IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205016 |

Automatic Framework for Android Malware Detection Using Machine Learning

Om Patil, Viraj Dalvi, Yash Chaudhari, Akshay Narkhede, Prof.Mrs. Suvarna Potdukhe

Department of Information Technology, RMD Sinhgad School of Engineering, Warje, Pune, India

ABSTRACT: Malware is one of the major issues regarding the operating systems or in the software world. Especially, to mention the Android operating system, reasons owning to compromisation by the developers. Hence, alot of malware detection techniques were developed for the same. But the techniques keeps lacking to detect newer versions or loathed variants of malwares. Despite numerous detection and analysis techniques being worked upon and developed to its full potentials , the detection accuracy of new malware is still a crucial issue. In the paper, we propose and study new technique which will be fabricated to an application for detecting malware in an android operating system more accurately and even informative variants of malwares. The main pivotal fraction of the methodology is having permission to access datasets of the respective operating system. The proposed methodology falls under the department of Machine Learning in the sky of Information Technology. The model will distinguish malicious files based on unique characters recognized while performing feature selection during training.

KEYWORDS: Mobile; Android; Malware; Operating system; Machine Learning; Malicious.

I. INTRODUCTION

Overview:

With the expansion of the Android market, as well as the increasing degree of dependence on mobile phones, malicious applications are growing rapidly. In the current situation, improving the efficiency of malicious application detection has become an urgent demand. Therefore, applying machine learning technology to malicious application detection which can reduce labor costs and improve detection efficiency has become a hot research direction. Various kind of malware attacks can occur because of the incompetency of current techniques The proposed model will be a leap ahead in detection of varied malware variants

Motivation:

Android has over one billion active users for all their mobile Devices with a market impact that is influencing an increase in the amount of information obtained from different users, facts that have motivated the development of malware by cybercriminals to solve the problems caused by malware. Android implements a different architecture and security controls, such as unique user ID For Each Application, System Permissions and Its Distribution platform Google play. The pivotal objectives to be achieved are increased accuracy in malware detection and saving cost and time. The proposed model will also solve the classification problem of whether a concerned file is malware or not .

Objective:

We use the Android phone Dataset to Detect Malicious applications. We give the Android phone data set as input Then data set goes to preprocessing, Segmentation phase after Both phases Done, we use SVM algorithm to classification the Detect the malicious Application.

II. REVIEW OF LITERATURE

Peng Tian and Xiaojun Huang, This paper proposes a new model for detecting Android malicious applications. The model obtains the API call sequences of APP runtime, and extracts features from them. These features have the highest correlation with malicious attributes detection and have the characteristics of small redundancy between each other. And noticed that API subsequences generated by normal behavior that may exist in a malicious application can interfere with the training of the detector. We use VSM, and K-means combined with the GBDT algorithm to

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205016 |

eliminate this interference and improve the detection accuracy. Experiments show that this method can effectively eliminate the influence of interference API sequence and obtain higher detection.[1].

Fei Chen, Yan Fu ,In this paper, we propose a new approach for the dynamic detection of malicious executables on the platform of Windows. Our approach extracts signatures of malicious executable's behaviors by using API (Application Program Interface) interception technique which makes possible the detection of unknown malicious executables. The dynamic detection of unknown malicious executables is achieved in three major steps: getting the sequence of API function calls of the executable, processing the API sequence to generate a vector, calculating the similarity between the vector and the feature library constructed by security policies to verify if the executable is malicious. The experiment confirms that this approach is effective in detection of unknown malicious executables.[2]

Yingbo Li, Jing Fang and Cheng Liu,With the increasing popularization of smartphones in life, malicious software targeting smartphones is emerging in an endless stream. As the phone system possesses the highest current market share, Android is facing a full-scale security challenge. This article focuses on analyzing the application of Dalvik injection technique in the detection of Android malware. Modify the system API (Application Program Interface) through Dalvik injection technique to detect the programs on an Android phone directly. Through the list of sensitive APIS called by malicious programs, eventually judge the target program as malicious or not.[3]

Yong Li, Due to the rapid growth of android malicious application samples, traditional detection methods need to spend a lot of time training, a detecting method for malicious mobile application based on incremental SVM was proposed to achieve incremental learning of the detection system. The method used the SVM as the classification and training algorithm and extracted sensitive permissions and APIs as application characteristics. Based on SVM, a dual weight function was designed to filter the historical training samples to avoid redundant samples, and the incremental learning method of SVM was implemented in combination with KKT conditions. Therefore, the training time could be reduced, and the learning efficiency of the malicious application detection system could be improved without reducing the training accuracy.[4]

Xie Bailin, Yu Shenzhen, Wang Tao,Today more and more network-based attacks occur at application layer. Observed from the network layer and transport layer, these attacks may not contain significant malicious activities, and generate abnormal network traffic. However, traditional security techniques usually detect attacks from those two layers. Although some security techniques can detect some application layer attacks, these techniques can only detect some known attacks and these techniques can't detect the unknown or novel attacks that happened on the application layer. In theory, application layer anomaly detection can detect unknown and novel attacks that happened on application layer, so the research of application layer anomaly detection is very important. This paper presents a new application layer anomaly detection method which is based on HSMM. The experimental results show that this method has high detection accuracy and low false positive ratio[5]

