International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 11 6170 - 6174

# Permission based Mobile Malware Detection System using Machine Learning Techniques

Mayuri Magdum
Computer Engineering
Modern Education Society's College of Engineering,
Pune, India
*mayurimagdum2006@gmail.com*

Prof.S.K.Wagh
Computer Engineering
Modern Education Society's College of Engineering,
Pune, India
*skwagh@mescoepune.org*

***Abstract*—**Mobile technology has grown dramatically around the world. Nowadays smart mobile devices are ubiquitous, i.e. they serve multiple purposes such as personal mobile communication, data storage, multimedia and entertainment etc. They have become important part of life. Implementing secure mobile and wireless networks is crucial for enterprises operating in the Internet-based business environment. Mobile market share has grown significantly in past few years so that we need to think about mobile security. Mobile security can be compromised due to design flaws, vulnerabilities, and protocol failures in any mobile applications, viruses, spyware, malware and other threats. In this paper we will more focus on mobile malware.

Many tools are available in the market to detect malware but new research trend in the mobile security is users should be aware of app before he/she installs from the app store. Hence we propose a novel approach for permission based mobile malware detection system. It is based on static analysis. It has 3 major parts in it 1) a signature database for storing analysis results of training and testing. 2) An Android client who is used by end users for making analysis requests, and 3) a central server plays important role as it communicates with both signature database and smartphone client. We can say that he is the manager of whole analysis process. It alerts user if the app is malicious or the benign based on it user can proceed whether to continue with it or not.

***Keywords*-*Mobile, Malware, Machine Learning, Static analysis, Dynamic Analysis, smartphone security, android***

*****

## I. INTRODUCTION

Now a Days Smartphones and tablets are becoming popular, effective and efficient malware threat analysis becomes important. According to study made by International Data Corporation, the smartphone market will grow four times faster than mobile phone market and the demand of Smartphones will rise considerably reaching the point where customers will replace their old mobile phones with smartphones.

Unfortunately,this growth in smartphone attracts malware authors too. Malware count has also grown significantly and they are capable to gain your personal information or data on your phone. They can slow mobile activity and Delete your contacts list. They can send spam to your mobile. They can send false Bluetooth connection signal, they can avail super user permissions and steal sensitive information. It has been reported that on Android, the most popular smartphone platform malware has been constantly attacked by new type of malware and thus the platform has also seen an evolution of anti-malware tools.

Security and privacy are two important issues for smartphones. From past decades it has been observed drastic growth in malware applications and for security analysts it has been challenge to track those and design antimalware tool.Our focus will be more on android operating system and as it shares highest market share from the past few years and also it is most widely used mobile operating system. There are many approaches for mobile malware detection and analysis 1) Static analysis 2) Dynamic analysis 3) Hybrid approach for malware detection. In static analysis we map the malicious behavior against the signature database which is already available. In dynamic analysis we compute the behavior of the system which is far deviated from the standard behavior and hybrid approach is combination of both static and dynamic analysis. If we combine these approaches with machine learning algorithms then good antimalware techniques are formed and vast research is going on this field for mobile security. In this paper, Section 1 gives introduction to mobile security. Section 2 gives the literature review and related work on mobile malware detection. Section 3 introduces various types of malware and mobile malware feature selection techniques. Section 4 introduces proposed work. In section 5 conclusion and future scope are given.

## II. RELATED WORK

The malware research has been started from the year 2005[1].An extensive research has been carried out in the field and to address the issue of malware Kevin Allix, Tegawendé F. Bissyandé, Quentin Jérome, Jacques Klein, RaduState,Yves Le Traon proposed a method called '10 fold

cross validation'[2].Younghee Park , Douglas S. Reeves , Mark Stamp derived a common behavior of malware using graph clustering and then their method generates one common behavioral graph by clustering a set of individual behavioral graphs, which represent kernel objects and their attributes based on system call traces[3].

Different features like the permission based features and the API call based features are considered in order to provide a better detection by training and combining their decisions using collaborative approach based on probability theory is given by Shina Sheen , R.Anitha, V.Natarajan[4].Kabakus Abdullah Talha , Dogru Ibrahim Alper , Cetin Aydin proposed a method based on permissions used in an application and static analysis is made using machine learning algorithm such as logistic regression[5].

