BORANG PENGESAHAN STATUS TESIS*

JUDUL: ANDROID MALWARE DETECTION THROUGH APPLICATION PERMISSION : AN ANALYSIS ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

SESI PENGAJIAN: 2014/2015

Saya MOHD SHAHRULAZAM BIN SAMSUDIN

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

_____ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

_____ TIDAK TERHAD

_____ _____

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat tetap: NO.24, JALAN ROTAN 5, TAMAN SRI PULAI, 81300 SKUDAI, JOHOR BAHRU, JOHOR

EN MOHD ZAKI BIN MAS'UD

Nama Penyelia

Tarikh: 24 August 2015 Tarikh: 24 August 2015

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

ANDROID MALWARE DETECTION THROUGH APPLICATION PERMISSION : AN ANALYSIS ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

MOHD SHAHRULAZAM BIN SAMSUDIN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS*

JUDUL: ANDROID MALWARE DETECTION THROUGH APPLICATION PERMISSION : AN ANALYSIS ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

SESI PENGAJIAN: 2014/2015

Saya MOHD SHAHRULAZAM BIN SAMSUDIN

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

_____ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

_____ TIDAK TERHAD

_____ _____

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat tetap: No.24, Jalan Rotan 5, _____

Taman Sri Pulai, 81300 Skudai, Nama Penyelia

Johor Bahru, Johor

Tarikh: 25 August 2015 Tarikh: _____

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

ANDROID MALWARE DETECTION THROUGH APPLICATION PERMISSION : AN ANALYSIS ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

MOHD SHAHRULAZAM BIN SAMSUDIN

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Computer Networking)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2015

DECLARATION

I here declare that this project report entitled

ANDROID MALWARE DETECTION THROUGH APPLICATION PERMISSION : AN ANALYSIS ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT :_____Date:_____

(MOHD SHAHRULAZAM BIN SAMSUDIN)

SUPERVISOR:_____Date:_____

(EN MOHD ZAKI BIN MAS'UD)

DEDICATION

This thesis is dedicated to my parent, who always taught me to perform and focus whatever work you been doing for the sake of your future and always remind me to gain knowledge as much as possible to be success in your life. They also taught me to finished whatever task or job that have you started by following to the schedule that have been created.

ACKNOWLEDGEMENTS

First and foremost, I would like to give a big appreciation to my supervisor, En Mohd Zaki Bin Mas'ud who assist, encourage and comment my project until the very end of the project. He is a very expert supervisor for me to carry out this project. Thank you for all your guidance, encouraged and comment till I able to finish my project.

My second appreciation to Saidah Mastura Binti Abdul Ghani who guide me during the first start-up of my final year project. You have become are my second supervisor that had helped me at the beginning of this research. I am appreciate with all your guide that have been taught.

Thanks to all my colleagues Nurfarah Athira Binti Muhammad Tuah, Nor Amira Binti Mohd Hurriff, Siti Roaiza Binti Mohd Yusoff and Fatin Athirah Binti Mohamed Nordin who had helped me a lot in completing and handle the big data. I thank all of them for their contribution and willingness to help me complete the project research.

ABSTRACT

In this technology era Internet of Thing(IoT), where everything or gadget are connected to the internet with no exception of android smartphone which had become vital accessory for everyone. As the number of android user grows up every year, more user are vulnerable to malware application on their smartphone which can widespread in the network within a seconds to perform malicious activity. Regardless of the new advancement technology shown in every new model introduce every year, the malware also evolve to achieve it target. Thus, to overcome this issue this research analysis will investigate the use of different feature selection through filter method and different classification algorithm to optimize the android permission to indicate malware application. These feature selection were evaluated through performance parameter to produce the highest accuracy of classification algorithm. Over 500 android application which consist of 250 malware and benign application each were tested with different feature length, feature selection and classification algorithm. For the purpose of studying the android manifest permission, the malware can be identifying by using data mining technique. In testing and validating the parameter of malware permission, this project have 3 types of performance parameter which are accuracy, true positive rate (TPR) and true negative rate (TNR). The result to be achieve the accuracy higher than 80% and lower of true negative rate.

