A Comprehensive Study for Mobile Android Malware Detection

Michael Yoseph Ricky and Robin Solala Gulo

***Abstract*—**Android Malware is a threat that needs to be considered by the Android OS user because it can interfere with the performance of the smartphone being used. Many applications that are scattered on the Google Store is a point of weakness for the user to determine which applications containing malware or not. This comprehesive study can explain about many applications that are scattered on the Google Store is a point of weakness for the user to determine which applications containing malware or not using several methods and display the comparison results between all previous study in the review of Android Mobile Malware. Till now there the optimal method suggested solution that could be considered as standard mechanism for Android Malware Detection is Bayesian classification algorithm based on the methods of previous research used in the detection of Android Malware. For this reason we intended to make our work as a survey paper to make it easier for Android user to determine installed application contains malicious program or not.

***Index Terms*—**Android, Malware Detection, Mobile Malware, Mobile Security.

I. Introduction

Mobile devices that exist in the world today which has become a popular as a smartphone term, a device that "smart" with a variety of operating systems. Operating systems that currently exist in the mobile devices are iOS, Android, Blackberry and others. Android is leading rapidly in its development, which until now being developed by Google and other developers since its open source operating system. The development of the Android operating system very rapidly starting from beta version (released in November 2007) and then popping the next version that is Cupcake (1.5), Donut (1.6), Eclair (2.0–2.1), Froyo (2.2–2.2.3), Gingerbread (2.3–2.3.7), Honeycomb (3.0–3.2.6), Ice Cream Sandwich (4.0–4.0.4), Jelly Bean (4.1–4.3.1), KitKat (4.4–4.4.4, 4.4W–4.4W.2), Lollipop (5.0–5.1.1) [1].

Play Store facilitate users as the online market where users can download a variety of Android OS application provided either free or paid. Android has a built-in feature that will check each application to be installed, either allowed to install applications from the Play Store alone or can install other applications developed by developers using the application apk file. Beginner user are advised to install only applications that have been authorized by Google.

In this paper will be discussed about a system that can protect android user from malicious application. The objective of this research is to present comparison previous research based on android mobile malware using K-means clustering, feature selection and profiling user-trigger dependence.

Manuscript received May 18, 2015. This work was supported in part by Doctorate of Computer Science Program Bina Nusantara University.

Ricky, Michael Yoseph. Author is with School of Computer Science Bina Nusantara University, Jl KH Syahdan no 9 Jakarta Barat (e-mail: mricky@binus.edu).

Gulo, Robin Solala. Author was with Bina Nusantara University, student of Master of Information Technology (e-mail: robin.gulo@binus.ac.id).

II. Malware

A. Vulnerability

Vulnerability is a key-hole weakness that allowed unauthorized user to have access in target device that caused either by error application setting or unintended left by administrator [2].

B. Malware Types

1) Adware: a malware which its objective is to advertise a non-harmful products or sites.
2) Moualbad: a malware that perform automatic calling secretly unnoticeable by user. This malware performs until device screen turned off and locked.
3) Botnet: a malware that controlled by attacker.

C. Attack Processes

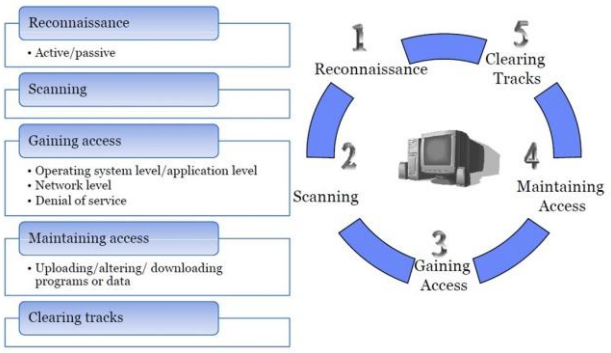

Fig. 1. Attack Processes.

Reconnaissance refers to the preparatory phase where an attacker seeks to gather as much information as possible about a target of evaluation prior to launching an attack. Active reconnaissance involves interacting with the target directly by any means. Passive reconnaissance involves acquiring information without directly interacting with the target.

