

**NDDOS – TCP SYN FLOODING DETECTION USING SVM**



**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**



NDDoS – TCP SYN Flooding detection using SVM

AZWAR HAFUZA BIN MOHD NASIR



This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Computer Networking) with Honours.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY  
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

## DECLARATION

I hereby declare that this project report entitled  
**[NDDoS – TCP SYN Flooding detection using SVM]**  
is written by me and is my own effort and that no part has been plagiarized  
without citations.

STUDENT : AZWAR HAFUZA BIN MOHD NASIR

Date : 10/9/2021



اوتنومر سته تیکنیکا ملایسا ملاک  
I hereby declare that I have read this project report and found

this project report is sufficient in term of the scope and quality for the award of  
Bachelor of Computer Science (Computer Networking) with Honours.

SUPERVISOR :  Date : 10/9/2021  
(TS. NOR AZMAN BIN MAT ARIFF)

## DEDICATION

Alhamdulillah. All praise to Allah in providing me a good surrounding upon completing this project. This dedication is for my family, my supportive members and for my supervisors Ts. Nor Azman Bin Mat Ariff for all the inspiration, motivation, support and always give the best to me while completing this project.



## ACKNOWLEDGEMENTS

At first, thank you Allah for all of the blessing and moral guidance while completing this project, and also for providing me a good and supportive surrounding. Because, if I'm doing alone, I will never succeed to complete this project.

Next, greatest thank you for my supervisor. Ts. Nor Azman Bin Mat Ariff. Because always give me moral support and not tired while answering my question. If not have the right guidance and knowledge, I sure that I will never complete this project.

Lastly, thank you for my family, members, classmates and for all who help me direct indirectly upon completing this project.



## ABSTRACT

Leading to Industrial Revolution (IR) 4.0, most of the services are depending on the technology. This changes also will lead to a war named cyber war. The most popular weapon that was used during cyber-attack is Denial of Service (DoS) or Distributed Denial of Service (DDoS). In order to control this problem, a good detection system must be implements in the network architecture. The purpose of doing this project is to propose a DoS TCP SYN flooding detection using machine learning algorithm specifically Support Vector Machine (SVM). In this study, a dataset that was used is gain from Canadian Institute for Cybersecurity named NSL-KDD dataset. In conclusion, hope this project will achieve its goals and can be used for all people as precaution step for securing its network.

## ABSTRAK

Menuju Revolusi Perindustrian 4.0, kebanyakan servis bergantung kepada teknologi. Perubahan ini boleh membawa kepada perang disebut sebagai perang siber. Senjata yang paling terkenal yang digunakan semasa perang siber ialah serangan DoS dan juga DDoS. Untuk mengekang masalah ini, system pendeteksi yang berkesan mestilah diwujudkan dan digunakan didalam seni bina rangkaian. Tujuan menjalankan projek ini adalah untuk mencadangkan 'DoS TCP SYN flooding detection' sebagai alat pendeteksi serangan DoS yang menggunakan algoritma pembelajaran mesin lebih spesifik ialah Mesin Sokongan Vektor. Semasa penyelidikan ini, set data yang digunakan ialah daripada Institusi Keselamatan Siber Kanada bernama NSL-KDD. Sebagai konklusi, semoga projek ini akan mencapai matlamatnya dan boleh dimanfaatkan oleh semua orang sebagai langkah awal untuk menyelamatkan rangkaian mereka.



