|| Volume 3 || Issue 5 || May 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Malware Detection on Android Smartphone's using Keywords Vectors and SVM

Vishakha Chilpipre[1], Prof. Varsha R. Dange[2]
Student, Department of Computer Science Dhole Patil College of Engineering, Pune, Maharashtra, India[1]
Assistant Professor, Department of Computer Science Dhole Patil College of Engineering, Pune, Maharashtra, India[2]

ABSTRACT: With the development of internet era Smartphone's have become more popular all around the world. Android is the most popular mobile operating system. With the increasing use of Smartphone's the number of malwares attacking these Smartphone's have also been increased. To effectively detect the new malwares and malicious software variants has been a difficult problem. Our method uses the Keywords Correlation Distance to compute the correlation between key codes such as API calls, Android permissions, the common parameters, and the common keywords in Android malware source code. Then Support Vector Machine is applied to make the system gain to accommodate the function of the new malicious software sample, so as to detect new malicious software and existing malwares. This method is different from the conventional methods which are based on the context of the text. This method combines the Characteristics Of The Malicious Software Categories And Operating Environment To Record The Behaviour Of The Malicious Software.

Keywords: *Android malware, Machine learning, Keywords Correlation Distance, SVM*

I INTRODUCTION

Malware is an abbreviation for two words malicious and software. Actually, it is software that included in the computer system for malicious purposes, without any knowledge from the computer owner. It may be used to collect important information, or gain access to computer systems. The seriousness of malicious software ranges from hurt the users with annoying Ads to steal important data. With the advent of the Internet era, the smart phones in the world are also getting more and more popular, especially the smart phone with Android operating system with its excellent performance. However, Android malwares have increased significantly in recent years. Most malware detection methods are based on traditional content signatures, such as a list of malware signature definitions, and compare each application against the database of known malware signatures. We propose feature extracted method based on the keywords vector. Every keywords vector is a set of keywords which can common complete a malicious attack. We know only some request may be no harm to users. Harm is often done by a series of malicious operations.

II LITERATURE REVIEW

With the development of smart phones, more and more mobile phone malwares have came out in the market especially on the popular platforms such as Android, which can potentially cause harm to users information. But how to effectively detect the new malwares and malicious software variants has been a difficult problem. In view of the traditional feature extraction method based on binary program, the author Junmei Sun was used a method for feature extraction of JAVA source code. The method uses the Keywords Correlation Distance to compute the correlation between key codes such as API calls, Android permissions, the common parameters, and the common key words in Android malware source code. Then SVM was applied to make the system gain to accommodate the function of the new malicious software sample, so as to detect new malicious software and existing malwares. This method is different from the conventional methods which are based on the context of the text. This method combines the characteristics of the malicious software categories and operating environment to record the behaviour of the malicious software. Experiments showed that the method is efficient and effective in detecting malwares on Android platform [1].

How to filter out malicious app is a serious problem for app markets. So Yongfeng Li(B) and Tong Shen proposed DroidADDMiner, an efficient and precise system to detect, classify and characterize Android malware. DroidADDMiner is a machine learning based system that extracts features based on data dependency between sensitive APIs. It extracts API data dependence paths embedded in app to construct feature vectors for machine learning. While DroidSIFT also attempts automated detection of Android applications according to data flow analysis, DroidADDMiner can not only reduce the run time but also characterize malwares behaviors automatically. They implement DroidADDMiner based on FlowDroid and

evaluate it using 5648 malware samples and 14280 benign apps [2].

Analyzing applications in order to identify malicious ones is a current major concern in information security; an additional problem connected with smart-phone applications is that their many advertising libraries can lead to loss of personal information. Lynn M. Batten, Veelasha Moonsamy and Moutaz Alazab are relate the current methods of detecting malware on smart phone devices and discuss the problems caused by malware as well as advertising [3].

Ahmed H. Mostafa and Marwa M. A. Elfattah, was presented an effective methodology to detect Android malware using static code analysis based models. The presented models are built to extract features relevant to malware based on extracted permissions from AndroidManifest.xml file as well as extracted methods and APIs from disassembled code to be used as features for training machine learning classifiers [4].

With the constantly increasing use of mobile devices, the need for effective malware detection algorithms is constantly growing. The research presented by M. Leeds, M. Keffeler and T. Atkison expands upon previous work that applied machine learning techniques to the area of Android malware detection by examining Java API call data as a method for malware detection. In addition to examining a new feature, a significant amount of work has been done in understanding how the model works and various ways of improving its accuracy. Ultimately a classification accuracy of around 80-85% was achieved using the JAVA API call feature [5].

