UNIVERSITI TEKNOLOGI MARA

POLYMORPHIC MALWARE DETECTION BASED ON DYNAMIC ANALYSIS AND SUPERVISED MACHINE LEARNING

NUR SYUHADA BINTI SELAMAT

MSc

February 2021

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Postgraduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Nur Syuhada Bt Selamat
Student I.D. No.	:	2014663648
Programme	:	Master of Science (Computer Science) -CS750
Faculty	:	Computer and Mathematical Sciences
Thesis Title	:	Polymorphic Malware Detection Based on Dynamic Analysis and Supervised Machine Learning
Signature of Student	:	……..
Date	:	February 2021

ABSTRACT

Currently, the size of malware grows faster each year and poses a thoughtful global security threat. The number of malware developed is increasing as computers became interconnected, at an alarming rate in the 1990s. This scenario caused a rising number of malware. It also caused many protections are developed to fight the malware. The most common method of detecting malware relies on signature-based detection. Unfortunately, this method is no longer effective to handle more advanced malware such as polymorphic malware that poses a thoughtful threat to the modern computing. Malware authors have created them to be more challenging to be evaded from anti-virus scanner. Extracting the behaviour of polymorphic malware is one of the major issues that affect the detection result.The main idea in this work is focus the behaviour(dynamic) of polymorphic malware infect in computer system and to extract feature selection and evaluate a limited set of dataset in order to improve detection of polymorphic malware.This study used dynamic analysis and machine learning to improve malware detection.This research demonstrated improved polymorphic malware detection can be achieved with machine learning.This research used four types of machine algorithm which are K-Nearest Neighbours, Decision Tree, Logistic Regression, and Random Forest. As with most studies,careful attention was paid to false positive and false negative rates which reduce their overall detection accuracy and effectiveness.The result showed that the Random Forest algorithm is the best detection accuracy compares to others classifier with 99 % on a relatively small dataset. The benefit of this work indicated that the implementation of a feature selection technique plays an important role in machine learning algorithms to increase the performance of detection.

ACKNOWLEDGEMENT

Firstly, I wish to thank God for giving me the opportunity to embark on my MSc and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Dr Fakariah Hani Hj Mohd Ali and also my co-supervisor Pn Noor Ashitah Abu Othman.

My appreciation goes to the F-Secure team members who provided the facilities and assistance during sampling. Special thanks to my colleagues and friends for helping me with this project.

Finally, this thesis is dedicated to my father, mother, beloved husband and daughters for the vision and determination to educate me. This piece of victory is dedicated to all of them. Alhamdulilah.

TABLE OF CONTENTS