



NEW P2P BOTNETS CLASSIFICATION AND DETECTION FRAMEWORK

RAIHANA SYAHIRAH BINTI ABDULLAH

DOCTOR OF PHILOSOPHY

2016



Faculty of Information and Communication Technology

**NEW P2P BOTNETS CLASSIFICATION AND DETECTION
FRAMEWORK**

Raihana Syahirah Binti Abdullah

Doctor of Philosophy

2016

NEW P2P BOTNETS CLASSIFICATION AND DETECTION FRAMEWORK

RAIHANA SYAHIRAH BINTI ABDULLAH

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

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2016

DECLARATION

I declare that this thesis entitled “New P2P Botnets Classification and Detection Framework” is the result of my own research 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 : **RAIHANA SYAHIRAH BINTI ABDULLAH**

Date :

APPROVAL

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

Signature :

Supervisor Name : **PM DR. MOHD FAIZAL BIN ABDOLLAH**

Date :

ABSTRACT

Botnets is a tool for high-profile cyber-attack. It is a collection of compromised computer infected with advance malware that allows an attacker to remotely control them. Some botnets used Peer to Peer (P2P) protocols and Peer to Peer (P2P) technology to control computers and exploits users. They are known as P2P Botnets. The unification of botnets and P2P technology make it more powerful and robust to be detected. Latest P2P botnets caused crisis and chaos to the network security. In order to deal with the issue, framework is needed to illustrate and explain the modules, terminologies and procedures as an important parts to implement the detection. But, the current P2P botnets detection frameworks are still not comprehensive enough to recognize the emergence of latest P2P botnets that cause financial loss and data damage to the network of the organization. Previous frameworks are incomplete and contained many of limitations which require some improvement. Lower detection rate and higher false alarms increase the failure of botnets detection. Hence, higher false alarm significantly causes ineffectiveness of detection. Due to the issues faced to identify the P2P botnets activities, the main objective of this research is to enhance P2P botnets detection framework using integrated approach. A complete analysis flow is performed to detect and classify the P2P botnets by adopting integrated analyser and integrated analysis. Besides developing a new framework, the research analysis classifies the behaviour of P2P botnets in order to differentiate between the P2P normal and P2P botnets. Through classification, this research introduces a generic P2P attack pattern and P2P behavioural model. Both generic P2P attack pattern and P2P behavioural model are then applied to develop the integrated approach that is used to validate the new P2P botnets detection. In evaluation and validation, the results showed that a new P2P botnets detection framework has effectively obtained high accuracy, high detection rates and lower false alarm. Significantly, the process of finding, identifying, classifying and detecting the P2P botnets is collaborated with Cyber Security Malaysia. Hence, this research introduces an enhancement framework to detect P2P botnets activities and validated by integrated approach that helps the network administrator to identify the existence of P2P botnets.

Comment [RSBA1]: 1.Botnets definition are added
2.Botnets and P2P are being relates

Comment [RSBA2]: Word 'improvisation' is changed to 'improvement'