ABSTRAK

Dalam era teknologi Internet of Thing(IoT) ini, di mana setiap alat yang disambungkan ke internet tidak terkecuali telefon pintar android yang telah menjadi aksesori penting untuk semua orang. Oleh kerana bilangan pengguna android membesar setiap tahun, banyak pengguna terdedah kepada aplikasi malware pada telefon pintar mereka yang boleh meluas dalam rangkaian dalam sesaat untuk melakukan aktiviti malicous. Tidak kira teknologi kemajuan baru yang ditunjukkan dalam setiap model baru diperkenalkan setiap tahun, malware yang juga berubah untuk mencapai target. Oleh itu, untuk mengatasi isu ini analisis kajian ini akan menyiasat penggunaan pemilihan ciri yang berbeza melalui kaedah penapis dan algoritma pengelasan yang berbeza untuk mengoptimumkan kebenaran android untuk menunjukkan aplikasi malware. Pemilihan ciri dinilai melalui parameter prestasi untuk menghasilkan ketepatan algoritma classfication yang tinggi. Lebih 500 android aplikasi yang terdiri daripada 250 malware dan aplikasi berbahaya setiap satu telah diuji dengan panjang ciri-ciri yang berbeza, pemilihan ciri dan algoritma pengelasan. Untuk tujuan pembelajaran kebenaran android aplikasi, malware boleh di kenal pasti dengan menggunakan teknik pengecilan data. Dalam ujian dan mengesahkan parameter kebenaran malware, projek ini mempunyai 3 jenis parameter prestasi yang tepat, kadar benar positif (TPR) dan kadar negatif benar (TNR). Hasilnya untuk mencapai ketepatan yang lebih tinggi daripada 80% dan rendah daripada kadar negatif benar.

TABLE OF CONTENTS

LIST OF TABLES

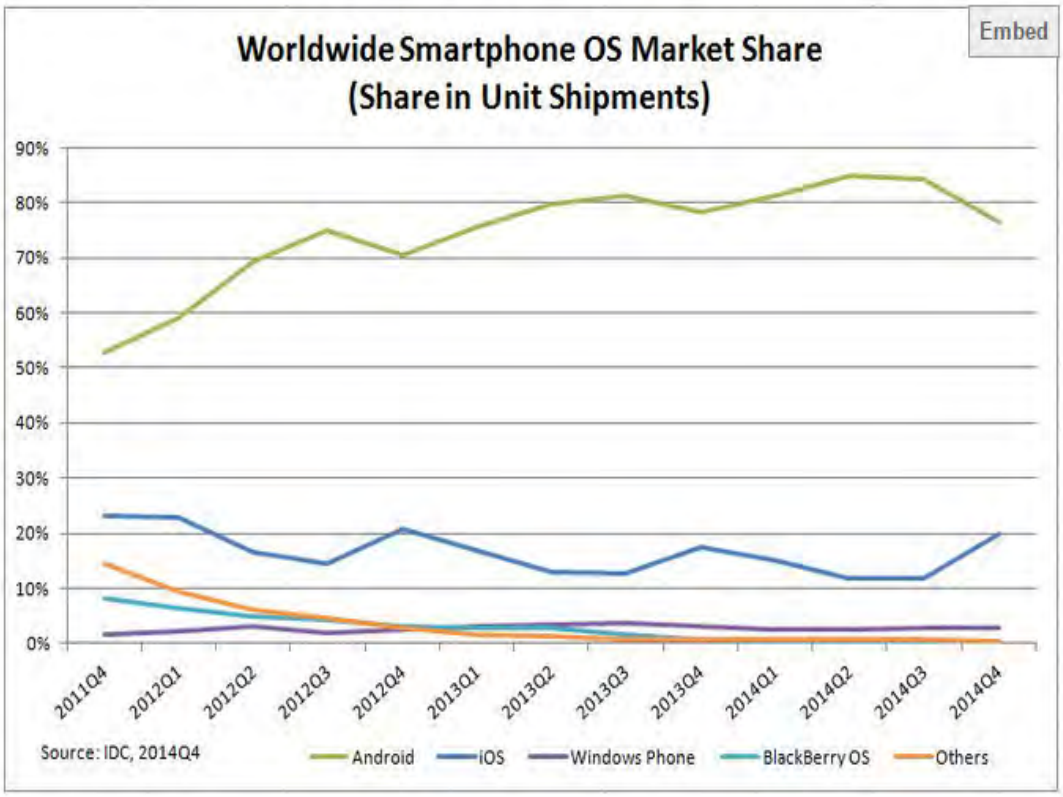
LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Project Background

