iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39813

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

A Survey on Malware detection using ML

Prof. Pritam Ahire[1], Mohanki Shreya[2], Shreya Shinde[3], Preeti Pisal[4], Manasi Manikumar[5]
[1, 2, 3, 4, 5]*Computer Engineering, Savitribai Phule Pune University*

***Abstract:** This Malware detection is a field of computer security that deals with the study and prevention of malicious software. It is not the only way to defend a company against a cyber- attack. In order to be effective, companies should analyse their risk and identify the vulnerabilities. In this paper, we will examine different techniques used to detect computer malware and malicious websites as well as future directives in this area of study and also, we will discuss the growth in computer malware and how traditional methods of detection are being replaced by innovative techniques like behavioural-based model and Signature-based model. Future directives involve developing better security products in order to fight against cyber fraud which is on a rise in recent years especially in Asia Pacific region. With this increase in cyber frauds and other malicious activities, traditional methods are not enough to block computers from it as this method has many drawbacks. In order to tackle these issues, researchers have been developing new techniques such as heuristic analysis, static & dynamic analysis which can detect more than 90% of malware samples without any false positives or negatives.*

***Keywords:** Behaviour-based approach, Dynamic analysis, Heuristic, Malware, Ransomware, Signature-based model, Static analysis, Vulnerability.*

I. INTRODUCTION

This As the number of malware threats is increasing, it becomes increasingly important to protect our computers and smartphones with anti-malware software. Machine learning is a powerful technology for detecting malicious software. It is trained on millions of samples so that it can learn to spot their characteristics at scale even when there are new types of malwares which have never been seen before. It is an approach to artificial intelligence that can be used to detect malware which uses pattern recognition, which extracts features from the file and compares them to known malware signatures. Also, it includes scanning the entire system or parts of it, extracting features of malicious software, comparing these features to known behaviors, and identifying the presence of malware. There is a very large competition between malware designer and examiner. Both research communities are working similarly, one of them designer is developing the malware detection system and other is designing the malicious software detect the software which will attack the computer and networks resources. Malware examiner examines the known malware and try to detect the malware to avoid the attack on the user's computer systems.

Malware found in the user's system are detected using either signature-based or behavior-based techniques. The signature-based malware detection system is quick and systematic but can be easily avoided by the obfuscated malware. On the other side, behavior-based techniques are stronger as compared to the obfuscation technique. Although, behavior-based techniques are very time-consuming. The onus for protection on the data is on the user. There are many different types of malwares which can attack in different ways. For example, spyware can record keystrokes and further make hacker's way much easier as the spyware aims to gather data about the victim or any organization and forward it to another system in such a way that it breaches the victim's privacy and benefits the cyber punks. While ransomware is a malware that encrypts user's computer files and demands ransom for data to be unlocked again. A user or organization's censorious data is encoded so that they cannot access databases, files, or applications. Hence, some amount of money is demanded to provide access for their data only. System virus is generally a program that is built outside the user's will and can cause damage to both the operating system and the hardware (physical) elements of a system. The various effects caused due the viruses are:

1) Destruction of files in the system.
2) Change in the file size.
3) Delete all data on the disc.
4) Damage in the file allocation table, which makes it impossible to read the information on the disk.
5) Various innocuous but disturbing graphic / sound effects.
6) Slowing down the working speed of the computer until it crashes worms. Computer worms are programs with damaging effects which use communication between computers to spread. Worms have similar features with viruses, worms are able spread like viruses in the system, but not locally, but on other computers. It uses computer networks to spread to other systems.

II. MOTIVATION

Our computers are vulnerable to malicious code. Malicious code is a type of software that seeks to damage or disrupt computer systems. Any code that is embedded in a program with the intent to destruct, corrupt, disable information, or gain access to sensitive information is considered malicious. Malware is continuously evolving and becoming more sophisticated, in order to exploit the power of our devices. Manufacturers, developers, and consumers need to pay attention to the security of their devices. This will reduce the probability of malware exploitation in the future. Malicious code is often categorized by the method used for delivery. The two most common types are Dropped Malware and Drive-by Malware. Dropped malware refers to programs that are sent through email messages, messages on Facebook chat, instant messages, Skype contacts and other social media platforms while Drive-by Malware refers to programs or malware that are delivered when a user accesses an infected website or downloads an infected file mistakenly thinking it's legitimate. When users download free software from the internet without double checking its reputation, they put themselves at risk of downloading malware which could invade their privacy and critical data.

The Malicious code needs to infiltrate the system and subvert the legitimate system before it can do any harm. A malicious program often downloads and executes other malicious programs on the compromised device. It does this through exploitation of vulnerabilities in the computer's operating system and installed applications. A recent study shows that more than 12% of top trending terms searched return corrupted results. This article gives a clear idea of how to design robust signature-based malware detection systems. The authors discuss the main disadvantages of this approach and propose various techniques which can help to improve the efficiency. The article talks about how machine learning is becoming popular for malware detection, but it is not the best solution in all cases. The authors discuss different approaches and provide examples on when they should be used

III. RELATED WORK

There are various approaches to malware detection. However, malware authors have used various techniques to evade detection and therefore this has led to a need for new and better approaches that can provide more accurate results.