ABSTRAK

Botnets atau lebih dikenali sebagai malware khusus merupakan serangan siber yang berprofil tinggi pada masa kini. Botnets merupakan gabungan komputer yang dijangkiti oleh malware khusus dan membenarkan penyerang mengawalnya secara jauh. Botnets ini juga menggunakan teknologi P2P sebagai protokol utama membolehkan pengawalan dan peneksplotasian berlaku terhadap pengguna pengguna. Ianya dikenali sebagai P2P botnets. Penggabungan botnets dengan teknologi P2P membuatkan botnets lebih sukar untuk dikesan. Jaringan botnets yang meluas menimbulkan fenomena krisis yang meruncing dalam keselamatan rangkaian. Rangka kerja pengesanan pada masa kini masih tidak begitu komprehensif untuk mengenalpasti kehadiran P2P botnets yang memberi impak yang negatif pada sistem kewangan dan rangkaian data dalam sesebuah organisasi. Hal ini menunjukkan rangka kerja terdahulu masih mempunyai kelemahan dan memerlukan penambahbaikan segera. Kadar kesilapan dalam proses pengesanan ditentukan melalui pengesanan kadar pengurangan amaran yang tinggi. Sekiranya kadar amaran melonjak pada angka yang tinggi, maka ini menunjukkan pengesanan tersebut adalah gagal. Berdasarkan masalah yang dihadapi, kajian ini mengusulkan idea baru bagi memperkenalkan rangka kerja baru yang lebih efektif untuk mengenalpasti aktiviti P2P botnets dalam sesebuah rangkaian. Justeru, objektif utama kajian ini adalah untuk memperkenalkan rangka kerja lengkap pengesanan P2P botnets menerusi penggabungan beberapa kaedah yang relevan secara hybrid. Satu analisis lengkap akan dipraktikkan untuk proses pengesanan dan pengecaman aktiviti P2P botnets ini dengan gabungan analisis dan gabungan penganalisis. Selain membangunkan rangka kerja baru, kajian ini akan mengklasifikasikan ciri-ciri dalam P2P botnets untuk membezakan antara P2P normal dan P2P botnets. Proses pengklasifikasian ini juga membolehkan kajian ini turut memperkenalkan paten serangan P2P dan model umum P2P. Kedua-dua paten dan model ini amatlah berguna untuk diaplikasikan dalam pembangunan kaedah gabungan pengesanan. Dalam proses penilaian dan pengesanan, keputusan yang ditunjukkan adalah baik iaitu kadar tinggi untuk proses pengesanan serangan dan kadar rendah untuk amaran. Secara hakikinya, dapatan dari proses kenalpasti dan klasifikasi pada P2P botnets ini dilaksanakan melalui kerjasama dengan pihak Cyber Security Malaysia. Oleh itu, kajian ini akan memperkenalkan satu rangka kerja baru untuk mengesan segala aktiviti P2P botnets ditentusahkan oleh pendekatan gabungan teknik hybrid yang dapat membantu pentadbir rangkaian untuk mengenalpasti kewujudan P2P botnets dalam sesebuah rangkaian.

ACKNOWLEDGEMENTS

Bismillahirrahmanirrahim... In the name of Allah, Most Gracious, Most Merciful

First and foremost, I would like to thank Allah Almighty for giving me excellent health, ideas and comfortable environment so that I can complete this thesis as scheduled.

My greatest thanks is to my mother (Norhayati), my father (Abdullah), my husband (Wan Sofhi) and my siblings (Hazwan Anwar, Syafiq Muzhafar, Idlan Muqri, 'Izzat Fahmi) for their continuous understanding, motivation, encouragement and patience throughout my PhD journey.

I would like to express my sincere appreciation to Assoc. Prof Dr Faizal Bin Abdollah for his excellent guidance, supervision, motivation, encouragement, patience and insight throughout the years of this PhD's endless journey.

I would like to extend my thanks to the staff of Fakulti Teknologi Maklumat dan Komunikasi (FTMK) and Pusat Pengajian Siswazah (PPS UTeM) for their time, guidance and support during my studies. Greatest appreciation goes to Universiti Teknikal Malaysia Melaka (UTeM) and Kementerian Pengajian Tinggi Malaysia for sponsoring this study.

Lastly, but in no sense the least, I am thankful to all colleagues and friends especially Nurulhuda Ahmad, Syazwani Yahya, Dr ZulAzri, Mohd Zaki Mas'ud, Dr Siti Rahayu, Dr Robiah for their valuable time, understanding, suggestions, comments and continuous motivation which made my PhD years a memorable and valuable experience.

