Faculty of Information and Communication Technology

AN EFFICIENT SIEVE TECHNIQUE IN MOBILE MALWARE DETECTION

Mohd Zaki bin Mas'ud

Doctor of Philosophy

2018

AN EFFICIENT SIEVE TECHNIQUE IN MOBILE MALWARE DETECTION

MOHD ZAKI BIN MAS'UD

**A thesis submitted
in fulfilment of the requirements for the degree of Doctor of Philosophy**

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

DECLARATION

I declare that this thesis entitled, "An Efficient Sieve Technique in Mobile Malware Detection" is the result of my own research work except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ………………………………

Name : Mohd Zaki Bin Mas'ud

Date : ………………………………

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : ……………………………………...……

Supervisor Name : Prof. Datuk Ts. Dr. Shahrin Bin Sahib

Date : ………………………………………..

DEDICATION

This thesis is dedicated with

Deepest love and affections to my beloved parents,

Hj. Mas'ud Taib and Hajjah Siti Juariah Hamdan

My son

Muhammad Hifzhan Irfan Mohd Zaki

Brother and sisters

Rozita, Roslina and Nazri

Their love, patience, guidance, wisdom and strength

Have inspired me throughout these years in

Universiti Teknikal Malaysia Melaka

To be the best that I can be.

ABSTRACT

Proliferation of mobile devices in the market has radically changed the way people handle their daily life activities. Rapid growth of mobile device technology has enabled users to use mobile device for various purposes such as web browsing, ubiquitous services, social networking, MMS and many more. Nowadays, Google's Android Operating System has become the most popular choice of operating system for mobile devices since Android is an open source and easy to use. This scenario has also ignited possibility of malicious programs to exploit mobile devices and consequently expose any sensitive transaction made by the user. A malware ability to quickly evolve has made mobile malware detection a more complex. Antivirus and signature based IDS require a constant signature database update to keep up with the new malware, thus exhausting a mobile device's resources. Even though, an anomaly-based detection can overcome this matter, an anomaly detection still produces a high amount of false alarms. Therefore, this research aims to improve Mobile Malware Detection by improving the accuracy, True Positive and True Negative as well as minimizing the False Positive rate using an n-gram system call sequence approach and a sieve technique. This research analyses the behaviour and traces of mobile malware application activity dynamically as mobile malware is executed on a mobile platform. Analysis done on mobile malware activity shows behaviour and traces of benign and malicious mobile applications are able to be distinctively classified through invocation of system call to a kernel level system by a mobile application. However, an n-gram system call sequence generated by this approach can contribute to a large amount of logged features that can consume a mobile device's memory and storage. Hence this research, introduces a sieve technique in Mobile Malware Detection process in order to search for an optimum set of n-gram system call. In order to evaluate the performance of the proposed approach Accuracy, True Positive Rate, True Negative Rate, False Positive Rate and Receiver Operating Characteristic curve are measured with dataset of mobile malware from Malware Gnome Project and benign mobile application from Google Play Store. The experiment finding indicates the 3-gram system call sequence is capable of improving Mobile Malware Detection performance in terms of accuracy as well as minimizing the false alert. Whereas the sieve technique is able to reduce number of n-gram system call features and providing an optimize 3-gram system call sequence features. The outcome indicate that a Mobile Malware Detection using 3-gram system call sequence as features and sieve technique is able to be used in improving a Mobile Malware Detection in classifying the benign and malicious mobile applications. The evaluation and validation shows that a Mobile Malware Detection using 3-gram system call sequence with sieve technique improve the classification performance. As a conclusion the 3-gram system call sequence Mobile Malware Detection with sieve technique is capable of classifying the benign and malicious mobile application more accurately and at the same time minimizing the false alarm.

