A systematic review on Android Malware Detection

Kezang Dema[1], Thinley Jamtsho[2]
[1]Information Technology Department, College of Science and Technology, Phuentsholing, Bhutan
[2]Department of Science and Mathematics, Phuentsholing Middle Secondary School, Phuentsholing
[1]kelden.dema@gmail.com, [2]thinlayjamtsho@gmail.com

Abstract—

Android malware is growing at alarming rate and spreading rapidly despite on-going mitigating efforts. This brings a necessity to find more effective solutions to detect those malwares and prevent users from any malicious threats. The aim of the systematic review is to summarize the situation that existed from 2010 to 2015 with regards to various android malware analysis approaches and detection methods. A total of 58 selected papers met the inclusion criteria based on title of articles, exclusion criteria, reading abstract and content of the selected 58 papers. Different data are extracted from these articles and recorded in an excel sheet for further analysis. Most of the paper discussed about the use of dynamic analysis approach to analyze malware and signature-based method for malware detection. The systematic review carried out would provide information to all researchers and further inform the requirements for future development of enhanced malware analysis and detection methods.

Keywords-Android malware, Dynamic analysis, Static analysis, Anomaly, Signature-based

I. Introduction

The advancement of mobile devices from a simple form of sending Short Message Service (SMS) and phone calls to smartphones particularly android is accelerating the mobile industry and device users are increasing exponentially (Ham and Choi, 2013). Since it is an open source, it allows programmers to make modifications even at system level. This leads to more serious security threats compared to others. With the gaining popularity of Android apps, there is also an increase in malwares targeting especially the android mobile devices (Aafer, Du and Yin, 2013). One of the security report by a security company in Finland F-Secure showed that 79% of the newly discovered 301 malware samples in 2012 target android system exponentially (Ham and Choi, 2013). This makes necessary to find effective methods to detect those malwares and help in protecting users against those malicious threats.

The common mobile malware detection methods are based on traditional computer virus detection method that is based on signature or behaviors. The new detection techniques that are introduced are machine learning based, semantic based and many others. All those methods apply various algorithms and classification methods in detecting android malwares. Most of the methods analyze malwares either statically or dynamically with the extraction of different features. In exponentially (Ham and Choi, 2013), it has proposed a machine learning methods for malware detection using various machine learning classifiers. Whereas in (Yerima, Sezer and McWilliams, 2014), the study on the behaviors of malwares were done using API calls to detect malwares in android.

Moreover, in (Arp, Spreitzenbarth, Hubner, Gascon and Rieck, 2014), they applied signature based methods using Support Vector Machine (SVM) algorithm. In (Aafer, Du and Yin, 2013), the android malware was detected by statically extracting API call functions and suggested the K-nearest classification method.

However, due to the lack of samples in study, most of the researches have compiled the malware themselves in order to validate their theories (Wu, Zhou and Xu, no date). Despite the introduction of various detection methods, malwares in android are growing at large.

II. Method

The review considered article types namely journals and conference proceedings that dealt with malware detection specific to android system. The articles were availed from two sources: Google Scholar and Abertay University's summon database using the search terms described in Table 1.

Table 1: Search terms

Keywords
1. Android malware detection 2. Malware detection techniques in android 3. Methods in detecting malware in Android 4. Android malware and detection methods 5. Anomaly-based android malware detection 6. Signature-based android malware detection 7. Machine learning-based android malware detection

The evaluation was conducted between the intervals 2010 to 2015. The selection of 6 year works was to reflect the situation of android malware during those period and its detection methods and procedures. All articles searched using Table 1 keywords displayed a huge number of papers that needed further sorting. The title of articles were used to sort papers in next phase. Some of the articles were discarded because the titles were completely irrelevant as shown below:

1. Who is tweeting on Twitter: human, bot, or cyborg?
2. The 17 Most Dangerous Places on the Web
3. Evading cellular data monitoring with human movement networks

The main review of articles were on android malware detection. A total of 1514 articles gathered using the title was furthered sorted to get the best articles for the review using the following exclusion criteria:

1. Non journal, white papers, newspapers.
2. Articles not written in English.
3. Articles evaluating malware detection not specific to android application.

