IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Malicious Android Application Detection Method using Machine Learning

Divya Chaudhari[1], Arati Chaure[2], Shreyash Dhadke[3], Tushar Dhanawate[4], Prof. Shraddha Shirsath[5]

Students, Department of Computer Engineering[1,2,3,4]

Professor, Department of Computer Engineering[5]

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract*:* *With the increasing popularity of the Android platform, we have seen the rapid growth of malicious Android applications recently. Considering that the heavy use of applications on mobile phones such as games, emails, and social network services has become a crucial part of our daily life, we have become more vulnerable to malicious applications running on mobile devices. This paper demonstrates on the problem of detecting malicious applications in the mobile internet, which is of great importance for personal information security and privacy security. We convert the android internet malicious application detection problem to a classification problem, and utilize the SVM classifier to solve it. Finally, we conduct an experiment to test the performance of the proposed method. Experimental results that the proposed can detect android internet malicious application with higher accuracy, true positive rate, and lower false positive rate.*

Keywords: SVM, Android internet, malicious application

I. INTRODUCTION

With the rapid development of the communication technology and information technology, mobile Internet can facilitate the free exchange between persons, and smart telephone has become indispensable tools in our daily life. In addition, the intelligent terminals play a key role in our daily life, and more and more privacy information are carried by them. Benefiting from mobile Internet and smart phones, people can do more things using their mobile devices, such as taking pictures and videos, playing games, communicating with friends. Mobile devices are faced with more security challenges in recent years.The first Android based smart device was developed in November 2008, and Android system rapidly goes into the smart system in the world. Two years later, Android acquired 48% world's smart device market .Meanwhile, as a famous development platform. Android has become the first choice of mobile application developers In 2017, the number of android devices will exceed one billon and its mobile application downloading may be exceed 5 billion times. We can find that more and more Android's t malicious applications will upload to various application platforms and bring great threats to users.However, smart devices may save rich privacy information and user's personal data. For example, mobile payment has been popular by more and more people, and then users' personal information are memorized in mobile devices. But, once the smart phone is lost, it will bring a great threaten at people's privacy. Therefore,security problems are becoming more and more serious, and it is for great importance to detect mobile Internet malicious applications.

With the expansion of the Android market, as well as the increasing degree of depen-dence 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 efficiencyhas become a hot research direction.

We use android phone dataset to detect malicious application. We give Android phone data set as input. Then data set go to preprocessing, Segmentation phase after both phase done we use SVM algorithm to classification the Detect the malicious application.

DOI: 10.48175/IJARSCT-10974

II. RELATED WORK

Yang Gao, "Anomaly Detection of Malicious Users' Behaviors for Web Applications Based on Web Logs"

With more and more online services developed into web applications, security problems based on web applications become more serious now. Most intrusion detection systems are based on every single request to find the cyber-attack insteadof users' behaviors, and these systems can only protect web application from known vulnerability rather than some zero-day attacks. In order to detect newly developed attacks, we analyze web logs from web servers and define users' behaviors to dividethem 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.

Wataru Matsuda, "Real-Time Detection System Against Malicious Tools by Monitoring DLL on Client Computers"

The targeted attacks cause severe damage worldwide. Detecting targeted attacks are 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 suspiciousfiles 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 intro- duce a practical detection method that utilizes Elastic Stack as Security Information and Event Management (SIEM). By using Elastic Stack, DLL information loaded oncomputers 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, Mimikatz, PowerShell Empire, and HUC Packet Transmitter. As a result, our method detectedChina Chopper, Mimikatz, PowerShell Empire with 100false positive occurred for HUC Packet Transmitter, and false positive rate was 0.55 lists are useful for detecting malicious tools in real-time using Elastic Stack.

ParnikaBhat,KamleshDutta, "MaplDroid: Malicious Android Application Detection based on Naive Bayes using Multiple Feature Set"

Android is currently the most popular operating system for mobile devices in the market. Android device is being used by every other person for everyday life activities and it has become a centre 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 malwareattacks is required. In this paper, an malware detection technique, MaplDroid has been proposed for detecting malware applications on Android platform. The pro- posed 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. MaplDroid achieved recall score 99.12

Dr. V. Lakshman Narayana1, "A Time Interval based Blockchain Model for Detection of Malicious Nodes in MANET Using Network Block Monitoring Node"

Mobile Ad Hoc Networks (MANETs) are infrastructure-less networks that are mainly used for establishing communication during the situation where wired network fails. Security related information collection is a fundamental part ofthe 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 15 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 node understands 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.

