

**A STATIC AND DYNAMIC HEURISTIC ANALYSIS TO CALCULATE
MALWARE THREAT SCORE**

By
Reischaga
11602015

BACHELOR'S DEGREE
in

INFORMATION TECHNOLOGY
ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY
Alam Sutera
Tangerang 15339
Indonesia

June 2020

Revision after the Thesis Defense on 14th July 2020

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in this thesis.

Reischaga

Student

Date

Approved by:

Dr. Charles Lim, M.Sc.

Thesis Advisor

Date

Mr. Yohanes Syailendra Kotualubun, M.Kom.

Thesis Co-Advisor

Date

Maulahikmah Galinium, S.Kom, M.Sc, PhD,

Dean

Date

Reischaga

ABSTRACT

A STATIC AND DYNAMIC HEURISTIC ANALYSIS TO CALCULATE MALWARE THREAT SCORE

By

Reischaga

Dr. Charles Lim, M.Sc., Advisor

Mr. Yohanes Syailendra Kotualubun, M.Kom., Co-Advisor

SWISS GERMAN UNIVERSITY

Malware can be defined as any software that performs malicious actions without the consent of the user. The volume of malware is increasing every year and the threat it poses is only becoming ever more prevalent. Currently, malware is involved in a large portion of security incidents, of which are currently handled without a threat level metric system. In this study, we demonstrate how static analysis and dynamic analysis could be utilized, both separately and combined, effectively in order to quantitatively assess malware threats by calculating their threat scores.

Keywords: Malware, Dynamic Analysis, Static Analysis, ATT&CK Framework, Threat Scoring



SWISS GERMAN UNIVERSITY

DEDICATION

I would like to dedicate this research project to my family and my beloved country Indonesia. Additionally, I wish to dedicate this work as a projection of the knowledge and mentor-ship given to me by my teachers and seniors, whom have tirelessly and selflessly taught me and answered my questions during the course of this research.



ACKNOWLEDGEMENT

First and foremost is to thank my mentors, family, friends, and God in giving relentless support until the completion of this thesis. I would like to express my gratitude to my advisor, as well as co-advisor for their support, patience and time. Dr. Charles Lim, M.Sc. was very helpful in guiding in analytical studies and gave very helpful insights and guidance in malware, as well as up to date situations and trends in the research world. Mr. Yohanes Syailendra Kotualubun, M.Kom. as my co-advisor had been very helpful in giving guidance and knowledge regarding static and dynamic analysis experimentations and scientific writing. Mr. Mario Marcello, S.Kom., B.Eng. and Alexander Suhandi, ST, M.Kom. had been helpful in providing insights into dynamic analysis. Last but not least, I would like to thank Mr. Ali Suwanda, S.Kom. as my research partner for his support throughout this thesis.



SWISS GERMAN UNIVERSITY

TABLE OF CONTENTS

STATEMENT BY THE AUTHOR	2
ABSTRACT	3
DEDICATION	5
ACKNOWLEDGEMENT	6
CONTENTS	11
LIST OF FIGURES	12
LIST OF TABLES	13
1 INTRODUCTION	14
1.1 Background	14
1.2 Problem Statement	16
1.3 Research Objective	16
1.4 Research Question	16
1.5 Hypothesis	17
1.6 Scope and Limitations	17
1.7 Significance of Study	17
1.8 Thesis Structure	18
2 LITERATURE REVIEW	19
2.1 Malware	19
2.1.1 Definition	19
2.1.2 Brief History	20
2.1.3 Malware Classification and Category	22
2.1.3.1 APT Malware	23
2.1.3.2 Non-APT Malware	23
2.1.3.3 Fileless Malware	23
2.1.3.4 File-Based Malware	24
2.1.4 Malware Evasion Techniques	24

