Malware Detection in Android platform

Venkatesh Nagadevara
9411206551
vena15@student.bth.se

Group Number: 45

Siva Prasad Patta
9312217137
sipa15@student.bth.se

GROUP MEMBERS PARTICIPATION

The work distribution in the group members is as follow

Group Member	Literature search	Research design	Report Writing
Venkatesh Nagadevara	45 %	55%	50 %
Siva Prasad Patta	55 %	45%	50 %

Table 1. Group Member's Participation

ABSTRACT

In this paper we present research proposal to find the problems in android malware detection. In this we conduct an background study on the topic and find the research gap in the field and try to give an appropriate solution. As android is the most used mobile operating system there is a increase in malware in many android applications. In this we discuss research question that are aroused and what are the approach that we use for solving problem.

Author Keywords

Android platform; malware detection; applications security; Malware attacks;

ACM Classification Keywords

K.6.5 [Security and Protection]: Authentication, Insurance, Invasive software (e.g., viruses, worms, Trojan horses), Physical security, Unauthorized access (e.g., hacking, phreaking).

INTRODUCTION

As android is one the most used mobile operating system. It has increased the user rate day by day. Improvement in the advancement in smart phones hardware capabilities such as GPS, high speed processors. The security is necessary for the device and underlying network. The security should be provided both on end user and service provider. Due to the increase in users it has created an open arena malware developers they target on valuable user data. Some of the well known malware attacks are information extraction, premium rate SMS, root exploits, search engine optimization, dynamically downloaded code, convert channels, botnets are some of the general attacks [1].

As smart phones are being attacked by malware. To find the downloaded find the downloaded software is verified by mobile anti viruses. Which use their own method to detect malware. Therefore applications are basically analyzed by static and dynamic analysis.

Static analysis is used to find that a application in malicious or not. That is done by extracting information from manifest file and see API calls, inter component communication, information passing through the application these information from the byte code are applied to k-means and EM-clustering algorithms to verify applications contains malware or not. Dynamic analysis in which app are run in a virtual simulator and test. The features of dynamic analysis is data and control flow analysis, emulated based analysis, logged behavior sequence [1].

Due to the android open source project we can install apps from third party markets and official android play store market. As the device consists of valuable user data such as contacts, bank details, message, email and many details. Due to

MOTIVATION

As android the most used worldwide operating system it being increasing day by day. Due to the increase in the android users and easy usage android is most popular mobile operating system. Android occupies more than 73% of market share in market. Due to the open source project any developer can develop android application. Android also supports third party software and markets app. Most android devices applications are downloaded from the official app store called Google play store. As android also supports third party applications. The android has become a target to the malware developers because of increase in the android users all over world. The android operating system consists of confidential data

such as contacts users data, data files, locations, and important bank details, IMEI details that have a high rate in black market. So the malware developers have developed techniques that can be used to retrieve user data without the intimation of the owner by creating malware applications. Even in Google app store application are sometimes malicious. Generally in Google play store the app is first introduced into a Google bouncer a static analysis emulator for malware and they are finally uploaded into app store. The hackers use many techniques to create malware such as information extraction in which information such as IMEI, location and personal data can e retrieved. Premium call and SMS malware in which user gets high cost premium unsubscribed SMS without initiation of owner. Both end user and malware user create Root exploit malware. Search engine optimization in which when we enter a URL a unknown malware application tries to install. Botnets are controlled by commands and network control. Due to this type increase in malware in daily usage applications we need to improve the malware detection techniques. Some of the malware detection methods are static analysis methods and dynamic analysis methods. The android malware can also be detected by method such as data mining and artificial intelligence. So due the increasing malware developers there have to be certain methods to decrease malicious applications. There are security issues on the user side and also security issues on the network. This why we are motivated in the topic android malware detection [13].

LITERATURE REVIEW

In this section the authors present the literature review based on the selected literature.

Search Strategy

At initial stage we have searched many research papers. Later searched some databases and selected some research papers that are relevant to the topic that we have selected. By eliminating unnecessary papers lead us to this papers.