Yang Gao,With more and more online services developed into web applications, security problems based on web applications are becoming more serious now. Most intrusion detection systems are based on every single request to find the cyber-attack instead of users' behaviors, and these systems can only protect web application from known vulnerability rather than some zero-day attacks. To detect newly developed attacks, we analyze web logs from web servers and define users' behaviors to divide them into normal and malicious ones. The result shows that by using the feature of web resources to define users' behaviors, a higher accuracy rate and lower false alarm rate of intrusion detection can be obtained[6]

Wataru Matsuda,The targeted attacks cause severe damage worldwide. Detecting targeted attacks is challenging because the attack methods are very sophisticated. Network-based solutions such as Firewall, Proxy Server, and Intrusion Detection System (IDS) have been widely used. In addition to this, recently, detection methods for malicious programs by monitoring behavior on the endpoints called Endpoint Detection and Response (EDR) have been proposed. Also, some researchers introduce detection methods using DLLs by analyzing suspicious files on the sandbox, such as Cuckoo. Using Cuckoo is one of the solutions for analyzing files that are already identified as malicious. In this research, we propose a real-time detection method of malicious tools using DLL information collected by System Monitor (Sysmon): a free logging tool provided by Microsoft. The purpose of our method is detecting new malicious processes in the production environment. We focus on DLLs commonly loaded by malicious tools regardless of the environments, then propose "the common DLL lists" for detection. Moreover, we introduce a practical detection method that utilizes Elastic Stack as Security Information and Event Management (SIEM). By

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205016 |

using Elastic Stack, DLL information loaded on computers can be uniformly monitored and enables real-time detection by comparing logs with the common DLL lists. We evaluate the effectivity of the proposed method using four free malicious tools introduced by US-CERT: China Chopper, Mimi Katz, PowerShell Empire, and HUC Packet Transmitter. As a result, our method detected China Chopper, Mimi Katz, PowerShell Empire with 100 false positive occurred for HUC Packet Transmitter, and false positive rate was 0.55lists are useful for detecting malicious tools in real-time using Elastic Stack.[7]

Parnika Bhat, Kamlesh Dutta,Android is currently the most popular operating system for mobile devices in the market. Android devices are being used by every other person for everyday life activities and it has become a center for storing personal information. Because of these reasons it attracts many hackers, who develop malicious software for attacking the platform; thus, a technique that can effectively prevent the system from malware attacks is required. In this paper, and malware detection technique, Mildrid has been proposed for detecting malware applications on Android platform. The proposed technique statically analyses the application files using features which are extracted from the manifest file. A supervised learning model based on Naive Bayes is used to classify the application as benign or malicious. Mildrid achieved Recall score 99.12[8]

Dr. V. Lakshman Narayana1,Mobile Ad Hoc Networks (MANETs) are infrastructure-less networks that are mainly used for establishing communication during the situation where a wired network fails. Security related information collection is a fundamental part of the identification of attacks in Mobile Ad Hoc Networks (MANETs). A node should find accessible routes to remaining nodes for information assortment and gather security related information during route discovery for choosing secured routes. During data communication, malicious nodes enter the network and cause disturbances during data transmission and reduce the performance of the system. In this manuscript, a Time Interval Based Blockchain Model (TIBBM) for security related information assortment that identifies malicious nodes in the MANET is proposed. The proposed model builds the Blockchain information structure which is utilized to distinguish malicious nodes at specified time intervals. To perform a malicious node identification process, a Network Block Monitoring Node (NBMN) is selected after route selection and this node will monitor the blocks created by the nodes in the routing table. At long last, NBMN nodes understand the location of malicious nodes by utilizing the Blocks created. The proposed model is compared with the traditional malicious node identification model and the results show that the proposed model exhibits better performance in malicious node detection[9].

III. PROPOSED METHODOLOGY

In this section, we present the MalDozer framework and its components. MalDozer has a simple design, where a minimalistic preprocessing is employed to get the assembly methods. As for the feature extraction (representation learning) and detection/attribution, they are based on the actual neural network. This permits MalDozer to be very efficient with fast preprocessing and neural network execution. Since MalDozer is based on supervised machine learning, we first need to train our model. Afterward, we deploy this model along with a preprocessing procedure on the targeted devices. Notice that the preprocessing procedure is common between the training and the deployment phases to ensure the correctness of the detection results

