Android Malware Detection

K Santhi Rani #1, G Sai Kumari#2, G Sireesha #3, N. Naga Vyshnavi #4, T Meghana #5

#1 Assistant Professor,Dept of CSE, Qis Institute of Technology,Ongole,Prakasam (Dt)
#2,3,4,5 B.Tech., Scholars, Dept of CSE, Qis Institute of Technology, Ongole, Prakasam (Dt)

Abstract: Malware is one of the major issues regarding the operating framework or in the software world. The android framework is also going through the same issues. We have seen other Signature based malware location strategies were utilized to recognize malware. Yet, the strategies couldn't recognize obscure malware. In spite of various discovery and analysis procedures are there, the discovery accuracy of new malware is as yet a crucial issue. In this paper, we study and feature the current identification and analysis techniques utilized for the android malicious code. Along with contemplating, we propose Machine learning algorithms that will be utilized to analyze such malware and also we will do semantic analysis. We will be having a data set of authorizations for malicious applications. Which will be compared with the consents extracted from the application which we want to analyze. Eventually, the client will actually want to perceive how much malicious authorization is there in the application and also we analyze the application through remarks.

1.Introduction

Malware is only the short name for malicious software, in general, alluded to many types of threatening or interruption creating software, spyware, Trojan horses, backdoor, and rootkits. The main aim of malware is to damage, steal, upset or then again do some bad actions. Malware is sufficiently incredible to contaminate any sort of figuring machine running application, and the counteraction of malware is in effect all around read for personal PCs (PC). A Smartphone gadget the location procedures utilized is lagging far behind as compared to the fast development of the versatile population is being Some new study has shown that there are about 2.1 million android applications are there in the market. Because of increase in usage of the android framework has prompted more rollout of android malware. This malware is spreading in the market by the outsiders creating applications. The Google android market also doesn't vow to guarantee that all the applications recorded are without threat. There are also such reports about Trojans applications that whenever downloaded, their malicious code is also installed and cannot be easily identified by Google's advancements during publication in the Google android market. The android threats incorporate banking Trojans, spyware, bots, root abuses, SMS fraud, phishing and fake installer. Android Apps are unreservedly available on Google Playstore, the official Android app store as well as outsider app stores for clients to download. Because of its open source nature and popularity, malware authors are increasingly zeroing in on creating malicious applications for Android operating framework. Despite various attempts by Google Playstore to ensure

against malicious apps, they actually discover their way to mass market and cause harm to clients by abusing personal information related to their telephone directory, mail accounts, GPS location information and others for abuse by outsiders or else take control of the telephones distantly. Consequently, there is need to perform malware analysis or figuring out of such malicious applications which present genuine threat to Android platforms. Broadly speaking, Android Malware analysis is of two types: Static Analysis and Dynamic Analysis. Static analysis basically includes analyzing the code structure without executing it while dynamic analysis is examination of the runtime behavior of Android Apps in constrained climate. Offered in to the always increasing variants of Android Malware presenting zero-day threats, a productive mechanism for discovery of Android malwares is required. In contrast to signature-based approach which requires regular update of signature database, machine-learning based approach in combination with static and dynamic analysis can be utilized to identify new variants of Android Malware presenting zero-day threats. In [5], broad yet lightweight static analysis has been performed achieving a good discovery accuracy of 94% utilizing Support Vector Machine algorithm. Nikola Milosevic et al. [6] introduced static analysis based classification through two strategies: one was consents based while the other included representation of the source code as a bag of words. Another approach based on recognizing most significant consents and applying machine learning on it for evaluation has been proposed in [7-11]. The main commitment of the work is decrease of feature measurement to not exactly half of original feature-set utilizing Genetic Algorithm with the end goal that it tends to be taken care of as contribution to machine learning classifiers for training with diminished intricacy while maintaining their accuracy in malware classification. In contrast to exhaustive technique for feature choice which requires testing for 2N various combinations, where N is the number of features, Genetic Algorithm, a heuristic searching approach based on wellness work has been utilized for feature choice. The upgraded feature set obtained utilizing Genetic algorithm is utilized to train two machine learning algorithms: Backing Vector Machine and Neural Network. It is noticed that a respectable classification accuracy of over 94% is maintained while working on a much lower feature measurement, along these lines, lessening the training time intricacy of classifiers.