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATION	
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF APPENDICES	vii
LIST OF ABBREVIATIONS	viii
LIST OF PUBLICATIONS	ix
CHAPTER	
1. INTRODUCTION	1
1.1 Introduction	1
1.2 Research Problems	4
1.3 Research Questions	9
1.4 Research Aim and Objectives	7
1.5 Research Scope	10
1.6 Research Significance	10
1.7 Thesis Organization	11
2. LITERATURE REVIEW	17
2.1 Introduction	17
2.2 Chapter Objective	17
2.3 Chapter Outline	18
2.4 Overview on Botnets and P2P Botnets Issues	19
2.4.1 Security Trends in Malaysia	22
2.4.2 Botnets and P2P Botnets Evolution	25
2.4.3 Botnets Taxonomy	29
2.5 P2P Botnets Intrusion Detection Problem	32
2.5.1 Overview on Intrusion Detection System (IDS)	32
2.5.2 Intrusion Detection System (IDS) Current Problem	34
2.6 Current Approach Used in Detecting Botnets and P2P Botnets	36
2.6.1 Overview on Botnets and P2P Botnets Detection Techniques	36
2.6.2 Overview on botnets and P2P Botnets Criterion	44
2.6.3 Overview on Botnets and P2P Botnets Analyser	45
2.6.4 Overview on Botnets and P2P Botnets Analysis Approach	48
2.7 Overview on P2P Botnets Detection Framework	50
2.7.1 Related Work on P2P Botnets Detection Framework	51
2.7.2 Analysis on Capability Criterion Current P2P Botnets	58
2.7.3 Analysis on P2P Botnets Framework Terminology	63
2.8 Technique Used for the Research	73
2.8.1 Data Mining-based	73

2.8.2	Signature-based	76
2.8.3	Anomaly-based	77
2.9	Overview on Evaluation and Validation Issues	80
2.9.1	Related Work on Metrics	81
2.9.2	Related Work on Datasets	83
2.9.3	Related Work on Comparison Approach	85
2.10	Summary	86
3.	RESEARCH METHODOLOGY	87
3.1	Introduction	87
3.2	Chapter Objective	87
3.3	Chapter Outline	87
3.4	Research Methodology	88
3.5	Research Activities	96
3.6	Summary	98
4.	ANALYSIS	99
4.1	Introduction	99
4.2	Chapter Objective	100
4.3	Chapter Outline	100
4.4	Analysis on Related Work on P2P Botnets	102
4.5	Data Preparation	105
4.5.1	Overview on Generic P2P Botnets Experimental Approach	105
4.5.2	P2P Botnets Data Collection	111
4.6	P2P Botnets Integrated Analyser Process Flow	112
4.6.1	P2P Botnets Host Analyser Process Flow	113
4.6.2	P2P Botnets Network Analyser Process Flow	115
4.7	Analysis and Findings on P2P Botnets Integrated Analyser	116
4.7.1	P2P Botnets Scenario	116
4.7.2	Attack Pattern Analysis: Host-Level Perspective	117
4.7.3	Attack Pattern Analysis: Network-Level Perspective	120
4.8	Proposed Generic P2P Botnets Attack Pattern	123
4.8.1	Generic P2P Botnets Attack Pattern for Host-Level	123
4.8.2	Generic P2P Botnets Attack Pattern for Network-Level	125
4.9	P2P Botnets Integrated Analysis Approach	126
4.10	Proposed General P2P Botnets Behaviour Model	136
4.11	Proposed Parameter for Detecting P2P Botnets	139
4.12	Summary	142
5.	DESIGN AND IMPLEMENTATION	144
5.1	Introduction	144
5.2	Chapter Objective	144
5.3	Chapter Outline	145
5.4	Model of P2P Botnets Detection	145
5.5	Proposed the Enhancement of P2P Botnets Detection Framework and Integrated Technique	147
5.5.1	Layer 1: Input Layer	148
5.5.2	Layer 2: Mining Data Layer	150