Malware detection is a difficult task to accomplish. Even after updates, new malware will continue to pop up, making it a never-ending battle. The traditional signature-based malware detection is not a good solution for this problem, because new samples can easily bypass these methods. A vision-based approach is a newer method which uses deep learning techniques in the AI process to identify the features of malicious software, this however does not work when there are no examples or when they are not available for training. To overcome this, researchers have proposed a selection and fusion approach which combines both these approaches in order to produce more accurate results. The copy-based technique relies on the signature of the malware. The detection is based on the comparison of the sample with signatures which are stored in an offline repository. The vision-based malware detection techniques are based on detecting anomalies in images taken from a webcam, or on detecting suspicious downloads. They have not been found to be as efficient as machine learning algorithms

IV. LITERATURE SURVEY

Sr No	Name Of Paper	Author	Existing System	Published On	Elements
1.	A survey on machine learning based malware detection in executable files	S.Geetha, Seifedine Kadry	IEEE	2020	The aim is to utilize the concept of machine learning and to build a model using algorithm.
2.	Analysis of Res Net and Google Net models for malware detection	R. U. Khan, X. Zhang and R. Kumar	IEEE	2019	As malicious Software increasing threats to computer software, so it has become key issues

					in today's era.
3.	A malware Detection method based on family behaviour graph.	Y.Ding, X. Xia, S.Chen and Y.Li	IEEE	2018	ML algorithms provide more options and space.
4.	How to make attention mechanisms more practical in malware classification	X. Ma, S.Guo, H, Li, Z. Pan, J.Qiu,Y.Ding	IEEE	2018	In ML malware samples are analysed and features set off extracted info
5.	Deep ground truth analysis of current Android malware	F.Wei, Y. Li, S.Roy, X.. Ou and W.Zhou	IEEE	2017	Proceeding International Conference Detection Intrusion Malware
6.	Deep learning at the shallow end: Malware classification for non-domain experts	Q. Le, O. Boydell, B. M. Name and M. Scanlon	IEEE	2018	ML is benefit in malware detection as it can develop a model to detect unknown malware.
7.	Novel feature extraction selection and fusion for effective malware	M. Ahmad, D. Ulyanov, S. Semenov, M. Trofimov and G. Giacinto		2018	In usability of ML algorithm, it has many advantages to detect malware.
8.	A Feature dependent Naïve-Bayes Approach and its application to software defect prediction problem	Over Farak Arar and Kursat Ayan	IEEE	2017	Naïve Bayes, features are assumed to be independent and have equal weight.

V. METHODOLOGY

The results of five different classification algorithms that are utilized for prediction are compared in this approach. An ML model is used to predict the class for a given file based on a previously trained model. Among the machine learning models examined were Ada-boost, decision tree, gradient boosting, and gaussian. To analyze data patterns, algorithms must be taught. Android was first released in 2008, and ML is showing signs of infiltration. Security issues were emphasized a few years later as the popularity of Android applications grew. In the last five years, there has been a greater focus on using machine learning for software security because many researchers are constantly identifying and proposing new ML-based solutions. We then generated numerous research questions based on the study's goal. The next step was to devise a search strategy for locating completed studies that may be used to address our research questions. At this point, the database's use, as well as the criteria for inclusion and exclusion, were determined. The study selection criteria were established in order to discover papers that aimed to answer the research objectives as articulated.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

VI. ALGORITHMS

A. *Naive Bayes*

It's probably the most powerful and accurate probabilistic machine learning algorithm. The advantages of using the Naive bayes algorithm is that it's fast and efficient to train and can generalize well from small amounts of training data. Naive Bayes algorithm is a probabilistic classification algorithm, which belongs to the group of expected-a-posteriori algorithms. In the probabilistic approach, each event is considered as a random variable, and the probability of each event is computed by dividing the number of occurrences of that event by the total number of occurrences. Naive Bayes classifier assumes that all features are independent of one another.

B. *Ada-boost*

Ada-boost is the market's top performing algorithm that has been developed for over 10 years. It is a proprietary, fast and accurate machine learning technique that improves results for machine learning problems. It has been tried and tested on an array of problems from data mining to natural language processing to computer vision problems. Ada-boost is a straightforward algorithm for designing a machine learning system. It works by incrementally improving a classifier's performance on a training dataset as additional data becomes available. It should be noted that the algorithm is designed to incrementally fit the parameters of the classifier to the training data. Ada-boost requires sequential training and cannot be applied to out-of-order data. The algorithm was originally developed by Bernard Widrow and Ted.

1) Mathematical Model

a) Naive Bayes:

P(b|y) = P(y|b) P(b) / P(y)

{P(b|Y) = P(y1|b) X P(y2|b) ……P(yn|c) X P(b)}

where,

P(b|y) is the posterior probability of certain class (b, target) given predictor (y, attributes).