ABSTRAK

Perkembangan peranti mudah alih di pasaran telah mengubah cara kita mengendalikan aktiviti kehidupan seharian. Pertumbuhan pesat teknologi mudah alih telah membolehkan pengguna menggunakannya untuk pelbagai perkara seperti melayari web, perkhidmatan merata, rangkaian sosial, khidmat pesanan multimedia dan banyak lagi. Kini, Sistem Pengendalian Android Google telah menjadi sistem operasi pilihan utama untuk peranti mudah alih disebabkan ia adalah dari sumber terbuka dan mudah digunakan. Senario ini juga memunculkan kemungkinan perisian hasad yang boleh mengeksploitasi peranti mudah alih dan seterusnya mendedahkan sebarang transaksi sensitif pengguna. Keupayaan perisian hasad untuk berkembang pantas telah menjadikan pengesanan perisian hasad mudah alih rumit. Sistem Pengesan Pencerobohan berasaskan antivirus dan kaedah tandatangan memerlukan kemas kini pangkalan data tandatangan secara tetap bagi setiap penemuan perisian hasad baru. Ini menyebabkan sumber peranti mudah alih cepat penuh. Walaupun pengesanan berasaskan anomali dapat mengatasi isu ini; ia masih menghasilkan jumlah penggeraan palsu yang tinggi. Oleh itu, penyelidikan ini bercadang menambahbaik Pengesanan Perisian Hasad Mudah Alih dengan meningkatkan ketepatan, Positif Benar dan Benar Negatif serta meminimumkan kadar Positif Palsu dengan menggunakan pendekatan urutan Sistem Panggilan n-gram dan teknik penyaringan. Penyelidikan ini menganalisis tingkah laku dan jejak aktiviti aplikasi Perisian Hasad Mudah Alih secara dinamik. Hasil analisis menunjukkan tingkah laku dan jejak aktiviti aplikasi mudah alih yang benigna dan hasad dapat diklasifikasikan melalui sistem panggilan yang dipanggil oleh aplikasi mudah alih dari sistem kernel. Walau bagaimanapun, urutan Sistem Panggilan n-gram yang dihasilkan menyumbang kepada pengumpulan log yang besar dan menyebabkan penggunaan sumber peranti memori dan storan yang tinggi. Oleh itu, teknik penyaringan diperkenalkan dalam Pengesanan Perisian Hasad Mudah Alih untuk mencari set ciri Sistem Panggilan n-gram yang optimum. Untuk menilai prestasi kaedah pendekatan yang dicadangkan, pengukuran penilaian Ketepatan, Kadar Positif Benar, Kadar Negatif Benar, Kadar Positif Palsu dan lengkung Ciri Pengendali Penerima digunakan diatas set data aplikasi perisian hasad mudah alih daripada Projek Gnome Malware dan aplikasi mudah alih yang bersih dari Google Play Store. Penemuan awal menunjukkan urutan sistem panggilan 3-gram mampu meningkatkan prestasi pengesanan Perisian Hasad Mudah Alih dari segi Ketepatan, serta meminimumkan Kadar Positif Palsu. Manakala teknik penyaringan dapat mengurangkan jumlah ciri yang perlu dilog seterusnya menyediakan urutan Sistem Panggilan 3-gram yang optimum. Hasil penemuan menunjukkan urutan sistem panggilan 3-gram Pengesanan Perisian Hasad Mudah Alih dengan teknik penyaringan dapat mempertingkatkan Pengesanan Perisian Mudah Alih dalam dalam mengelaskan aplikasi mudah alih yang benigna dan hasad. Ujian Penilaian dan pengesahan menunjukkan yang urutan sistem panggilan 3-gram Pengesanan Perisian Hasad Mudah Alih dengan teknik penyaringan dapat mempertingkatkan prestasi pengelasan. Sebagai kesimpulan urutan sistem panggilan 3-gram Pengesanan Perisian Hasad Mudah Alih dengan teknik penyaringan mampu mengelaskan aplikasi benigna dan hasad dengan lebih tepat dan pada masa yang sama meminimumkan penggera palsu.

ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, all praises to Allah for His strength and blessing in completing this thesis. I would like to express my gratitude to my respectful supervisor, Professor Datuk Ts. Dr. Shahrin Sahib, whose expertise, understanding, and patience significantly enhanced my graduate experience.

I would like to express my appreciation to my co-supervisor Professor Madya Dr. Mohd Faizal Abdollah, my mentors Dr. Siti Rahayu Selamat and Dr. Robiah Yusof for their support and aid in making my PhD journey a success. My appreciation also go to Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Education Malaysia for sponsoring this research. My deepest thanks to all the people who had given their support and motivation to make this journey a success especially to all my colleagues in the department of Sistem Komputer dan Komunikasi (SKK) and generally in Fakulti Teknologi Maklumat dan Komunikasi (FTMK) for their constructive discussions and help with the analysis and in thesis writing during the course of this research.

