Malicious Application Detection and Classification System for Android Mobiles

Sapna Malik, Department of Computer Science Engineering, Maharaja Surajmal Institute of Technology, Delhi, India
Kiran Khattar, School of Engineering and Technology, Ansal University, Gurgaon, India

ABSTRACT

The Android Mobiles constitute a large portion of mobile market which also attracts the malware developer for malicious gains. Every year hundreds of malwares are detected in the Android market. Unofficial and Official Android market such as Google Play Store are infested with fake and malicious apps which is a warning alarm for naive user. Guided by this insight, this paper presents the malicious application detection and classification system using machine learning techniques by extracting and analyzing the Android Permission Feature of the Android applications. For the feature extraction, the authors of this work have developed the AndroData tool written in shell script and analyzed the extracted features of 1060 Android applications with machine learning algorithms. They have achieved the malicious application detection and classification accuracy of 98.2% and 87.3%, respectively with machine learning techniques.

KEYWORDS
Android Mobiles, Android Permissions, Machine Learning Techniques, Malware Detection

INTRODUCTION

The Smartphones are getting immensely popular all over the world. It has become the Personal Remote of Life than just being a medium to communicate owing to the huge functionality it offers. The Smartphones have transformed to everyone’s online bank, online shopping mall, and online tutor along with the traditional voice communication facility. It is a Go- to-Device for various day to day activities such as clicking pictures, watching movies, shopping, calling, chatting and many more. Mobile devices are equipped with advanced user interfaces, processing capability and adequate memory. It holds lots of personal information like contact list, online banking passwords, credit card details and location of the proprietor. According to Expedia Signal Survey (Mishra, 2014), the Indians are becoming addicted to Smartphones for their daily needs. These capabilities are provided to the Smartphones through mobile applications.

Android phones are one of the popular Smartphones nowadays. According to a survey conducted by Nielsen and Informate Mobile Intelligence, 62% of Indians prefer Android mobile phones over other Smartphones (Brindaalakshmi, 2013). The Android applications are downloaded by the user from official markets such as Google Play Store and also from the unofficial markets. The unofficial market is full of malicious applications which lure the customers to download its apps with a heavy discount on products or payback offers. Even the Official Market Google Play Store has not been left untouched by the intruder. 13 infected applications have been removed from the Google Play Store in January 2016 (Acharya, 2016). In the recent report of April 2016 (Bisson, 2016), 100 applications have been found infected on Google Play Store.

DOI: 10.4018/IJACI.2018010106
Intrusion detection and Prevention systems are systems which detect intrusion in the mobile, PC and in the network and prevent the stakeholder from being harmed by these malicious apps. The intrusion in a mobile phone is possible through mobile applications. There are three types of Intrusion detection techniques, Signature based, Behavior-based and Anomaly based. In Anomaly based intrusion detection system, the system detects the change in the behavior of the mobile application in comparison to the normal behavior pattern of benign application.

The anomaly intrusion detection uses two types of behavior pattern analysis techniques, Static analysis, and Dynamic analysis. In Static analysis technique, the static features of Android applications are extracted such as permission requested, method call and API call sequence from the source code without executing the application. In Dynamic analysis technique, the features of Android applications are extracted by executing the application.

This research paper presents the Anomaly-based Intrusion Detection System based on the static feature analysis of Android Permission for detection and Classification of the malicious applications using Machine Learning Techniques.

The contributions made by this work are as followed:

- 533 Benign applications samples and 527 Malicious Android applications samples from 81 malware families have been taken for training the machine learning algorithms for malicious application detection and classification;
- Feature Android Permission Requested of Android Application has been extracted with proposed feature extraction tool AndroData, a tool written in shell scripting language;
- Rigorous analysis of Android Permission request pattern of sample Android applications has been done and dataset has been refined based on the analysis.
- Machine Learning Models have been trained with the dataset and performance of various machine learning algorithms are evaluated for malicious application detection and classification of malicious applications from 81 malware families.

RELATED WORKS

Sarma et al. (2012) proposed Category-based Rare Critical Permission signal, denoted CRCP, a warning signal based on the rules which make the correlation between the permission requested and the category of requesting an application. Sanz et al. (2012) used multiple classification algorithms like K-Nearest Neighbor, Tree Augmented Naive, SVM, J48 and Random Forest for classification of Android applications into several categories like entertainment, tools, games and multimedia and communication, based on the permission features and have maximum ROC 0.93 with TAN. Sato, Chiba and Goto (2013) proposed the method based on the malignancy score of Android application for classification of benign and malicious applications and have the accuracy of 90%. They had used 30 benign and 30 malicious applications for this experiment. Aung and Zaw (2013) experimented the different machine learning algorithms like J48, K-Means, CART over a dataset of 500 Android applications for the malware detection based on the permission features. MAMA (Sanz, Santos, Laorden, Ugarte-Pedrero, Bringas, 2013) a machine learning malware detection system based on permission and used features of the Android application and have the accuracy of 94.83%. Huang, Tsai and Hsu (2013) have experimented with Android Permission and 20 Uses Feature of Android application with different machine learning algorithms and have 81% accuracy with J48. DREBIN (Arp, Spreitzenbarth, Hubner, Gascon & Rieck, 2014) uses six features including Android Permission for detecting malware with SVM and have the accuracy of 94%. Liu and Liu (2014) proposed Two-Layered Permission-Based Android Malware Detection Scheme using machine learning techniques for classification of benign and malicious applications. Kang, Jang, Mohaisen, and Kim (2015) an Android malware detection and classification system based on a serial number of certificate, the likelihood of Permission and the similarity score between the Android applications. They have used
Bi-Objective Supply Chain Optimization With Supplier Selection
[www.igi-global.com/chapter/bi-objective-supply-chain-optimization-with-supplier-selection/208050?camid=4v1a](www.igi-global.com/chapter/bi-objective-supply-chain-optimization-with-supplier-selection/208050?camid=4v1a)

Reassessing the Foundations of Semiotics: Preliminaries
[www.igi-global.com/article/reassessing-foundations-semiotics/64637?camid=4v1a](www.igi-global.com/article/reassessing-foundations-semiotics/64637?camid=4v1a)