P(b) is the prior probability of given class.

P(y|b) is the likelihood which is the probability of the predictor given class.

P(y) is the prior probability of predictor.

2) *Decision Tree:* A decision tree is a tree structure in which an internal node represents a property (or attribute), branch represents a decision rule, and each leaf node indicates the result. The root node is the topmost node of the tree. Recursive partitioning is a method of partitioning the tree in a recursive manner. This flowchart-like structure assists you in making decisions.

It measures impurity in the node of the tree. Its value lies between 0 and 1. Hence, the Gini index of value 0 means sample are perfectly equivalent, and all elements are similar, whereas Gini index of value 1 means maximal inequality among elements. It is sum of the square of the probabilities of each class.

Mathematically it is written as:

$$Entropy = -\sum_{i=1}^{n} p_i * \log(p_i)$$

3) *Gini index/Gini Impurity:* It measures impurity in the node of the tree. Its value lies between 0 and 1. Hence, the Gini index of value 0 means sample are perfectly equivalent, and all elements are similar, whereas Gini index of value 1 means maximal inequality among elements. It is sum of the square of the probabilities of each class.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

It is illustrated as,

$$Gini\,index = 1 - \sum_{i=1}^{n} p_i^2$$

a) Firstly, assign a sample weight for each sample

$$sample\ weight = \frac{1}{\#\ of\ samples}$$

b) Then we have to calculate the Gini Impurity for each given variable

$$Gini\ Impurity = 1 - (the\ probability\ of\ True)^2 - (the\ probability\ of\ False)^2$$

c) Calculate the Amount of Say for the stump that was created

$$Amount\ of\ say = \frac{1}{2}\log\left(\frac{1 - total\ error}{total\ error}\right)$$

d) Then we have to calculate the new sample weights for the next stump.

VII. SYSTEM ARCHITECTURE

A. *Architecture of Hypervisor*

Hypervisor it is a crucial piece of software that makes the virtualization possible. It theoretical guest the machines and the operating system they run on from the effective hardware. In this architecture, there are two hypervisor Type-I hypervisor and Type-II hypervisor. The Type-I is known as bare metal or native metal whereas Type-II is known as hosted hypervisors. Hypervisor is used in this project, to abstract operating system and application from their essential hardware. It creates a virtualization that separates the CPU and then the RAM and other hardware resources from the virtual machine you create.

1) A bare metal hypervisor (Type-I) has a direct access to hardware resources. It consists of various VMM (Virtual *Type-I Hypervisor:* Machine Monitors) as per requirement and hardware. This hypervisor is also considered to be secure Examples of Type-I hypervisor are VMware, ESXi, XEN, AHV.
2) *Type-II Hypervisor:* A hosted hypervisor (Type-II) runs inside an operating system then as per the requirement of hardware. It consists of VMM (Virtual Machine Monitors), examples of Type-II hypervisor are VirtualBox, VMware Player, QEMU. The main difference between the Type-I and Type-II hypervisor is that Type-II hypervisor are generally installed on a pre-existing OS.

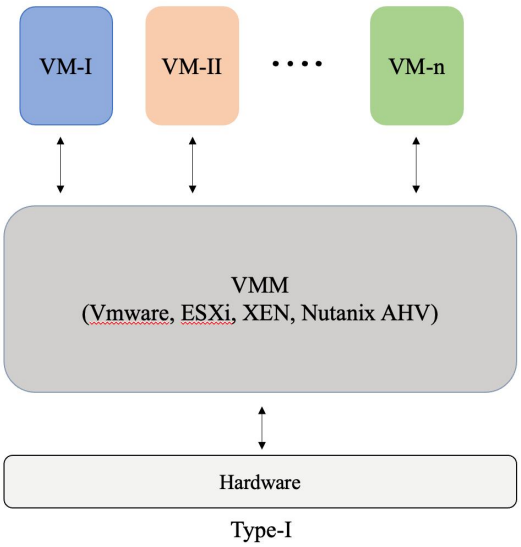


Fig: Type I Hypervisor

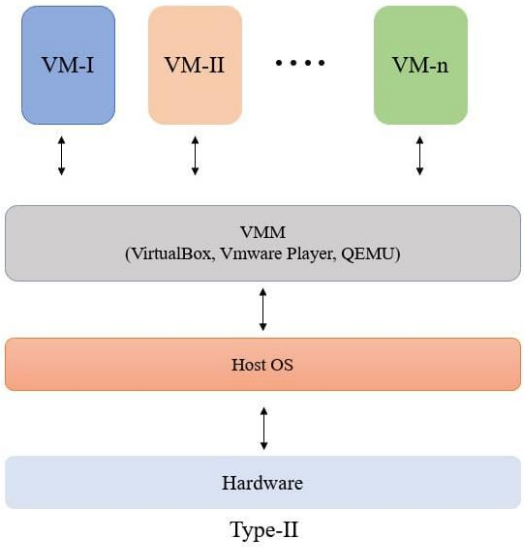


Fig: Type II Hypervisor

B. Architecture of Signature-based Model

Signature based techniques are fast and abstract. The one simple way of creating signature- based malware files is using a hash algorithm. Hash algorithm is an encryption algorithm and is used to verify integrity of data. Some commonly used hash algorithms are MD5, SHA-1, SHA-2, NTLM, LANMAM. In this signature-based approach the malware is detected based on general pattern of files. In this approach, first the data samples are collected in Malware samples and Benign samples. Malware samples are used to resolve hazard technique and develop defenses.[15]