Jae-wookJang ,Hyunjae Kang , Jiyoung Woo , Aziz Mohaisen ,Huy Kang Kim invented a system an anti-malware system based on similarity matching of malware-centric and malware creator-centric information which is able to detect classify malware samples in to similar sugroups[6].DONG Hang1, HE Neng-qiang, HU Ge, LI Qi, ZHANG Miao proposed a method to detect variants of known malware families in Android devices using simplify Dalvik instructions. This method is based on the sequence of instructions [7].

DoaaHassana, Matthew Might, and VivekSrikumar found the similarity based approach to detect the malware. In their technique similarity between methods is computed by using the normalized compression distance (NCD).with the help of either zlib or bz2 compressors or then similarity measure is computed. That computed similarity measure is then used for training and then afterword's predicting result whether the app is malicious or benign [8].

Another way to find malicious apps is discovered by Yajin Zhou, Zhi Wang, Wu Zhou, Xuxian Jiang. They proposed a method in which they used permission based behavioral footprinting scheme to detect new types of known Android malware families. Then they applied a heuristics-based filtering scheme to identify certain type of behaviors of unknown malicious families [9].

Seung-Hyun Seo, Aditi Gupta, Asmaa Mohamed Sallam, ElisaBertino, KangbinYim proposed a method to detect mobile malware threats to homeland security. In their proposed approach they define characteristics inherent in mobile malware and show mobile attack scenarios which are feasible against Homeland Security. They derived a static analysis tool, DroidAnalyzer, which identifies potential Vulnerabilities of Android apps and the presence of root exploits[10].A. Shabtai, L. Tenenboim-Chekina, D. Mimran, L. Rokach, B. Shapira, Y. Elovici, discovered a method to find mobile malware based on semi supervised machine learning despite of regular static and dynamic based analysis.[11].

Ping Wang, Yu-Shih Wang proposed a technique based on signature based analysis and SVM to detect malware. They also used a cross validation scheme for improving accuracy malware detection [12].

Karim O. Elish, XiaokuiShu, Danfeng (Daphne) Yao, Barbara G. Ryder, Xuxian Jiang described a method which is based on classification approach to detect malicious android app. Analysis and Results showed that the method proposed is highly accurate[13].

Jehyun Lee, Suyeon Lee, Heejo Lee proposed an method for malware screening which includes mechanism to extract a set of family representative binary patterns from evaluated family members as a signature and to classify each set of variants into a malware family via an estimation of similarity to the signatures.This similarity they used detect malware in their proposed method [14]

.Wanqing You, Kai Qian, MinzheGuo, Prabir Bhattacharya proposed a hybrid approach for mobile threat analysis, The key of this approach is the unification of data states and software execution on critical test paths conditions. The outcome leads to combine the benefit of static and dynamic analysis. This is the main benefit of their technique that is they used a hybrid approach for analysis[15].

## III. TYPES OF MALWARE AND FEATURE SELECTION TECHNIQUE

### *A. Types of Malware*

There are variety of attacks available for android ranging from adware to the most sophisticated and dangerous ones. We will just go through main malware attacks which happened in last few years and drawn attention of the world.

- Android Dowgin is one example of an adware. It works such that it installs itself on Android device in with respect to other application. After successful installation it starts displaying advertisements in the notification area of the device which cannot be easily removed.
- Some of the malware are having financial intentions. Upon installation, some applications send expensive short message service (SMS) to premium numbers without user's knowledge that reflects itself in user's bills. Based on a report by Sophos, a security firm, a malicious version of the popular Angry Bird game secretly sends premium SMS that costs GBP 15.
- Some of malware have crossed these limits too, as they were able to make the call in background without users notice. It then starts calling premium numbers. As soon as the user interacts with the device, the malware ends the

call.