Refining Research Question and Search Strings

The literature review was structured by analyzing research papers from the scientific research databases like inspec and IEEE explorer. We use keywords as android malware detection. Then selected advanced search and entered the keyword. This resulted in 114 papers among them 101 were conference publications 11 are magazines and journals and 2 are early access articles.

Include/Exclude criteria

The research papers are selected from IEEE explorer database based on the criteria described as

Inclusive Criteria

- Researches done from last three years that is from (2012-2015) are selected and refined.
- Theoretical study and experimental facts that are possible are selected.
- Papers describing about malware detection methods for android are considered.
- Only papers that have vocabulary that is relevant to research question are selected.

Exclusion Criteria

- Papers that do not have relevant data are excluding from resulted papers.
- Same type papers that are selected excluded from the study.
- Articles that have only abstracts but no full text are eliminated.
- Articles which are not in English are eliminated from selected papers

Malware issues in Android Platform

The malware issues based on literature review is as follows

As the increase in android platform is increasing it has become an arena for malware developers. As a result there growth in malware. All Antivirus cannot identify all malware they can only identify only to a extent [1]. As applications are present in Google app store there are increase in malware, which will cause financial loss and also user data vulnerability. As android is an open project it can also install applications from third party app stores. This may increase the growth malware because the third party app store may not check for malware [6]. Actually malware is a general term used to refer the hostile software such as botnots, spybots whose major target is user data. As the malware detection in android devices are far behind than compared to the increase in the malware today [4]. As previous study says malware is detected by malicious dataset analysis, dynamic analysis, similarity and heuristics based malware detection, signature detection and many [1]. Even the erased data can be gained by the malware in Linux operating systems.

Android Operating System Structure

Android is one the most popular mobile operating system.

The malware detection is important because of valuable user data. Traditional malware is detected by signature-based analysis. In this we compare the application files with the already list malware. This method helps only a little because it only identifies already recognized malware only. The all applications that are uploaded to the Google play undergo scan under a Bouncer. And another main type detection is behavior based detection.[10]"The behavior based detection method is a detection method through machine learning classifiers for attack pattern during the performance of Application package file is a jar file mainly consists of manifest file, source code and other resources. There contains permissions that allow accessing android operating system. As android is a open source project. We can give root access to the device and install third party application and software. The malware attacks most common are repacking attack. In which malware developer unpacks the application and repacks it with malware. Metamorphism is a new method by which we can bypass the new anti-malware software. To achieve this they insert dummy code, API and code manipulation. Therefore the appearance of application can false detect.

Malware and process behavior in a normal situation that occurs in devices". Generally these are classified by system calls and the data that it is accessing. Shabtai et al. used [10] KBTA (Knowledge-based temporal abstraction) methodology observes the patterns for any malicious behavior. Generally android conduits of five layers application layer that is in contact with user and runs applications. Framework layout it is written in c language it quickly recognizes graphics and other. The Android runtime consists of core libraries and dalvik virtual machine. And other main layers are libraries and Linux kernel.

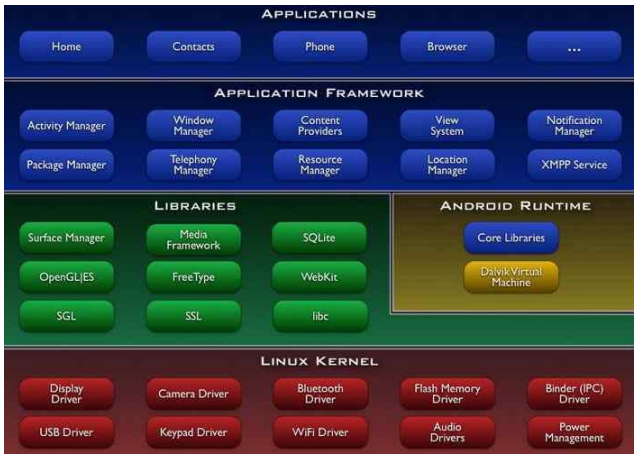

Figure 1: Android Architecture[3]

Security Issues in Android Platform

As android has become most popular mobile operating system. Now android is best or most used operating system though it has many security issues, which may lead to vulnerability to malware. Malware stealing useful data from user. The malware is mainly coming from applications. The malware developers can steal valuable information such as banking information, private SMS [3]. The applications are not only downloaded from Google play store. They also download from third party app stores [4]. In this app stores they are not verified as apple store manually [2]. Some malware access with user location access without permission of user. As android security is based on permission based it controls all the permission in the system. The permission checks API and intent, and then allows android system. Some of threats of android are phishing pages, Trojans, spyware, bots, root explorer, fraud, premium dialer fake installer [4]. These are some of the malware attacks commonly seen in android platform.