These two types of samples are extraction of software in a static features form. In Signature-based model there are many tools that have been developed in order to find and identify malware. The signature may be a combination of strings, bytes, or an executable file or something else entirely. One of the most popular, is using PE View, Hash Generator or, PS Studio which allow extracted applications to be executed in an isolated environment so that the risks of running unknown code are minimized with the use of assemblers that are IDA Pro, Capstone, which further extract n-gram bytes which can detect malicious files and the four algorithms applied over the n-byte feature sets are Naïve Bayes (NB), Ada boost, Decision tree (DT) and Artificial Neutral Networks (ANN).So basically, this technique identifies malicious code based on a pattern matching approach. This technique uses the scan for sequence within code to identify malicious.

However, the disadvantage with this approach is that many common malicious behaviors may be missed because of their short execution time or simple operation. Although this technique can detect malicious files significantly to some extent. It is important to note that this method does not work if the pattern has been changed by a virus, so it may be difficult to detect new viruses with this method alone since it uses a database of known patterns shared among many anti-malware programs. This technique is often used by antivirus or anti-malware software, which relies on the signature database to identify what virus or other malicious code is being installed on a computer.

This technique has become less effective due to the following reasons: i) The development of polymorphic code. ii) The increased use of encryption and encoding techniques in malware injection to prevent detection during transmission. iii) The increasing number of variants, both known and unknown, with which attackers can sign their malicious code with legitimate keys, thus preventing it from being detected as such. Grouping of un-labelled examples in machine learning are called as clustering. It depends on an unsupervised machine learning. On the other side, the examples that are labelled are used for clustering then it becomes classification. Thus, by using this method the algorithms used are K-means, SVM, RF, k-NN.

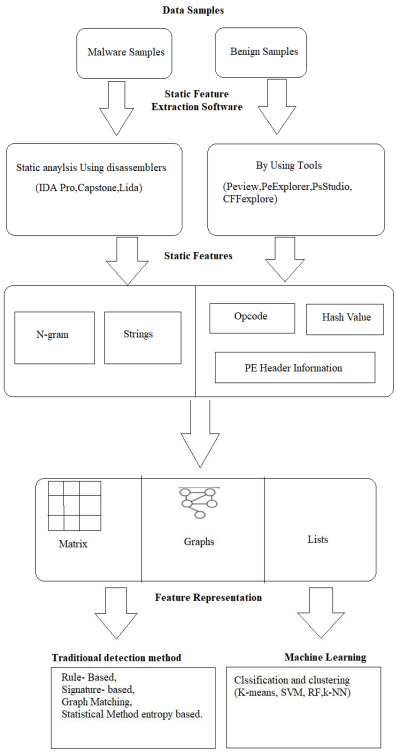


Fig: Signature- Based model

C. Architecture of Behavior based model

There are many ways that malware can execute on a machine. Allowing the detection of malware to be based on behavior would provide better protection against new or mutated viruses. Behavior based malware detection is also known as heuristic detection in the computing industry, and it deviates from the traditional method of scanning executables. For example, when scanning executables that are executed in memory only, like scripts or web pages, there is no need to scan them for malicious activity since they cannot be executed without an executable present. It can detect new mutants of malware and differentiate whether a file is benign or malicious. Detecting malware on the surface is easier as it will show itself as a process or a file on the system. But behavior-based detection methods look for more subtle indicators of malware such as changes in the frequency and duration of start-up sequences.

Behavior based techniques are more resilient and are very time consuming. The behavior of malware files depends to many categories. Behavior depends on various features such as API's, browser, system and network events. In this approach, it provides a way to solve the obfuscated malware. For solving this method, the obfuscation technique is used. It is the technique that makes the textual and binary data difficult to understand, that is, to make attackers hard antivirus to detect the malware files. The data samples are collected in Malware samples and Benign samples. These two types of samples are extracted in a behavioral features form. In Malware samples, it uses the tools, so the tools used are Process Explorer, Wireshark, Reg shot, T Dump. In Benign samples, they are analyzed by using static or dynamic analysis. It uses the sandboxes which detects suspicious activities in VM. The sandboxes used are Cuckoo, CW Sandbox, Anubis, Norman. The features are extracted and converted into File operations, Registry Changes, Network artefacts, System calls. It is the best approach to differentiate between malware and benign files. The converted features are done in traditional detection method and machine learning. In traditional method, the tools used are Rule-based, Graph Matching that is on API calls and Statistical Method. On machine learning algorithm, the method used are classification and clustering, in this method, the algorithm used are SVM, DT, KNN, NB, Ensemble, CNN, RNN, K-means, etc.