2.1.4.1	Obfuscation	25
2.1.4.2	Fragmentation and Session Splicing	25
2.1.4.3	Protocol Violations	25
2.1.4.4	Inserting Traffic at IDS	25
2.1.4.5	Denial of Service (DOS)	26
2.1.4.6	Code Reuse Attacks	26
2.1.4.7	Packers	26
2.1.5	The Importance and Necessity of Accurate Malware Analysis and Threat Scoring	26
2.1.6	Malicious Software vs. Suspicious Software	27
2.2	Static Analysis	27
2.2.1	Entropy-Based Static Analysis	29
2.2.2	Portable Executable (PE) File Header-Based Static Analysis	29
2.3	Dynamic Analysis	31
2.3.1	Function Call Monitoring	31
2.3.1.1	Application Programming Interface (API) Calls	32
2.3.1.2	System Calls	32
2.3.2	Function Parameter Analysis	32
2.3.3	Information Flow Tracking	33
2.3.4	Instruction Trace	33
2.3.5	Autostart Extensibility Points (ASEPs)	34
2.4	Sandbox Evasion Techniques	34
2.4.1	Detection-Dependant Evasion	34
2.4.1.1	Fingerprinting	34
2.4.1.2	Reverse Turing Test	36
2.4.1.3	Targeted	36
2.4.2	Detection-Independent Evasion	36
2.4.2.1	Stalling	36
2.4.2.2	Trigger Based	37
2.4.2.3	Fileless Malware	38
2.5	Heuristic Analysis	38
2.6	Graph Theory	39
2.7	Cyber Situational Awareness (SA)	39
2.8	MITRE ATT&CK Framework	40
2.9	Malware Analysis Tools and Frameworks	41

2.9.1	Static Analysis Tools	41
2.9.2	Dynamic Analysis Engine	44
2.9.2.1	Malware Dynamic Analysis in a Virtual Machine (VM)	44
2.9.2.2	Malware Dynamic Analysis in an Emulator	45
2.9.2.3	Cuckoo Sandbox	46
2.10	Our Approach and Position of our Study	47
2.11	Summary	49
3	RESEARCH METHODOLOGY	50
3.1	Overview	50
3.2	Framework of the Methodology	50
3.3	General Architecture	51
3.3.1	Static Analysis	52
3.3.1.1	Static Analysis Features	53
3.3.1.2	Features Extraction and Validation	55
3.3.2	Dynamic Analysis	55
3.3.2.1	Dynamic Analysis Features	57
3.3.2.2	Features Extraction	58
3.4	Data Collection	58
3.4.1	Malicious Software Dataset	58
3.4.2	Benign Software Dataset	58
3.5	Evaluation	58
3.5.1	Static Analysis Result Evaluation	59
3.5.2	Dynamic Analysis Result Evaluation	60
3.5.3	Final Evaluation	61
3.5.4	Classification	61
3.6	Validation and Accuracy	62
4	EXPERIMENTAL RESULTS	63
4.1	Experiment Setup	63
4.1.1	Hardware Setup	63
4.1.2	Software Setup	65
4.1.3	Malware Datasets	65
4.1.4	Benign Datasets	66
4.2	Preliminary Experiment	66
4.2.1	Packers Effects on Malware	66

4.2.2	Preliminary MITRE ATT&CK Mapping and Experiment	71
4.2.3	Imphash Experiment	77
4.2.4	Malware Similarity based on Static Analysis vs. Dynamic Analysis	79
4.2.4.1	IAT-based Similarity Analysis	79
4.2.4.2	Graph-based Similarity Analysis	80
4.2.4.3	Observation	81
4.3	Static Analysis Result	82
4.3.1	Static Threat Scoring Formula	82
4.3.2	Experiment Result	82
4.3.3	Experiment Result Analysis	83
4.4	Dynamic Analysis Result	84
4.4.1	Dynamic Threat Scoring Formula	84
4.4.2	Experiment Result	84
4.4.3	Experiment Result Analysis	85
4.5	Combining Static Analysis and Dynamic Analysis	86
4.5.1	Final Threat Scoring Formula	86
4.5.2	Experiment Result	88
4.5.3	Validation	90
4.5.4	Experiment Result Analysis	90
4.6	Performance Measurement	91
4.7	Summary	91
5	CONCLUSION	92
5.1	Accomplishments	92
5.2	Limitations	93
5.2.1	Static Analysis	93
5.2.2	Dynamic Analysis	93
5.3	Future Works	93
5.3.1	Malware Dataset	93
5.3.2	Static Analysis-based Features	93
5.3.3	Kernel Level Data for Dynamic Analysis	94
5.3.4	MITRE ATT&CK Mapping	94
5.3.5	Graph-based Analysis and Database	94
	GLOSSARY	95

