ENHANCEMENT OF MALWARE DETECTION TECHNIQUE

TAN TING QIAN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS*

JUDUL: ENHANCEMENT OF MALWARE DETECTION TECHNIQUE

SESI PENGAJIAN: 2015/2016

Saya TAN TING QIAN
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)
Alamat tetap: 12-3-11 BLOK 12,
JALAN 1/1B,
TAMAN INTAN BAIDURI,
52100 KUALA LUMPUR.
Tarikh: 19/8/2016

(TANDATANGAN PENYELIA)
DR. SITI RAHAYU BINTI SELAMAT

Tarikh: 20 AUG 2016

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

ENHANCEMENT OF MALWARE DETECTION TECHNIQUE

TAN TING QIAN

This report is submitted in partial fulfilment of the requirements for the
Bachelor of Computer Science (Computer Security)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2016

DECLARATION

I hereby declare that this project report entitled
ENHANCEMENT OF MALWARE DETECTION TECHNIQUE
is written by me and is my own effort and that no part has been plagiarized
without citations.

STUDENT : ______________________ Date : 19/8/2016
(TAN TING QIAN)

I hereby declare that I have read this project report and found
this project report is sufficient in term of the scope and quality for the award of
Bachelor of Computer Science (Computer Security) With Honours.

SUPERVISOR : ______________________ Date : 20 AUG 2016
(DR. SITI RAHAYU BINTI SELAMAT)

DEDICATION

I would like to dedicate this project to my beloved family for giving me their continuous support, love and affection while I am developing this project. Besides, I would like to dedicate this project to my respected supervisor, Dr Siti Rahayu Binti Selamat who has been a constant source of knowledge and inspiration. Her guidance and expertise have helped me a lot in the process of developing this project.

ACKNOWLEDGEMENT

Thank God for giving me the opportunity and strength to complete this Final Year Project titled Enhancement of Malware Detection Technique.

I would like to express my deepest thanks to my supervisor, Dr. Siti Rahayu binti Selamat for her guidance, advice, support and valuable time in supervising and guiding me to complete this project.

Besides, I would like to thank everyone that helped me by giving me suggestions and advice when I am doing this project. I would like to express my deepest appreciation to my family and friends for their continuous support, understanding and cooperation throughout the final year project.

ABSTRACT

Recently, the number of malware has been growing exponentially. As the number of malware is growing rapidly and they can evade from detection, the difficulty in detecting malware becomes a crucial problem that is faced in the malware detection field. This problem leads to a need of an improved malware detection technique. Thus, the aim of this project is to enhance the malware detection technique. The objectives of this project are to identify the parameter for detecting malware attack and to enhance the malware detection technique. The methodology used to carry out the project involves seven main phases namely literature review phase, data collection phase, analysis phase, design phase, algorithm development phase, testing phase and documentation phase. The main processes involved are such as conducting research on malware detection, collecting network traffic data, analysing the network traffic data collected, designing the experiment and detection algorithm, testing the detection algorithm developed and doing the project documentation. In the end of the project, a detection algorithm that uses the detection parameters identified is developed. As the future works for this project, the detection algorithm may be developed to a more robust system and improved further to cover on the malware prevention.

ABSTRAK

Kebelakangan ini, bilangan *malware* telah berkembang dengan pesat. Oleh kerana bilangan *malware* berkembang pesat dan berkemungkin akan terlepas daripada pengesanan, kesukaran dalam mengesan *malware* telah menjadi masalah yang dihadapi dalam bidang pengesanan *malware*. Masalah ini membawa kepada keperluan untuk teknik pengesanan *malware* yang lebih baik. Oleh itu, tujuan projek ini adalah untuk menambahbaik teknik pengesanan *malware*. Objektif projek ini adalah untuk mengenalpasti *parameter* untuk mengesan serangan *malware* dan menambahbaikan teknik pengesanan *malware*. Kaedah yang digunakan untuk menjalankan projek ini melibatkan tujuh fasa utama iaitu fasa kajian literatur, fasa pengumpulan data, fasa analisis, fasa reka bentuk, fasa pembangunan algoritma, fasa pengujian dan fasa dokumentasi. Proses-proses utama yang terlibat adalah seperti membuat penyelidikan tentang pengesanan *malware*, mengumpul data trafik rangkaian, menganalisis data trafik rangkaian yang telah dikumpul, mereka bentuk eksperimen dan algoritma pengesanan, menguji algoritma pengesanan dan membuat dokumentasi projek. Di akhir projek, algoritma pengesanan yang menggunakan *parameter* pengesanan yang dikenalpasti telah dibangunkan. Sebagai kerja-kerja masa depan untuk projek ini, algoritma pengesanan ini boleh dibangunkan kepada sistem yang lebih mantap dan dipertingkatkan lagi untuk meliputi pencegahan *malware*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIXES

CHAPTER I

INTRODUCTION

1.1 Introduction

Malware or malicious software is a program that is designed to purposely infiltrate or destruct a computer system without the owner's knowledge. It can appear in the form of code, scripts, active content and other software (Ashwini Mujumdar *et al.*, 2013). Malware disrupts computer operations, gathers sensitive information or gains access to private computer systems by infecting the system with malicious code. Malware are software that are created with malicious intent, it does not include software that causes accidental harm due to its vulnerability or deficiency.