4. Articles evaluating about the malware families in android applications.

All articles were written in English. The exclusion criteria did not look into either the quantity or the quality of papers. All the exclusion criteria were applied to articles identified and the resultant 398 articles were selected based on those set criteria. Furthermore, articles were sorted based on abstract and a total of 129 articles were selected. The final set of 58 papers were selected for the review after reading its full contents. Most of the selected 58 articles discuss about the methods and techniques for android malware detection, the analysis of malwares and algorithms used for the detection. All those articles were recorded in excel sheet in order to extract data from the content for the review. The types of information gathered from those 58 selected articles are based on following contents as shown in Table 2.

Table 2: Types of data gathered from selected articles

Data
• Malware detection approaches • Malware analysis approaches • Algorithm used • Features used • Evaluation scale • Successful detection rate • Paper type • Publisher • Malware types

III. Result

The keyword search in Google Scholar and Summon database of Abertay University identified a total of 4485 articles. The articles was then screened based on title, resulting to 1514 articles. Next, based on exclusion criteria, a total of 398 articles were selected, followed by 129 articles screened out based on articles' abstract. The final 58 set of articles were selected for the review based on its content. A total of 47 articles were specifically obtained from Google Scholar. There were 11 duplicates between Google Scholar and Summon database. Selecting 11 articles from any of these sources, a total of 58 unique articles were selected for the full review. All the articles were published between the intervals of 2010 to 2015. The distribution of these 58 articles based on year from 2010 to 2015 is shown in Figure 1.

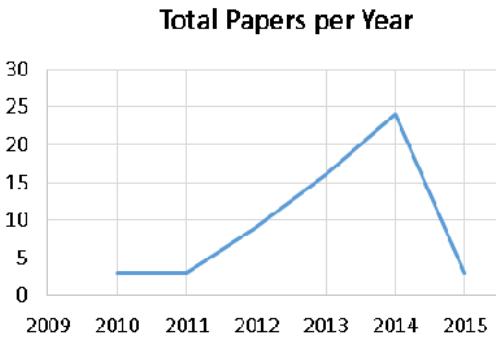

Figure 1 Number of publications identified for review and their year of publication

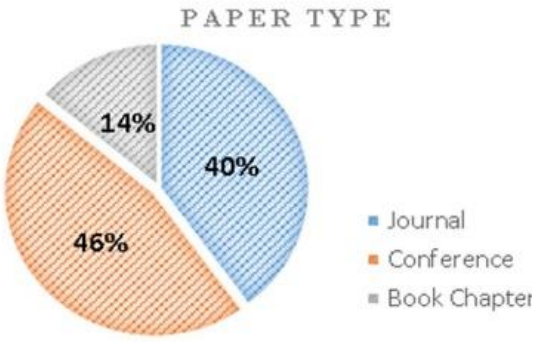

Figure 2 Different paper types of reviewed articles

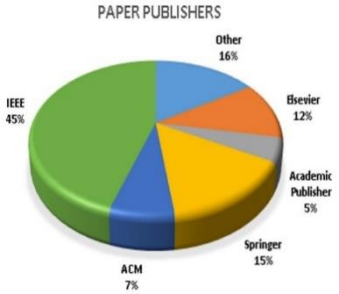

Figure 3 Publishers of the reviewed articles

The 58 articles have 46% consisting of conference proceedings, 40% journal articles and 14% book chapters as shown in Figure 2. The leading publishers of these articles was IEEE with 45% of articles published by them. The percentage distribution of different publishers are shown in Figure 3.

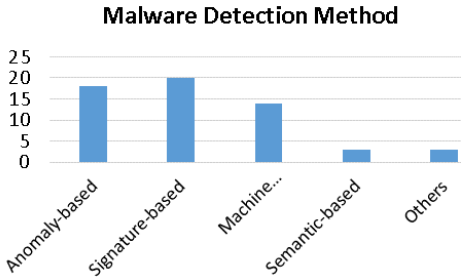

Figure 4 Android malware detection methods used in reviewed articles

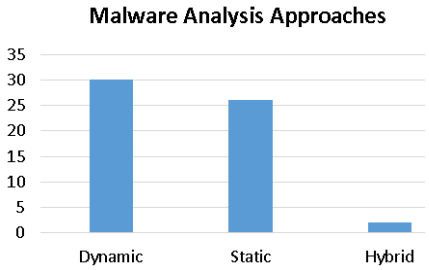

Figure 5 Android malware analysis approaches

The signature-based method dominates in the detection of android malware as depicted in Figure. 4. 20 articles from 58 articles applied the signature-based method, 18 used the anomaly based method, 14 with machine learning, 3 used the semantic-based and 3 used other different methods. For the malware detection method, the malwares need to be analyzed. Majority of the articles used the dynamic analysis approach followed by static and hybrid analysis. The calculated detection rates and some of the detected android malware types from 58 articles are shown in Figure 6 and Figure 7

