International Conference on
Contemporary Technological Solutions towards fulfillment of Social Needs

Review of Malware Analysis, Classification and Detection Techniques

Manish Kumar Sahu [1], Ravi Singh Pippal [2],
[1,2]Department of Computer Science & Engineering,
RKDF University, Bhopal, India

Abstract: The technological advances in computing devices have created great impact for wide range of application and these application are basically software oriented system. Since computing devices may be coupled with third-party software applications, so, many security and privacy difficulties can be stimulated by malwares. However, current anti-malware programs are still ineffective in some cases, imperfect and have limited functionalities. A lots of malware attacks can happen on computing device applications due Internet, and it is installed or executed. In this papers, some expects practical with effective malware detection methods is reviewed comprehensively. We introduce malware definition, classification, working models, evolution and security threats of computing devices. Finally, we evaluate open issues and challenges in this research area and also motivate future research direction.

Keywords: Malware Analysis, Classification, Malware Detection, Anti-Malware.

I Introduction

With the rapid development of concurrent software, wireless communications, ubiquitous networking and enhanced sensing capabilities, mobile devices have raised lately, particularly smartphones, wearable body sensor network devices and portable tablets. Modern smartphone users in thw world reached more than 2.53 billion according to Statista (The statistical portal) and in will reach 2.87 billion in 2020. Based on the statistics report of mobile operating systems (OSs) in 2015, the Android based system accounted for 77% among them, while *i*OS obtained 18% ranking second and Windows ranked third with 3%. A mobile device becomes an open concurrent software platform that can run various mobile apps developed not only by mobile device manufactures but also by many third parties. However, the third-party app developers cannot ensure the security and integrity of their shipped apps [1].

At the same time of fast growth of mobile apps, mobile malware is developing quickly. Mobile malware is a malicious program targeting mobile devices. Based on Nokia Threat Intelligence Report (Motive Security Labs Malware Reports Nokia Networks), the total growth of Android malware samples in their database was 342% in 2015, so was the growth of iOS and Windows malware. Mobile malware holds similar purposes to computer malware and intends to launch attacks to a mobile device to induce various threats, e.g., system resource occupation, user behavior surveillance, and user privacy intrusion. Mobile malware pays special attention to the typical properties of the mobile devices, such as mobility and networkconnectivity, to gain specific profits, e.g., tracking user trajectory, disturbing or blackmailing users via short message fraud, forcing users to pay for extra mobile service fees, and disclosing user credentials [2].

Mobile malware is evolving quickly in recent years. In the beginning, Khan et al. (2015) showed that mobile malware mimicked the strategies used by Personal Computer (PC) malware, which is apt to cause system corruption or divulge private user information. Later on, when multifunctional mobile devices became mainstream in the market, mobile malware evolved accordingly. For example, with the ability of wireless networking, attackers can easily intrude a mobile device via air interfaces instead of a physical connection. Mobile malware could make use of mobile devices to send premium SMS messages to increase profit and subscribe to extra paid mobile services secretly (Zhou and Jiang 2012) [3]. In recent years, the mobile devices enhanced with sensing and networking capabilities face novel threats and malware, which can gain super privilege to manipulate user information, e.g., getting access to accelerometers and gyroscopes and leaking private user information to a remote server. Nowadays, malware can rely on advanced camouflage techniques to produce metamorphoses and heteromorphic versions. It also uses evade techniques to circumvent regular detection [4]. Besides that, it can broadcast itself using social networks based on social engineering attacks by making use of the curiosity and credulity of mobile users, which is relatively difficult to prevent. The company Proofpoint pointed out that in 2015 users unknowingly downloaded over two billion times malicious applications with data-stealing functions (Threat Operations Center Proofpoint) [5]. It is undeniable that with wearable smart devices, portable hospital devices (Calderon et al. 2015), and other emerging devices, there will be more security threats targeting mobile devices. In a word, attackers are sensitive at grasping unique mobile ecosystem opportunities to develop malware [6].

In the review, existing malware detection methods encompass two different approaches while collecting feature.