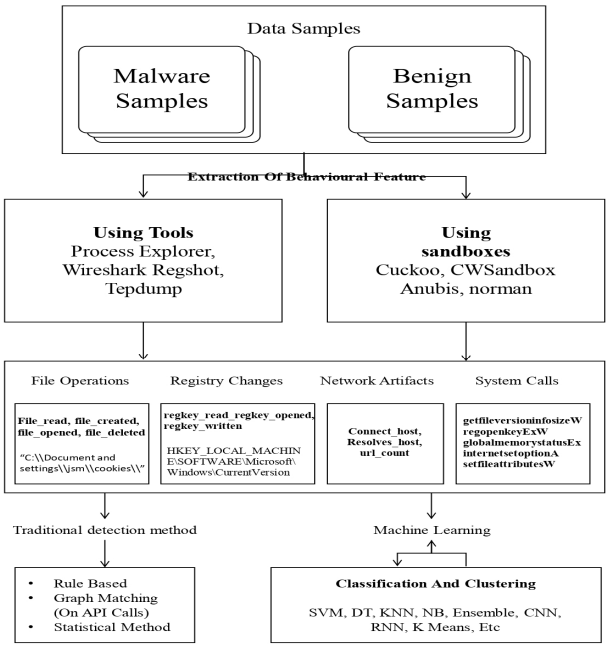


Figure. Behavioural – based Model

Fig: Behaviour-based Model

D. Architecture of Hybrid Based Malware Detection

The signature and behaviour-based techniques have both do's and don'ts, so to overcome this issues researcher have developed hybrid based malware detection. In this architecture, the files that are malicious and benign are analysed in static and dynamic ways. These files are transferred into training malware classifiers. It is the method of assigning a malware sample to a distinct malware family. These classifiers are used for many important security problems. At last, the testing is done in both the analysis that is in signature based and behaviour -based technique. These techniques are done for both the files where finally the malware detector database is updated. This both detection is used to detect the known and unknown malware files. So, by using this technique the scanning of malware time is less and a few false positives. It also shows best results in detecting the polymorphic malware. Polymorphic malware means many forms of malware that are viruses, worms, bots, Trojans. This hybrid approach is more accurate, faster, and more robust to evolving malware threats. Another method is by using data samples which consists of a set of known malicious files and normal program files in order to detect patterns that can identify malware accurately.

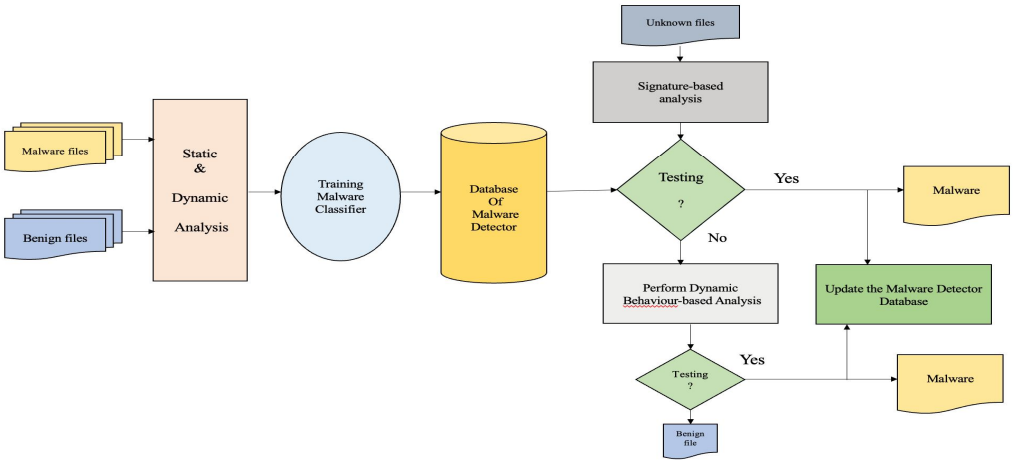


Fig: Hybrid Based Malware Detection

VIII. APPLICATIONS

The operation of this malware detection is to analyze the important research conducted on cybersecurity and understand how a researcher tackled the problem. The researchers together find the fastest alternate method of detection and analysis of malware that is by using machine learning. So, by using ML it is easy to determine the best extraction features, representation and classification methods for the malware detection. Machine Learning consist of various fields that are subdivided into supervised and unsupervised learning, which are further used f used for malware detection. So, the techniques that are used are Naïve Bayes and Neural Networks