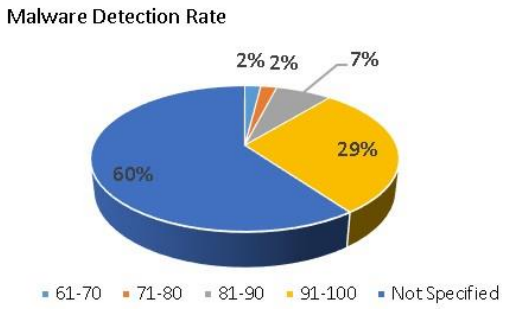

Figure 6 Android malware detection rates specified in the reviewed articles

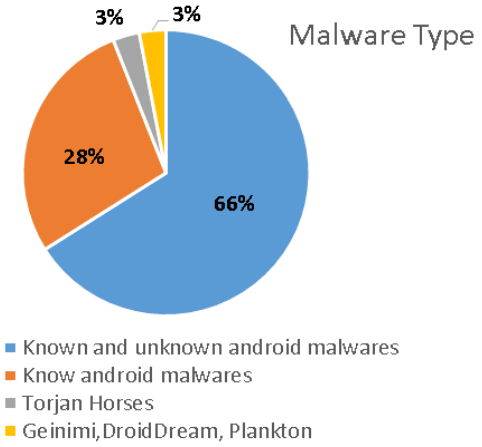

Figure 7 Types of android malware detected in reviewed articles

IV. Discussion

The android malwares are spreading rapidly and if the android phones are infected by malwares, the users' face serious threats such as sensitive information leakage, getting root privileges and many others (Isohara, Takemori, and Kubota, 2011). Hence, there is a need of effective methods to detect those malwares and protect against its impacts. Figure 1 suggests that there has been a steady increase in publications on android malware detection every year with the increase of different malwares and android users annually.

Among 58 full reviewed articles, majority of the articles were from the conference proceedings followed by journals and book chapters as depicted in Figure 2. This would in some way provide insights to researchers when deciding which type of papers to select for their new publications. It is clear from Figure 3 that the leading publishers contributing to this field of area is IEEE. Other popular publishers involve ACM, Springer, Elsevier and Academic publishers. The majority of papers provided data related to methods that were applied for malware detection, the malware analysis approach, different kinds of algorithms applied and the features that were used for the malware analysis.

Detection system includes two tasks - analysis and detection (Landage, and Wankhande, 2013). The malware analysis is necessary to build effective malware detection methods. In most of the articles, the dynamic analysis is commonly used which analyses the file during its execution. This would help in understanding behaviors of the malwares when in action and further increase the malware detection. The other analysis that is mostly used is static analysis which examines the software codes. The least used analysis technique is the hybrid analysis which is a combination of both static and dynamic analysis as depicted in Figure 5. For the malware analysis, different authors applied various algorithms like K-mean clustering, Naive Bayes, Support Vector Machine (SVM) and many others. Moreover, various features were used for the malware analysis. The Table 3 shows that the permission feature is used more compared to other features for the purpose of malware analysis and detection.

Table 3: Number of studies employing different features for malware analysis and detection.