III. SYSTEM ARCHITECTURE

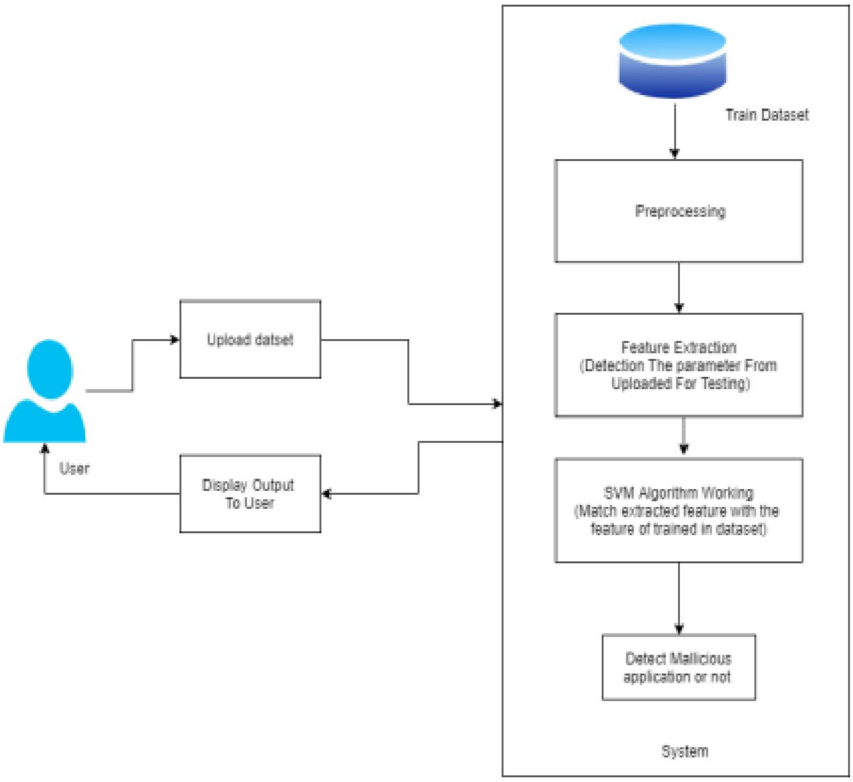


Figure 5.1: system architecture1

IV. METHODOLOGY

The current section presents a methodology for classifying Android applications as either malicious or benign based on permission analysis. To gather data for our study, we obtained a sample of 2,500 malicious Android applications from sources such as VirusShare and Zeltseretc, and 1,500 benign applications from Google Play and other trusted websites. Our feature extraction process involved studying malware and the different categories of permissions that an application requires for its functioning. We identified four types of Android permissions, namely Normal, Signature, Special, and Dangerous permissions, based on their level of risk to the user's privacy. We extracted permissions from the manifest.xml file of each .apk package using a Python script that compiled them into a CSV file. To prepare our dataset, we used a machine learning approach that aimed to detect both known malware families and unknown malware, with the goal of reducing the likelihood of malware evading detection by scanners in the Android community. Our dataset comprised 4,000 samples of both malicious and benign Android applications. We employed supervised and unsupervised machine learning algorithms, including Logistic Regression, K-Nearest Neighbors, Decision Tree, and Gaussian Naive Bayes, to classify the applications as either malicious or benign. Logistic Regression is a statistical algorithm that models the relationship between input and output numerical values. K-Nearest Neighbors is a classification algorithm that uses Euclidean distance to measure the K nearest neighbors of a data point and predict the output. Decision Tree is a supervised learning algorithm that uses a data structure to solve problems, with leaf nodes representing class labels and internal nodes representing attributes. Gaussian Naive Bayes is a classification algorithm based on probabilities that can be used for both binary and multi-class classification problems.

