

**DETECTING ADVANCED PERSISTENT THREAT ATTACK BASED ON DNS  
NETWORK TRAFFIC USING BAYESNET ALGORITHM**

By

Irfan Husein Al Darodjat

12112016

BACHELOR'S DEGREE

in

INFORMATION TECHNOLOGY

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY

EduTown BSDCity

Tangerang 15339

Indonesia

February 2017

**DETECTING ADVANCED PERSISTENT THREAT ATTACK BASED ON DNS  
NETWORK TRAFFIC USING BAYESNET ALGORITHM**

By

Irfan Husein Al Darodjat

12112016

BACHELOR'S DEGREE

in

INFORMATION TECHNOLOGY

FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY



SWISS GERMAN UNIVERSITY

EduTown BSDCity

Tangerang 15339

Indonesia

February 2017

**Revision after the Thesis Defense on 26 January 2017**

## 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.

Irfan Husein Al Darodjat

Student

Date

Approved by:

Charles Lim, MSc., ECSA, ECSP, ECIH, CEH, CEI

Thesis Advisor

Date

Dr. Ir. Lukas, MAI, CISA, IPM

Thesis Co-Advisor

Date

Dr. Ir. Gembong Baskoro, M.Sc

Dean

Date

Irfan Husein Al Darodjat

## ABSTRACT

### DETECTING ADVANCED PERSISTENT THREAT ATTACK BASED ON DNS NETWORK TRAFFIC USING BAYESNET ALGORITHM

By

Irfan Husein Al Darodjat

Charles Lim, MSc., ECSA, ECSP, ECIH, CEH, CEI, Advisor

Dr. Ir. Lukas, MAI, CISA, IPM, Co-Advisor

SWISS GERMAN UNIVERSITY

A new class of threats, known as Advanced Persistent Threats (APTs), has drawn increasing attention from researchers, primarily from the industrial security sector. APTs are cyber attacks executed by sophisticated and well-resourced adversaries targeting specific information in high-profile companies and governments. This research proposed a mechanism to detect APT threat based on DNS traffic using BayesNet classification algorithm. The validation of the classification is performed. The system successfully achieve 99.6% correctly classified instance. From 4 weeks of student and staff traffic, 223 true APT was found. This result means APT Threat exist in Swiss German University (SGU) DNS server. Feature of APT also can be found in DNS traffic. This research is a precursor in SGU highlighting the directions for future research of APT detection.

*Keywords:* advanced persistent threat, APT, sophisticated attacks, classification, DNS, machine learning, data mining



## DEDICATION

To my parents, my campus and for the future of my country: Indonesia



## ACKNOWLEDGEMENTS

In the name of God, the Most Gracious, the Most Merciful and may Allah send blessings and peace upon Prophet Muhammad SAWW and the family of Muhammad.

First, I would like to deliver my sincere gratitude to my advisor, Mr. Charles Lim, and my co advisor Mr. Lukas, for the limitless support and time during my thesis research, for their patience, motivation, and immense knowledge. Their knowledge and guidance help me in all the time of research and writing of this thesis and related researches.

Then, I would like to thank all of IT Department of Swiss German University. Lecturers and staffs, Mr. Kho, Mr. Maula, Mr James, Michael who helped me to pass OFSE, Mrs. Mawar, and many others, for all their helps during my study in Swiss German University.

Also, I would like to thank all of seniors and friends in IT SGU, during this research development. For the support and knowledge that help me complete this research. Especially Kris, Mario, Aldi, and Garuda Solusi Kreatif company help me a lot providing the resources for this research.

Lastly, I would like to express my wholehearted thanks to my family for their generous support throughout my entire life. Because of their unconditional love and prayers, i have the strength to complete my bachelor degree.