Effect of Malware on Devices

These are some ways that malware is effecting android devices.

We are facing serious technical challenges from existing malware. As android is open source there is access to official app store and third party stores [1]. First the application is downloaded and contacted to verification. No application is downloaded without approval of Google services. Burt in third party app stores they are downloaded and installed without any verification. The app asks application for permissions user usually accepts permission with seeing details. The applications contain malware and affect the device in many ways. The malware access sensitive information of user without his permission. The malware mainly slows the performance of device. It utilizes the device memory without the permission of kernel. The malware is growing day by day and signature based detection has become useless for malware detection.

Security Features in android

Following are some of the security features of android.

As android has become most popular mobile operating system. But due to the lack of security features it has become vulnerable to many malware. Many application markets browse the applications and presents it. Thus to prevent illegal data mining they created authentication by giving access to goggle account [1]. And also after we download application we see application asking for permission to install. They usually

base on certificates, which contain signature-based verification before app is installed. Depending on antivirus vendors for malware detection.

Common Attacks in Android Platform

As per all papers mainly all attacks target on user data, files, banking details, text messages and phone calls related to malware.

There are different types [7] of malware most them focus on premium SMS and financial loss. Malware mainly focus on software of android device.

According to papers [2,7] malware can be divided to Trojan, virus, worm, adware, and spyware.

- Trojan: If a device is infected with Trojan the hacker can control Trojan remotely and receive user data secretly.
- Virus: If a device in infected with virus files in the device will be duplicated and cannot be used again. This may also slow device.
- Worm: It replicates the user files and also moves to other devices and do the same processes.
- Spyware: It is software installed on device without the permission of user. It helps the attacker to remotely spy on the device and receive user data such as text message and hear user calls.
- Adware: When a device is infected with adware it pops up ad, which we press, can direct to URLS or subscrir to offers without notification to user.

In articles [2,7] some of the common attacks used by hackers are

- Sniffing: In this hacker searches for Bluetooth enabled devices and try to infect them with harmful malware.
- Spam: Span is unwanted email that come without subscription and direct to some phishing page.
- Phishing: This attack is mostly seen in social media in which attacker provide a link which exactly looks like authentic website but extracts information from user.
- Premium SMS: In this attack the device will subscribe to high cost SMS that user unwillingly subscribes. This attack is most seen in android devices.

Results and Methods for malware detection

The papers we have selected some methods for malware detection they are as follows.

Forensic analysis it is a process in which we consider some set of applications and analyze the data. Then we will find some dataset and some pattern by which the attackers commonly attack [1]. Machine learning can also be used in malware detection [2,3,4]. We use clustering techniques in machine learning. In this we use the clustering gathers all the data and store in database. We abstract useful information from application for malware detection. We use xml files, which contains the permission logs. Which can be useful for malware detection [2].

Machine learning classifiers can also be used in malware detection. We analyze the application system calls and log information; API calls and decides whether application is behaving maliciously or not [2,3]. The advantage of this method is we can access the permissions of applications without dynamic analysis for malware detection. As the android is a open platform the is more risk. We can also use parallel machine learning classifiers to solve this problem. By using malware samples and benign applications with help of machine learning classifiers we can strengthen malware detection in android [10]. Data mining techniques can also be used in android malware detection.

Android malware detection can be permission combination malware detection technique. This android malware detection is based on permission combination that is requested by malware that are present in manifest file. By this method there is 96% of increase in malware detection than compared to begin application recognition rate up to 88%[6]. Another method described is online scheme for detecting malware. As malware in android is increasing such as fishing websites, premium SMS and other. They introduced a method for malware detection by analyzing the UID based online detection. It also increases the malware detection rate [7].