The new era of technology has become tremendously arise in smartphone usage among people as a medium of a communication and social networking where everyone is keep busying typing a message to their loving one. Google's Android Operating System(OS) is one of a very well-known and popular smartphone operating system which cover major smartphone operating system instead of another popular smartphone iOS by apple. From the statistic gained from International Data Corporation(IDC) in quarter of 2014, Android OS was the highest worldwide smartphone OS market share among the other OS smartphone as shown in Figure 1.1. Moreover, according to trend now Android OS has become most vulnerable smartphone OS to malware which increase year after year as shown in Figure 1.2.



Period	Android	iOS	Windows Phone	BlackBerry OS	Others
Q4 2014	76.6%	19.7%	2.8%	0.4%	0.5%
Q4 2013	78.2%	17.5%	3.0%	0.6%	0.8%
Q4 2012	70.4%	20.9%	2.6%	3.2%	2.9%
Q4 2011	52.8%	23.0%	1.5%	8.1%	14.6%

Figure 1.1: Worldwide Smartphone OS Market Share (IDC, 2014)

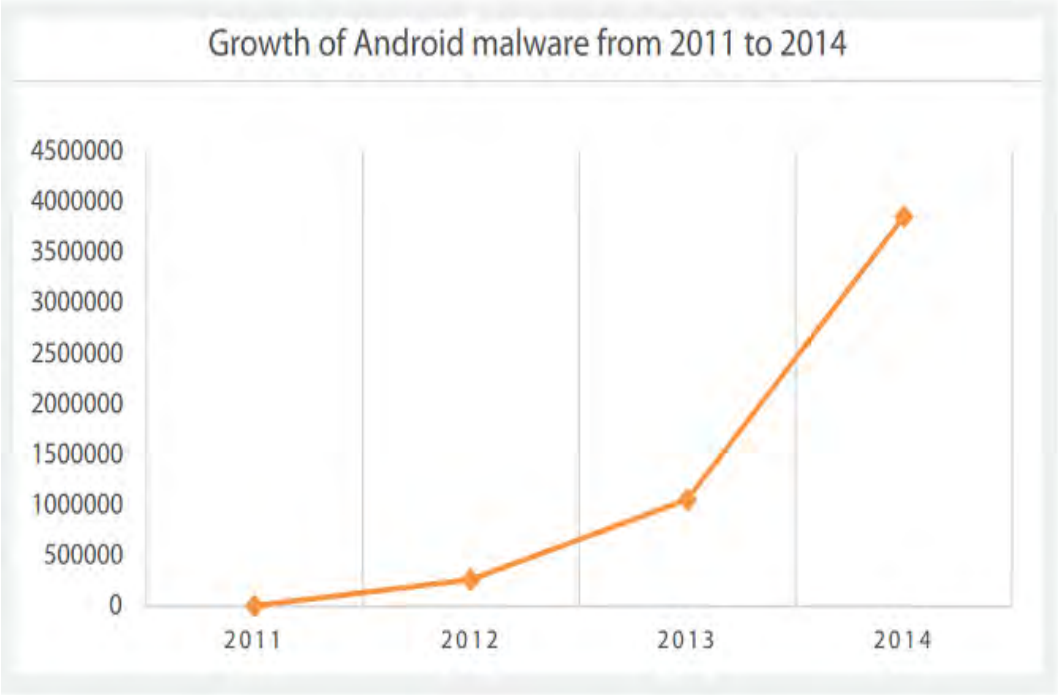


Figure 1.2: The Growth of Android Malware from 2011 to 2014 (Quick Heal Threat Research Labs, 2014)