2. Background Work

There are various ways and techniques through malware or any malicious record can enter your framework or application. A portion of the basic strategies of malware getting interfered into the framework are as per the following: -

- Penetration: Penetration procedures generally utilized for malware applications for installation activation and running on the android framework are repackaging, updating and downloading.

- Repackaging: It is among the regular procedures for malware engineers to install malicious applications on an android platform.Repackaging approach for popular applications and abuse them as malware. The engineer downloads such sorts of application and recodes them and adds their own malicious code and uploads that application to the official Android app store or on the various markets.
- Updating: This procedure is considerably more hard for distinguishing malware.The malware engineer may in any case utilize repackaging however instead of encoding deliver code to the application the designer may incorporate an update part that will able to download malicious code at the run time.
- Downloading: This is the most traditional attacking strategy. The malware designer needs to attract the client to download intriguing and attractive applications

3.Literature Review

In "Android Malware Detection Using Machine Learning on ImagePatterns "[1], the paper was distributed in the year 2018. They have played out the malware discovery with the assistance of 300 malware records and 300 kindhearted apk documents, also they managed to generate just 183 malware and 300 favorable gray-scale images. The other 117 malware samples were unable to generate into images because the apk documents were tainted or either that records didn't contain classes.dex document. Also, the accuracy was a lot less in all the algorithms they utilized. They have recognized with the help of three diverse classifier strategies namely the knearest neighbor(KNN), Random Forest (RF), and Decision Tree(DT).

In "Android mobilesecurity by detecting and classification of malware based on permissions usingmachine learning algorithms "[2], the paper was distributed in the year 2017. They had utilized diverse machine learning algorithms like Naive Bayes,j48, random woodland, Multiclass classifier and multilayer perceptron to recognize android malware and evaluate the performance of each algorithm. Here they executed a framework for classifying android applications with the assistance of the machine learning strategies to check whether it is a malware or normal application. For validating their framework they have gathered 3258 samples of android apps and those have to be extracted for each application, extract their features and have to train the models going to be evaluated with the assistance of classification accuracy and time taken for the model.

In "An Android Behaviour-Based Malware Detection Method using Machine Learning"[3], the paper was distributed in the year 2016. They have proposed a Robotium program in an Android sandbox that can trigger any android application automatically and screen its behavior. The program has a UI Identification automatic trigger program that can tap the portable

applications in a meaningful request. The program was able to perform largerscale tests. They also attempted to construct a choice model utilizing behavior that has gathered with the assistance of the random woodland algorithm. It has had the option to decide if the obscure application is malware and also shows its certainty value. They could store the outcome and also the certainty value of the obscure apk document in their database.

In "RanDroid: Android malware detection using random machinelearning classifiers"[4], the paper was distributed in the year 2018. They have proposed the android malware identification framework with the help of consents, APIs, and also with the presence of distinctive key apps information, for example, the dynamic code, Reaction code, native code, cryptographic code, database, and so on as the feature to train and fabricate classification model just by utilizing various machine learning methods which can automatically recognize malicious Android apps(Malware) from the legitimate ones.