III SYSTEM ARCHITECTURE

The proposed system uses a feature extraction method based on keywords correlation distance which is different from the traditional method based on binary program. In this method Java code is extracted from apk file and keyword extraction is done also, permissions in android manifest file are checked. Second, we use feature vector to describe malicious software feature including not only APIs, but also the common parameters and common package etc. Third, we give a malware detection method through SVM based on the feature vector set, which can detect new malwares and malicious software variants.

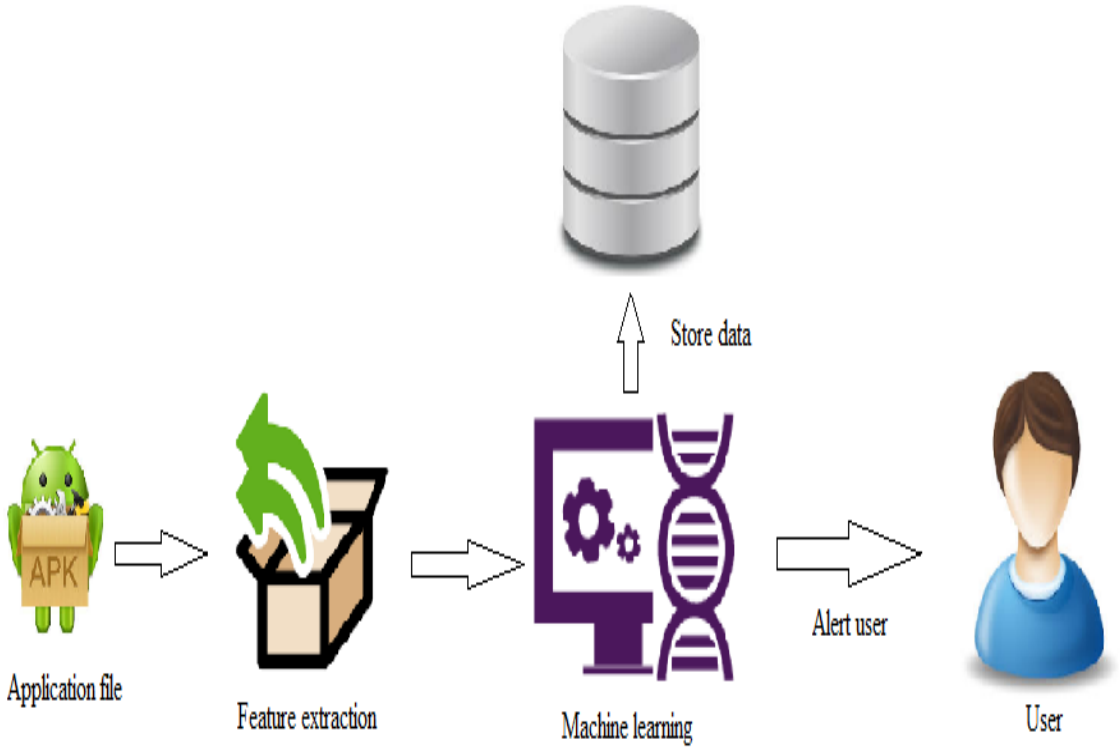

Figure 1: Architecture Diagram

IV MODULES

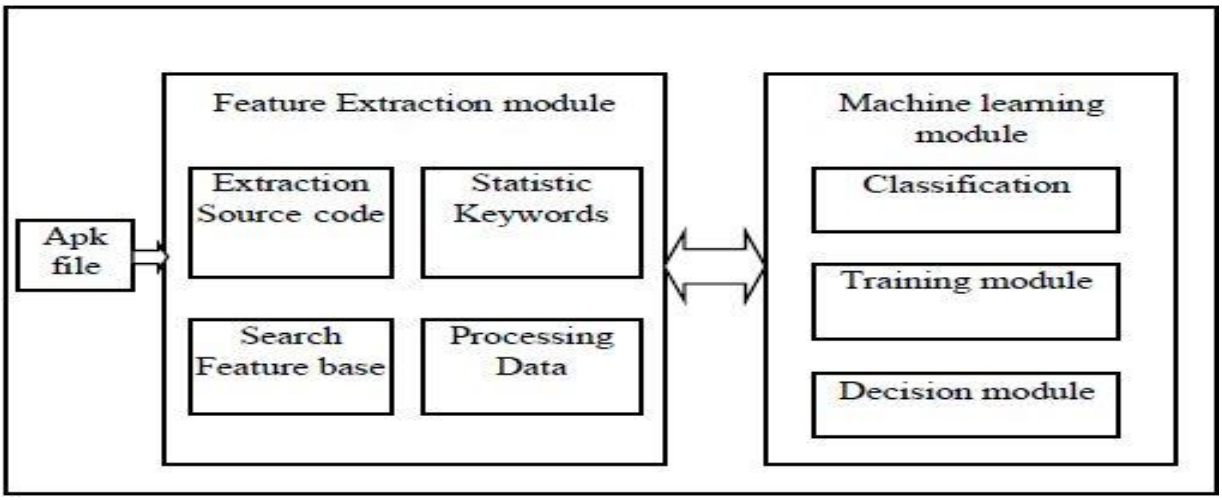

Figure 2 Block Diagram

Proposed system mainly consist of two modules,

1) **Feature Extraction module:** The feature extraction module is responsible for the feature extraction