V. RESULT

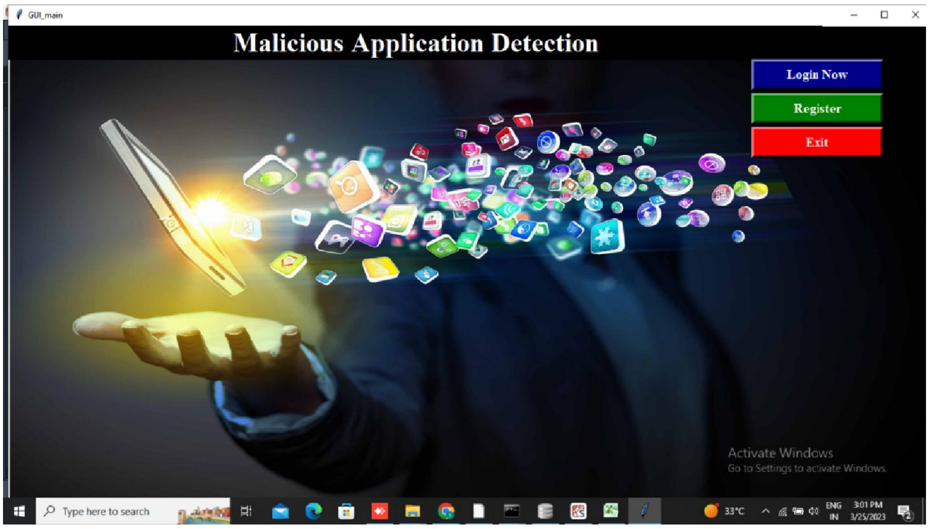

GUI_main
Malicious Application Detection
Login Now
Register
Exit
Activate Windows
Go to Settings to activate Windows.
Type here to search

Fig:- Home Page

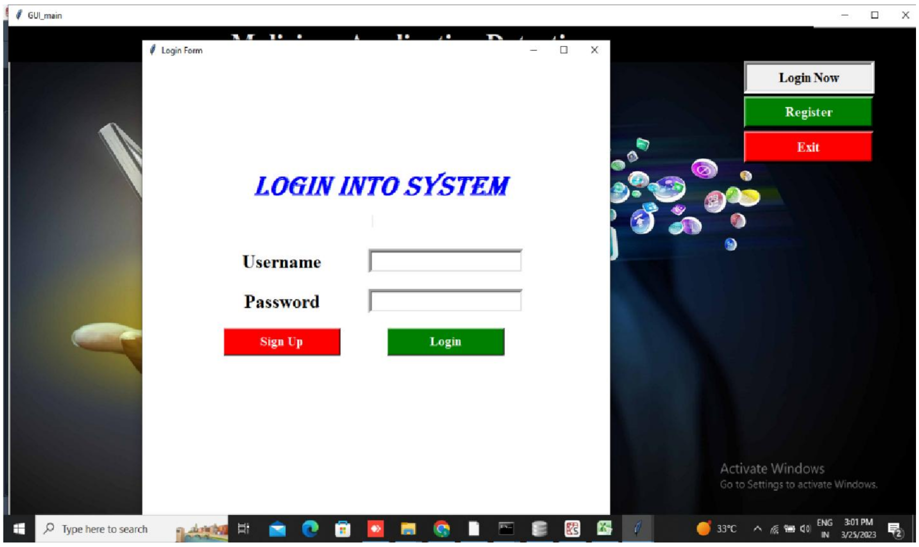
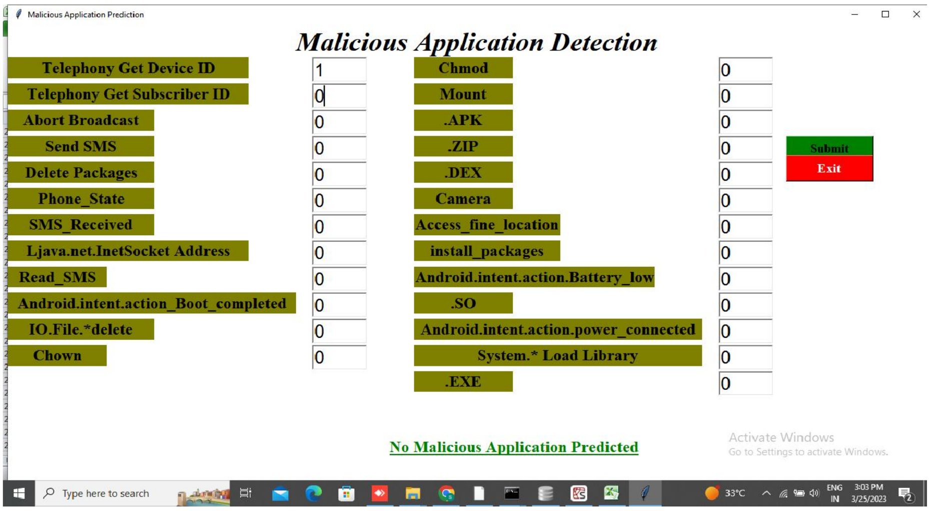


