DETECTION OF MALWARE IN ANDROID USING MACHINE LEARNING

PROF.V.VINAY KRISHNA
[1]Assistant Professor, CSE,Chalapathi Institute of Technology,Guntur, India
[2]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[3]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[4]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India
[5]UG Student,CSE,Chalapathi Institute of Technology,Guntur, India

Abstract Malware is one of the major issues regarding the operating system or in the software world. The android system is also going through the same problems. We have seen other Signature-based malware detection techniques were used to detect malware. But the techniques were not able to detect unknown malware. Despite numerous detection and analysis techniques are there, the detection accuracy of new malware is still a crucial issue. In this paper, we study and highlight the existing detection and analysis methods used for the android malicious code. Along with studying, we propose Machine learning algorithms that will be used to analyze such malware and also we will be doing semantic analysis. We will be having a data set of permissions for malicious applications. Which will be compared with the permissions extracted from the application which we want to analyze? In the end, the user will be able to see how much malicious permission is there in the application and also we analyze the application through comments.

1. INTRODUCTION

In this technological era, smartphone usage and its associated applications are rapidly increasing due to the convenience and efficiency in various applications and the growing improvement in the hardware and software on smart devices. It is predicted that there will be 4.3 billion smartphone users by 2023. Android is the most widely used mobile operating system (OS). As of May 2021, its market share was 72.2% .The second highest market share of 26.99% is owned by Apple iOS, while the rest of the 0.81% is shared among Samsung, KaiOS, and other small vendors. Google Play is the official app store for Android-based devices. The number of apps published on it was over 2.9 million as of May 2021. Of these, more than 2.5 million apps are classified as regular apps, while 0.4 million apps are classified as low-quality apps by AppBrain. Android's worldwide popularity makes it a more attractive target for cybercriminals and is more at risk from malware and viruses.

Studies have proposed various methods of detecting these attacks, and ML is one of the most prominent techniques among them. This is because ML techniques are able to derive a classifier from a limited set of training examples. The use of examples thus avoids the need to explicitly define signatures in developing malware detectors. Defining signatures requires expertise and tedious human involvement and for some attack scenarios explicit rules signatures do not exist, but examples can be obtained easily. Numerous industrial and academic research has been carried out on ML-based malware detection on Android, which is the focus of this review paper.

The taxonomical classification of the review is presented. Android users and developers are known to make mistakes that expose them to unnecessary dangers and risks of infecting their devices with malware. Therefore, in addition to malware detection techniques, methods to identify these mistakes are important and covered in this paper.

Detecting malware with ML involves two main phases, which are analyzing Android Application Packages (APKs) to derive a suitable set of features and then trainingmachine and deep learning (DL) methods on derived features to recognize malicious
APKs. Hence, a review of the methods available for APK analysis is included, which consists of static, dynamic, and hybrid analysis. Similar to malware detection, vulnerability detection in software code involves two main phases, namely feature generation through code analysis and training ML on derived features to detect vulnerable code segments. Hence, these two aspects are included in the review's taxonomy.

2. LITERATURE REVIEW

Android Malware Detection Using Machine Learning on Image Patterns.
Darus,FauziMohd,Salleh Noor Azurati Ahmad, AswamiFadillahMohdAriffin.

Android platform has been targeted by cyber-criminals due to the increase number of Android users in 2017. More than 8,000 Android malware were identified everyday making it is difficult for the malware analyst to detect them. Traditional malware detection techniques are no longer reliable to detect newly created malware in short period of time. In this paper, we use a different approach to detect Android malware. The Android malware will be visualized into gray scale images and their image features will be extracted using GIST descriptor. The detection will be done and compare using three different classifiers namely k-nearest neighbor (KNN), Random Forest (RF), and Decision Tree (DT).

Attackers who make malicious applications have come up with new methods of targeting victims of Android users.So, many researchers have verified the effectiveness of the available tools or suggested new tools that may be moreeffective in the process of accurately detecting malicious applications.Abdulrahman et al.offers a new mechanism for detecting malicious Android applications by using pseudo-dynamic analysis and constructing an API call graph for each execution path, based on a deep learning model builtand trained on a data set consisting of approximately 30 thousand malicious apps and 25 thousand benign apps.

Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms.
Vrama,P.Ravi Kiran,KotariPrudvi raj,KV Subba Raju.
Android occupies a major share in the mobile application market. Android mobiles have become an easy target for the attackers. The main reason is the user ignorance in the process of installing and usage of the apps. Android malware can be detected based on the permissions it requests from the user. Several machine learning algorithms are being used in the detection of android malware based on the list of permissions enabled for each app. This paper makes an attempt to study the performance of some of the machine learning algorithms, viz., naïve Bayes, J48, Random Forest, Multi-class classifier and Multi-layer perceptron. Google play store 2015 and 2016 app data are used for normal apps and standard malware data sets are used in the evaluation. Multi-class classifier was found to be outperforming the other algorithms in terms of classification accuracy. Naïve Bayes classifier has outperformed as far as model construction time is concerned.

An Android BehaviourBased Malware Detection Method using Machine Learning.
Chang,Wei-Ling, Hung-Min Sun, Wei Wu.
Android OS has become the leader of the global smart phone market. One of the superiority of Android is its open source

licenses, and it has designed its software developer kit (SDK) to work across as many platforms as possible. The open source Android platform is more flexible than iOS and allows developers to take full advantage of the mobile operation system. The popularity of Android system attracts many developers to develop not only useful and creative applications, but also some malicious software. They insert malware on Android Google play or other third party platform, pretending to be normal applications. Malware on Android can not merely steal users privacy information such as device information, phone number, IMEI, contact list, credit card number, bank account numbers
and even make the device be a C&C server botnet.

Android malware detection using random machine learning classifiers.

Koli, J. D. RanDroid

The growing popularity of Android based smartphone attracted the distribution of malicious applications developed by attackers which resulted the need for sophisticated malware detection techniques. Several techniques are proposed which use static and/or dynamic features extracted from android application to detect malware. The use of machine learning is adapted in various malware detection techniques to overcome the manualupdating overhead. Machine learning classifiers are widely used to model Android malware patterns based on their static features and dynamic behaviour. To address the problem of malware detection, in this paper we have proposed a machine learning-based malware detection system for Android platform. Our proposed system utilizes the features of collected random samples of good ware and malware apps to train the classifiers. The system extracts requested permissions, vulnerable API calls along with the existence of app's key information such as; dynamic code, reflection code, native code, cryptographic code anddatabase from applications, which was missing in previous proposed solutions and uses them as features in various machine learning classifiers to build classification model. To validate the performance of proposed system, "Ran Droid" various experiments have been carried out, which show that the Ran Droid is capable to achieve a high classification accuracy of 97.7 percent.

3. EXISTING SYSTEM

In the existing system, the application permissions are extracted to detect the malware and executed through the command prompt. A proper GUI was not provided to execute the tasks. All the commands were run through the command prompt. It was difficult for the non-technical user to use the system. And also Semantic analysis was not implemented.

4. PROPOSED SYSTEM

In this paper, we study and highlight the existing detection and analysis methods used for the android malicious code. Along with studying, we propose Machine learning algorithms that will be used to analyze such malware and also we will be doing semantic analysis. We will be having a data set of permissions for malicious applications. Which will be compared with the permissions extracted from the application which we want to analyze.

The user will be able to see how much malicious permission is there in the application and also we analyze the application through comments.

5. SYSTEM ARCHITECTURE:

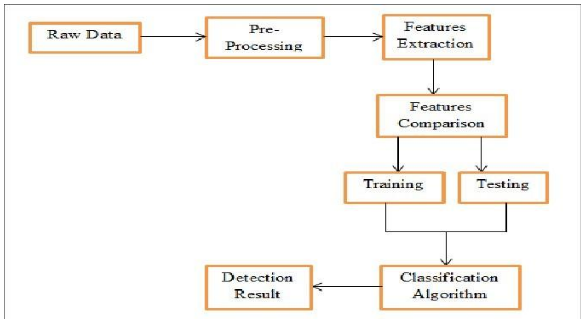

Fig 1 System Architecture

6. IMPLEMENTATION

Load Dataset:

First, load the dataset using pandas read_csv() method. It is an important pandas function to read csv files and do operations on it. We can read csv file not only locally, but from a URL through read_csv() or we can choose what columns needed to export so that we don't have to edit the array later.

