Smartphone Malware's Detection using Data Mining and Machine Learning Techniques

[1] Prof. Priyanka Kokare [2] Prof. Shraddha Mankar [3] Prof. Ashvini Nikhade[4] Priyanka Tambat
Assistant Professor, Computer Department
ZE'S ZCOER, Narhe, Pune-411041

Abstract Nowadays, use of Mobile platforms are increasing rapidly. Users are capable of executing increasingly complex tasks very easily by using android apps. Banking is an area where lots of people are doing their banking tasks by using banking apps provided by the respective banks. Mobile devices are targeted by malware's due to rising use of android apps. Due to openness feature of android it became favorite for users and developers alike.Users are downloading number of apps from Play store any time. During downloading, the number of malware's incurred with apps, behind the scene; Malware's performs the various activities like hacking the authentication process which contains various confidential data of Users and Login activities. Due to this increasing use of mobile apps for work, Users are losing their Smartphone's Integrity and Confidentiality. Developers are free to develop any kind of apps and without doing any scrutiny of their apps, developers are publishing their own created apps in a play store. Developers, who are hackers, are taking advantage of user illiteracy about this malware attacks through the apps.This paper, presents the system which would help to provide protection to users android phones by analyzing and removing such malicious apps. The proposed system would work by analyzing the permission's as features which are taken by users during installation. Clustering and Classification techniques are used to analyze the apps. The main motivation of this paper is how to find and remove malicious apps which are present in user's android device.

Keywords: Benign, Clustering, Data Mining, Apps, Classification, Malicious

I. Introduction

Due to increasing use of Smartphone's and rapid enhancements in Smartphone brands, Many Mobile application developers inspired to develop more apps. Number of operating systems has been developed for mobiles, such as Android OS(Google), iOS (Apple) and Palm OS (Garnet OS), Symbian OS (Nokia), etc. Android is the most popular one, due to its features as Light weight, Cost effective, Open source in all mobile operating systems. Android built on Open Linux Kernel. Attackers are targeting more on Smartphone's running Android. According to the threat report of 2015 of Kaspersky the 48.35% of attacks occurred on banking applications. The banking details are hacked by hackers by sending Trojan-SMS and Trojan-Banker malware, Most Risky threat were Scare wares, Ad wares and Ransom ware. As compared to the threat attack report of 2014,the number of android phones infected due to malwares is increased by three times.

There are various techniques available to detect malware's from Smartphone's running android. They are Signature Based Techniques, Machine Learning techniques ,De –compilation based Techniques and Rule Based Techniques. Signature based malware detection technique work by using signatures of malicious codes or data. Signatures are made with properties, characteristics of malware's.

In this signature based technique, signatures are made by human beings manually. But that will be limitation for this technique as if any new malware will come and attack, then it become impossible to detect that new malware, because there is no signature present in cluster to find it. More time needed to make signatures manually. As signatures are made manually, so it may happen wrong signature will be made. Time required to detect unknown malware's using this technique will be more.

De-compilation based technique will work by decompiling the android app which user is using. The decompiled code of app will be used for recovery of source code. Next after De-compilation, This technique applies the structure flow, semantic patterns, data flow analysis and control flow analysis methods to detect maliciousness in app. But due to missing of data and control flow, it will increase the false alarm. As a limitation, this technique failed to identify benign apps and malicious apps.

Third technique, Rule based Technique is a data mining technique to detect malicious app. It finds out whether there any unwanted operations going on due to presence of malicious apps in phones.

Extracted features from the apps which are labeled as features are given to machine learning algorithms. Features are extracted from manifest.xml (definition file) and .dex (code based) file. Once the features are captured, they are learned by algorithms. These learned features are used as indictor of

malicious applications. Classifier will classify the benign features and malicious features. These classifiers are used to measure accuracy This The proposed system is automated will create features at runtime and help to differentiate between good and bad apps.

In order to use an application, user has to grant all permissions that an application needs and then installed application runs under granted permissions. A user who accepts all permissions and install the apps, didn't even understand the meaning of those words. So unwillingly, user had accepted the permissions so that the malwares will enter in his/her phone and start changing the system files or any authenticated details from device.

In this paper, we propose a system which will protect the android users from malicious applications. As a first step, the permissions which were accepted by users are stored as Features, Clusters were made from this collected features by applying k-means clustering algorithm on it. Secondly The classification algorithm Naive Bayesian is applied on clusters to classify whether the app is good or harmful.

II.Literature Survey

In this related work, it gives the Literature survey of which technologies had developed to detect this malicious apps and how this technologies implemented using which methods is explained.