Android malware has not only become a threat to end-user. It has become a threat to network operator. So, they proposed a method to detect malicious behavior in android devices. We device an network traffic monitoring technique. In this method every packet that sent and received are observed for malware behavior [8]. An other method is also mentioned in selected paper that is observing the data flow in the source code and detect the malicious behavior of the application [9]. The methods and results are in following table.

Reference articles	Problem description	Methods used	Result
[1]	Android malware detection with digital forensic	Digital forensic and datasets	Positive
[2]	Malware detection using Clustering techniques	Clustering techniques	Positive
[3]	Malware detection using Machine learning classifiers	Machine learning classifiers	Positive
[4]	Malware detection using Permissions and API calls	Machine learning using Permissions and API calls	Positive
[5]	Android malware detection Markov logic network	Markov logic network Using API	Positive
[6]	Malware detection of mobile phones	Malware detection permissions	Positive
[7]	Online malware detection	Online scheme	Positive
[8]	Malware detection using Network traffic monitoring	Network monitoring system architecture	Positive
[9]	Malware in source code	Static analysis and data flow	Positive
[10]	Malware detection using Parallel machine	Static analysis and parallel classifiers	Positive
[11]	Malware detection using Analyzing and detection	Call flow graph	Positive
[12]	Malware analyzing semantic Behavior	Semantic analysis	Positive
[13]	Malware detection by Sensitivity analysis	Static analysis and sensitivity Analysis	Positive
[14]	Static detection of android Malware	Permissions and API calls	Positive
[15]	Malware detection using Manifest analysis	Manilyzer	Positive

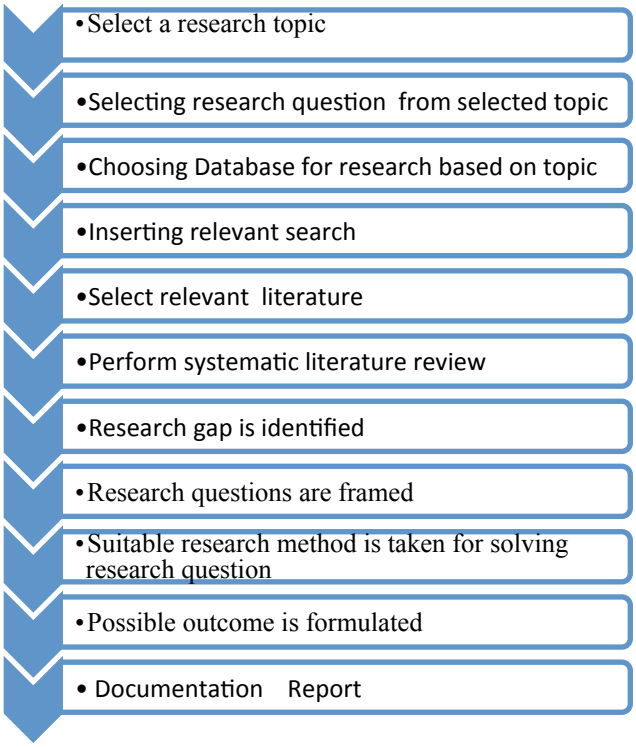

Figure2. Research Process

QUALITY ASSESSMENT CRITERIA

The research papers that chosen for the can be graded on the scale of (1-5) where 1 is poor and 5 is being excellent.

- Q1. How clearly the malware detection was proposed in the paper?
- Q2. How did the author succeed in malware detection?

RESEARCH QUESTION	Q1	Q2
[1]	5	3
[2]	4	3
[3]	3	2
[4]	4	4
[5]	5	3
[6]	3	4
[7]	5	3
[8]	3	3
[9]	5	3
[10]	5	3
[11]	2	1
[12]	4	4
[13]	3	3
[14]	3	3
[15]	5	4

Table2. Assessment of papers

RESEARCH PROPOSAL

Motivation & Objectives

As android is one fast growing operating system. Due to which many malware users are attacking by introducing malware applications. There are some methods for detecting android malware. As it is open source project and also supports third party software. There is a high risk of malware application development. So it is essential to employ malware detection methods .The objective this survey is to know what methods are useful for malware detection and what methods are more efficient. And to find defects in other malware analysis method. The research questions are mentioned below.