SWISS GERMAN UNIVERSITY

## TABLE OF CONTENTS

<b>STATEMENT BY THE AUTHOR</b>	<b>2</b>
<b>ABSTRACT</b>	<b>3</b>
<b>DEDICATION</b>	<b>5</b>
<b>ACKNOWLEDGEMENTS</b>	<b>6</b>
<b>TABLE OF CONTENTS</b>	<b>9</b>
<b>LIST OF FIGURES</b>	<b>10</b>
<b>LIST OF TABLES</b>	<b>11</b>
<b>1 INTRODUCTION</b>	<b>12</b>
1.1 Research Background . . . . .	12
1.2 Research Problem . . . . .	14
1.3 Research Objective . . . . .	14
1.4 Research Questions . . . . .	15
1.5 Hypothesis . . . . .	15
1.6 Significance of Study . . . . .	15
1.7 Research Limitation . . . . .	15
1.8 Document Structure . . . . .	15
<b>2 LITERATURE REVIEW</b>	<b>16</b>
2.1 Internet and Domain Name System . . . . .	16
2.2 Internet and Security Threats . . . . .	16
2.2.1 External Threats . . . . .	17
2.2.1.1 Phishing . . . . .	17
2.2.1.2 Denial of Service Attack . . . . .	17
2.2.2 Internal Threats . . . . .	17
2.2.2.1 Internal Accident . . . . .	17
2.2.2.2 Insider Threat . . . . .	18
2.2.3 Advanced Persistent Threat . . . . .	18
2.2.3.1 APT Taxonomy . . . . .	21
2.2.3.2 Case Study of APT Attacks . . . . .	21

2.3	Domain Name System . . . . .	22
2.3.1	Domain Name Space . . . . .	22
2.3.2	Domain Name Registration . . . . .	23
2.3.3	Domain Name Resolution . . . . .	23
2.4	Anomaly Traffic . . . . .	24
2.4.1	Anomaly DNS Traffic . . . . .	25
2.4.2	APT Traffic . . . . .	26
2.4.2.1	Daily Similarity . . . . .	26
2.4.2.2	Number of distinct IP address . . . . .	27
2.4.2.3	Number of distinct countries . . . . .	28
2.4.2.4	Average TTL value . . . . .	28
2.4.2.5	Silent IP . . . . .	28
2.4.2.6	Very low frequency query . . . . .	28
2.4.2.7	Number of Domain Share the Same IP With . . . . .	29
2.5	Non-Malicious Traffic . . . . .	29
2.5.1	Content Delivery Network . . . . .	29
2.5.2	Normal DNS Traffic . . . . .	30
2.5.2.1	Time To Live . . . . .	30
2.5.2.2	IP Address . . . . .	30
2.5.2.3	Country . . . . .	31
2.6	Data Mining . . . . .	31
2.6.1	Data Mining Process . . . . .	31
2.6.2	Data Mining Tools . . . . .	33
2.6.2.1	Weka . . . . .	33
2.6.2.2	Scikit-Learn . . . . .	34
2.6.2.3	Rapid Miner . . . . .	34
2.7	Machine Learning . . . . .	34
2.7.1	Clustering . . . . .	35
2.7.2	Classification . . . . .	35
2.7.2.1	J48 . . . . .	36
2.7.2.2	Multilayer Perceptron . . . . .	37
2.7.2.3	BayesNet . . . . .	38
2.7.2.4	NaiveBayes . . . . .	39
2.7.2.5	IBk . . . . .	39
2.8	Related Works . . . . .	40

<b>3</b>	<b>RESEARCH METHODOLOGY</b>	<b>46</b>
3.1	Research Methodology . . . . .	46
3.2	Research Framework . . . . .	46
3.3	Data Collection . . . . .	47
3.3.1	Network Architecture . . . . .	47
3.3.2	Flow of Traffic . . . . .	47
3.3.3	Data Collection Time . . . . .	48
3.4	Pre Processing . . . . .	48
3.5	Classification . . . . .	49
3.6	Evaluation . . . . .	50
<b>4</b>	<b>EXPERIMENT RESULTS</b>	<b>51</b>
4.1	Environment Setup . . . . .	51
4.2	Data Collection . . . . .	52
4.3	Pre-processing . . . . .	54
4.3.1	First Filter . . . . .	54
4.3.2	Second Filter . . . . .	54
4.3.3	Feature Extraction . . . . .	54
4.4	Classification . . . . .	55
4.5	Results . . . . .	57
4.5.1	Classifier Performance . . . . .	57
4.5.2	Student Traffic . . . . .	58
4.5.3	Staff Traffic . . . . .	59
4.6	Analysis . . . . .	59
4.6.1	Case 1 . . . . .	59
4.6.2	Case 2 . . . . .	60
4.6.3	Case 3 . . . . .	60
4.7	Mitigation . . . . .	60
<b>5</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>62</b>
5.1	Conclusions . . . . .	62
5.2	Recommendation . . . . .	62
5.3	Future Work . . . . .	62
	<b>GLOSSARY</b>	<b>64</b>
	<b>REFERENCES</b>	<b>70</b>
	<b>CURRICULUM VITAE</b>	<b>71</b>