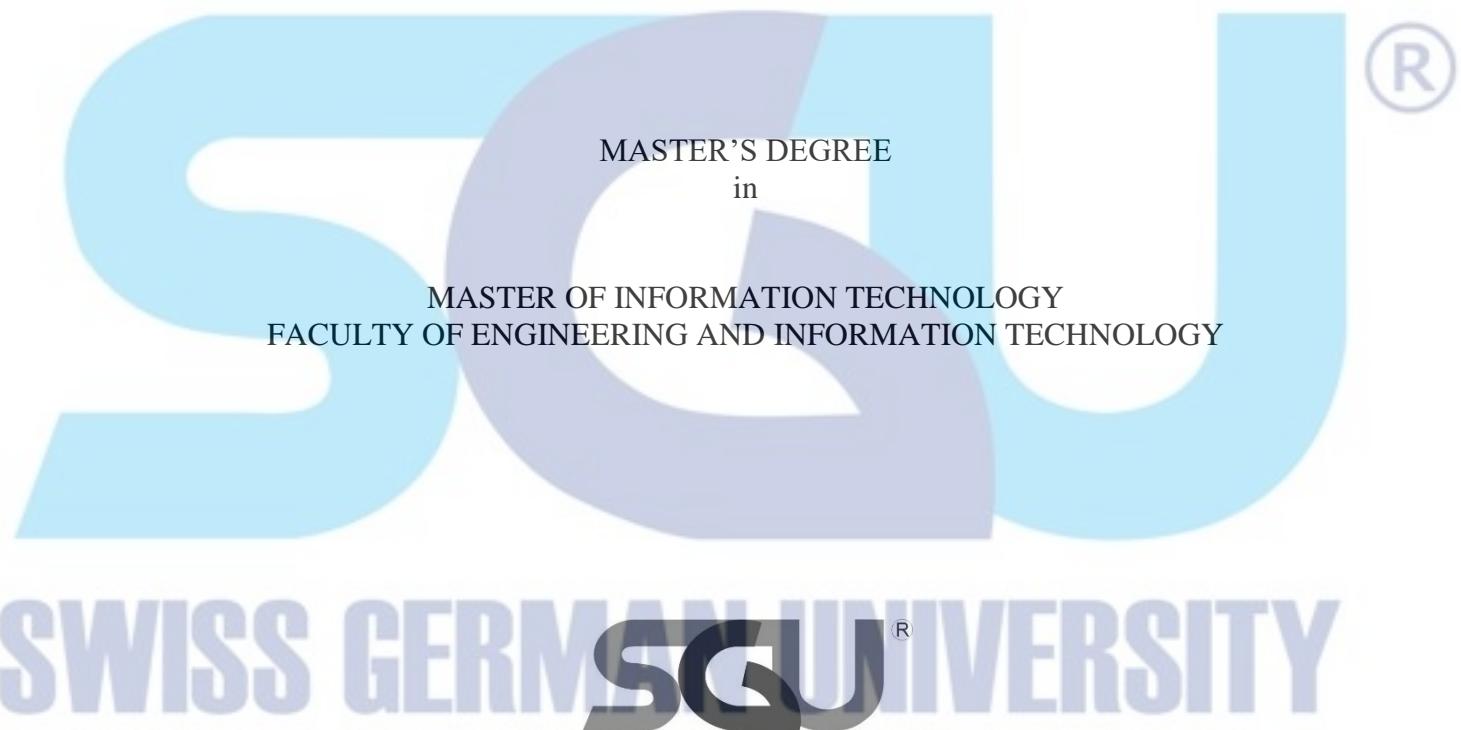


**THE DETECTION OF BOTNET THREATS USING NETFLOW AND
NETWORK RAW TRAFFIC**

By

Anton Purba
2-2013-102



SWISS GERMAN UNIVERSITY[®]

SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

August 2018
Revision after Thesis Defense on 2 August 2018

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 the thesis.

Anton Purba

Student

Date

Approved by:

Dr. Lukas, S.T., MAI

Thesis Advisor

Date

Charles Lim, M.Sc.

Thesis Co-Advisor

Date

Dr. Irvan Setiadi Kartawiria, S.T., M.Sc.