Features	Occurrences
API calls	6
Permissions	9
API calls and permissions	6
System calls	8
Resource Consuption (CPU, memory, battery, audio, Wi-Fi)	8

The Figure 4 clearly shows that the signature-based detection, also referred to as misuse detection is commonly used in most of the reviewed articles, followed by the anomaly-based detection. Signature-based method detects malware using sets of rules and policies and one of its advantage is the precise detection of android malware based on the match of signatures (Wu, Zhou, and Xu, no date). And, anomaly detection method detects based on changes in patterns of signatures and its advantage lies in the prediction of unknown malwares. The other new method commonly used was the machine learning-based. The least used was semantic-based along with other new methods. The new techniques like machine learning-based and others performs better compared to predominant detection techniques namely signature and anomaly based. However, most of the reviewed articles used those two old detection techniques. Some of the articles stated their specific malware detection rate which signifies the accuracy of their proposed techniques as shown in Figure 6. Almost 60% of articles did not stated or talked about the malware detection rate and it becomes difficult to draw conclusion on the effectiveness of those mentioned detection techniques.

Malware comes in different forms such as Trojan horse, spyware, virus, scareware, adware, trapdoor and many others (Landage, and Wankhande, 2013). Not all the articles provided the actual type of android malware detected. It was difficult to draw conclusion on it but based on the detection methods, it was easy to get the types of malware detected. The common android malware types that were detected were Trojan horses, Geinimi, DoridDream and Plankton. A wide range of both known and unknown android malwares were detected. The Figure 7 shows the clear distribution of various types of malware detected.

All those articles are evaluating different approaches in detecting android malwares. But it is noted that most of the articles explained about the evaluation of methods using various algorithms rather than its use in an operational environment. The existing signature and anomaly-based detection method still dominates in android malware detection though some of the new detection methods gives better solution. The review showed that the detection of both known and unknown malwares has higher percentage rate. This result provides a positive effect along with the growth of android system and emergence of different undesired malwares.

V. Conclusion

The systematic review was conducted to review and analyze the android malware analysis methods and detection methods within the year of 2010 to 2015. A total of final 58 papers from 1514 papers were sorted and selected for the review after excluding those papers that didn't met the inclusion criteria. The review provided better knowledge of the status with respect to android malware detection like the common methods used, the malware analysis techniques, various features used for malware analysis, algorithms used to differentiate between malwares and non-malwares and the malware detection rates of all those proposed methods. The review suggests that this field has potential opportunities in times to come. Hence, it will help in providing the noble researchers working in this particular field in giving ideas and informing requirements for future development of such systems with better techniques to tackle with the rising of new android malwares

VI. Acknowledgement

The study was a research undertaken in the year 2015 especially to bring notice about the status on android malware detection methods and techniques between the intervals of 2010 to 2015. The views expressed related to any topics in this publication are those of the authors.

References

[1] Aafer, Y., Du, W. and Yin, H. (2013) 'DroidAPIMiner: mining API-level features for robust malware detection in android', *International Conference on Security and Privacy in Communication Networks*, pp. 86-103.

[2] Arp, D., Spreitzenbarth, M., Hubner, M., Gascon, H. and Rieck, K. (2014) 'Drebin: effective and explainable detection of android malware in your pocket', *in Proc. NDSS*.

[3] Chan, P. P. K. and Song, W. (2014) 'Static detection of android malware by using permissions and API calls', *IEEE International Conference on Machine Learning and Cybernetics*, pp. 82-87.

[4] Ham, H. and Choi, M. (2013) 'Analysis of android malware detection performance using machine learning classifiers', *IEEE International Conference on ICT Convergence*, pp. 490-495.

[5] Isohara, T., Takemori, K. and Kubota, A. (2011) 'Kernel-based behavior analysis for android malware detection', *IEEE Seventh International Conference on Computational Intelligence and Security (CIS)*, pp. 1011-1015.

[6] Landage, J. and Wankhande, M. P. (2013) 'Malware and malware detection techniques: a survey', *International Journal of Engineering Research and Technology*, 2(12), pp. 61-68.

[7] Wu, D., Mao, C., Wei, T., Lee, H. and Wu, K. (2012) 'Droidmat: android malware detection through manifest and API calls tracing', *IEEE Seventh Asia Joint Conference on Information Security*, pp. 62-69.

[8] Wu, Z., Zhou, X. and Xu, J. (no date) 'A result fusion based distributed anomaly detection system for android smartphones', *Journal of Networks*, 8(2), pp. 273-282.

[9] Yerima, S. Y., Sezer, S. and McWilliams, G. (2014) 'Analysis of Bayesian classification-based approaches for Android malware detection', IET *Information Security*, 8(1), pp. 25-36.