LITERATURE REVIEWS

Malware detection techniques are currently in use, although studies have revealed that they also have certain drawbacks. Because it is impossible to manage databases for each version, signature-based approaches are no longer useful. Only signatures that are already stored in the database can be detected. A rather simple obscurity mechanism can be used to avoid this strategy. Behavior-based techniques are only effective when malicious and benign files are executed together. Attacks that are well communicated have the ability to deceive anomaly-based strategy. It has a high rate of false positives. Malware detection using PE-header based methods has also been successful, but the challenge with these approaches is picking the optimal attributes for categorization. Also, when the executables are compressed, PE-header based techniques fail.

Arkajit Datta et al. [4] talked about the malware analysis methods i.e. static and dynamic. Limitations of static analysis and concluded that dynamic analysis is better.

Mahajan et al. [3], the malwares could be divided into various families and then the resulting dataset should be used for machine learning, various ML algorithms were compared on various tools and the results state that Random Forest is the best algorithm for analyzing the dataset.

A. Wallenstein et al. [1], there are two ways to group metamorphic malware, based on how they communicate and transform themselves. In terms of communication, it can be categorized into open-world or closed-world malware. Openworld malware can mutate through connection with other sites over Internet, just as Conficker worm did in 2008. As for closed-world malware, on the other hand, it reprograms itself by referring to a pseudo-code representation that it carries.

Rezaei et al. [2] proposed a method based on detection circle using logarithm probabilities on string occurrence, specifically located character occurrence, and the amount of virus similarities. They claimed the proposed method can achieved more than 91% viruses which outperformed anti-viruses scanner. When applying

encryption, anti-virus failed to detect the viruses but their method can detect 70% of them. This implies applying heuristic method is better than tradisional anti-virus scanner.

Isadora P. Possebon et al. [5] proposed ensemble learning technique for network traffic classification. How to combine individual algorithms using meta-learning techniques in order to obtain more robust traffic classification metrics. They presented a comparative analysis among meta-learning approaches and individual classifiers to classify network traffic. They investigated and evaluated a range of meta-learning techniques, including Voting, Stacking, Bagging and Boosting. Then proposed a new experimental analysis of different meta-learning techniques - also known as ensemble learners - and compared them with their own base classifiers when used individually.

Liao Yibin [6] By examining the MS Windows PE header attributes, a method to identify between safe and malicious exe files was proposed in 2012. Utilizing the standardized structure information in Windows, the distinctive traits are retrieved. Three actions (1) A sizable dataset of malicious and legitimate executables was gathered, (2) the characteristics of each header file were extracted and compared to identify the distinctions between legitimate and malicious exe files, and (3) icons from the PE file were extracted in an effort to identify the malware files' most common icons. Yibin Liao in 2012 provided a method for separating malware from benign executable files by examining the MS Windows PE header information.

Dhruwajita Devi et al. [7] (2012) suggested that the properties of the PE file play a crucial part in identifying packed executables in order to understand the significance of the PE file's attributes. During packing, the PE file's internal structure will change significantly. Due to this modification, it is impossible to do reverse engineering to analyse the file and for antivirus software to determine whether it is malware or benign. In order to determine if a given binary is packed or not, the task is divided into two phases: extracting

different aspects of portable executables, assessing the features extracted, and finally selecting the best set of features.

In 2015, Mohaddeseh Zakeri et al. [8] focused on important static heuristic features and fuzzy classification algorithms to develop a test to detect malware and packed files. To prevent being discovered, this was done by using a variety of obfuscation tactics. Unfortunately, classic malware detection techniques like signature scanning are no longer reliable, as has been asserted. Researchers have shown that these obfuscations affect the PE File and result in a number of abnormalities.

Park et al. [9] proposed technique to create a common behavioural graph that represents the execution behaviour of a family of malware instances was put forth by Park and others. By grouping a collection of distinct behavioural graphs, which represent kernel objects and their properties based on system call traces, the approach creates a single common behavioural graph. The resulting common behavioural graph has a path known as HotPath that is shared by all instances of malware from the same family. No matter how many new instances are introduced, the resultant common behavioural graph is very scalable. Additionally, it is resistant to system call attacks.