Aim and Background

Android is one the most popular mobile operating system. The malware detection is important because of valuable user data. Traditional malware is detected by signature-based analysis. In this we compare the application files with the already list malware. This method helps only a little because it only identifies already recognized malware only. The all applications that are uploaded to the Google play undergo scan under a Bouncer. And another main type detection is behavior based detection.[10]"The behavior based detection method is a detection method through machine learning classifiers for attack pattern during the performance of malware and

process behavior in a normal situation that occurs in devices". Application package file is a jar file mainly consists of manifest file, source code and other resources. There contains permissions that allow accessing android operating system. As android is an open source project. We can give root access to the device and install third party application and software. The malware attacks most common are repacking attack. In which malware developer unpacks the application and repacks it with malware. Metamorphism is a new method by which we can bypass the new anti-malware software. To achieve this they insert dummy code, API and code manipulation.

Research Questions

These are research question are found by seeing research gap in literature review

Q1: Does the developer think that his data is secure in android smart phone?

Q2: Does developer know the malware threats of malware?

Q3: What methods used for malware detection in android?

Q4: Which methods do you think is more efficient in malware detection?

Motivation

The Q1 is about what developers think that about data is security of their valuable information in android operating system.

The Q2 is about how extent does the developer know about the threats about the vulnerability of android.

The Q3 is about what method are used that can be used for detection of malware in android devices.

The Q4 is about what method do the developer think is more efficient for malware detection.

Research method

The above research question can be solved by conducting a research. The research methods available are case study, experiment method, post-mortem, survey.

- We did not select Case Study because the research questions that we framed are based on developer's point of view. It is mainly based on developer's opinion. So we cannot conduct case study.
- Experiment is not selected because authors are not proposing any new method for implementation.
- Post-mortem is not selected because we are only dealing with one subject area and it may prejudge the results of research questions.
- Survey is selected because the question that we proposed are only suitable for multiple opinion of experts.

Sample selection

In this sampling we select a group of whom you want to send your survey. As we are selecting android malware detection which is a much more technical aspect that everyone cannot answer the questions. So we select a technical professionals as software developers, analysts, testers, UI developers And also the student with technical skills. And we also want to target small scale companies, large scale companies that have idea over the android. We should take care of the results that are obtained.

DATA COLLECTION METHOD

Data collection methods can be done in many ways online questions, interviews and many. Among them we are selecting online survey. We create questioners use Google for creation. We send the links to people through mail. The questioners consist of both open-ended questions and closed ended questions. Open ended questions are something that participant can answer to a direct question. In closed ended questions we give direct answers to questioners. The question will be posted in universities and companies. Answers and data are collected. We have to mail to the professional and keep them reminders so that we can collect both quantitative and qualitative data. We should collect data and organize them in a order.

- Questionnaire
- Interviews

For gathering information there are many ways, we choose to conduct a survey as a data collection method. By developing a questioner, which can be distributed among the android developers. We choose android developers because they work on the android platform very closely. According to the article "The information is directly obtained from the people" [16]. As we are given a limited amount of time surveys can give us good hand in data collection process. As we have decided to distribute the questioner among the android developers we can get good amount of results, which will help us to answer our research questions.

Our survey includes the following procedural methods:

- The list of respondents is prepared.
- Then the questionnaire form is developed.
- The questions prepared for the research questions were once again rechecked.
- The questions are forwarded to the respondents using the web service.
- The data is extracted and then analyzed once the responses were all collected.

Risk management & Limitations

Deceptive Data:

The research questions from the feedback. As we choose the survey method to gather the information, sometimes these surveys may mislead to the false data as many of them my not give the proper feedback. In order to overcome the problem we have to exclude all the irrelevant data.

Inelegant questions

As this is our first survey to conduct in the android platform, we may not frame the questions in a proper manner and the developer may confuse while giving the feedback. This risk can be handled by conducting a pilot survey and also by reading and implementing the survey guide lines [17].