Last but not least, from the bottom of my heart a highest gratitude to my family for their love and caring. Especially to my late father, Haji Mas'ud Hj. Taib and my mother, Hajjah Siti Juariah Hj. Hamdan for their encouragement and blessing, my eternal love to my son, Muhammad Hifzhan Irfan, who has been the pillar of strength in all my endeavours. Finally, to those who indirectly contributed to this research, your kindness has inspired me to embark on this journey

TABLES OF CONTENTS

Comment [JD1]: Formatting note:

Page 113 should be on right margin of table (subheading 4.5.4)

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Comment [JD2]: This list should be in alphabetical sequence, beginning with:
AB Anomaly Base

Ending with:
TPR True Positive Rate

API	-	Application Program Interface
RP	-	Research Problem
RQ	-	Research Question
RO	-	Research Objective
IDS	-	Intrusion Detection System
MMD	-	Mobile Malware Detection
TP	-	True Positive
TN	-	True Negative
FP	-	False Positive
FN	-	False Negative
SB	-	Signature Base
AB	-	Anomaly Base
SPB	-	Specification Base
SVM	-	Support Vector Machine
TPR	-	True Positive Rate
FPR	-	False Positive Rate
TNR	-	True Negative Rate
ROC	-	Receiver operating characteristic
IG	-	Information Gain
CHI	-	Chi-Square test,
CFS	-	Correlation-based feature
BF	-	Best First
GA	-	Genetic Algorithm
EA	-	Evaluation Algorithm
PSO	-	Particle Swam Optimization
CNC	-	Command and Control

LIST OF PUBLICATIONS

Mohd Zaki Mas'ud, Shahrin Sahib, Mohd Faizal Abdollah, Siti Rahayu Selamat and Choo Yun Huoy, 2017. A Comparative Study on Feature Selection Method for n-gram Mobile Malware Detection. *International Journal of Network Security, 19*(5), pp. 727-733.

Mohd Zaki Mas'ud, Shahrin Sahib, Mohd Faizal Abdollah, Siti Rahayu Selamat and Robiah Yusof, 2016. An Evaluation of n-gram System Call Sequence in Mobile Malware Detection. *ARPN Journal of Engineering and Applied Sciences, 11*(5), pp. 3122-3126.

Mohd Zaki Mas'ud, Shahrin Sahib, Mohd Faizal Abdollah, Siti Rahayu Selamat and Robiah Yusof, 2014, May. Analysis of Features Selection and Machine Learning Classifier in Android Malware Detection. *In 2014 IEEE International Conference on Information Science and Applications* (ICISA), pp. 001-005.

Mohd Zaki Mas'ud, Shahrin Sahib, Mohd Faizal Abdollah, Siti Rahayu Selamat and Robiah Yusof, 2014. Android Malware Detection System Classification. *Research Journal of Information Technology, 6*(4), pp. 325-341.

Mohd Zaki Mas'ud, Shahrin Sahib, Mohd Faizal Abdollah, Siti Rahayu Selamat, Robiah Yusof & Rabiah Ahmad, 2013. Profiling Mobile Malware Behaviour through Hybrid Malware Analysis Approach. *In 9th IEEE International Conference on Information Assurance and Security* (IAS), 2013, pp. 78-84

Comment [JD3]: See above

Comment [JD4]: See above

Comment [JD5]: Article titles should have minimal capitalisation

CHAPTER 1

INTRODUCTION

1.1 Introduction

Comment [JD6]: Zaki Masud – it appears that the subheading numbering was updated while I was proofreading (this was not intentional on my part). Apologies for this; however, it should be corrected after all other changes are updated (after you accept or reject each change throughout the thesis) and Track Changes is then turned off. Then click on update Table of Contents – this should correct it.

The popularity of mobile devices over recent years has been continuously growing; with functionality similar to a personal computer. Mobile device users can do more than just making calls and handling Short Message Service (SMS). According to the International Telecommunication Union (ITU) (2016), at the end of 2016 there are almost 7.5 billion mobile users with more than 3.8 billion mobile-broadband subscriptions worldwide. The rise of mobile devices which have full functionality of a personal computers and support of latest communication technology has enabled users to always get connected to the Internet anywhere at any time. A mobile device can be used for various purposes such as web browsing, ubiquitous services, social networking, Multimedia Messaging Service (MMS) and many more. Robust Operating System (OS) Technology supporting mobile devices has also contributed to the rapid development of mobile applications on the mobile devices.