Dean

Date

Anton Purba

ABSTRACT

MASTER OF INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY

By

Anton Purba

Dr. Lukas, S.T., MAI, Advisor
Charles Lim, M.Sc., Co-Advisor

SWISS GERMAN UNIVERSITY

Nowadays botnets are used for most of cybercrime attack. After successfully infected the local machine, typically botnets will run on hidden process and use masquerade channel to communicate with its command and control server. HTTP and IRC are common legitimate protocols that used by attacker to covert the communication. This behavior makes the detection of botnet activity is the challenging problem. To detect botnet communication need a detail network traffic information to analyzed. Unfortunately detecting botnet using a network raw traffic also challenging and need more resource to do the process. NetFlow is an option to use for botnet detection. However, using sampling traffic such like NetFlow data also challenging to botnet detection accuracy. Meanwhile, the availability of raw network data is limited comparing to NetFlow data, that widely available. In this work we explores how accurate is detection can be achieved using NetFlow. To perform the evaluation, supervised machine learning algorithm is used, and two type of dataset; NetFlow and network raw traffic will be evaluated. This thesis work experiment found 99,4% the accuracy of botnet detection using NetFlow, therefore flow-based detection system in high-speed bandwidth environment is recommended.

Keywords: netflow, botnet, machine learning, malware, c&c.



SWISS GERMAN UNIVERSITY

DEDICATION

I dedicate this works to my lovely family, my wife, my son and my daughter.



ACKNOWLEDGEMENTS

First, the researcher wishes to give thanks to Jesus Christ who guards me and my lovely family in this research because without Jesus Christ, I couldn't complete this thesis. Second, thanks for my advisor, Dr. Lukas, S.T., MAI, my co-advisor Charles Lim, M.Sc, and MIT Dean Dr. Eka Budiarto, S.T., M.Sc., who helpful in guiding me from beginning until the end of thesis process.



TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR	2
ABSTRACT	3
DEDICATION	5
ACKNOWLEDGEMENTS	6
LIST OF FIGURES	10
LIST OF TABLES	11
CHAPTER 1 - INTRODUCTION	12
1.1 Research Background	12
1.2 Problem Statement	13
1.3 Research Objectives	13
1.4 Research Question	14
1.5 Hypothesis	14
1.6 Scope of Study	14
1.7 Significance of Study	15
1.8 Thesis Outline	15
CHAPTER 2 - LITERATURE REVIEW	16
2.1 Botnet Overview	16
2.1.1 Botnets and Bots Definition	17
2.1.2 Botnets Evolution	18
2.1.3 Technique and Capabilities	20
2.1.4 Botnet Prevention	22
2.1.5 Command & Control Channels (C&C Channels)	24
2.2 Network Traffic Monitoring	28
2.2.1 Active Network Traffic Monitoring	28
2.2.2 Passive Network Traffic Monitoring	29
2.2.3 Flow Monitoring	30
2.3 Machine Learning	32
2.3.1 Definition	32
2.3.2 How Machine learning is used today	34
2.3.3 Machine Learning Methods	35
2.3.4 Algorithm	36
2.4 Related Work	36

2.4.1	An efficient flow-based botnet detection using supervised machine learning.....	36
2.4.2	Flow-based identification of botnet traffic by mining multiple log files	37
2.4.3	Detecting P2P botnets through network behavior analysis and machine learning	37
2.4.4	Botnet detection based on network behavior	38
2.4.5	BotTrack: Tracking Botnets Using NetFlow and PageRank	38
2.4.6	Disclosure: detecting botnet command and control servers through large-scale NetFlow analysis	38
2.5	Our Approach on Botnet Detection	43
2.6	Summary	43
CHAPTER 3 – RESEARCH METHODS		46
3.1	Overview.....	46
3.2	Framework	46
3.3	General System Architecture	46
3.4	Methodology.....	47
3.4.1	Dataset	47
3.4.2	Data Preparation	48
3.4.3	Data Classification.....	57
3.4.4	Evaluation	58
3.4.5	Validation	58
CHAPTER 4 – RESULTS AND DISCUSSIONS		60
4.1	Environment Setup.....	60
4.1.1	Sever Specification	60
4.1.2	Software.....	60
4.2	Data Set Collection	60
4.2.1	NetFlow Dataset.....	62
4.2.2	Raw Traffic Dataset	62
4.3	Data Pre-Processing	63
4.3.1	NetFlow Data	63
4.3.2	Raw Traffic Data	63
4.4	Classification	64
4.5	Evaluation and Result	64
4.5.1	Feature Importance Result.....	64
4.5.2	Classifier Performance.....	66
4.6	Experiment Summary	68
CHAPTER 5 – CONCLUSIONS AND RECCOMENDATIONS		70
5.1	Conclusions	70
5.2	Recommendation	70
5.3	Future Work	70
GLOSSARY.....		72

REFERENCES.....	73
CURRICULUM VITAE.....	77