Scanning refers to the pre-attack phase when the hacker scans the network for specific information on the basis of information gathered during reconnaissance. Scanning can include use of dialers, port scanners, network mapping, sweeping, vulnerability scanners, etc.

Gaining access refers to the penetration phase. The attacker exploits the vulnerability in the system. The exploit can occur over a LAN, the Internet, or as a deception, or theft. Examples include buffer overflows, denial of service, session hijacking, and password cracking.

Maintaining access refers to the phase when the attacker tries to retain his/her ownership of the system. Attackers may

prevent the system from being owned by other attackers by securing their exclusive access with Backdoors, RootKits, or Trojans. Attackers can upload, download, or manipulate data, applications, and configurations on the owned system.

Covering tracks refer to the activities carried out by an attacker to hide his misdeeds. Reasons include the need for prolonged stay, continued use of resources, removing evidence of hacking, or avoiding legal action.

III. RELATED WORK

Android OS is vulnerable for Malware attack, therefore a lot of effort in Android security that perform earlier detection. An Android Antivirus has been developed to solve this problem. The previous research of Android Malware detection can be implemented to expand user knowledge which android application can be trusted installed on their devices and to prevent unintended malware installed in devices.

Almin, Shaikh Busra et al. [3] proposed a system will help Android user either to make decision to proceed the ongoing installation of applications or to remove application containing malicious application that has been installed previously. In the proposed system, k-means clustering algorithm used when performing authorization checking process to determine that application contain malicious program or not. The results will used naïve Bayesian classification algorithm to check accurately either the application contains a malicious program or not.

The process steps are described below:

1. Identify the applications which already installed
2. Retrieve application information (application name, package, version, and other data) that is significantly important for the user by using permission extraction.
3. Clustering of permissions will classify the permissions of each application using k-Means algorithm and splits it into two clusters are malicious cluster and safe cluster
4. The cluster will be checked for its accuracy malicious using Naive Bayesian algorithm to ensure truly malicious applications or secure.
5. The last stage to provide recommendations to the user either to remove the malicious application or to ignore the malicious application.

The result of their experiment are the proposed system called Android Application Analyzer will perform an analysis of the applications that are installed whether harmful or not that need to be uninstalled. Proposed system has been tested on Android OS Jelly Bean version 4.4.2 using malicious application DogWar, iCalendar, and SuperSolo.

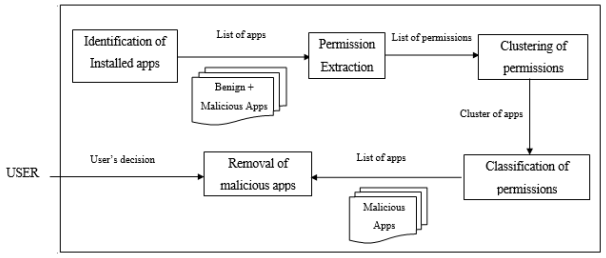

Fig. 2 Android Application Analyzer Model

Experiment is done with existing anti-virus in the Android OS that already exist today and compared with proposed system that has been successfully detects experiment application as malware with the following results:

Sr.No	Malwares	360, Avast, McAfee	Lookout	Karspersky	Android Application Analyzer
1	SuperSolo	✓	✓	✓	✓
2	Dogwar	✓	x	✓	✓
3	ICalendar	✓	✓	x	✓

Fig. 3 Android Application Analyzer Result Testing

Android Application Analyzer has been able to detect malicious application on Android OS and provides recommendations to remove them. The system can also grouping either an application is malicious application or not, even better than the anti-virus that already exists today in the Android OS.

Takayuki et al. conducted research on a system that perform risk checking level of an Android application during the initial installation [4].

The proposed system will summarize application information on google play (the number of downloads, user ratings, user reviews). Then analyzed the combination of permissions malicious application to be installed, so that the user can decide either to continue the installation or terminate installation based on information on the risk level indicator.