II Taxonomy of Malware

malware can be theoretically divided into several classes according to their malicious goals and behaviors. As a supplementary condition, distribution technique is another reference standard. Normally, there are two main distribution strategies: self-propagating and social engineering. The first approach uses different strategies to automatically install malware into mobile devices, like worms, while the second one takes advantage of user curiosity and unawareness of security to allure them to manually install apps (e.g., adwares). Herein, we summarize several basic types of mobile malware based on their malicious behaviors.

III Related Work

Sahu *et al.* [7] proposed a hybrid method, based on DAG and Gaussian Support Vector Machines, for malware classification. Experiments with the KDD Cup 1999 Data show that SVM-DAG can provide, the better generalization ability and effectively classified malware data. Moreover, the modified algorithms proposed in this desecration outperform conventional CIMDS and ISMCS in terms of precision and recall. Specifically, accuracy of the modified algorithms can be increased due to future allocation of DAG, and reduces feature subset increases the accuracy of classification. From their experiments, the DAG-SVM has ability to detect the predefined types of attecks which were commonly known containing high accuracy and low false positive rate which is measured as below 1%,the method used in our system produces the result as it is capable of classifying the attack and normal data of KDDCUP99 with high accuracy.The learning process used in our proposed work in which SVM training process whose task is of grouping the attecks is very well efficient and capable of producing accurate result of malware data. Our empirical result shows better performance in compression of ISMCS and other data mining technique for malware detection.

Khalilian *et al.* [8] presented G3MD, a novel approach for detection of metamorphic malwares, specifically for viruses and worms. A current issue concerning metamorphic detection is the case where an entire block of a benign file is inserted into a malware. We addressed this issue by mining frequent sub-graphs from the opcode graph of metamorphic malware. Then, a characteristic vector of binary entries is composed in which the presence or absence of each frequent sub-graph is specified for each file. Next, these vectors are used to train a binary classifier that allows for determining the type of any incoming suspicious file. We conducted several experiments on three families of viruses, namely NGVCK, G2, and MPCGEN and a large family of worms namely MWOR. The latter malwares consist of worms with different ratios of padding blocks taken from benign files. The high precision we obtained in our results demonstrates the effectiveness of G3MD and is evidence to corroborate our hypothesis.

Du *et al.* [9] formulated the measurement of malware detection metrics in the absence of ground truth as a statistical estimation problem. We presented a statistical methodology to tackle this problem by designing naive estimators and adjusted estimators. We validated these estimators based on numerical experiments of synthetic data with known ground truth. We learned useful insights in terms of the accuracy (or usefulness) of these estimators in various parameter regimes. We applied these estimators to analyze a real dataset collected from Virus Total.

IV Open Issues and Proposed Approach

Open issues motivate and direct future research. Considering the situation of current development and the constraint resources in the mobile devices, we suggest several promising future research directions in the field of dynamic mobile malware detection. Except for the future research trends listed below, high detection accuracy and efficient detection algorithms are always highly expected.

- Information privacy protection should be investigated with regard to cloud-based mobile malware detection. Due to the constraint resources of the mobile devices, cloud-based detection becomes an inevitable tendency. But, how to avoid the unexpected leakage of user privacy in the process of collecting data for malware detection is a practical and crucial issue. Privacy-preserving mobile malware detection over the cloud is an interesting and significant research topic. In this research, secure transmission protocols, safe data storage with flexible access control, and most importantly secure data processing should be studied.

- Real-time detection with lightweight detection algorithms is another interesting research direction. Malware detection by monitoring behaviors, evidence, and features can identify zero-day attacks and detect malware after malware infection. But, there is a serious time delay. Most existing methods are either server-based or cloud-based systems. They are pretty heavy and complicated with a long time delay. The methods that are both lightweight and efficient thus can be installed in the mobile devices to realize real-time detection are highly expected. This kind of method can find newly generated malware initiated at app runtime during its execution. Although designing such a detection method is not easy, it is a very promising research direction.