a) **Decompiling**

In this module we unzip Android application package (APK) to get the Manifest.xml file in the root directory then we use the open source software dex2jar and jadnt158 to decompile the classes.dex in the directory

b) Selection of Keywords

We select five representative keywords set according to the observation of malwares

1. Android Permission

2. Activity Action Intent Parameter
3. Broadcast Intent Action Constant
4. The commonly Package Name
5. API Call

c) Statistic keywords

In this module we record the frequency and location of every keyword in class of APK and in the configuration file, to storage the information using matrix then use Keywords Correlation Distance algorithm calculated the distance between the two keywords.

2) Machine learning module

Machine learning module is responsible for classification and decision making.

a) Classification

We present a classification method based on SVM(Support Vector Machine).SVM is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis

b) Training module

We using LIBSVM is a library for Support Vector Machines to train. The steps of training are shown as follows:

a) Prepare training samples and testing samples.

b) Build java project, import LibSVM jars.

c) Put the training samples and testing samples under the project directory, also you can build a directory by yourself.

c) Decision module

In this module we classify the keyword sets as seven types: NETWORK, PHONE_STATE, SMS, BULETOOTN, SYSTEM_INFO, GPS_LOCATION, WRITE_STORAGE. We give different keyword the different weight.

V ALGORITHM

- **Input:** D data set, on-demand features, aggregation-based features
- **Output:** Classification of Application

1. **for** each application App id in D **do**
2. Get on-demand features and stored on vector x for App id
3. x.add (Get Features(app id));
4. **end for**
5. **for** each application in x vector **do**
6. Fetch first feature and stored in b, and other features in w.
7. hw, b (x) = g (z) here z=(wTx + b)
8. **if**($z \geq 0$)
9. assign g(z)= 1;
10. **else** g(z) = -1;
11. **end if**
12. **end for**

VI RESULT AND CONCLUSION

In this paper we proposed an extraction method of Android malware detection based on KCD. Then we combine the feature into keywords feature vector. Finally, learn and decision by SVM for detecting new malware and malicious variant. This system is different from conventional methods. Experiments show the method is effective and efficient in detecting malwares on Android platforms.

REFERENCES

[1] Junmei Sun, Kai Yan, Xuejiao Liu , Chunlei Yang, Yaoyin Fu, "Malware Detection on Android Smartphones using Keywords Vector and SVM", Hangzhou Institute of Service Engineering Hangzhou Normal University Hangzhou, China $978-1-5090-5507-4/17/31.00 2017 IEEE ICIS 2017, May 24-26, 2017.

[2] Yongfeng Li(B), Tong Shen, Xin Sun, Xuerui Pan, and Bing Mao, "Detection, Classification and Characterization of Android Malware Using API Data Dependency", Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2015.

[3] Lynn M. Batten, Veelasha Moonsamy and Moutaz Alazab, "Smartphone Applications, Malware and Data Theft" Springer Science Business Media Singapore 2016.

[4] Ahmed H. Mostafa, Marwa M. A. Elfattah and Aliaa A. A. Youssif, "An Intelligent Methodology for Malware Detection in Android Smartphones Based Static Analysis", International Journal of Communication 2016.

[5] M. Leeds, M. Keffeler, T. Atkison, "Examining Features for Android Malware Detection", Computer Science Department, University of Alabama, Tuscaloosa, AL, USA Intel Conf. Security and Management SAM'17 ISBN: 1-60132-467-7, 2017.

[6] Saba Arshad, Abid Khan, Munam Ali Shah, Mansoor Ahmed, ''Android Malware Detection & Protection: A Survey", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.

[7] Hossein Hassani, How to do the final year project. https://zodml.org/sites/default/files/, ''Malware Detection on Android Smartphones using Keywords Vectors and SVM", How to do the Final Year Projects - A Practical Guideline for Computer Science and IT Students.pdf, ISBN 978-87-403-0277-6, 2012.

[8] Plagarism checker (SEO Tools): Plagiarism checker is a tool that can search billions of documents, and find matches even if they are only a few words in length, finding plagiarism has become as easy as detecting information in google.

[9] https://en.wikipedia.org/wiki/Support Vector Machine.