Suleiman Y. et al. conducted research and find effective approaches in detecting malicious application based on Bayesian classification model that identify codes and characteristics of application, checking for suspicious activity such as taking sensitive information from the phone, run a malicious embedded at runtime or in an external folder or other places [5]. Running code used Java-based Android package profiling tool to automatically reverse engineering process of APK file checked.

Agematsu, H et al. create a security manager module to define a knowledge database that can determine the malicious behavior application in Android, event notification module that will provide running activities notification to the security manager. If security manager finds the installed applications do not have a notification codes, then remove all applications that do not have event notification code [6].

Ghorbanian, M. et al. conduct research intrusion detection model based on Log Files derived from command logcat via the Android logging system, then the data was analyzed based on pattern matching to detect whether there is malicious rule sets used in an Android application [7].

Dong-Jie Wu et al. conduct research using DroidMat application that can provide statistical data to detect Android malware using the K-means algorithm and kNN algorithm to classify the application if it is safe or contains malicious [8].

Wei Tang et al. conduct research on the Security Distance models in mitigating malware. The model is looking for differences between permissions and combination of permissions with a security problem. The higher the value of threat point indicated an application has a large security threat, otherwise declared as safe from malicious apps [9].

Hamandi, K et al. conduct research on Malware SMS and proposed an application that notified user whether the received SMS contains malicious program or not and advised user to specify the settings to grant or block the sender of the received SMS [10].

H. Thanh conduct research about Android Malware variation and methods for analyzing them based on data that has been collected by using tools such as smali, blacksmali and others to do the reverse engineering process of a apk file

and displays the results of its detection and suggestions to help user recognizes malware [11].

Heloise et al. conduct research about trends and characteristics of Android botnet which helps enable Android users to identify botnet activity as Android bot. As a result, today trend are SMS Trojans, root exploits, and receives commands from the remote server bot [12].

Enck et al. conduct research to create lightweight certification used as security rules to match the malicious properties of Malware. The system will block all applications that have possible risky permissions combination or containing suspicious action. Method used is machine learning that will recognize Malware in the Android OS [13].

Feizollah, Ali et al. review 100 papers published from 2010 to 2014 with the perspective of feature selection in the detection of malware on mobile devices [14]. To develop an effective detection system, in this research will select only a promising subset of the features from hundreds of features available. In this research that categorize different features into 4 groups, namely static features, dynamic features, hybrid features, and application metadata. In addition, in this paper discussed about datasets used in the previous study and how to measure evaluation.

The methods used are feature selection such as android permissions, java code, certification, the behavior of the application on the device, etc. These features selection is the first step and a very important, and checking each group features category in detection of malware on mobile devices (static, dynamic, hybrid, and application metadata).

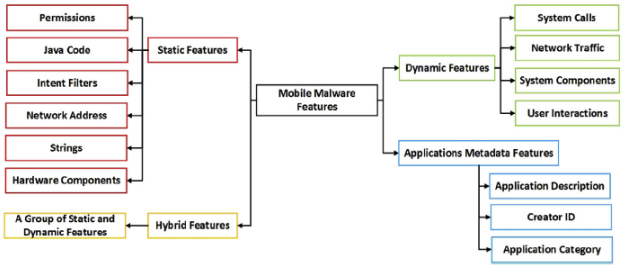

Fig. 3. Taxonomy of Mobile Malware Features

Feature selection will conducted in two ways: selection based on rationalizing and selection based on feature ranking algorithms. In this study authors only classify category of malware on mobile devices as a result of the above. The previous studies is lack of tested samples of malware for mobile devices.

Elish, Karim et al. conduct research to classify accurately an android application whether harmful or not [15]. The methods used are extraction of data-flow feature when a user trigger sensitive API, known as user-trigger dependence, and classification is done by the data dependence graph, trigger metric tuple per operation (trigger metric feature captures static dependence relationship with the user input/action and sensitive operations that provide critical system functions in the program), aggregated metrics and program analysis for feature extraction.