4.System Analysis

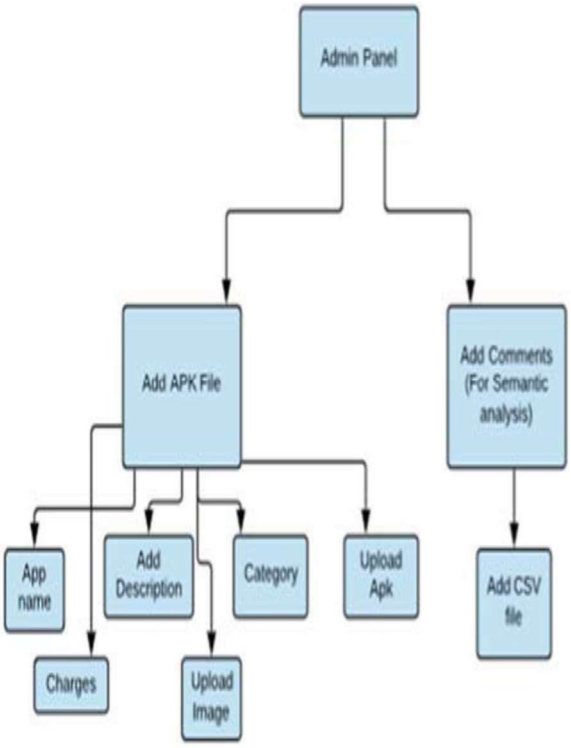
Existing System

The main contribution of the work is reduction of feature dimension to less than half of original feature-set using Genetic Algorithm such that it can be fed as input to machine learning classifiers for training with reduced complexity while maintaining their accuracy in malware classification. In contrast to exhaustive method of feature selection which requires testing for 2N different combinations, where N is the number of features, Genetic Algorithm, a heuristic searching approach based on fitness function has been used for feature selection. The optimized feature set obtained using Genetic algorithm is used to train two machine learning algorithms: Support Vector Machine and Neural Network. It is observed that a decent classification accuracy of more than 94% is maintained while working on a much lower feature dimension, thereby, reducing the training time complexity of classifiers.

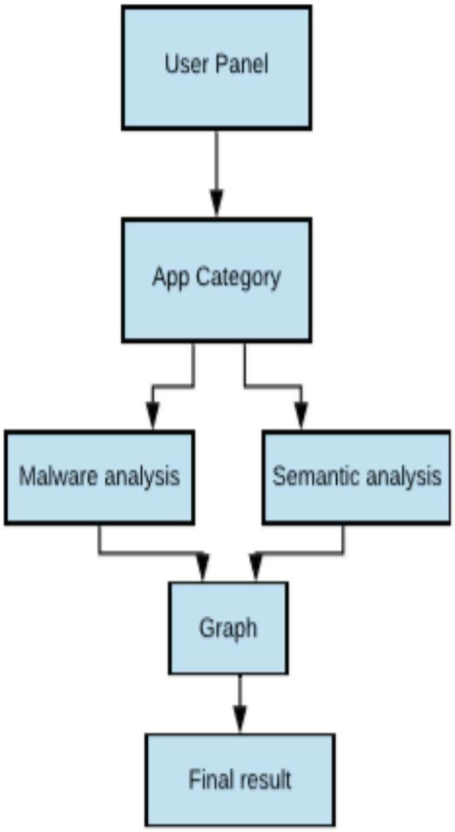
Proposed system :-

Two set of Android Apps or APKs: Malware/Goodware are reverse engineered to extract features such as permissions and count of App Components such as Activity, Services, Content Providers, etc. These features are used as featurevector with class labels as Malware and Goodware represented by 0 and 1 respectively in CSV format.To reduce dimensionality of feature-set, the CSV is fed to Genetic Algorithm to select the most optimized set of features. The optimized set of features obtained is used for training two machine learning classifiers: Support Vector Machine and Neural Network. In the proposed methodology, static features are obtained from AndroidManifest.xml which contains all the important information needed by any Android platform about the Apps. Androguard tool has been used for disassembling of the APKs and getting the static features. In our system, we have implemented an admin panel as well asa user panel. In the admin panel admin have the access to uploadthe apk files and its details along with its categorization and

alsothe admin can upload the comment that can be used forsemantic analysis.In the user-panel the user can see the select the category of theapplication and can see its details like pricing description name.User can see the malicious percentage of the application. Andthe processed output of the semantic analysis will be displayedto the user in the form of graph and the user will get a properreview of the application.



Admin Panel 1



User Panel 1

5. Results

The Malware Detection can recognize a wide range of consents in view of the which it has been asked and furthermore which of the consents which it has been taken of course. Likewise, the semantic investigation is been utilized to get the appropriate remarks resultant it in to get if the application is been appropriate or not. The consent based examination and furthermore the semantic investigation gives the appropriate yield with the goal that the client can utilize those specific applications or on the other hand not.