IX. ADVANTAGES AND LIMITATIONS

A. Advantages of ML

Advantage of machine learning is to find the best solution, and we will find the best solution for you! advantages is a data-driven podcast that explains complex topics in a way that is easy to understand. Malware threats can come from anywhere at any time. Keep your devices safe and secure with the Advantages malware detection software. Detect malware that is polymorphic. advantages is the best way to learn and teach. It's a learning system where you can search for videos and lessons on anything, create your own content, and help others learn with you and can spot patterns and protect against similar attack, could tell the difference between harmful and non-malicious files. An attacker can penetrate, seize control, and carry out objectives via non-malware attacks by exploiting insecure software that a normal end user would use on a daily basis (think web browsers or Office-suite applications). Attackers may also utilize the successful exploit to obtain access to native operating system utilities (such as PowerShell or Windows Management Instrumentation – WMI) or other applications that give them execution freedom. These native tools give users extraordinary rights and privileges, allowing them to carry out of the most -simple network commands that lead to important data. It serves as a computer security early warning system. Detects malware threats that haven't been seen before. Detecting previously unknown sorts of malware attacks. With advantage's malware security suite, you'll be safe in the digital age. Not only does this protect your computer and mobile devices from the latest in malware and phishing attacks, but also from the oldest. Advantage also comes with a handy bot for online shopping protection. Detector of data flow dependencies. The Advantage tool is a data flow dependencies visualization and editing application. It is a set of filters, modified data sets, and a visualization for the data flow dependencies of a model. The tool can help you to analyze and understand your model's data flow dependencies.

B. Limitations of ML

The ever-growing list of malware is becoming harder to spot. This software prevents harmful software from running on your computer and is compatible with any antivirus installed on the machine. The machine learning disadvantage is a system that learns from its mistakes. It is an algorithm which employs clinical testing to different treatments, in order to find one that is more effective, for a given set of inputs.

Ada-boost is an advanced online advertising platform that lets you quickly and easily craft ads to be the most persuasive. It's simple, easy-to-use interface is perfect for small businesses or entrepreneurs without any coding experience. The disadvantages decision tree is a decision-making method that will allow you to do a more systematic analysis of the pros and cons of two or more choices. It works as a strategy for making decisions by considering the positives and negative of each option. Naïve Bayes is a simple, but powerful probability-based classifier that automatically does some of the heavy-lifting for you. You can also use it to calculate some useful parameters for some other machine learning algorithms. To analyse data patterns, algorithms must be taught. When it comes to machine learning, there is a risk of using insufficient algorithms and creating limiting predictions.

X. FUTURE SCOPE

Other data sets can be added to improve accuracy, and more algorithms with greater performance can be added to improve accuracy. It can be hosted on the web for real-time analysis of exe files in the cloud. One could try categorizing data into different malware categories. One can further make a valid set, experiment with different methods. Development of a hybrid approach with two heads. Dynamic analysis will also be carried out, utilizing both automatic and boring methods, as well as a variety of dynamic instruments. Naïve Bayes is a simple, but powerful probability-based classifier that automatically does some of the heavy-lifting for you. You can also use it to calculate some useful parameters for some other machine learning algorithms. Ada-boost is a set of tools to make marketing more predictive with insights from machine learning. It helps with marketing attribution, predicting marketing spend & ROI, and showing marketing ROI over time.

XI. CONCLUSION

The goal of this research is to find a machine learning solution to the malware problem. Due to the increasing rise of malware, we require automatic solutions to detect infected files. The data set was constructed using infested and clean executables in the first portion of the investigation, and we used a Python script to extract the data needed for the data set generation. After it has been created, the data collection must be ready to train machine learning algorithms. The algorithms that were used were decision trees, Naïve Bayes, and ADA-Boost. After applying the best accuracy methods, it has a Random Forest algorithm with an accuracy of 99.406012 percent. According to this study, the best method for spotting risky apps is Random Forest. If we add a significantly larger number of files to the data set in the future, we can improve the accuracy. This research gives a thorough examination of the topic. Naïve Bayes is often used in text classification, spam filtering, target recognition, or medical diagnosis. Informational classification techniques assume mutually independent and identically distributed attributes. In other words, each attribute is independent from the others, and the values of the attributes are all drawn from a single distribution. The Naïve Bayes classifier is a probability-based classification technique. It uses Bayes' theorem to calculate the posterior probability of each class given a particular piece of evidence or feature. Ada-boost is a data processor that improves the performance of people who are searching for things in their own personal or corporate information.

