INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 1735-1738

0e-ISSN :
2583-1062

Impact
Factor:
5.725

SEDMDROID: AN ENHANCED STACKING ENSEMBLE FRAMEWORK FOR ANDROID MALWARE DETECTION

Saleha landage[1], Premsai kalekar[2], Rachana dudam[3], Preeti shinge[4], Prof. P. P Kalyankar[5]

[1,2,3,4]Student, Dept of Computer Science & Engineeing, BMIT"s, Solapur, India,

[5]Prof. Dept of Computer Science & Engineering, BMIT"s, Solapur, India

DOI: https://www.doi.org/10.58257/IJPREMS33580

ABSTRACT

The Popularity of the Android platform in smartphones and other Internet-of-Things devices has resulted in the explosive of malware attacks against it. To fight against the explosive growth of Android malware, we propose a static malware detection framework, known as SEDMDroid. Malware presents a serious threat to the security of devices and the services they provided, e.g. stealing the privacy sensitive data stored in mobile devices. The main aim of the study is to explore the malware prediction in android. Data pre-processing and model selection is the first two faces. In model selection that, the data is divided into two portions, train set and test set, in a ratio of 80% and 20%, respectively. In the third phase Principle Component Analysis is implemented for feature reduction, in classification the most prominent prediction models are implemented, such as Support Vector Machine and Multi-Layer Perceptron algorithms are used to evaluate how they impacted model accuracy. Finally, the predicted result in the form of AUC/RUC curve was used to analyse the findings obtained on the test set.

1. INTRODUCTION

General Introduction

Remote sensing

Android malware is often deceptive. A mobile app called Ads Blocker, for example, promised to remove pesky ads from your phone, which sometimes pop up to cover your screen just when you're about to access something important. But people quickly found the app was nothing less than malware that served up more ads, according to security researchers.

It's just one example of malware that can frustrate Android phone users, plaguing them with ads that the creators get paid to display, even when they're looking at unrelated apps. Malware often also harvests fake clicks on the ads, doubling up on the value for the makers.

Project Introduction

Mobile devices have become an integral part of daily life. They provide many useful functions such as the ability to read and write e-mails, surf the Internet, indicate nearby facilities, video conferencing, and voice recognition, to name but a few. However, the popularity and adoption of mobile devices also attract malware writers to develop mobile malware in order to harm these devices. According to a Kaspersky security report, 884,774 new malware was introduced in 2015, three times more compared to 2014. Symantec also reported that one zero day was attack per week on average discovered in 2015. Moreover, they emphasized the large increase in the volume of Android variants (40%) besides new Android malware families added in 2015 (6%). Hence, in order to protect mobile devices from such threats, researchers and security companies work to develop effective and efficient anti-malware systems. There are some techniques available for malware analysis and detection with varying strengths and weaknesses. Two common types of malware detection techniques, according to how the code is analysed, are static and dynamic analyses. The past few years have witnessed the drastic increase of mobile apps providing various facilities for personal and business use. The proliferation of mobile apps is due to billions of users who enable developers to earn revenue through advertisements, in-app purchases, etc. A multitude of apps developed by many independent developers, involving unfriendly ones, can be hard for users to determine the trustworthiness of these apps. Whenever users install a new app, they are under the risk of installing malware. Unlike desktop apps, mobile apps can have the privilege, after declared (e.g., in Manifest file of Android platform), to access sensitive information such as contact lists, SMS messages, GPS, etc. To make full use of resources of mobile devices and support abundant functionalities of mobile apps, such mechanism of permission declaration remarkably fulfils its job. Hence, smartphone has become an attractive target for cyber-attacks. Android, an open source Operating System (OS), currently ranks first in the smartphone market with a share of more than 85%. Android platform, allows users to install and run multifarious applications (Apps), which are partially released by many unorganized developers through diverse third-party App markets. Owing to its popularity and openness, Android is the main target for attacking smart phones, accounting for about 98.5%. In fact, the rapid development and popularity of Android system is not only on mobile phones, but also on smart TVs, car navigation systems and intelligent house system. Therefore, Android is also regarded as one of the most promising platforms in the developing ecosystem of the Internet of Things. In such context, it is easy to predict that Android will continue to be an extremely attractive target for attackers to generate and disseminate more powerful and trickier malicious software (malware).

2. LITERATURE REVIEW

The ubiquity of Android devices and the exponential growth of mobile applications have exposed users to an increased risk of malware threats.

This literature review explores existing approaches, disadvantages and proposed approach and advantages in android malware detection with a specific focus on leveraging machine learning for enhanced security.

2.1 EXISTING SYSTEM

To fight against the explosive growth of Android malware, we Propose a static malware detection framework, known as SEDMDroid.