Fig:- Login Page


Malicious Application Prediction
Malicious Application Detection
Submit
Exit
Activate Windows
Go to Settings to activate Windows.
Type here to search

Telephony Get Device ID	1	Chmod	0
Telephony Get Subscriber ID	0	Mount	0
Abort Broadcast	0	.APK	0
Send SMS	0	.ZIP	0
Delete Packages	0	.DEX	0
Phone_State	0	Camera	0
SMS_Received	0	Access_fine_location	0
Ljava.net.InetSocket Address	0	install_packages	0
Read_SMS	0	Android.intent.action.Battery_low	0
Android.intent.action_Boot_completed	0	.SO	0
IO.File.*delete	0	Android.intent.action.power_connected	0
Chown	0	System.* Load Library	0
		.EXE	0

No Malicious Application Predicted

Fig:-Prediction Page

DOI: 10.48175/IJARSCT-10974

VI. 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 call sequence. On the basis of the existing research, the feature 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.

VII. FUTURE SCOPE

Investigate the use of ensemble methods to combine multiple machine learning models, each with its own strengths and weaknesses, to improve overall detection accuracy. Explore the application of deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), to capture more intricate patterns and behaviors in Android applications. Design algorithms that can continuously learn from new samples and adapt the detection system in real-time to keep up with the evolving landscape of Android malware.

REFERENCES

[1]. G. Jacob, H. Debar, and E. Filiol, "Behavior detection of mal ware: from a survey toward an established taxonomy", ComputVirol, pp: 251-226, 2008.

[2]. M. Schultz, E. Eskin, E. Zadok and S. Stolfo, "Data mining methods for detection of new malicious executables", Proceedings of the 2001 IEEE Symposium on Security and Privacy, Washington : IEEE Computer Society, pp. 38-49, 2001.

[3]. P. Joshi, C. Jindal, M. Chowkwale, R. Shethia, S. A. Shaikh, and D. Ved, "Protego: A passive intrusion detection system for android smartphones," in 2016 International Conference on Computing, Analytics and Security Trends (CAST), Dec 2016, pp. 232–237.

[4]. A. Kapratwar, F. Di Troia, and M. Stamp, "Static and dynamic analysis of android malware," pp. 653–662, 01 2017.

[5]. Raja, L.; Baboo, S.S. An overview of MANET: Applications, attacks and challenges. Int. J. Comput. Sci. Mob. Comput. 2014, 3, 408–417

[6]. Sivakami, R.; Nawaz, G.K. Secured communication for MANETS in military. In Proceedings of the 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), Tirunelveli, Tamilnadu, India, 18–19 March 2011; pp. 146–151.

[7]. Cho, J.-H.; Swami, A.; Chen, R. A survey on trust management for mobile ad hoc networks.IEEE CommunSurv. Tutor. 2010, 13, 562–583.

[8]. VenkataRaoMaddumala, (2020), "Enhanced Morphological Operations for Improving the Pixel Intensity Level", International Journal of Advanced Science and Technology, Vol. 29, No. 03, (2020), pp. 9191 - 9201.

[9]. Schweitzer, N.; Stulman, A.; Shabtai, A.; Margalit, R.D. Mitigating denial of service attacks in olsr protocol using fictitious nodes. IEEE Trans. Mob. Comput. 2015, 15, 163–172.

[10]. Bharathi C R ,(2018),"Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs", Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13-1921- 1_63.

DOI: 10.48175/IJARSCT-10974