Split Data Set:

Split the data set into two types. One is train dataset and another one is test dataset. However, having surplus data at hand still does not solve the problem. For ML models to give reasonable results, we have to ensure the quality of data and use it properly.

Train Dataset:

Train dataset will train our data using fit method. A model that is well-fitted produces more accurate outcomes. It contains the data which will be fed into the model. In simple terms, our model would learn from this data.

Test Dataset:

Test dataset will test the dataset using algorithms like Support Vector Machine, Neural Networks and the combination of Genetic algorithm on both.The test dataset contains the data on which we test the trained and validated model.

Predict data set:

Predict() method will predict the results. We use it to predict the values based on the previous data behaviors and thus by fitting that data to the model.

Predict() function will perform a prediction for each test instance and it usually accepts only a single argument or input which is usually the data to be tested.

7. SCREENSHOTS

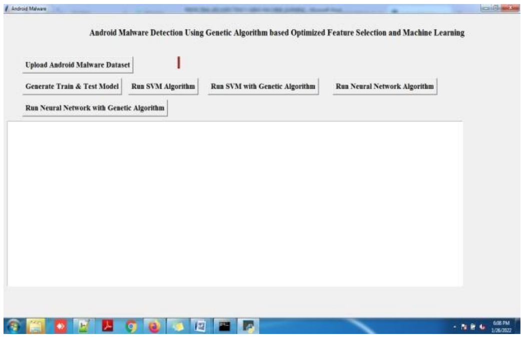
Android Malware
Android Malware Detection Using Genetic Algorithm based Optimized Feature Selection and Machine Learning
Upload Android Malware Dataset
Generate Train & Test Model
Run SVM Algorithm
Run SVM with Genetic Algorithm
Run Neural Network Algorithm
Run Neural Network with Genetic Algorithm

Fig 9.1 Selecting Android Malware Dataset

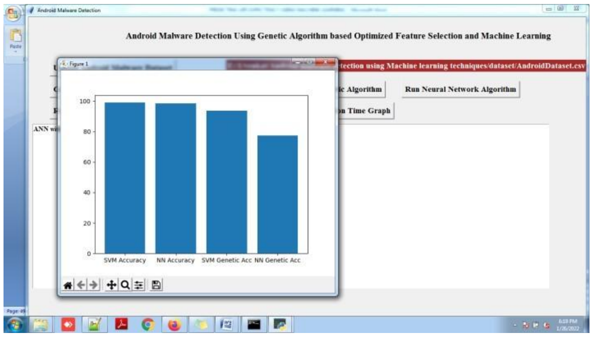
Android Malware Detection
Android Malware Detection Using Genetic Algorithm based Optimized Feature Selection and Machine Learning
Figure 1
tection using Machine learning techniques/dataset/AndroidDataset.csv
ic Algorithm
Run Neural Network Algorithm
on Time Graph
ANN wi
100
80
60
40
20
0
SVM Accuracy
NN Accuracy
SVM Genetic Acc
NN Genetic Acc

Fig 9.9 Graph representing the Accuracy

8. CONCLUSION

Malware is a critical threat to the user computer system in terms of stealing confidential information or disabling security. This project present some of the existing machine learning algorithms directly applied on the datasets of malware. In our work, we proposed a system for permission analysis and semantic analysis. Our system is also used to detect malware permissions based on an application by comparing it with a dataset. This proposed system can be applied in the fields of the security system and also for the n users like a malware detection software. It explains how the algorithms will play a role in detecting malware with high accuracy and predictions.