REFERENCES	105
APPENDICES	63
Appendix A Malware Datasets	106
Appendix B Benignware Dataset	120
Appendix C Source Codes	122
C.1 Static Analysis-based Threat Scoring Source Code	122
C.2 Dynamic Analysis-based Threat Scoring Source Code - Features Ex- traction	129
C.3 Dynamic Analysis-based Threat Scoring Source Code - Evaluating Ex- tracted Features	131
C.4 Dynamic Analysis - High-level Features Extraction	136
C.5 Static Analysis - Extract Imphash	139
C.6 Static Analysis - Extract Content of IAT	140
C.7 Static Analysis - IAT-based Similarity Analysis and Construct Heat Map	141
C.8 Dynamic Analysis - Graph-based Similarity Analysis and Construct Heat Map	143
C.9 Preliminary MITRE Experiment - MITRE Techniques Mapping	145
Appendix D Submitted Paper based on this Thesis	151
CURRICULUM VITAE	158

List of Figures

1.1	Total number of consumer and business detections in 2019 vs. 2018(Malwarebytes Labs, 2020)	14
1.2	Extrapolated average malware alerts for participating organizations (Ponemon Institute LLC, 2015)	15
2.1	Windows PE File Format	29
2.2	Cmd.exe PE header (Dubyk, 2020)	30
2.3	System call visualization	32
3.1	Overall Framework	50
3.2	Flow of analysis and threat Scoring framework	51
3.3	The static analysis architecture	53
3.4	The dynamic analysis architecture	56
4.1	Custom sections found on DNSChanger	70
4.2	Imphash experiment result on malware dataset	78
4.3	Imphash experiment result on benign dataset	78
4.4	IAT-based similarity experiment result of samples from family Emotet.PB	80
4.5	Graph constructed from the dynamic analysis result extracted from cmd.exe	80
4.6	High level dynamic analysis-based similarity experiment result of samples from family Emotet.PB	81
4.7	Threat scoring result on malware samples based on static analysis	83
4.8	Threat scoring result on benign samples based on static analysis	83
4.9	Threat scoring result on malware samples based on dynamic analysis	85
4.10	Threat scoring result on benign samples based on dynamic analysis	85
4.11	Final Threat Scoring Result on Malware Samples	89
4.12	Threat scoring result on benign samples based on hybrid analysis	89

List of Tables

2.1	Static features tools comparison	43
2.2	Comparison of our approach and the analysis methods of other works	48
3.1	Static features criteria	60
3.2	Static features criteria explanation	60
4.1	Hardware setup	64
4.2	Software setup	65
4.3	Malware dataset families and quantities	66
4.4	Static analysis on malware samples packed by UPX	68
4.5	Static analysis on malware unpacked samples	69
4.6	SHA256 hash of samples pre-packing operation and post-packing operation	70
4.7	MITRE techniques mapping to Cuckoo report	72
4.8	MITRE techniques mapping to Cuckoo report (contd.)	73
4.9	Tactics and techniques used by a Ransom/Win32:Cerber.J sample	74
4.10	Tactics and techniques used by a Ransom/Win32:Cerber!rfn sample	75
4.11	Tactics and techniques used by multiple benign samples	76
4.12	Example of Static Analysis-based Threat Score Result	82
4.13	Example of Dynamic Analysis-based Threat Scoring	84
4.14	Weighting and Threshold Experiment Result	87
4.15	Applying Different Weighting on Samples	88
4.16	Example of Final Threat Scoring	89
4.17	Final Threat Scoring Classification Result	90
4.18	Final Threat Scoring Classification Accuracy	90