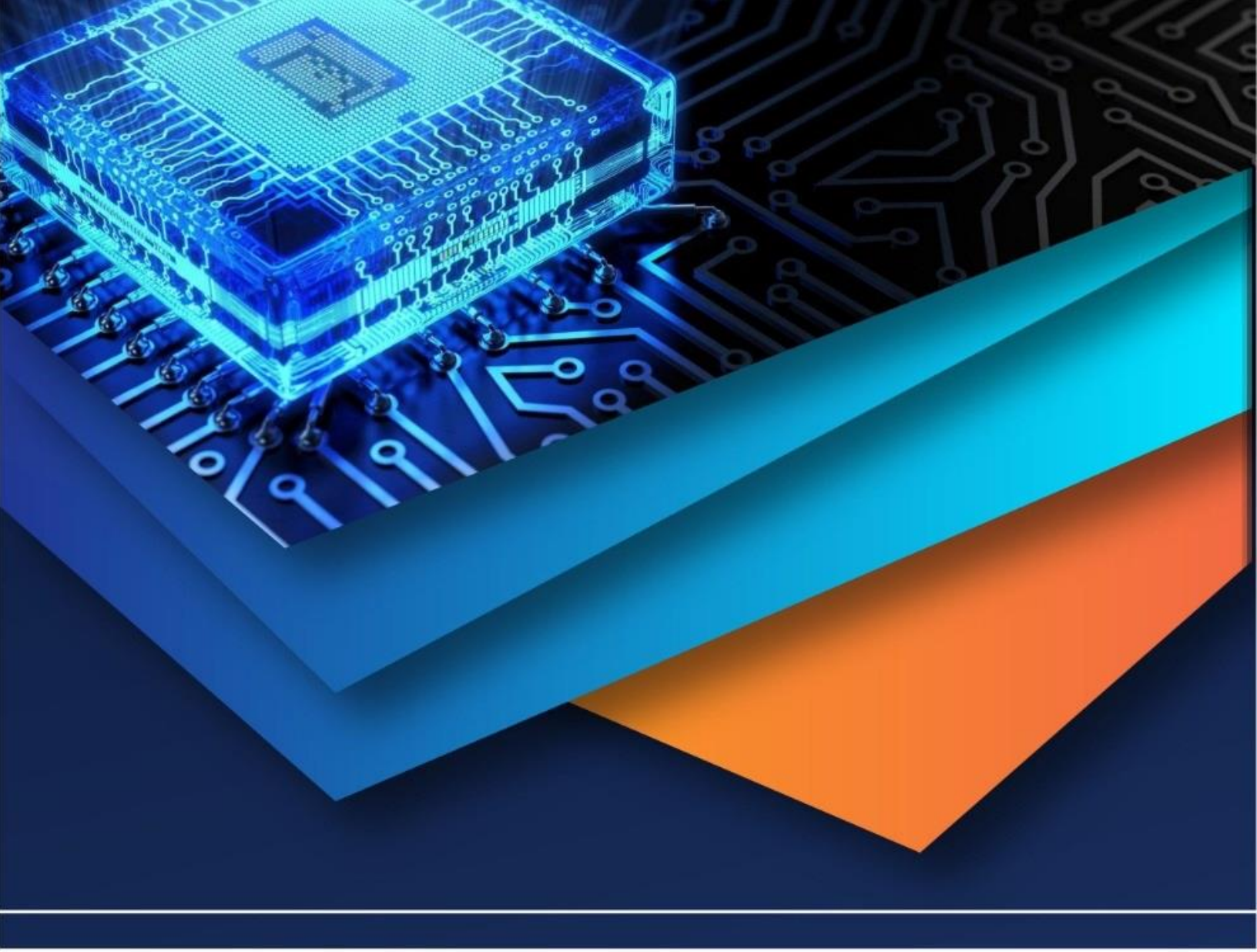
XII. ACKNOWLEDGMENT

The authors would like to thank the publishers and researchers for making their resource available and also, we are thankful to our teachers for their guidance. We would like to express our gratitude to our guide Professor Pritam Ahire for his encouraging support and guidance in carrying out this work. We express our sincere thanks to Dr DY Patil Institute of Engineering and Technology, Ambi, Pune for permitting us to take our work this further.

