- mance. Results analysis showed that the system properly differentiated between benign and malware sam- ples under various network environments. The discussion indicates the significance of feature extrac- tion and

model adjustment in achieving optimization on detection accuracy. Although RF of- fered quicker training and prediction times, ANN consumed greater computational power but pro- vided greater insights into complex malware behaviors. In general, the system has good potential for real- time deployment, though more work is required to deal with new malware variants and minimize false alarms.

7. conclusion

The explosive expansion of computer networks amplified the demands for efficient malware detection sys- tems to defend against adaptive cyber threats. This study introduced a new malware detection method by in- tegrating Random Forest (RF) and Artificial Neural Network (ANN) algorithms to improve detection accu- racy. Both algorithms exhibited robust performance in distinguishing malware with high precision and re- call. The employment of a systematic taxonomy assisted in detection principle and operational organiza- tion, enhancing system flexibility. Experimental results showed that RF provides quicker pro- cessing, whereas ANN performs well when it comes to detecting intricate patterns within data. The comple- mentary capabilities of both methods imply that the use of hybrid systems could yield more effective secu- rity solutions. The suggested system is promising for real-time malware detection in heterogeneous network environments. Future research will involve increasing the size of the dataset, enhancing feature extraction, and combining other machine learning methods to decrease false positives further. Overall, this work helps advance intelligent and scalable cybersecurity defenses.

8. REFERENCES.

1. F. Akyildiz et al., “Wireless Sensor Networks: A Survey, “Elsevier Comp. Networks, vol. 3, no. , 2019, pp. 393–422
2. G.Li, J.He, Y. Fu. “Group-based Malware detection system in wireless sensor networks” Com- puter Communications, Volume 31, Issue 18 (December 2019)
3. Michael Brownfield, “Wireless Sensor Network Denial of Sleep Attack”, Proceedings of the 2019 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.
4. FarooqAnjum, DhanantSubhadrabandhu, SaswatiSarkar *, Rahul Shetty, “On Optimal Placement of Malware Detection Modules in Sensor Networks”, Proceedings of the First International Conference on Broadband Networks (BROADNETS19).
5. Parveen Sadotra et al, International Journal of Computer Science and Mobile Computing,Vol.5 Issue.9, September- 2019, pg. 23-28.