- A botnet is more dangerous type of attack than any other infected applications listed above. Upon infecting the device, the attacker gains the whole access to the devices and it can perform malicious activities. A botnet attack then becomes racket of such infected devices. An attacker gains the power to access hundreds of infected devices in a single botnet attack. e.g. security analysts discovered an infected version of the Angry Birds Space in April 2012.
- As per reports of Symantec 2014,TrojanDroidpak malware uses hybrid threats for infecting mobile devices. It first gains access to a personal computer and automatically downloads Android application package (APK) file which is malicious. When the normal user connects his Android device to the computer, then malicious file attempts to install itself on the device. After the successful installation, it attempts to convince the user to download and install the infected version of Korean banking application.

*B. Mobile Feature selection Technique*

*1) Static Feature*

Static feature contains features available in the android apk file such as AndroidManifest.xml file and Java code file. Android operating system has base of Linux core; among which it comprises important part of Linux security architecture. Before installing any application, it gives the list of requested permissions to the user. When user grants the permissions, the application is installed successfully on the device. It is found that there are 145 official Android permissions in Android 4.4(kitkat). Google categorized them into four groups namely normal, dangerous, signature, and signature or system.

Advantage of working with static features is that they are easy to extract and disadvantage features can be added when the new version of android is launched in the market then signature database should be updated every time when new versions are released.

*2) Dynamic features*

There are two main types of dynamic features used in recent works: system calls and network traffic. Every application demands resources and services from operating system through system calls, such as read, write and open. Network traffic is another kind of dynamic feature used by scientists and researchers. Applications are usually prompted to connect to network to send and receive data, receive updates, and subsequently they are maliciously leaking personal data to attackers. There are more than 250 system calls in a Linux kernel which are also available in Android.

*3) Hybrid Features*

Hybrid features are group of static and dynamic features that are used together in detection systems.

They are the most comprehensive features, as they involve vetting Android application installation file as well as analyzing behavior of the application at runtime. (Blasing 2010) developed AASandbox which analyzes both static and dynamic features. It extracts permissions at first and then Java code from the APK file and this tool uses them as static features. It then installs the application; logs system calls, and uses it as dynamic feature [16]. All the features can be summarized diagrammatically as below fig 1.

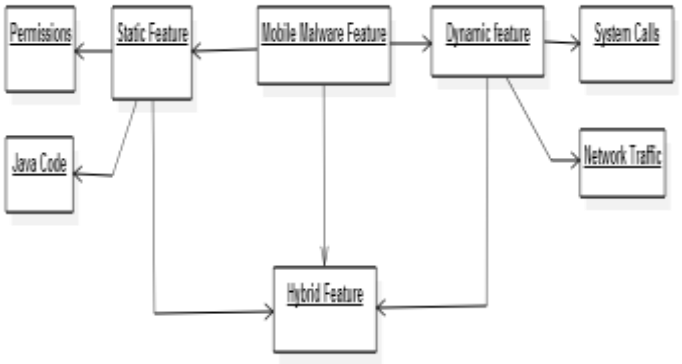


Fig 1. Taxonomy of Mobile Malware Features

*C. Feature share in applications to detect Malware*

If we review the antimalware applications [2011-2014] then we will find that 45% of applications are based on static features.42% of applications have used dynamic features and 10% of applications have used hybrid features. Only 3%of applications have used other features than these 3 and we will name those as application metadata. Diagrammatically it can be represented as below in fig 2.

## IV. PROPOSED WORK

We propose a novel architecture which will analyze the android application and will give alert user whether the particular app contains malware or not [5].This architecture is based upon static analysis method. The architecture of proposed system consists of 3 parts as mentioned below. (1) A signature database is used to store extracted information about applications and analysis results, (2) an Android client which can be used by end users to grant application analysis requests, and (3) a central server responsible for communicating with both signature database and smartphone client and managing whole analysis process [5].

A signature database contains the extracted permissions from malicious dataset as well as benign dataset. The sample set of permissions can be represented as below in ER diagram format as in fig 3.