REFERENCES

[1] W. Han, J. Xue, Y. Wang, L. Huang, Z. Kong, MalDAE : Detected as well as explained malware based on correlation and fusion of static and dynamic characteristics, (2019) 208–233, http://dx.doi.org/10.1016/j.cose.2019.02. 007.
[2] P. Burnap, R. French, F. Turner, K. Jones, Malware classified using machine activity data, and self-organising feature maps Comput. Secur. 73 (2017) 399–410, http://dx.doi.org/10.1016/j.cose.2017.11.016:
[3] A. Damodaran, F.D. Troia, C.A. Visaggio, T.H. Austin, M. Stamp, A comparison between static, dynamic, and hybrid analysis for malware detection is done, J. Comput. Virol. Hacking Tech. (2017) 1–24, http://dx.doi.org/10.1007/s11416-015-0261-z.
[4] E.M. Dovom, A. Azmoodeh, A. Dehghantanha, D.E. Newton, R.M. Parizi, H. Karimipour, Fuzzy pattern tree for edge malicious files detection and categorization in Iot, J. Syst. Archit. 97 (March) (2019) 1–7, http://dx.doi.org/10.1016/j.sysarc. 2019.01.017.
[5] M. Ficco, F. Palmieri, Leaf : cybersecurity training platform for realistic edge-iot scenarios,J.Syst.Archit.97(September2018)(2019)107–129, http://dx.doi.org/10.1016/j.sysarc.2019.04.004.
[6] K. Khan, A. Mehmood, S. Khan, M.A. Khan, Z. Iqbal, W.K. Mashwani, A survey on intrusion detection and prevention in wireless ad - hoc networks, J. Syst. Archit. (2019) 101701, http://dx.doi.org/10.1016/j.sysarc.2019.101701.
[7] AV-TEST, Malware statistics and trends report, AV-TEST, 2020, https://www. avtest.org/en/statistics/malware.
[8] A. Bushby, F. Cybersecurity, How deception can change cyber security defences, Comput. Fraud Secur. Bull. 2019 (1) (2019) 12–14, http://dx.doi.org/10.1016/ S1361-3723(19)30008-9.
[9] E. Gandotra, D. Bansal, S. Sofat, Malware analysis as well as classification: A survey, J. Inf. Secur. 05 (02) (2014) 56–64, http://dx.doi.org/10.4236/jis.2014.52006, http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/jis.2014. 52006.
[10] S.M. Muzammal, M.A. Shah, S.J. Zhang, H.J. Yang, security risks and authentication techniques for smart devices: Int. J. Autom. Comput. 13 (2016) 350–363, http://dx.doi.org/10.1007/s11633-016-1011-5.
[11] J. Singh, J. Singh, Challenges of malware analysis: techniques for obfuscation Int. J. Inf. Secur. Sci. 7 (2018).
[12] Y. Gao, Z. Lu, Y. Luo, Survey on malware anti-analysis, in: 5th International Conference on Intelligent Control and Information Processing, ICICIP 2014 - Proceedings, 2015, pp. 270–275, http://dx.doi.org/10.1109/ICICIP. 2014.7010353.
[13] S. Alam, R. Horspool, I. Traore, I. Sogukpinar, made framework for metamorphic malware analysis and real- time analysis, Comput. Secur. 48 (2015) 212–233, http://dx.doi.org/10.1016/j.cose.2014.10.011,arXiv:arXiv:1011.1669v3, http: //linkinghub.elsevier.com/retrieve/pii/S0167404814001576.
[14] J. Singh, J. Singh, Ransomware: an illustration of malicious cryptography (2) (2019), 1608–1611, http://dx.doi.org/10.35940/ijrte.B2327.078219.
[15] X. Hu, Large-scale malware analysis, detection, and signature generation (ProQuest Dissertations and Theses), 2011, p. 190, http://search.proquest.com/ docview/918832186?accountid=44888.
[16] P. Coogan, speyeye bot v/s zeus bot, 2010, http://www.symantec.com/ connect/blogs/spyeye-bot-versus-zeus-bot.
[17] N. Eltayieb, R. Elhabob, A. Hassan, F. Li, A blockchain-based signcryption scheme to secure data sharing in the cloud, J. Syst. Archit. (August 2019) (2020) http://dx.doi.org/10.1016/j.sysarc.2019.101653.
[18] D. Jang, Y. Jeong, S. Lee, M. Park, K. Kwak, D. Kim, B.B. Kang, reusing anti-emulation techniques for huge-scale software deployment, Comput. Secur. (2019) http://dx.doi.org/10.1016/j.cose.2019.02.005.
[19] S. Mahdavifar, A.A. Ghorbani, Application of deep learning to cybersecurity: A survey, Neurocomputing 347 (2019) 149–176, http://dx.doi.org/ 10.1016/j.neucom.2019.02.056, http://www.sciencedirect.com/science/article/ pii/S0925231219302954.

[20] W. Zhang, H. Wang, H. He, P. Liu, DAMBA: Detecting android malware by ORGB analysis, IEEE Trans. Reliab. 69 (1) (2020) 55–69, http://dx.doi.org/10.1109/TR.2019.2924677.
[21] L. Liu, B.-s. Wang, B. Yu, Q.-x. Zhong, malware classification and new malware detection using ML, Front. Inf. Technol. Electron. Eng. 18 (9) (2016) 1–12, http://dx.doi.org/10.1631/FITEE.1601325.
[22] R. Kaur, M. Singh, Hybrid real-time zero-day malware analysis and reporting system, Int. J. Inf. Technol. Comput. Sci. 8 (4) (2016) 63–73, http://dx.doi.org/10.5815/ijitcs.2016.04.08, http://www.mecs-press.org/ijitcs/ijitcs-v8-n4/v8n4- 8.html.
[23] J. Milosevic, M. Malek, A. Ferrante, Time, accuracy and power consumption tradeoff in mobile malware detection systems, Computers & Security 82 (2019) 314–328, http://dx.doi.org/https://doi.org/10.1016/j.cose.2019.01.001, http:// www.sciencedirect.com/science/article/pii/S0167404818307880.
[24] H. Studiawan, F. Sohel, C. Payne, A survey on forensic investigation of operating system logs, Digital Investigation 29 (2019) 1–20, http://dx.doi.org/https://doi.org/10.1016/j.diin.2019.02.005,http://www.sciencedirect.com/science/article/pii/S1742287618303980.
[25] B. Ndibanje, K.H. Kim, Y.J. Kang, H.H. Kim, T.Y. Kim, H.J. Lee, Applied sciences cross-method-based analysis and classification of malicious behavior by API calls extraction, 2019, http://dx.doi.org/10.3390/app9020239.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY