Abstract

For Malware Detection Machine Learning methods are applied extensively in ascertaining if the given APK file is malware or not. Machine learning methods are found to be less time consuming and less resource consuming compared to non-machine learning-based techniques. We have focused on Machine Learning methods for detecting unknown malware. For detecting the malware a researcher needs to create a dataset of its own. Our dataset generation process includes Android File Collection, Decompilation, and Feature Mining phases. We have already discussed the Android File Collection phase in our previous paper [1]. We have collected 15508 Malware files and 4000 Benign Files using Android File Collection. Android Files contains unstructured data in the form of text and XML files which are complex to process and store. Here our goal is to perform the decompilation of these collected Android files such that we get all the resources as well as the source code in a single instance. We aim to handle the big data in terms of Android Files and process them properly performing the Decompilation. In this paper, we have proposed an available automated solution for decompiling the files that also solves the complexity of handling and processing the big data. We have also discussed our
Decomposition phase and presented the structure of the reverse-engineered APK file. We have used an online JADX decompiler [5] for performing the reverse engineering of the APK files.

References

1. Prerna Agrawal, Bhushan Trivedi, "Automating the process of browsing and downloading APK Files as a prerequisite for the Malware Detection process ", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol 9, Issue 2, March - April 2020, pp. 013-017, ISSN 2278-6856
2. Prerna Agrawal, Bhushan Trivedi, “Machine Learning Classifiers for Android Malware Detection”, 4th International Conference on Data Management, Analytics and Innovation (ICDMAI) Springer AISC Series, New Delhi, Jan 2020. (Paper to be Published)
3. Prerna Agrawal, Bhushan Trivedi, “Analysis of Android Malware Scanning Tools”, International Journal of Computer Sciences and Engineering (IJCSE), Vol.7, Issue.3, pp.807-810, Mar 2019.
4. Prerna Agrawal, Bhushan Trivedi, “A Survey on Android Malware and their Detection Techniques”, Third International Conference on Electrical, Computer and Communication Technologies (ICECCT) IEEE, Feb 2019.
5. Decompilation of APK Files, Online Link: http://www.javadecompilers.com/APK
6. Meet Kanwal, Sanjeev Thakur, “An App Based on Static Analysis for Android Ransomware”, International Conference on Communication and Automation (ICCCA), 2017.
7. Neeraj Chavan, Fabio Di Troia, Mark Stamp, “A Comparative Analysis of Android Malware”, 3rd International Workshop on Formal Methods for Security Engineering (ForSE), 2019.
8. Suleiman Yerima, Sakir Sezer,” Android Malware Detection Using Parallel Machine Learning Classifiers”, 8th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST), Sept 2014.
9. Zi Wang, JuronCai “DroidDeepLearner: Identifying Android Malware Using Deep Learning” Sarnoff Symposium IEEE, Sep 2016.
10. J.D. Koli, “Randroid: Android Malware Detection using Random Machine Learning Classifiers”, International Conference on Technologies for Smart City Energy Security and Power (ICSESP) IEEE, Mar 2018.
11. JADX Decompiler Download Files and Download Instructions, Online Link: https://github.com/skylot/jadx
12. Androguard Tool Project Download and Description, Online Link: https://pypi.org/project/androguard/
13. Androguard tool API Docs, Online Link: https://androguard.readthedocs.io/en/latest/api/androguard.html
14. K.V.Kanimozhi, Dr.M.Venkatesan, “Unstructured Data Analysis-A Survey”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2015
15. Min Chen, Shiwen Mao, Yunhao Liu, “Big Data: A Survey”, Mobile Network Applications Springer, 2014, DOI: 10.1007/s11036-013-0489-0
16. Xindong Wu, Xingquan Zhu, Gong-Qing Wu, Wei Ding, “Data Mining with Big Data”, IEEE Transactions on Knowledge and Data Engineering Vol 26, Issue 1, Jan 2014.
17. Jinchuan Chen, Yueguo Chen, Xiaoyong Du, Cuiping LI, Jiaheng LU, “Big data
challenge: a data management perspective”, Frontiers of Computer Science Springer-Verlag Berlin Heidelberg, April 2013.

18. Fan W, Bifet A, “Mining big data: current status, and forecast to the future”, ACM SIGKDD Explor Newsletter, Vol 4, Issue 2, 2013, pp.1–5.

**Index Terms**

Computer Science  Security

**Keywords**

Malware, APK files, Decompilation, Reverse Engineering, Machine Learning, Malware Detection