Eskandari et al. [10] in order to increase the accuracy of the malware analysis procedure while keeping its speed at a tolerable level, Eskandari introduced a revolutionary strategy that combines machine learning techniques with utilising hybrid analysis methodologies. They referred to their method as HDM-Analyzer, which is short for a hybrid analyzer built using data mining techniques. The API call sequences were extracted using dynamic analysis whereas the enriched control flow graph (ECFG), which contains information about API calls, was extracted using static analysis. After extracting the features, they employed a matching engine to merge the features with the appropriate ECFG. Each conditional jump was then given a label based on the dynamic information. Now, a machine learning technique is used to create a learning model using the labeled ECFG nodes. When scanning for unknown executable files, HDM-Analyzer employs this learning model.

Mahawer and Nagaraju in 2013 used Opcode with a histogram intersection kernel and Support Vector Machine were employed. by. An efficient metamorphic malware detection system had been created using histogram intersection. This method increased the detection rate by using code normalisation to base the detection method on Most Frequently Occurred (MFO) Opcodes in disassembled files. The Euclidean Distance Equation was used to calculate the distance in opcode between malicious and benign files.

Research Gap

Static analysis is not possible when malware is encrypted or packaged. Dynamic analysis appears to be the most typical solution in such instances. However, the problem is figuring out how to evaluate behavior and quantify it in order to detect metamorphic malware in an automated manner. Currently, there is no method to detect metamorphic malware with high positive rate. The main problem with the detection of metamorphic malwares is availability of limited resources and dataset. Due to its highly complex nature the detection rate is very low. Our main objective is to study the behavior and properties of metamorphic malware for its detection with high positive rate.

REFERENCES

1. A. Wallenstein, R. Mathur, M. R. Chouchane, and A. Lakhotia, “The design space of metamorphic malware,” in 2nd International Conference on i-Warfare and Security (ICIW 2007), pp. 241–248.
2. F. Rezaei, M. K. Nezhad, S. Rezaei, and A. Payandeh, “Detecting encrypted metamorphic viruses by hidden markov models,” in Fuzzy Systems and Knowledge Discovery (FSKD), 2014 11th International Conference on, pp. 973–977.
3. Mahajan, G., Saini, B., & Anand, S. (2019, February). Malware Classification Using Machine Learning Algorithms and Tools. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-8). IEEE.
4. Arkajit Datta, Kakelli Anil Kumar, Aju. D. International Journal of Engineering Research & Technology (IJERT). ISSN: 2278-0181 Vol. 10 Issue 04, April-2021.
5. Isadora P. Possebon , Anderson S. Silva , Lisandro Z. Granville , Alberto Schaeffer-Filho , Angelos Marnerides in IEEE 2019 Symposium on Computers and Communications (ISCC) “Improved Network Traffic Classification Using Ensemble Learning ” DOI: 10.1109/ISCC47284.2019.8969637.
6. Yibin Liao. “PE-Header-Based Malware Study and Detection.” Corpus ID : 16132156. Computer Science, 2012.
7. Dhruwajita Devi and Sukumar Nandi. “PE File Features in Detection of Packed Executables”. International Journal of Computer Theory and Engineering, Vol. 4, No. 3, June 2012.
8. Mohaddeseh Zakeri, Fatemeh Faraji Daneshgar, Maghsoud Abbaspour. “A static heuristic approach to detecting malware targets.” Security Comm. Networks 2015; 8:3015–3027. 2015 John Wiley & Sons, Ltd.
9. Park, Y., Reeves, D.S., Stamp, M.: Deriving common malware behavior through graph clustering. Comput. Secur. 39, 419–430 (2013)

10. Eskandari, M., Khorshidpour, Z., Hashemi, S.: Hdm-analyser: a hybrid analysis approach based on data mining techniques for malware detection. J. Comput. Virol. Hacking Tech. 9(2), 77–93 (2013)