This framework is a two-tier architecture, including the ensemble of base learners MLP and the fusion of base leaner output by SVM. At the first stage, the double disturbance of feature space and sample space ensures the diversity of the training subsets, and PCA is run on these subsets separately. For each branch, keeping all principal components achieved by the PCA and transforming the whole training dataset into a totally new set, MLP is run on this new set, to guarantee the accuracy of the base leaner.

- Evaluates the classical MLAs and deep learning architectures for malware detection, classification, and categorization using differentpublic and private datasets
- Our major contribution is in proposing a novel image

  processing technique with optimal parameters for MLAs and deep learning architectures to arrive at an effective

  zero-day malware detectionmodel.
- Overall, this paper paves way for an effective visual

  detection of malware using a scalable and hybrid deep learning framework forreal-time deployments

Disadvantages

- ➢ Low accuracy.
- ➢ The performance is considerably very low
- ➢ Lower learning rate was found to be good in identifying the executable as either benign or malware

2.2 PROPOSED SYSTEM:-

The proposed model is introduced to overcome all the disadvantages that arise in the existing system. This system will increase the accuracy of the machine learning results by detecting malware from android dataset using machine learning algorithm.

It enhances the performance of the overall classification results. Predict the malware from android data is to find the accuracy more reliable.

- In this system, the malware dataset as input was taken from a dataset repositorylike the UCI repository. Then, we have to implement the data pre-processing step such as checking any missing values to avoid wrong prediction, label encoding is, to encode the data into numeric binary integer values.
- Then, we have to split the dataset into test and train. Test data is used to predictthe model and train data is used to evaluate the model.
- Then, we have to implement the feature selection for selecting the best featuresfrom the split data.
- Then, we have to implement the classification algorithm (i.e.) machine learningsuch as Random forest and KNN. Finally, the experimental results show the performance metrics such as accuracy, precision, and recall

Advantages

- High performance.
- It can handle packed malware, and can work on various malwares irrespective of the operating system.
- ALGORITHMS

In an Android Malware Detection using machine learning, several algorithms

can be employed to analyze and classify applications as either benign or malicious.

Here are some commonly used algorithms for this purpose:

1. Support Vector Machines (SVM):

- SVM is a supervised learning algorithm that can classify applications by finding the hyperplane that best separates the feature space into distinct classes. It is effective in handling high-dimensional data, making it suitable for feature-rich representations of Android apps.

2. multilayer perceptron (MLP)

- A multilayer perceptron (MLP) is a feed forward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses back propogation for training the network.

3. Random Forest:

- Random Forest is an ensemble learning method that builds a multitude of decision trees during training and outputs the class that is the mode of the classes of the individual trees. It is robust and less prone to overfitting, providing a reliable approch for Android malware detection.

4. K-Nearest Neighbors (KNN):

- KNN classifies applications based on the majority class of their neighboring dat points in feature space. It is a non-parametric and instance-based learning algori suitable for identifying similarities in behavior patterns.
- Principle Component Analysis (PCA):
- Principal Component Analysis (PCA) is one of the most commonly used unsupervised machine learning algorithms across a variety of applications: exploratory data analysis, dimensionality reduction, information compression, data de-noising, and plenty more.

Architecture diagram

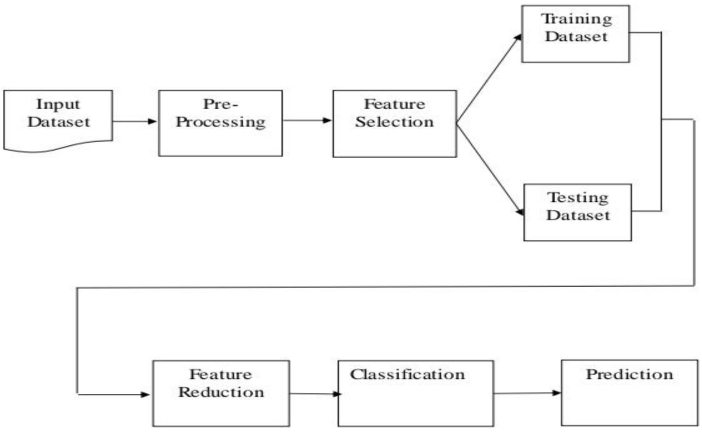
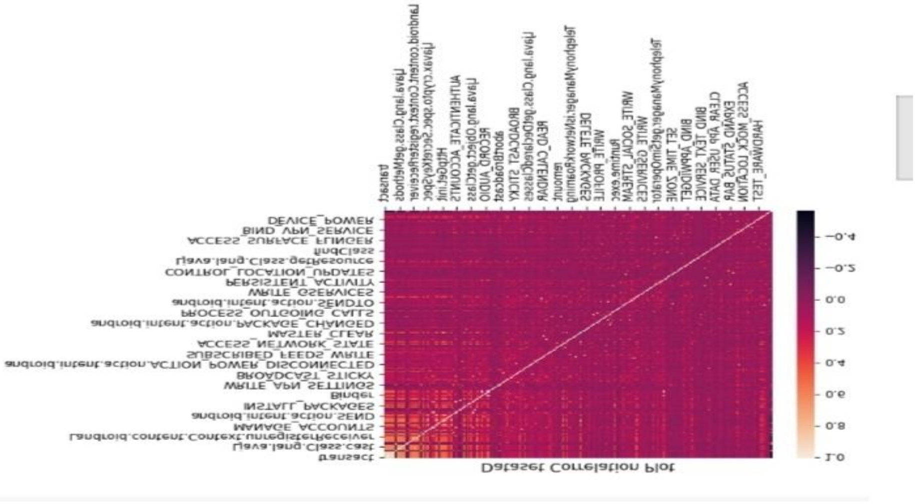