Malicious software or application which is define as malware is written by unknown people are purposely release mostly for mischievous act such as stealing android user credential information, sending text without the user knowledge, spamming and web browser search optimization. Malware is general name where can be in the form of Trojan, Virus, Botnet and Worm which are unusually found in android device but are synonym to personal computer desktop. In this advance mobile android technology where every year has the new reveal up to date smartphone by competitive smartphone company such as Samsung, Lenovo and Asus try to attract many customer with sophisticated functions meanwhile malware also evolve and not to miss but become the next target of android smartphone device.

1.2 Problem Statements

Android smartphone nowadays are widely vulnerable to malware where it can be easily widespread rapidly within a seconds in the network. Thus the study research about malware permission need to carry out to identify this malware. The research problem (RP) is summarized into Table 1.1

Table 1.1 Summary of Research Problem

No	Research Problem
RP1	Malware are updated with new unknown signature gives difficulty in malware detection and prevention.

Thus, one Research Question (RQ) is built to identify the research problem as discussed in previous section is explain in table 1.2

Table 1.2 Summary of Research Question

RP	RQ	Research Question
RP1	RQ1	How to identify android malware permission?

This research question is done by considering the malware's permission which is highlighted in RP1 in Table 1.1. This research question (RQ) is the primary guides to formulate the research objective (RO) of this project.

1.3 Objectives

According on the research questions formulated in previous section, appropriate Research Objectives (RO) are developed as follows and the Research Objective (RO) are summarized as shown below in table 1.3.

RO 1: To study about the parameter of android malware's permission.

- ✓ The static analysis of reverse engineering need to be carry out in order to identify and compare the permission contain in benign and malware APK file.

RO 2: To compare the permission of android in benign and malware APK file.

- ✓ To detect the permission comparison, we need to know what type of permission usually contain in benign and malware APK file.

RO 3: To generate the graph of android permission in benign and malware APK file.

- ✓ The graph show containing android permission in benign and malware APK file which will be analyze through investigation of static analysis reverse engineering.

Table 1.3 Summary of Research Objective

RP	RQ	RO	Research Objective
		RO1	To study about the parameter of android malware permission.
		RO2	To compare the android permission contain in benign and malware APK file of android
		RO3	To generate the graph of android permission contain in benign and malware APK file of android

1.4 Scopes

1. Type of analyze android malware permission are specific which is static analysis instead of dynamic analysis
2. Feature selection of filter is choose as a technique to classify malware
3. Classification is use to evaluate feature selection performance

1.5 Project Significance

1. APK file extraction using Androguard is used
2. Benign and malware permission graph is generated
3. Type permission used in benign and malware are identify and analyzed

1.6 Report Organization

To assure this research is run in progressing smoothly and successfully, report organization is constructed in order to arrange chapter by chapter respectively. The summarization and description of each chapter are been depicted below:

Chapter 1: Introduction

This chapter review on introduction, project background, research problems, research questions, research objectives, scopes, project significant and report organization.

Chapter 2: Literature Review

In this chapter, all related study about malware permission, static analysis technique and feature selection were done. The output of the study will be utilized in the following chapter which in the methodology

Chapter 3: Methodology

In this chapter, project methodology will be discuss according to activities, step taken and stage followed in order to make sure this project run smoothly in sequence and priority.

Chapter 4: Design and Implementation

This chapter show the selected parameter are collected and analysed to identify android permission contain in benign and malware APK file. The output of the data, will be used to construct and generate a graph of benign and malware android permission. Then, the graph can be interpret into percentage of type android permission contain in benign and malware APK file.

Chapter 5: Testing and Analysis

The chapter 5 will be showing the steps and methods in testing and analyzing the data collected. After that, the result of comparative analysis will be discuss and explain in detail.

Chapter 6: Conclusion

The conclusion are make based on summarization of whole work that have done.

