An Efficient Detection of Android Ransomware Applications Using Logistic Regression

[1]Merina Y, [2]Carolene Praveen R
Department of Computer Science, SJSV CAS, Coimbatore 641005

***Abstract* – Mobile devices have occupied all characteristics of personal and professional life like as smart watches are used to track our daily actions and the quality of our sleeping, mobile applications are used to make financial transactions; vehicles comprise a significant amount of interrelated computational fundamentals that are used to control their functionality, extending from fuel injection to infotainment. Dangerous infrastructures, such as smart grids and community transport, also make use, at different levels, of mobile and battery-operated devices. In this paper we have analyzed ransomware applications to identify the significant permissions for applied to the logistic regression classifier for the efficient detection.**

Keywords: **Ransomware, Android, logistic regression, accuracy.**

I. INTRODUCTION

Prevalent adoption of smart mobile devices and their augmented usage for personal and business purposes, attracted the consideration of attackers, mainly criminals, and augmented their interest in mistreating these devices in order to gain profit, collect isolated and sensitive data, or disturb users [1]. This reproduces in an increase of malware, by which we reflect any malicious software that improvements access to a device for the purpose of theft data, damaging the device, or annoying the user. Malware is currently one of the most pertinent security problems of mobile devices, since number of come across malicious samples is continually on the rise, and, according to McAfee Labs Threats Report, total mobile malware grew by 79% in the past four quarters to reach 16.8 million samples. Furthermore, as specified in 2016 Trend Micro Security Predictions Micro, 3 out of 5 applications used in China are supposed to contain malware.

The upsurge of malware, calls for a growth in the effectiveness of malware detection systems [2]. Although mobile malware recognition systems are used in surroundings with limited computational resources, in order to be accepted in practice and to help users to protect against malicious infections, they have to achieve a set of requirements. First requirement is high recognition accuracy, without creating too many false positives and without worrying regular usage of the device. Then, a contagion with malware should be identified as early as possible in its execution, in order to minimalize potential damage. Moreover, the overhead on battery and computational resources ingesting due to detection system should be insignificant. Finally, the discovery system should provide information on which of the numerous types of malware are

executing, in order to allow better sympathetic of severity of the occurrence and to propose conceivable countermeasures [3]. Though, with the augmented number of malware samples and a currently present overabundance of malicious behaviors, these are stimulating tasks, that are moreover complicated with the fact that forced resources and battery-operated nature of mobile devices meaningfully limits their aptitude to run complex malware uncovering systems, so in most applied scenarios the trade-offs among the mentioned requirements have to be measured.

II. RELATED WORK

In [4] offered an detection system which detects malware through analyzing AndroidManifest.xml and tracing systems calls. It works on the static analysis of the applications permissions, services, fragments, components, intent messages, and API calls. First it extracts the different features from the apps' Android- Manifest.xml namely, permissions and intention messages. Then, it inscriptions the applications components namely, activity, service, and receiver as initial points to trace the API calls that are related to the permissions. The collected features are related to permissions, components (activity, receiver, and service), intents, and usage of the API calls with different kind of components. Next, it applies K-means clustering algorithm to model malware while the number of clusters are determined by singular value decomposition (SVD) measures to identify the malware in the given applications.

In [5] suggested a detection system that applies a set of predefined security procedures at installation time to discover any match between the benign application and templates of malicious application patterns. First it declares a specific permutation of permissions could be dangerous and used to takeoff malicious actions. The system involves of three components namely, installer, security service, and database of security instructions. The installer excerpts the security configuration from the applications AndroidManifest.xml file. Next, the security service applies the security procedures against the extracted configuration information, if there is a match then the application fails in passing confirmation process; in this scenario, the system offers two choices either dismissing the installation or providing the results of investigating the risk of granting the required permissions to the user to make the decision of the particular application. On the other side, the proposed system functions only at installation time and doesn't offer any support at runtime. Also, it is inadequate to the available information in the package's AndroidManifest.xml file itself.

III. PROPOSED SYSTEM

The architecture framework consists of data collection which contains both benign applications and malicious ransomware applications. Once the data has been collected, the collected data will undergo into reverse engineering phase where the each application has disassembled and got the manifest file from each application. From the manifest file the

significant permission are extracted to create a dataset for classification purpose. The permissions are the main security features of android operating system [20]. Since, the permission are taken as a features for this research work. In this research work, logistic regression classification algorithm has been applied to classify the benign and ransomware applications. The proposed framework architecture is given in the figure 1.

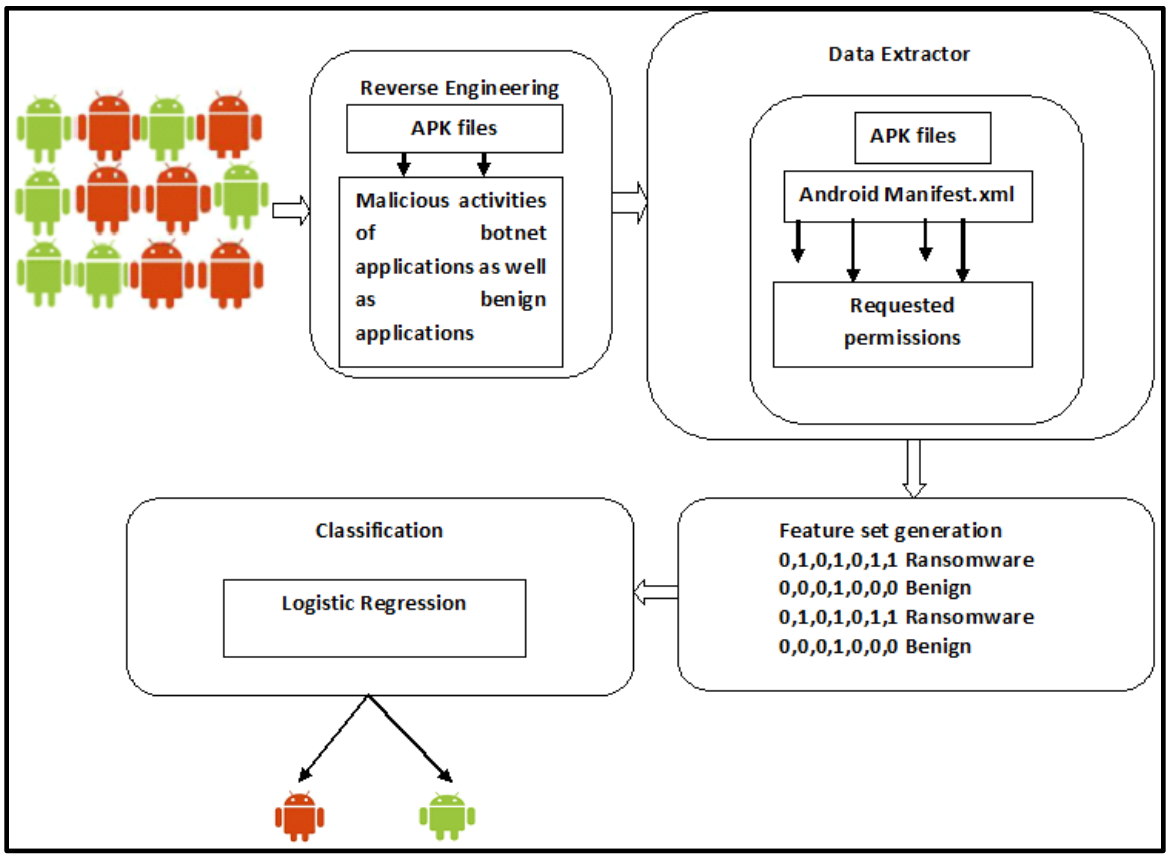

Figure 1. Proposed framework architecture

Logistic regression is the suitable regression analysis to conduct when the dependent variable is binary. Comparable all regression analyses, the logistic regression is an extrapolative analysis [6]. Logistic regression is used to define data and to explain the relationship among one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. It calculates the probability of an outcome that can only have two values that is binary values. The logistic regression will identify the values outside the acceptable range.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed research work utilizing the logistic regression classification algorithm for identification of the ransomware applications. To evaluate the performance of the proposed work utilized four different measures namely, precision, recall, accuracy and false positive rate.

Evaluated our classifier with various evaluation measures, such as accuracy, F-measure and false positive rate.

Accuracy is percentage of correctly identified ransomware applications.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

True Positive (TP) =Number of samples correctly predicted as ransomware.
False Positive (FP) = Number of samples incorrectly predicted as ransomware.
True Negative (TN) = Number of samples correctly predicted as benign.
False Negative (FN) = Number of samples incorrectly predicted as benign.

Precision is a measure of what fraction of test data is detected as ransomware are actually from the ransomware classes.

$$\text{Precision (P)} = \frac{TP}{TP+FP}$$

Recall measures the fraction of ransomware class that was correctly detected.

$$\text{Recall (R)} = \frac{TP}{TP+FN}$$

False Positive Rate (FPR) is percentage of wrongly identified benign classes.

$$\text{Positive Rate (FPR)} = \frac{FP}{FP+TN}$$

The experimental results are given in table1.

Measures	Values
Precision	0.970
Recall	0.970
Accuracy	97.01
False Positive Rate	0.042

Table 1. Experimental results of logistics regression Classifier

From the table, one can observe that the logistic regression classifier achieves highest accuracy of 97.01% and false positive rate of 0.042 when detecting the ransomware applications.

V. CONCLUSION

In this research, we have introduced a machine learning based android ransomware applications detection system. The system starts with analyzing the android ransomware applications, then recognition of important features and applied machine learning method to categorize the data into android ransomware applications and benign application. Throughout logistic regression classifier achieved 97.01% accurateness.

REFERENCES

[1] Alhawi, O. M., Baldwin, J., & Dehghantanha, A. (2018). Leveraging machine learning techniques for windows ransomware network traffic detection. Cyber Threat Intelligence, 93-106.

[2] Alzahrani, A., Alshehri, A., Alshahrani, H., Alharthi, R., Fu, H., Liu, A., & Zhu, Y. (2018, May). RanDroid: Structural Similarity Approach for Detecting Ransomware Applications in Android Platform. In 2018 IEEE International Conference on Electro/Information Technology (EIT) (pp. 0892-0897). IEEE.

[3] Andronio, N., Zanero, S., & Maggi, F. (2015, November). Heldroid: Dissecting and detecting mobile ransomware. In International Workshop on Recent Advances in Intrusion Detection (pp. 382-404). Springer, Cham.

[4] Azmoodeh, A., Dehghantanha, A., Conti, M., & Choo, K. K. R. (2017). Detecting crypto-ransomware in IoT networks based on energy consumption footprint. Journal of Ambient Intelligence and Humanized Computing, 1-12.

[5] Canfora, G., De Lorenzo, A., Medvet, E., Mercaldo, F., & Visaggio, C. A. (2015, August). Effectiveness of opcode ngrams for detection of multi family android malware. In Availability, Reliability and Security (ARES), 2015 10th International Conference on (pp. 333-340). IEEE.

[6] Chen, J., Wang, C., Zhao, Z., Chen, K., Du, R., & Ahn, G. J. (2018). Uncovering the face of android ransomware: Characterization and real-time detection. IEEE Transactions on Information Forensics and Security, 13(5), 1286-1300.