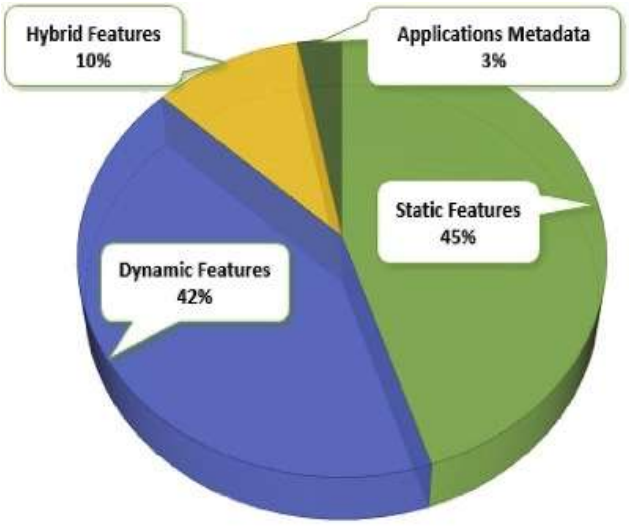


Fig 2. Feature share in antimalware applications.

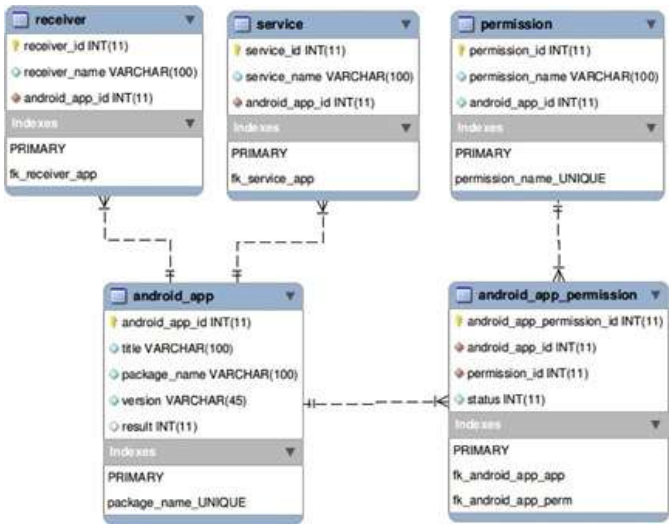


Fig 3. Permissions representation in E-R diagram format.

We will briefly summarize that how malware score calculations are done.

After permission extractions we will calculate malware score for each application. Before that we need to compute permission malware score (PMS).PMS is nothing but no of malwares that uses the permissions from all of the malware dataset that we have, and then next we calculate application malware score (AMS)[5]. AMS is simply summing up permission malware score.

$$PMS=\sum \text{No.of Malware that Uses the permission/No.of all Malware} \quad (1)$$

$$AMS=\sum PMS \quad (2)$$

Client is only responsible for making analysis request. Central server plays important role in the app analysis. Server is first trained by Logistic regression and then some threshold is set at first in the training phase itself and in testing phase the client will make the request. If the permission count extracted for the testing app falls above the threshold then it is definitely malicious and if it falls down then definitely it is benign.

The whole process is depicted below in the fig 4.

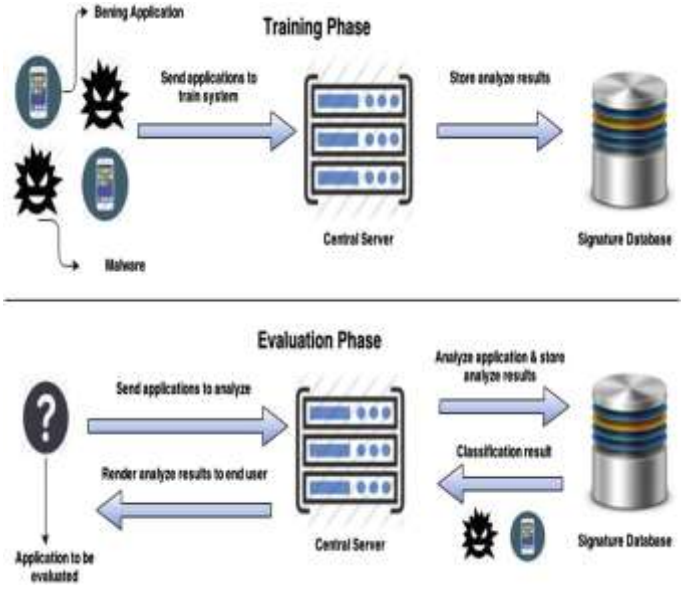


Fig.4Software architecture for mobile malware detection.