A. Overview of Android

The operating system Android is built on Linux as its base. Developers are developing android apps in java and Java Libraries designed by Google are used for controlling their operations. Android architecture is shown below in Figure 1.1.Software stack is the combination of Linux kernel and c/c++ libraries that are exposed through the app framework. Different services and management will be provided at run time by this app framework. Linux kernel handles the core services. Abstraction is also provided. Dalvikvirtual machine and core libraries are included in Android Run Time.

Android Malware Survey

In this paper [1],Eiichiro Kodama and Takayuki Matsudo have proposed a system that assesses and presents risk level of an android application. Two parameters are involved in Risk assessment. Initial factor is how many downloaded apps are there and second is the user ratings from android market.Second is to analyse the permissions which are collected from user. In this paper ,Rule are generated to identify whether app is malicious. Risk calculation formulae is given in it to calculate risks.System then present risk information of application to user using various colors as risk indicators.

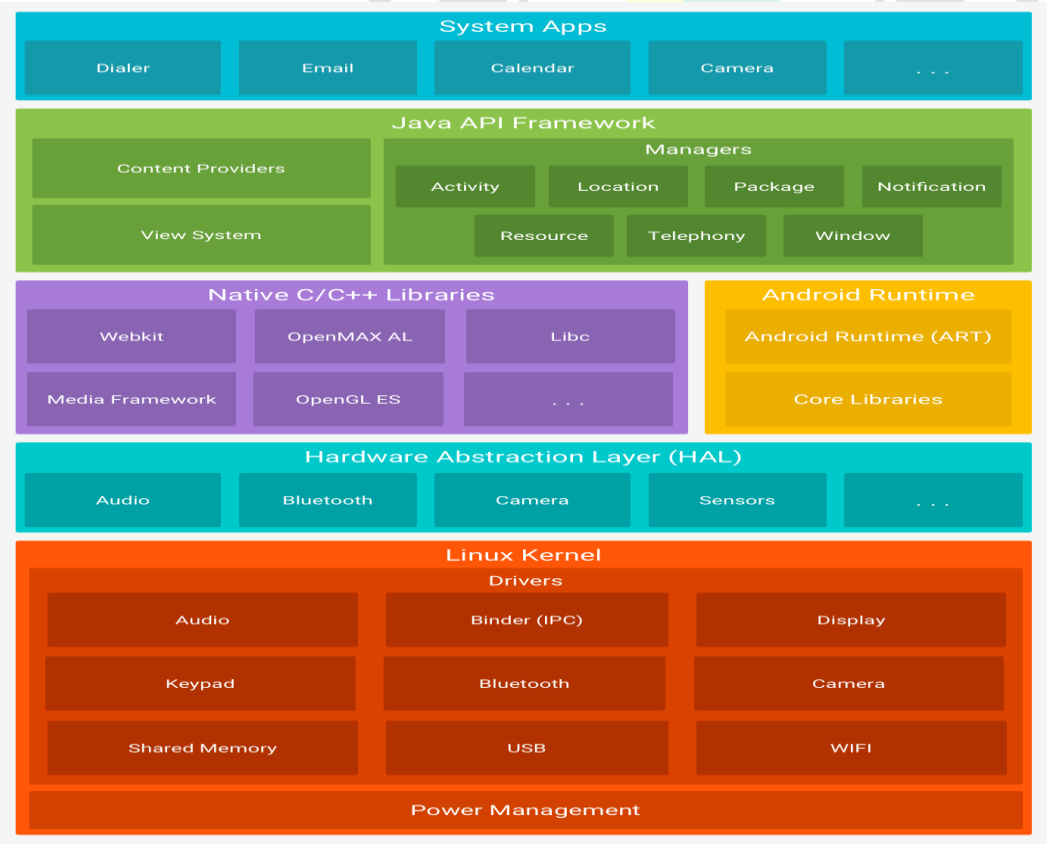

Fig. 1.1Android Architecture

In this paper [2], S.Y.Yerima has proposed a system using Bayesian classification. The system is made with collection of code and app characteristics. Bayesian classification is used to classify.

In this paper [3], Agamas has proposed the system consisting of 3 main entities or modules such as event notification,Security manager and market manager. It requires application developers to insert an event notification code into their applications which would inform every event to security manager, whenever an application launches a security related events. If security manager finds any application without an event notification code, it sends the message to the market manager that resides at android market place. The market manger then excludes all such applications that don't include event notification code. A security manager is build inside android OS and it uses a dedicated API to create a knowledge database which it could use to judge malicious behavior.

In this paper[4],Ghorbanian Metal ,have proposed a host base Intrusion Detection model, in which the log files are first inserted by logicat command through android logging system and are give to analyzing module .Analyzing module then involves matching engine to detect intrusion .Matching engine performs the pattern matching with the malicious rule set that is designed to detect malwares from benign apps.