- The huge number of malware samples and big size of collected data for malware detection cause a big data problem, which challenges the detection and forces it to be much efficient to handle big data. The speed

of malware growth has never slowed down. The sample database is becoming enormous. So, finding a way to dramatically and efficiently reduce the sample space and effectively detect malware is urgent. To some possible extent, distributed detection systems and cloud-based solutions can make this problem easier. Besides, data mining and fusion methods and some other strategies used in big data processing can also be applied in solving the big data issues in this research field.

- Hardware attack detection and persisting malware removal need serious exploration. A new developing trend of malware is to pose threats to hardware and even virtual machine. So far, there are no effective solutions to handle hardware infections. How to remove malware residing in hardware thoroughly without changing other hardware is becoming a hot issue. What is more, as we have discussed above, an easy and effective method for removing and cleaning persisting malware should also be studied.

V Conclusion

Malware detection is significant in ensuring the quality of mobile concurrent software. This paper gave a thorough review on dynamic mobile malware detection. We reviewed existing work with regard to technical categories, applied classification algorithms, the features used for detection, and the target mobile operating systems. We also considered the places of detection analysis, real-time detection support, privacy preservation support, the threats that can be revealed and overcome, the measures used for detection performance evaluation, and performance test results. In addition, we also commented on their pros and cons. Based on the review, we pointed out open research issues in order to direct future research trends.

REFERENCES

[1] B. Miller, A. Kantchelian, M. C. Tschantz, S. Afroz, R. Bachwani, R. Faizullabhoy, L. Huang, V. Shankar, T. Wu, G. Yiu, A. D. Joseph, and J. D. Tygar, "Reviewer integration and performance measurement for malware detection," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, J. Caballero, U. Zurutuza, and R. J. Rodríguez, Eds. Cham: Springer International Publishing, 2016, pp. 122–141.

[2] T. Hadad, R. Puzis, B. Sidik, N. Ofek, and L. Rokach, "Application marketplace malware detection by user feedback analysis," in *Information Systems Security and Privacy*, P. Mori, S. Furnell, and O. Camp, Eds. Cham: Springer International Publishing, 2018, pp. 1–19.

[3] L. Liu, B.-s. Wang, B. Yu, and Q.-x. Zhong, "Automatic malware classification and new malware detection using machine learning," *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 9, pp. 1336–1347, Sep 2017. [Online]. Available: https://doi.org/10.1631/FITEE.1601325

[4] P. Yan and Z. Yan, "A survey on dynamic mobile malware detection," *Software Quality Journal*, vol. 26, no. 3, pp. 891–919, Sep 2018. [Online]. Available: https://doi.org/10.1007/s11219-017-9368-4

[5] J. Chen, M. H. Alalfi, T. R. Dean, and Y. Zou, "Detecting android malware using clone detection," *Journal of Computer Science and Technology*, vol. 30, no. 5, pp. 942–956, Sep 2015. [Online]. Available: https://doi.org/10.1007/s11390-015-1573-7

[6] A. Souri and R. Hosseini, "A state-of-the-art survey of malware detection approaches using data mining techniques," *Human-centric Computing and Information Sciences*, vol. 8, no. 1, p. 3, Jan 2018. [Online]. Available: https://doi.org/10.1186/s13673-018-0125-x

[7] M. K. Sahu, M. Ahirwar, and P. K. Shukla, "Improved malware detection technique using ensemble based classifier and graph theory," in *2015 IEEE International Conference on Computational Intelligence Communication Technology*, Feb 2015, pp. 150–154.

[8] A. Khalilian, A. Nourazar, M. Vahidi-Asl, and H. Haghighi, "G3md: Mining frequent opcode sub-graphs for metamorphic malware detection of existing families," *Expert Systems with Applications*, vol. 112, pp. 15 – 33, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417418303580

[9] P. Du, Z. Sun, H. Chen, J. Cho, and S. Xu, "Statistical estimation of malware detection metrics in the absence of ground truth," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 2965–2980, Dec 2018.