## V. CONCLUSION AND FUTURE SCOPE

The analysis tool proposed above will enhance the security of the application which is to be installed on the mobile phone. It will alert the user before installing any application that whether it contains malware or not.In future if we can combine with another machine learning algorithm to improve accuracy of the classification and of course it will ultimately ensure security too.

## REFERAANCES

[1] MOBILITY 2015 - The Fifth International Conference on Mobile Services, Resources, and Users http://www.researchgate.net/publication/278968819

[2] Kevin Allix, Tegawendé F. Bissyandé, Quentin Jérome, Jacques Klein, Radu State,Yves Le Traon, "Large-Scale Machine Learning-based Malware Detection", March 2014, ACM, ACM 978-1-4503-2278-2/14/03.

[3] Younghee Park, Douglas S. Reeves , Mark Stamp, "Deriving common malware behavior through graph clustering", September 2013, Elsevier, computers & security 39 ( 2013) 419 -430

[4] Shina Sheen, R.Anitha,V.Natarajan, "Android based malware detection using a multifeature collaborative decision fusion approach",October 2014, Elsevier, Neurocomputing151(2015)905–912.

[5] Kabakus Abdullah Talha , Dogru Ibrahim Alper , Cetin Aydin , "APK Auditor: Permission-based Android malware

detection system", March 2015, Elsevier, Digital Investigation 13 (2015) 1-14.

[6] Jae-wook Jang , Hyunjae Kang , Jiyoung Woo , Aziz Moh aisen ,Huy Kang Kim , "Andro-AutoPsy: Anti-malware system based on similarity matching of malware and malware creator-centric information", June 2015, Elsevier, Digital Investigation 14 (2015) 17-35.

[7] DONG Hang, HE Neng-qiang , HU Ge, LI Qi, ZHANG Miao, "Malware detection method of android application based on simplification instructions", July 2014,Elsevier, July 2014, 21(Suppl. 1): 94–100

[8] Doaa Hassana, Matthew Might, and Vivek Srikumar,"A Similarity-Based Machine Learning Approach for Detecting Adversarial Android Malware".

[9] Yajin Zhou, Zhi Wang, Wu Zhou, Xuxian Jiang, "Detecting Malicious Apps in Official and Alternative Android Markets".

[10] Seung-Hyun Seo, Aditi Gupta, Asmaa Mohamed Sallam, ElisaBertino, KangbinYim, "Detecting mobile malware threats to home land security through static analysis", June 2013,Elsevier, Journal of Network and Computer Applications 38(2014)43–53

[11] A. Shabtai, L. Tenenboim-Chekina, D. Mimran, L. Rokach, B. Shapira, Y. Elovici, "Mobile malware detection through analysis of deviations in application network behavior" , Feb 2014,Elsevier, computers & security 43(2014)1-18

[12] PingWang, Yu-ShihWang, "Malware behavioural detection and vaccine development byusing a support vector model classifier", Dec2014, Elsevier, Journal of Computer and System Sciences 81(2015)1012–1026

[13] Karim O. Elish, Xiaokui Shu, Danfeng (Daphne) Yao, Barbara G. Ryder, Xuxian Jiang, "Profiling user-trigger dependence for Android malware detection", November 2014,Elsevier, computers & security 49(2015)255-273

[14] Jehyun Lee, Suyeon Lee, Heejo Lee, "Screening smartphone applications using malware family signatures", Article in press,Elsevier

[15] Wanqing You, Kai Qian, Minzhe Guo, Prabir Bhattacharya, "POSTER: A Hybrid Approach for Mobile Security Threat Analysis",June 2015,ACM, ACM 978-1-4503-3623-9/15/06.

[16] Blasing T, Batyuk L, Schmidt A-D, Camtepe SA, Albayrak S. An android application sandbox system for suspicious software detection. In: 5th International Conference on Malicious and Unwanted Software (MALWARE); 2010. p. 55-62. http://dx.doi.org/10.1109/MALWARE.2010.5665792.