1.7 Summary

In conclusion, this research are mainly to identify and analyze the behavior of android malware through static analysis. In the next chapter, research about the malware behavior will be in the form of literature review

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

In this chapter 2, literature review of regarding topics malware shall be discussed. All the findings from the literature review which tell about the malware issues will be gathered and compiled resources from any relevant literatures and thesis, journals, article, books, internet, report or other knowledge resources that are used to write this literature review. Figure 2.1 below show the Literature Review Operational Framework of this research.

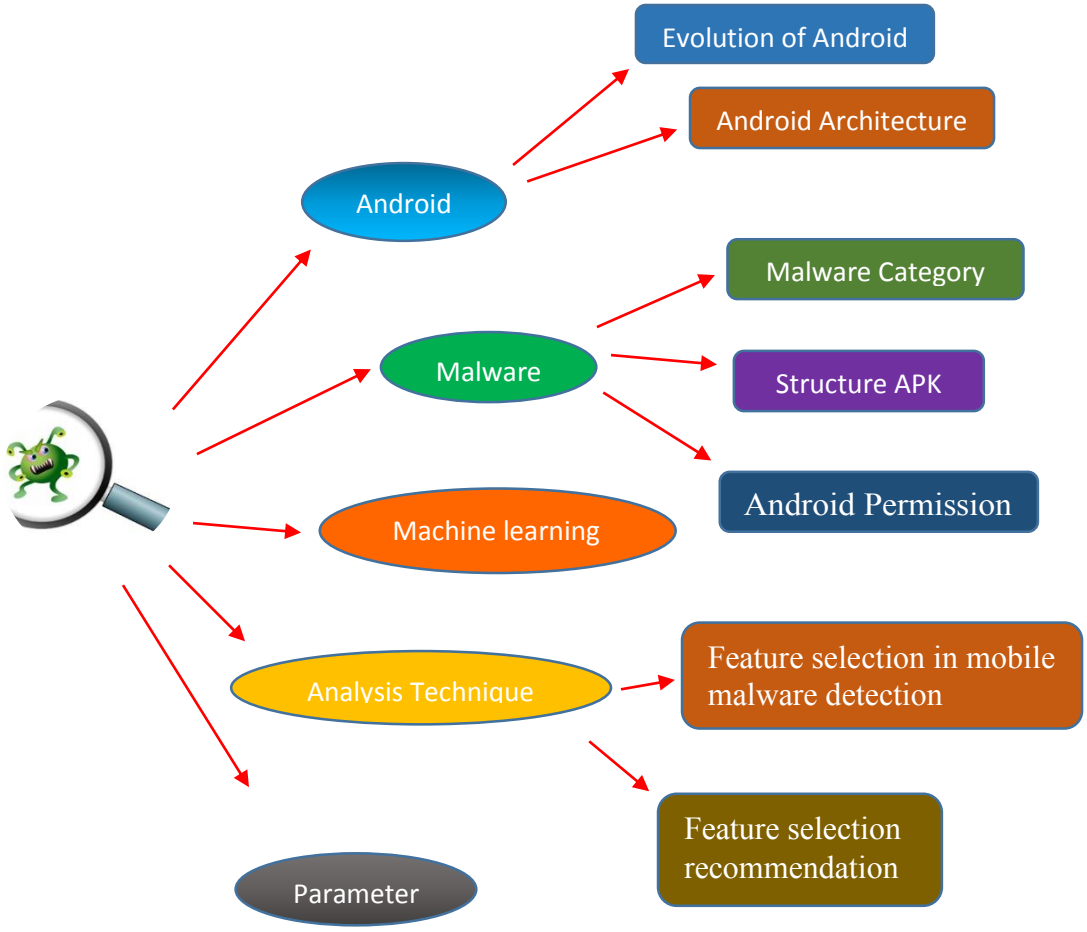


Figure 2.1: Literature Review Operational Framework

2.2 Related Work

Significant amount of research has been done in past to detect malware using windows PE file format. Moreover, many anti-malware vendors have adopted different methods to identify the malicious executables. Researcher and malware analysts have applied different approaches in determining the process to detect malware by using static or dynamic malware analysis technique. Additionally various data mining technique have also been used. The following sections discuss the related work.