5.5.3	Layer 3: Analyser Layer	155
5.5.4	Layer 4: Analysis Layer	160
5.5.5	Layer 5: Detection Layer	162
5.5.6	Layer 6: Output Layer	192
5.6	Summary	192
6.	TESTING AND RESULT VALIDATION	194
6.1	Introduction	194
6.2	Chapter Objectives	194
6.3	Chapter Outline	195
6.4	Data Preparation	196
6.5	Testing Design	197
6.6	Result Validation	199
6.6.1	Proposed P2P Botnets Detection Technique	199
6.6.2	Previous Technique and Framework	202
6.6.3	Snort as Commercial Product	202
6.7	Summary of Result Testing	203
6.8	Summary	208
7.	CONCLUSION	210
7.1	Introduction	210
7.2	Research Summarization	210
7.3	Research Contribution	213
7.4	Research Limitation	216
7.5	Future Research	217
7.6	Discussion	218
	REFERENCES	220
	APPENDICES	251

LIST OF TABLES

TABLE	TITLE	PAGE
1.1	Number of Security Incidents for 2005 - 2015 excluding Spam	2
1.2	Summary of Research Problems	8
2.1	Comparison of Security Incidents between Q3/2012, Q1/2013 and Q2/2013	24
2.2	Description of Botnets Attack Type	25
2.3	Evolution of Botnets Generation Landscape	29
2.4	Summary of Botnets and P2P Botnets Techniques Classification	41
2.5	Detection Criteria	45
2.6	Summary of Analyser	47
2.7	Summary of P2P Botnets Detection Approach	49
2.8	P2P Botnets Detection Criterion	60
2.9	General Terminology to Describe the Module in P2P Botnets Detection Framework	64
2.10	Analysis of Modules Involved in P2P Botnets Detection Framework	67
2.11	Example of Classification Table	74
2.12	Classification Table	81
2.13	Analysis of Metrics Used by Previous Researcher	82

2.14	List of Six P2P Botnets Files	83
2.15	Summary of Snort	85
3.1	The Relationship between Research Questions (RQ), Research Objectives (RO) and Research Contribution	96
4.1	Summary of Hardware and Software Used in the Experiment	108
4.2	P2P Traffic Dataset Summary	111
4.3	Summary on P2P Botnets Attack Pattern for Host-Level	118
4.4	Summary on P2P Botnets Attack Pattern for Network-Level	120
4.5	Summary on Generic P2P Botnets Attack Pattern for Host-Level	123
4.6	Summary on Generic P2P Botnets Attack Pattern for Network- Level	125
4.7	Operation Process by P2P Botnets	131
4.8	Parameters Description	142
5.1	List of P2P Botnets Variants in Monitoring Module	150
5.2	Selected Attributes	151
5.3	Summary of Classification Model for Each Variant	154
5.4	Summary of Detection Rate for Each Variant	155
5.5	Classified of Botnets Attack in Host Analyser	156
5.6	List of Favourite Locations and Path of P2P Botnets Attack	157
5.7	Classified of P2P Botnets Impact/Effect in Network Analyser	158
5.8	List of Suspicious Port and Suspicious IP Address of P2P Botnets Attack	159
5.9	Integration of Analysis Approach	161

5.10	Hardware Requirement	167
5.11	Software Requirement	167
5.12	Description of Control Caption in Signature-based Module	169
5.13	Signature-based Module Detection Result	170
5.14	Time Distribution in Palevo Dataset	177
5.15	Average Packet Distribution and Relative Frequencies in Palevo Dataset	178
5.16	Computation of Chi-Square Test Statistic for the First Time Slot in Palevo Dataset	179
5.17	Computation of Chi-Square Test Statistic for the Last Time Slot in Palevo Dataset	180
5.18	Time Distribution in Invalid Hash Dataset	180
5.19	Average Packet Distribution and Relative Frequencies in Invalid Hash Dataset	181
5.20	Computation of Chi-Square Statistic for the First Time Slot in Invalid Hash Dataset	182
5.21	Computation of Chi-Square Statistic for the Last Time Slot in Invalid Hash Dataset	182
5.22	Time Distribution in RBot Dataset	183
5.23	Average Packet Distribution and Relative Frequencies in RBot Dataset	183
5.24	Computation of Chi-Square Test Statistic for the Test of First Time Slot in RBot Dataset	185