Based on experiment conducted, the evaluation of the 1433 malware application and 2684 free popular application showed that the level of classification accuracy is 2.1% false negative rate and false positive rate of 2.0%. This method also discovered a new malicious applications in Google play that cannot be detected by antivirus software. The method used in this research and from previous studies is the use of benign properties in the program. These benign properties are observed on a trusted program is not in a dangerous application.

IV. RESULTS AND DISCUSSIONS

In this section, we make a comparison between all papers in the review of Android Mobile Malware.

TABLE I: ANDROID MALWARE COMPARISON

Paper	Method	Detection	Result
A Novel Approach to Detect Android Malware	k-means clustering algorithm & naïve Bayesian classification algorithm	100% detected malicious application DogWar, iCalendar, and SuperSolo	Android Application Analyzer has been able to detect malicious application on Android OS and provides recommendations to remove them.
A Proposal of Security Advisory System at the time of installation of Applications on android OS	unknown	has been analyze the combination of permissions malicious application to be installed, so that the user can decide either to continue the installation or terminate installation based on information on the risk level indicator	user can decide either to continue the installation or terminate installation based on information on the risk level indicator
A new Anroid Malware Detection Approach using Bayesian Classification	Bayesian classification model	has been identify codes and characteristics of application, checking for suspicious activity	system will run a code to automatically reverse engineering of APK file
A proposal to Realize the Provision of Secure Android Applications	unknown	define a knowledge database that can determine the malicious behavior application in Android	succesfully remove all applications that do not have event notification code

Signature-based Hybrid Intrusion detection System (HIDS) for Android Devices	pattern matching	intrusion detection model based on Log Files	succesfully detect whether there is malicious rule sets used in an Android application
Droidmat: Android Malware Detection through manifest and API calls Tracing	K-means algorithm and kNN algorithm	provide statistical data to detect Android malware	succesfully classify the application if it is safe or contains malicious
Extending android Security Enforcement with a security Distance Model	unknown	looking for differences between permissions and combinatio n of permissions with a security problem	higher the value of threat point indicated an application has a large security threat
Advanced Information networking and Applications Workshops	unknown	notified user whether the received SMS contains malicious program or not	successfull y advised user to specify the settings to grant or block the sender of the received SMS
Analysis of Malware Families on Android mobiles: Detection Characteristics Recognizable by Ordinary Phone Users and How to fix it	unknown	analyzing Android Malware based on data that has been collected by using tools	displays the results of its detection and suggestions to help user recognizes malware
Android botnets on the rise: Trends and Characteristics	unknown	enable Android users to identify botnet activity as Android bot	today trend are SMS Trojans, root exploits, and receives commands from the remote server bot
On lightweight mobile phone application certification	machine learning	create lightweight certificatio n used as security rules to match the malicious properties of Malware	system will block all application s that have possible risky permissions combinatio n or containing suspicious action
A review on feature selection in mobile malware detection	feature selection	classify category of malware on mobile devices	effective detection system

V. CONCLUSION

As we make a comprehensive study in this paper, we noted that Android Malware detection can be applied in android OS to detect an application is harmful or not. Thus the overall methods used are pattern matching, K-means algorithm and kNN algorithm, naïve Bayesian classification algorithm, machine learning, and feature selection.

Till now there the optimal method suggested solution that could be considered as standard mechanism for Android Malware Detection is Bayesian classification algorithm based on the methods of previous research used in the detection of Android Malware. For this reason and others we intended to make our work as a survey paper to make it easier for android user to determine installed application contains malicious program and decide to remove the application or ignore the results. Future works will be conducted in the implementation of android detection system in android market (Google Play), hence user will be provided trusted applications.

REFERENCES

[1] Android Developer. http://developer.android.com/reference/packages.html retrieved at May 1, 2015

[2] Raiyn, Jamal. "A Survey of Cyber Attack Detection Strategies", *International Journal of Security and Its Application*, vol. 8, no. 1, pp. 247-256, 2014. http://dx.doi.org/10.14257/ijsia.2014.8.1.23

[3] Almin, Shaikh Bushra., Chatterjee, Madhumita. "A Novel Approach to Detect Android Malware", *International Conference on Advanced Computing Technologies and Applications, Procedia Comptuer Science 45, pp. 4017-417, 2015.*

[4] Takayuki Matsudo, Eiichiro Kodama, Jiahong Wang, and Toyoo Takata. "A Proposal of Security Advisory System at the time of installation of Applications on android OS", *15th IEEE International Conference on network-based Information System (NBiS)*, 2012.