REFERENCES

1. J. Li, L. Sun, Q.Yan, Z. Li, W. Srisa-an and H. Ye, "Significant permission identification for

Machine-Learning-Based Android malware detection", IEEE Trans. Ind. Informat., vol. 14, pp. 3216-3225, Jul. 2018.

2. D. J. J. Tan,T.W.Chua and V. L. L. Thing, "Securing Android: A survey taxonomy and challenges", ACM Comput. Surv., vol. 47, no. 4, pp. 58, May 2015.
3. S. H. Qing, "Research progress on Android security", J. Softw., vol. 27, no. 1, pp. 45-71, Jan. 2016.
4. J. Lopes, C. Serrao, L.Nunes, A.Almeida and J. Oliveira, "Overview of machine learning methods for Android malware identification", Proc. 7th Int. Symp. Digit. Forensics Secur. (ISDFS), pp. 1-6, Jun. 2019.
5. M. Choudhary and B. Kishore, "HAAMD: Hybrid analysis for Android malware detection", Proc. Int. Conf. ComputerCommunalityInformant. (ICCCI), pp. 1-4, Jan. 2018.
6. M. Taleby, Q. Li, M. Rabbani and A. Raza, "A survey on smartphones security: Software vulnerabilities malware and attacks", Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 30-45, 2017.
7. M. Choudhary and B. Kishore, "HAAMD: Hybrid analysis for Android malware detection", Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), pp. 1-4, Jan. 2018.
8. M. Taleby, Q. Li, M. Rabbani and A. Raza, "A survey on smartphones security: Software vulnerabilities malware and attacks", Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 30-45, 2017.
9. M. Choudhary and B. Kishore, "HAAMD: Hybrid analysis for Android malware detection", Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), pp. 1-4, Jan. 2018.
10. M. Taleby, Q. Li, M. Rabbani and A. Raza, "A survey on smartphones security: Software vulnerabilities malware and attacks", Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 30-45, 2017.
11. A. Souri and R. Hosseini, "A state-of-the-art survey of malware detection approaches using data mining techniques", Hum.-Centric Comput. Inf. Sci., vol. 8, no. 1, pp. 22, Jan. 2018.
12. A. Amamra, C. Talhi and J.-M. Robert, "Smartphone malware detection: From a survey towards taxonomy", Proc. 7th Int. Conf. Malicious Unwanted Softw., pp. 79-86, Oct. 2012.
13. H. Lubuva, Q. Huang and G. C. Msonde, "A review of static malware detection for Android apps permission based on deep learning", Int. J. Comput. Netw. Appl., vol. 6, no. 5, pp. 80-91, Sep./Oct. 2019.
14. E. J. Alqahtani, R. Zagrouba and A. Almuhaideb, "A survey on Android malware detection techniques using machine learning algorithms", Proc. 6th Int. Conf. Softw. Defined Syst. (SDS), pp. 110-117, Jun. 2019.
15. A. A. A. Samra, H. N. Qunoo, F. Al-Rubaie and H. El-Talli, "A survey of static Android malware detection techniques", Proc. IEEE 7th Palestinian Int. Conf. Electr. Comput. Eng. (PICECE), pp. 1-6, Mar. 2019.
16. A. Qamar, A. Karim and V. Chang, "Mobile malware attacks: Review taxonomy & future directions", Future Gener. Comput. Syst., vol. 97, pp. 887-909, Aug. 2019.
17. Y. S. I. Hamed, S. N. A. AbdulKader and M. M. Mostafa, "Mobile malware detection: A survey", Int. J. Comput. Sci. Inf. Secur., vol. 17, no. 1, pp. 56-65, Jan. 2019.

18. P. Yan and Z. Yan, "A survey on dynamic mobile malware detection", Softw. Qual. J., vol. 26, no. 3, pp. 891-919, Sep. 2018.
19. M. Odusami, O. Abayomi-Alli, S. Misra, O. Shobayo, R. Damasevicius and R. Maskeliunas, "Android malware detection: A survey", Proc. Int. Conf. Appl. Inform. (ICAI), pp. 255-266, 2018.