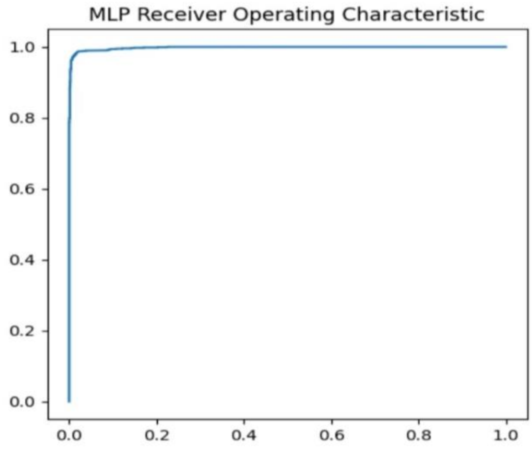
Fig.3.1 Architecture diagram

3. RESULTS

Dataset co-relation plot



MLP Receiver operating characteristic :



4. CONCLUSION

In this study, the machine learning classifiers are predict the android malware. The Android Malware data is taken as input data and applied into pre-processingmethod. In pre-processing method clean the dataset and apply the label encoding. Then it processed into feature selection method, in this method the dataset is splitinto training dataset and testing dataset. After that PCA algorithm is implemented and it will apply the feature reduction. Finally the classification method machinelearning algorithm is used to predict the malware in android and the result basedon accuracy and roc accuracy.

We conclude that, a machine-learning based method for the detection of malwareattacks in the software

The research in the paper adopted an approach based on the random forest and KNN which was classify the attacks effectively.

The experimental results indicate that the proposed approach outperformed the machine learning algorithms and achieved the highest performance in terms of Accuracy, Precision and F1-score.

5. REFERENCES

[1] Y. Mirsky, A. Shabtai, L. Rokach, B. Shapira, and Y. Elovici, "SherLock vs Moriarty: A Smartphone Dataset for Cybersecurity Research."

[2] A. Saracino, D. Sgandurra, G. Dini, and F. Martinelli, “MADAM: Effective and Efficient Behavior-based Android Malware Detection and Prevention,” IEEE Transactions on Dependable & Secure Computing, vol. PP, no. 99, pp.1-1, 2018.

[3] M. Xu, C. Song, Y. Ji, M. W. Shih, K. Lu, C. Zheng, R. Duan, Y. Jang, B. Lee, and C. Qian, “Toward engineering a secure android ecosystem: A surveyof existing techniques,” Acm Computing Surveys, vol. 49, no. 2, pp. 38, 2016.

[4] T. Lei, Z. Qin, Z. Wang, Q. Li, and D. Ye, “EveDroid: Event-Aware AndroidMalware Detection Against Model Degrading for IoT Devices,” IEEE Internet of Things Journal, 2019.

[5] G. Tao, Z. Zheng, Z. Guo, and M. R. Lyu, “MalPat: Mining Patterns of Malicious and Benign Android Apps via Permission-Related APIs,” IEEE Transactions on Reliability, vol. 67, no. 1, pp. 355-369, 2018.

[6] Y. Sun, H. Song, A. J. Jara and R. Bie, "Internet of Things and Big Data Analytics for Smart and Connected Communities," in IEEE Access, vol. 4, pp. 766-773, 2016, doi: 10.1109/ACCESS.2016.2529723.

[7] S. Sen, E. Aydogan, and A. I. Aysan, “Coevolution of Mobile Malware and Anti-Malware,” IEEE Transactions on Information Forensics & Security, vol.13, no. 10, pp. 2563-2574, 2018.

[8] X. Ke, Y. Li, and R. Deng, “ICCDetector: ICC-Based Malware Detection onAndroid,” IEEE Transactions on Information Forensics & Security, vol. 11, no. 6, pp. 1252- 1264, 2017.