Comment [JD7]: recent

Currently, there are several mobile device OSs namely iOS from Apple, Blackberry, Symbian, Windows mobile and Android by Google. Among these OSs, Google's Android OS is widely consume in the mobile devices market shares;, Gartner Inc. stated that 84.1% smartphone sales during the first quarter of 2017 is on Android platform (Forni and Meulen, 2017). Android OS open source nature, credibility, performance and ease of customizing has made most mobile users choose mobile devices supported by Android OS from the others. Despite a rapid growth of Android-based mobile devices in the market, ahead of the other competitors, it also has

Comment [JD8]: …ahead of…

become an ideal place for malware writers. An increase in mobile applications in Android has also ignited the possibility of malicious programs which can exploit mobile devices. These malicious program are targeting the mobile devices because of the devices are used for online banking, online shopping or any sensitive transaction.

In early 2000, malicious software or malware has been only associated mainly with Desktop Computers but as the mobile technology evolved the malware has now proliferated the mobile space. Proliferation of malware on mobile technology exposed mobile user's sensitive information to malicious actions. Since 2010 new mobile malware is appearing at a regular interval. In 2012, Kaspersky Security Bulletin (Denis and Yuri, 2012) has reported that Android-based malware is growing at an alarming rate. Figure 1.1 shows that 98.96% of newly found mobile malware is targeting the Android-based platform. Mobile malware effect is lethal, mobile malware can steal credential information from the device, sniffed user activity and location, overbilled users by sending random SMS and MMS to contacts, launched denial of services attack from user devices and overloaded device resources such as memory, battery and storage (La Polla et al., 2013).

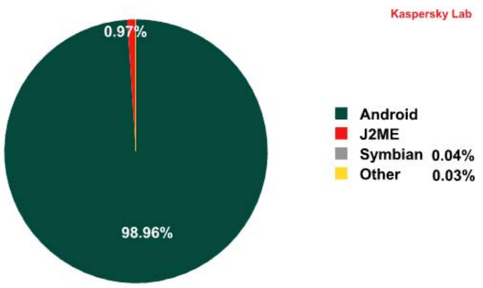

Figure 1.1: Distribution of Mobile Malware by Platform, 2012 (Denis and Yuri, 2012)

In 2010, Malware in general has cost consumer in United States USD 2.3 billion and caused 1.3 million personal computers to be replaced. In addition, in the same year, malware infection has cause USD 118 billion financial impact worldwide. As mobile devices technology is now adapting all the personal computer capabilities, mobile devices are going to have similar effects. According to Juniper Network Mobile Threat Center (2012), the effect of mobile malware includes exploiting vulnerabilities in mobile payment gateway that can provide the attacker an immediate USD 10 million profit. Mobile malware has become an emerging threat in cyber security and some countermeasures need to be taken to overcome mobile malware infections. Therefore, developing and improving mobile devices security to the same level as computer security is important, especially in finding a mechanism to protect the system and data resources from any kind of intrusion (Sundaram, 1996).

1.2 Research Problem

Malicious software or called malware, is purposely written to exploit the vulnerabilities found in a computer system. Malware developers write malware code for different purposes which mostly are used for malicious intention (Robiah et al., 2009). The rapid evolution of malware signature and behaviour have made it difficult to stop. Previously, malware such as Virus, Trojan, Worm and Botnet are synonym to personal computers and rarely found in a mobile device. However, as the mobile devices are become increasingly complex and can support complex OS, mobile devices has become the malware's next target. The worldwide epidemic of malware infections has given malware authors a generous financial benefit through their activities in stealing credential information and gaining access to financial accounts. At present, in response to the emergence of mobile malware, security companies have released mobile antivirus applications as a defence mechanism.

Anti-malware applications, known as Antivirus for mobile, have a similar function as the one on the Desktop version; mobile version antivirus still detects malware based on the known malware signature and is useful for cleaning up the device after it has been infected. With a more advanced malware introduced, the signature is kept on changing from one variant to

Comment [JD9]: ADD TO REFERENCE LIST

Comment [JD10]: Do you mean IT ? If so, add IT to list of abbreviations. Otherwise, replace with specific noun.

Comment [JD11]: Revise word choice (associated with?)

Comment [JD12]: reward / benefit