5.25	Computation of Chi-Square Test Statistic for the Test of Last Time Slot in RBot Dataset	185
5.26	Time Distribution in Allapple.L Dataset	186
5.27	Average Packet Distribution and Relative Frequencies in Allapple.L Dataset	186
5.28	Computation of Chi-Square Test Statistic for the First Time Slot in Allapple.L Dataset	187
5.29	Computation of Chi-Square Test Statistic for the Last Time Slot in Allapple.L Dataset	187
5.30	Time Distribution in Srvcp Dataset	188
5.31	Average Packet Distribution and Relative Frequencies in Srvcp Dataset	188
5.32	Computation of Chi-Square Test Statistic for the First Time Slot in Srvcp Dataset	190
5.33	Computation of Chi-Square Test Statistic for the Last Time Slot in Srvcp Dataset	190
5.34	Combination of Signature-based and Anomaly-based Detection Result	191
6.1	Summary of Datasets for Evaluation	196
6.2	Testing Result for Proposed Technique	200
6.3	Testing Result on Integrated-based Detection	201
6.4	Testing Result for a Commercial Product (Snort)	203

6.5	Comparison between Preliminary Research Result and Testing Result	203
6.6	Summary of Accuracy Detection	204
6.7	Summary of Overall Detection Rate	205
6.8	Summary of False Alarm Rate	207
6.9	Summary of Evaluation using Metric	208

LIST OF FIGURES

FIGURE	TITLE	PAGE
2.1	Structure of Chapter 2	18
2.2	Global Botnet Infections from McAfee Threat	19
2.3	Global Internet Traffic Trends 2006-2012 by IBM	21
2.4	Comparison of Botnets Drones Pattern in 2009 and 2011	23
2.5	Timeline of Botnets Evolution	27
2.6	Evolution of P2P Botnet Technology	28
2.7	Botnets Scenarios	30
2.8	General Botnets Taxonomy	31
2.9	Botnets Architecture	32
2.10	Components of IDS	33
2.11	Botnets Detection Technique	37
2.12	Types of Hybrid Approach	48
2.13	P2P-based Detection Framework	51
2.14	C&C Protocol-Independent Detection Framework	52
2.15	Botnets Detection Approach Architecture	54
2.16	General P2P Botnets Detection Framework	55
2.17	P2P Botnets Detection System	55
2.18	P2P Detection Framework	56

2.19	IPS-Based P2P Botnets Detection	57
2.20	Proposed Integrated P2P Botnets Detection Technique	72
3.1	The Structure of Chapter 3	88
3.2	Main Phases of Research Methodology	88
3.3	Preliminary Analysis Stage in Analysis Phase	90
3.4	Identifying Attack Pattern Experiment Stage in Analysis Phase	91
3.5	Design Phase	91
3.6	Development Phase	92
3.7	Implementation Phase	93
3.8	Testing and Evaluation Phase	94
3.9	Summary of Overall Detail Research Methodology	95
4.1	The Structure of Chapter 4	101
4.2	Testbed Experiment Approach for P2P Botnets Analysis	106
4.3	P2P Testbed Environment	107
4.4	Integrated Analyser Process	113
4.5	Host-Level Analysis Process	114
4.6	Network-Level Analysis Process	115
4.7	P2P Botnets Attack Scenario	117
4.8	P2P Botnets Attack Pattern in Host-Level	119
4.9	P2P Botnets Attack Pattern in Network-Level	121
4.10	General P2P Botnets Attack Pattern in Host-Level	124
4.11	General P2P Botnets Attack Pattern in Network-Level	126
4.12	P2P Testbed Environment Analysis Design	127