Malware detection technique is the technique used to identify or detect malware. It is crucial to detect malware as soon as possible to prevent a system from being used by the malware for malicious purposes. Hence, it is important to have a robust detection technique that can detect as much malware as possible. Signature based detection technique works on the byte sequences and hashes of the known malware. These signatures are saved in a large database of the Antivirus products (Farid Daryabar *et al.*, 2012). Signature based detection detects a malware by matching the signature of the suspected malware to the signature in the database.

On the other hand, behaviour based detection technique is a type of detection technique that concentrates on the behaviour of a program to determine whether it is malicious or not. Since the signature based anti-malware systems are built on the basis of known malware, they cannot detect new, unknown malware or even variants of known malware. Thus, without accurate signatures, this technique cannot effectively detect polymorphic malware (Ashwini Mujumdar *et al.*, 2013). Behaviour based detection can overcome the signature based detection's weakness of unable to detect unknown malware that do not have a signature stored in the database. Therefore, the signature based detection should be enhanced so that it can be more effective in detecting malware.

1.2 Problem Statement (PS)

As the number of malware increases rapidly nowadays, the signature based detection has difficulty in detecting the unknown malware as the signature of the malware might not be created yet when it starts to attack. Due to this problem, the malware can evade the detection technique easily. Thus, the signature based detection should be improved so that it can be more effective in detecting malware.

Table 1.1: Summary of Problem Statement

PS	Problem Statement
PS_1	Difficulty in detecting malware as malware can evade detection technique.

1.3 Project Question (PQ)

The question of the project is how to detect a malware. There are new or unknown malware variants being created every day. Besides, there are polymorphic and metamorphic malware. Polymorphic malware mutates while keeping the original code in contact and metamorphic malware reprogram itself using obfuscation technique where the children malware is different from the parent malware (Mathur and Hiranwal, 2013). These malware can easily evade from signature based detection because for each new variant, it has a different signature. Thus, signature based detection cannot identify them.

Table 1.2: Summary of Project Question

PS	PQ	Project Question
PS_1	PQ_1	How to detect a malware?

1.4 Project Objective (PO)

The objective of this project is to identify the parameter for detecting malware attack. Besides, the other objective of this project is to enhance the malware detection technique so that it can detect both known and unknown attack efficiently. The enhanced malware detection technique will be able to facilitate in the malware detection.

Table 1.3: Summary of Project Objective

PS	PQ	PO	Project Objective
PS_1	PQ_1	PO_1	To identify the parameter for detecting malware attack.
		PO_2	To enhance malware detection technique.

1.5 Project Scope

The scope of this project includes the network traffic data that is used in the analysis phase to get the malware's attributes that will be used as the detection parameters. The malware type that is used is Sasser worm. Tcpdump is used to collect the Sasser worm's network traffic while Wireshark is used to analyse the network traffic collected.

Java programming language is used to develop the detection algorithm. SQLite database is used to store the detection parameters and detection results. This project covers only on detection while prevention is not covered.

1.6 Project Contribution (PC)

Antivirus and Intrusion Detection System companies may benefit from this project because the determined parameter may help these companies to enhance the current detection technique they use in their products. Besides, it might help them to detect malware attack especially unknown malware attack more easily. The second contribution is the enhancement of malware detection technique. As the enhanced detection technique can detect known and unknown attack, it may contribute in helping the investigator on detecting malware more effectively.

Table 1.4: Summary of Project Contribution

PS	PQ	PO	PC	Project Contribution
PS_1	PQ_1	PO_1	PC_1	Parameter for detecting malware attack.
		PO_2	PC_2	Enhancement of malware detection technique.

1.7 Project Methodology

The project methodology includes doing the literature review, data collection, analysis, design, algorithm development, testing and documentation. In literature review phase, the malware, detection process, detection techniques, parameter and network traffic will be analysed and analysis will be done to determine which detection technique should be enhanced so that it can be more effective in detecting malware. The second phase is data collection, the network traffic data of the Sasser worm will be captured and collected.

Then, the data collected from the second phase will be analysed in the third phase which is the analysis phase. The fourth phase is design. The experiment and detection algorithm will be designed in this phase. Next, the detection algorithm will be developed in the fifth phase which is the implementation phase. Then, the sixth phase is testing, the detection algorithm will be tested to determine whether it can detect the Sasser worm’s attack. The last phase is the documentation such as producing final report and final presentation.

1.8 Thesis Organization

Chapter 1: Introduction

This chapter discusses the introduction of the project, problem statement, project question, project objective, project scope and project contribution.

Chapter 2: Literature Review

This chapter consists of the literature review of the main topics in this project. These include the analysis of the malware, detection process, detection techniques, parameter and network traffic data. Justification of the selected malware, detection process, detection technique and parameter used in the project will be presented.

Chapter 3: Project Methodology

This chapter discusses the project methodology and describes the activities that are carried out in each phase in details. A milestone of the project is represented using a table.

Chapter 4: Design

This chapter discusses on the experiment approach that will be used, data analysis process, analysis of Sasser worm attack and the detection algorithm design. The network design will be created and the flowchart of the detection algorithm design will be produced.