6. Conclusion

In our work, we propose a framework for authorization examination and semantic investigation. Our framework is additionally used to identify malware authorizations dependent on an application by contrasting it and a dataset. This proposed framework can be applied in the fields of the security framework and furthermore for the n clients like a malware discovery programming. Nonetheless, there are limits in our framework. The authorizations which we are characterizing are according to our however it can vary from clients to clients. The consents which the client likes that it's anything but a malware-based can be malware for some other client. Future works will contain the improvement of that.

References

[1] Darus,Fauzi Mohd,Salleh Noor Azurati Ahmad,and Aswami Fadillah MohdAriffin."Android Malware Detection Using Machine Learning on ImagePatterns"2018 Cyber Resilience Conference(CRC).IEEE,2018.

[2] Vrama,P.Ravi Kiran,Kotari Prudvi raj,and KV Subba Raju."Android mobilesecurity by detecting and classification of malware based on permissions usingmachine learning algorithms."2017 InternationalConference on I-SMAC (IoTin Social,Mobile,Analytics and Cloud)(I-SMAC).IEEE2017.

[3] Chang,Wei-Ling, Hung-Min Sun,and Wei Wu."An Android Behaviour-Based Malware Detection Method using Machine Learning." 2016 IEEEInternational Conference on Signal Processing, Communications andComputing (ICSPCC).IEEE, 2016.

[4] Koli, J. D. "RanDroid: Android malware detection using random machinelearning classifiers." 2018 Technologies for Smart-City Energy Security andPower (ICSESP).IEEE,2018.

[5] S. Arshad, M. A. Shah, A. Wahid, A. Mehmood, H. Song, and H. Yu,"SAMADroid: A Novel 3-Level Hybrid Malware Detection Model forAndroid Operating System," *IEEE Access*, vol. 6, pp. 4321–4339, 2018.

[6] T. Kim, B. Kang, M. Rho, S. Sezer, and E. G. Im, "A Multimodal DeepLearning Method for Android Malware Detection using VariousFeatures," vol. 6013, no. c, 2018

[7] A. Martin, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "EvolvingDeep Neural Networks architectures for Android malwareclassification," *2017 IEEE Congr. Evol. Comput. CEC 2017 - Proc.*, pp.1659–1666, 2017.

[8] X. Su, D. Zhang, W. Li, and K. Zhao, "A Deep Learning Approach toAndroid Malware Feature Learning and Detection," 2016 IEEE Trust.,pp. 244–251, 2016.

[9] K. Zhao, D. Zhang, X. Su, and W. Li, "Fest : A Feature Extraction andSelection Tool for Android Malware Detection," *2015 IEEE Symp.Comput. Commun.*, pp. 714–720, 4893.

[10] A. Feizollah, N. B. Anuar, R. Salleh, and A. W. A. Wahab, "A reviewon feature selection in mobile malware detection," *Digit. Investig.*, vol.13, pp. 22–37, 2015.

[11] A. Firdaus, N. B. Anuar, A. Karim, M. Faizal, and A. Razak,"Discovering optimal features using static analysis and a genetic searchbased method for Android malware detection *," vol. 19, no. 6, pp. 712–736,

Author's Profile

K.Santhi Rani is currently working as a assistant professor in CSE Department, QIS institute of technology, Ongole, Andhrapradesh, India. She has fifteen years of experience in teaching undergraduate students. Her research interests are in the areas of image segmentation and machine learning

G Sai Kumari pursuing B.Tech in Computer Science Engineering from QIS institute of technology, Ponduru road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2017-21 respectively.

G. Srireesha pursuing B.Tech in Computer Science Engineering from QIS institute of technology, Ponduru road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University kakkinada in 2017-2021 respectively.

N. Naga vyshnavi pursuing B.Tech in Computer Science Engineering from QIS institute of technology, Ponduru road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2017-21 respectively.

T. Meghana pursuing B.Tech in Computer Science Engineering from QIS institute of technology, Ponduru road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2017-21 respectively.