4.13	Palevo Operation Process in Static Analysis	132
4.14	Palevo Dynamic Analysis	133
4.15	Correlation Process in RBot Summary	134
4.16	Proposed P2P Botnets Integrated Analysis Approach	135
4.17	Basic P2P Botnets Behaviour Model	137
4.18	General P2P Botnets Behaviour Model	138
5.1	The Structure of Chapter 5	145
5.2	Process Flow of P2P Botnets Detection	146
5.3	Proposed P2P Botnets Detection Framework Architecture	147
5.4	P2P Activity Capturing Process	149
5.5	Process Flow of SVM Data Mining Module	152
5.6	P2P Botnets Detection Module	162
5.7	Process Flow of Overall P2P Botnets Detection	163
5.8	Process Flow of Signature Module	165
5.9	Design of Signature-based Module	168
5.10	Process Flow of Statistical Test	174
6.1	The Structure of Chapter 6	195
6.2	Testing and Validation Procedure	197
6.3	Conclusion of Result Analysis against P2P botnets dataset	199

LIST OF ABBREVIATIONS

A	- Accuracy
ACK	- Acknowledge
AIS	- Artificial Immune System
C & C	- Command and Control
CEO	- Chief Executive Officer
CERT-RO	- Computer Emergency Response Team Romania
CI	- Computational Intelligence
CSI	- Computer Security Institute
DDNS	- Dynamic Domain Name Systems
DDoS	- Distributed Denial of Service
DNS	- Domain Name System
DR	- Detection Rate
ECE	- ECN Echo
FAR	- False Alarm Rate
FIN	- Finish
FN	- False Negative
FP	- False Positive
FTP	- File Transfer Protocol
GTBot	- Global Threat Botnets

HTTP	- Hypertext Transfer Protocol
Integrated P2P_DT	- Integrated Peer to Peer Detection Technique
Integrated_SAM	- Integrated Signature-based Anomaly-based Mining-based
IBM	- International Business Machines
ICMP	- Internet Control Message Protocol
IDS	- Intrusion Detection Systems
IM	- Instant Messaging
IP	- Internet Protocol
IRC	- Internet Relay Chat
ISS	- Internet Security Systems
MITM	- Man in the Middle
MyCERT	- Malaysia Computer Emergency Response Team
NetBEUI	- NetBIOS Extended User Interface
OSI	- Open System Interconnection
P2P	- Peer-to-Peer
PC	- Personal Computer
PDH	- Push
QoS	- Quality of Service
RFC	- Requests for Comments
RST	- Reset
SVM	- Support Vector Machine
SYN	- Synchronize
TCP	- Transmission Control Protocol

TCP/IP	-	Transmission Control Protocol and Internet Protocol
Td	-	Delay Time
TN	-	True Negative
TP	-	True Positive
UDP	-	User Datagram Protocol
URG	-	Urgent

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Literature Review on P2P_DT	211
B	Literature Review on Detection Criteria in P2P_DT	215
C	Static Code Analysis	216
D	Chi-Square Table	224

LIST OF PUBLICATIONS

Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh, 2016: P2P Botnets Detection Module through Hybrid Approach. *Proceedings of the 5th International Cryptology and Information Security Conference (CRYPTOLOGY)*, Kota Kinabalu, Sabah.

Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh, Mohd Zaki Mas'ud, Siti Rahayu Selamat, Shahrin Sahib, 2013. Enhanced P2P Botnets Detection Framework Architecture with Hybrid Analyzer: Host-based and Network-based. *Information and Assurance Conference (AIS)*, Tunisia, pp. 72-77.

Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh, Mohd Zaki Mas'ud, Robiah Yusof, Shahrin Sahib, 2013. Preliminary Study of Host and Network-based Analysis on P2P Botnets Detection. *International Conference on Technology, Informatics, Management, Engineering and Environment (TIME-E)*, Bandung, Indonesia, pp. 105-109.

Raihana Syahirah Abdullah, Faizal M.A., Zul Azri Muhamad Noh, 2016: Automated Simulation P2P Botnets Signature Detection by Rule-based Approach. *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 7, No. 8, pp. 131-135.