In this paper[5],Dong -Jie Wu et al implemented a system called Droid Mat.It uses static malware deetction method to detect malicious apps.In static analysis, the system extract various information such as permissions, deployment of components, Intent Message passing and API Calls for characterizing android application behaviour. It then applies k-means algorithm that enhances malware modelling capability and uses K-NN algorithm to classify the application as benign or malicious.

In this paper[6], G Dini have proposed a MADAM-Multilevel anomaly detector for android malware uses features to detect android malware for both kernel level and user levels

III.Proposed System

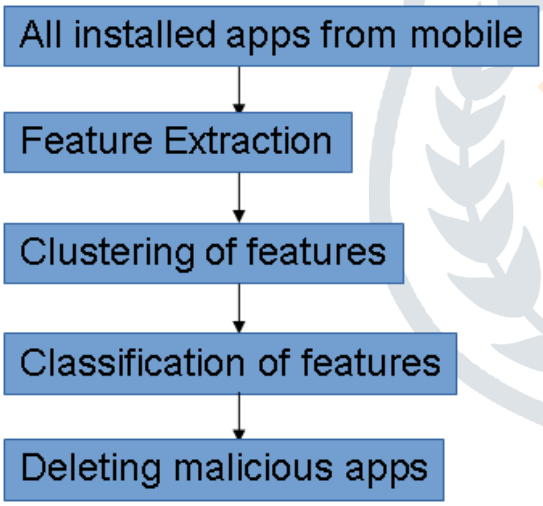

Fig 3.1 System Architecture

1. Identification of Installed Apps:

Initially identify the list of installed apps.PackageManager class of android is used to identify installed apps.PackageManager provides methods for querying and manipulating installed packages.GetPackageManager () function can be used to call Package Manager

2. Feature Extraction

Extract permissions given at time of installation as Featuress.Add more parameters as app name, pacakge name, version of app. This all information would be useful to the user.PackageInfo class is used to extract these information about contents of package. The information collected through the Package Info class correspondences to all the information present in AndroidManifest.xml.

3. Clustering of Features

K-means algorithm is used for clustering. Two clusters are formed one as malicious permission cluster and other benign permission cluster. Permissions of each app are taken as input and each individual app is added to one of this cluster. With some known families of malwares the malware cluster is made. List of Permissions represent one family of malware. Hence each cluster will be having set of known harmful permission.

4. Classification of Features

Since clustering could result in benign app declared as malicious. There is need to accurately classify whether the app belongs to benign app or malicious app. It is achieved using Naive Bayesian Classification Algorithm

5. Deleting Malicious apps:

This module gives user the result of classification which apps are benign and which apps are malicious. It allows user to give input whether to delete malicious app or not from the phone.

IV.CONCLUSION

The Android Application Analyzer (AAA)will allow users to identify malicious apps installed in the phone and provide facility to uninstall them. This proposed system will be implemented using classification and clustering algorithm, but current Antimalware apps using Signature based detection method to detect malicious apps which had limitations that they unable to detect newly coming malicious apps.Zero day attacks are not handled using signature based technique. This proposed system does not require updating of malwares online. It can be used offline by users to detect malicious apps without having updating of antimalware tool. Limitation of this proposed system will be, it needs to create a new cluster if any new malware app comes in the phone.

This proposed system will give more accuracy as compared to existing antimalware apps which in turn provides confidentionality, Integrity, Availablity measures to users. Users will be more secured by using this system.

V. REFERENCE

[1] Takayuki Matsudo, Eiichiro Kodama(2012) "A proposal of security advisory system at the time of installation of applications on android os"

[2] S.Y.Yerima (28March 2013),"A new android malware detection approach using Bayesian classification"

[3] Agematsu H,Kani J, Nasaka K,.Kawabatta H,ADMS(July 2012):"A proposal to realize the provision of secure android applications(An application development and Management System)".

[4] Ghorbanian M,Shanmugan B,Narayansamy G, (April 2013)"Signature based hybrid intrusion detection system for android devices(HIDS)"

[5] Dong-Jie Wu ,Ching Hao Mao(Aug2012), "DroidMat-Android Malware detection through Manifest and API Call Tracing"

[6] Gianluca Dini , Fabio Martinelli , Andrea Saracino , and Daniele Sgandurra,F Martinelli (2012)," MADAM: a Multi-Level Anomaly Detector for Android Malware"

[7] J. Oberheide and C. Miller, "Dissecting the Android Bouncer," presented at the SummerCon2012, New York, NY, USA, 2012.

[8] K.Shi and K.Ali. Getjar mobile application recommendations with very sparse datasets. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '12, pages 204–212, 2012.

[9] J.Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in Proc. 27th Annu. ACM Symp. Theory Comput., 1995, pp. 209–218.

[10] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. SIGKDD Explor. Newsl.,13 (2):50–64,May2012.

.