Improper Responses

There is a high chance of no response from the developers in the given time since many of them are busy in developing the system and some of them are not interested in giving feedbacks.

TIME AND ACTIVITY PLAN

Days	Activity plan
2015-20-05	Submit research proposal
10 days	Prepare questions to answered research questions
15 days	Recheck the prepared questions
20 days	Conduct survey
20 days	Gather and analyze data received
25 days	Document the survey report and submit

Table3. Time and Activity Plan

Threats and validity

Researcher partiality: The selected research may contain researcher partiality.

Subject bias: Subject partiality is one of the threat as the responses.

Conclusion

We consider to determine the malware threat to android devices through this paper. Malware has become a major threat in android market. As malware is increasing day by day. There is need of sophisticated techniques for malware detection. In this paper authors conducted literature review which may help in solving problem in future.

REFERENCES

[1] K. Allix and Q. Jerome, "A Forensic Analysis of Android Malware How is Malware Written and How it Could be Detected ?," *pp. 384–393, 2014*

[2] a a a Samra, Y. Kangbin, and O. a Ghanem, "Analysis of Clustering Technique in Android Malware Detection," *2013 Seventh Int. Conf.*

Innov. Mob. Internet Serv. Ubiquitous Comput., pp. 729–733, 2013.

[3] Hyo-Sik Ham and Mi-Jung Choi, "Analysis of Android malware detection performance using machine learning classifiers," *2013 Int. Conf. ICT Converg., pp. 490–495, 2013.*

[4] N. Peiravian and X. Zhu, "Machine learning for Android malware detection using permission and API calls," *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI, pp. 300–305, 2013.*

[5] M. Rahman, "DroidMLN: A markov logic network approach to detect android malware," *Proc. - 2013 12th Int. Conf. Mach. Learn. Appl. ICMLA 2013, vol. 2, pp. 166–169, 2013*.

[6] S. Liang and X. Du, "Permission-Combination-based Scheme for Android Mobile Malware Detection," *2014.*

[7] S. Liang, X. Du, and C. C. Tan, "An Effective Online Scheme for Detecting Android Malware," *2014.*

[8] J. Li, L. Zhai, X. Zhang, and D. Quan, "Research of android malware detection based on network traffic monitoring," *Ind. Electron. ..., no. 2011, pp. 1739–1744, 2014.*

[9] C.-M. Chen, J.-M. Lin, and G.-H. Lai, "Detecting Mobile Application Malicious Behaviors Based on Data Flow of Source Code," *2014 Int. Conf. Trust. Syst. their Appl., pp. 1–6, 2014*.

[10] S. Y. Yerima, S. Sezer, and I. Muttik, "Android Malware Detection Using Parallel Machine Learzning Classifiers," *2014 Eighth Int. Conf. Next Gener. Mob. Apps, Serv. Technol., pp. 37–42, 2014.*

[11] W. Park, K. Lee, K. Cho, and W. Ryu, "Analyzing and Detecting method of Android Malware via Disassembling and Visualization," *pp. 817–818, 2014.*

[12] J. Kwon, J. Jeong, J. Lee, and H. Lee, "DroidGraph : Discovering Android Malware by Analyzing Semantic Behavior," *vol. 1, pp. 498–499, 2014.*

[13] S. H. Moghaddam and M. Abbaspour, "Sensitivity Analysis of Static Features for Android Malware Detection," *no. Icee, pp. 920–924, 2014.*

[14] P. P. K. Chan and W. Song, "S TATIC DE TECTION OF ANDROID MALWARE BY USING PERMISSIONS AND API CALLS," *pp. 13–16, 2014*.

[15] S. Feldman, D. Stadther, and B. Wang, "Manilyzer: Automated Android Malware Detection through Manifest Analysis," *2014 IEEE 11th Int. Conf. Mob. Ad Hoc Sens. Syst., pp. 767–772, 2014.*

[16] Survey Procedures, "Quick Tips," Development, *pp. 10–11, 2002.*

[17] E. D. De Leeuw, J. J. Hox, and D. A. Dillman, International Handbook of Survey Methodology, *vol. 21. 2008*.