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205016 |

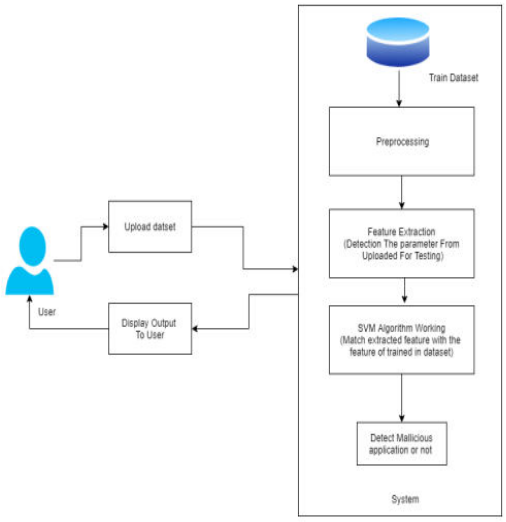


Fig. 1. Proposed System Architecture

- Admin:In this module, the admin must log in by using a valid user name and password. After login successful he can do some operations such as View All Users and Authorize, View All E-Commerce Website and Authorize, View All Products and Reviews, View All Products Early Reviews, View All Keyword Search Details, View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.
- View and Authorize Users:

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as username, email, address and admin authorize the users.

- View Charts Results:

View All Products Search Ratio, View All Keyword Search Results, View All Product Review Rank Results.

- Ecommerce User:

In this module, there are n number of users present. Users should register before doing any operations. Once the user registers, their details will be stored in the database. After registration successful, he has to login by using authorized username and password Once Login is successful user will do some operations like Add Products, View All Products with reviews, View All Early Product's reviews, View All Purchased Transactions.

- End User:

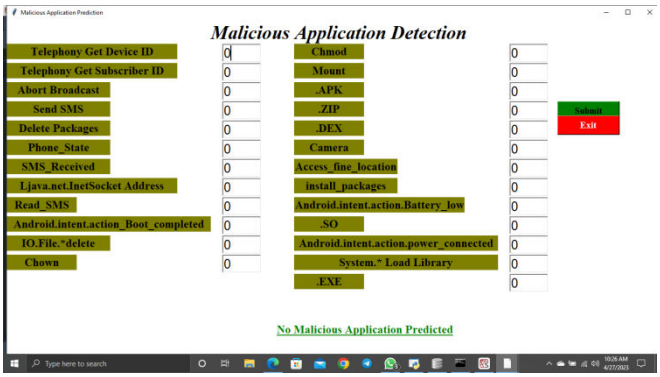
In this module, there are n number of users present. Users should register before doing any operations. Once the user registers, their details will best or to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Manage Account, Search Products by keyword and Purchase, View Your Search Transactions, View.

IV. RESULTS AND DISCUSSION

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2023 ||

| DOI:10.15680/IJIRSET.2023.1205016 |



V. CONCLUSION

The principle of application monitoring is studied deeply, and the dynamic detection of application based on Hook technology is used to obtain the system API callsequence. On the basis of the existing research, thefeature selection algorithm is extended and the most helpful features of detection are extracted. Aiming at the problem of interference in the application API sequence, a detection model is designed and implemented. A scheme of using the vector space model to remove the interference API sequence is proposed.

REFERENCES

1 Stephen Feldman ,Dillon Stadther ,Bing Wang, “Manilyzer: Automated Android Malware Detectionthrough Manifest Analysis ,” in 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems.

2 WILLIAM ENCK,PETER GILBERT,SEUNGYEOP HAN,VASANT TENDULKAR,BYUNGON CHUN,LANDON P. COX,JAEYEON JUNG,PATRICK MCDANIEL,ANMOL N. SHETH,”TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones,” ACM Transactions on Computer Systems (TOCS) TOCS Homepage archive Volume 32 Issue 2, June 2014

3 Corrado Aaron Visaggio,Eric Medvet,Gerardo Canfora,Francesco Mercaldo “Detecting Android Malware using Sequences ofSystem Calls,” in Third International Workshop on Software Development lifecycle for Mobile ESEC/FSE 2015

4 Alberto Ferrante,Eric Medvet,Francesco Mercaldo,Jelena Milosevic,Corrado Aaron Visaggio, “Spotting the Malicious Moment: CharacterizingMalware Behavior Using Dynamic Features,” 2016 11th International Conference on Availability, Reliability and Security

5 Zhenyu Ni, Ming Yang, Zhen Ling, Jia-nan Wu, Junzhou Luo , ”Real-time Detection of Malicious Behavior in Android Apps,” 2016 International Conference on Advanced Cloud and Big Data 6 Quan Hu, Xiuliang Mo, Chundong Wang ,”Based on the Markov Chain Model of Android Platform Malicious APP Detection Research,” 2016.4 Journal of Tianjin University of Technology Vol. 32, No. 2

7 (2017) The Xposed webpage .[Online]. Available: http://repo.xposed.info

8 Li Wang Yuanchun Wan , Weidong Hao ,”Based on Honeypot Android Malicious Code Dynamic Analysis,” 2016 Microcomputer Application Volume 32, No. ll

9 (2017) The malicious application set webpage . 10 (2017) The virustotal webpage .[Online]. Available: https://www.virustotal.com

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details