[5] Yerima, S.Y., Sexer,S., McWilliams, G. "A new Anroid Malware Detection Approach using Bayesian Classification", *Advanced Information networking and Applications (AINA), 2013, IEEE 27th International Conference*, Vol, no, pp 121-128, 25-28 March 2013.

[6] Agematsu, H., Kani, J., Nasaka, K., Kawabata, H., Isohara, T., Takemori, K., Nishigaki, M. "A proposal to Realize the Provision of Secure Android Applications--ADMS: An Application Development and Management System: Innovative Mobile and Internet Service in Ubiquitous Computing (IMIS), *2012 Sixth International Conference*, vol, no, pp.677,682,4-6 July 2012.

[7] Ghorbanian, M., Shanmugam, B., Narayansamy, G., Idris, N.B. "Signature-based Hybrid Intrusion detection System (HIDS) for Android Devices," *Business Engineering and Industrial Applications Colloquium (BEIAC), 2013 IEEE*, vol,no,pp.827,831,7-9 April 2013. http://dx.doi.org/10.1109/beiac.2013.6560251

[8] Dong-Jie Wu, Ching-Hao Mao, te-En Wei, Hahn-Ming Lee, Kuo-Ping Wu. "Droidmat: Android Malware Detection through manifest and API calls Tracing. "*Information Security (Asia JCIS), 2012 seventh Asia Joint Conference*, vol, no, pp.62,69, 9-10 Aug 2012.

[9] Wei Tang, Gunag jin, Jiaming He, Xinaliang jiang, "Extending android Security Enforcement with a security Distance Model", Internet Technology and Applications (iTAP), *2011 International Conference*, vol., no., pp.1,4, 16-18 Aug. 2011.

[10] Hamandi K., Chebab, A., Elhajj, I.H., Kayssi, A., "Android SMS Malware: vulnerability and Mitigation, *"Advanced Information networking and Applications Workshops (WAINA), 2013 27th International Conference*, vol,no.,pp.1004,1009, 25-28 March 2013. http://dx.doi.org/10.1109/waina.2013.134

[11] H, Thanh, "Analysis of Malware Families on Android mobiles: Detection Characteristics Recognizable by Ordinary Phone Users and How to fix it.," *Journal of Information Security*, vol 4 no 4, pp.213-224, 2013. http://dx.doi.org/10.4236/jis.2013.44024

[12] Pieterse, H., Olivier, M.S., "Android botnets on the rise: Trends and Characteristics," *Information Security for South Africa* (ISSA), 2012, vol.,no.,pp.1,5, 15-17 Aug. 2012

[13] Enck, W., Ongtang, M., McDaniel, P," On lightweight mobile phone application certification." *Proceeding of the 16th ACM conference on Computer and Communication Security*, New York, NY, USA, ACM(2009) 235-245. http://dx.doi.org/10.1145/1653662.1653691

[14] Feizollah, Ali, Nor Badrul Anuar, Rosli Salleh, Ainuddin Wahid Abdul Wahab, "A review on feature selection in mobile malware detection", *Digital Investigation 13*,pp. 22-37, 2015. http://dx.doi.org/10.1016/j.diin.2015.02.001

[15] Elish, Karim O., Xiaokui Shu, Danfeng (Daphne) Yao, Barbara G. Ryder a, Xuxian Jiang. "Profiling user-trigger dependence for Android Malware detection". computers and Security 49, pp. 255-273, 2015.

Michael Yoseph Ricky is a student of Doctorate of Computer Science Program in Bina Nusantara University. His area covers artificial intelligence, game technology, personal agent and computer science.

Robin Solala Gulo graduated in computer engineering and now pursuing his master degree in computer science.