## TABLE OF CONTENTS

	<b>PAGE</b>
<b>DECLARATION.....</b>	<b>II</b>
<b>DEDICATION.....</b>	<b>III</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>IV</b>
<b>ABSTRACT .....</b>	<b>V</b>
<b>ABSTRAK .....</b>	<b>VI</b>
<b>TABLE OF CONTENTS.....</b>	<b>VII</b>
<b>LIST OF TABLES .....</b>	<b>XIII</b>
<b>LIST OF FIGURES .....</b>	<b>XIV</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XVII</b>
<b>LIST OF ATTACHMENTS.....</b>	<b>XVIII</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
1.1 INTRODUCTION.....	1
1.2 PROBLEM STATEMENT (PS).....	2
1.3 PROJECT QUESTION (PQ).....	4
1.4 PROJECT OBJECTIVE (PO).....	4
1.5 PROJECT SCOPE .....	5
1.6 PROJECT CONTRIBUTION (PC).....	5
1.7 REPORT ORGANISATION.....	6
1.8 CONCLUSION.....	6
<b>CHAPTER 2: LITERATURE REVIEW.....</b>	<b>7</b>

2.1	INTRODUCTION .....	7
2.2	INTRUSION DETECTION SYSTEM (IDS).....	9
2.2.1	IDS DEFINITION .....	9
2.2.2	IDS DETECTION METHOD .....	9
2.2.2.1	SIGNATURE-BASED .....	10
2.2.2.2	ANOMALY-BASED .....	10
2.2.3	TYPE OF IDS .....	10
2.2.3.1	HOST INTRUSION DETECTION SYSTEM (HIDS) .....	11
2.2.3.2	NETWORK INTRUSION DETECTION SYSTEM (NIDS).....	11
2.3	DENIAL-OF-SERVICE (DOS).....	11
2.3.1	DOS DEFINITION.....	11
2.3.2	DISTRIBUTED DENIAL-OF-SERVICE (DDOS) .....	12
2.3.3	CATEGORIES OF ATTACK .....	12
2.3.4	TYPE OF ATTACK.....	13
2.3.4.1	SYN FLOOD ATTACK.....	13
2.3.4.2	ICMP FLOOD ATTACK.....	14
2.3.5	DOS ATTACK TOOLS .....	15
2.3.6	DOS DETECTION TECHNIQUE .....	16
2.3.7	DOS PREVENTION TECHNIQUE .....	16
2.4	MACHINE LEARNING .....	17
2.4.1	MACHINE LEARNING DEFINITION.....	17
2.4.2	DATASET .....	18
2.4.3	FEATURE EXTRACTION.....	18
2.4.4	FEATURE SELECTION.....	19

2.4.5	CLASSIFIER VS MODEL.....	20
2.5	CRITICAL REVIEW.....	20
2.5.1	A STUDY ON NSL-KDD DATASET FOR INTRUSION DETECTION SYSTEM BASED ON CLASSIFICATION ALGORITHMS .....	20
2.5.2	APPLICATION-LAYER DDOS DETECTION BASED ON ONE- CLASS SUPPORT VECTOR MACHINE.....	23
2.5.3	DDOS ATTACK MODELING AND DETECTION USING SMO .....	26
2.5.4	MACHINE LEARNING DDOS DETECTION USING STOCHASTIC GRADIENT BOOSTING .....	28
2.5.5	SYN FLOOD ATTACK DETECTION IN CLOUD COMPUTING USING SUPPORT VECTOR MACHINE.....	30
2.5.6	A MACHINE LEARNING APPROACH FOR DDOS (DISTRIBUTED DENIAL OF SERVICE) ATTACK DETECTION USING MULTIPLE LINEAR REGRESSION .....	31
2.6	PROPOSED SOLUTION .....	35
2.7	CONCLUSION.....	35
<b>CHAPTER 3: DESIGN</b> .....		<b>36</b>
3.1	INTRODUCTION.....	36
3.2	METHODOLOGY .....	37
3.2.1	PREVIOUS RESEARCH.....	38
3.2.2	INFORMATION GATHERING.....	38
3.2.3	DEFINE SCOPE.....	38
3.2.4	DESIGN AND IMPLEMENTATION .....	38
3.2.5	TESTING AND EVALUATION OF MODEL.....	39
3.2.6	DOCUMENTATION .....	39
3.3	PROJECT GANTT CHART .....	39
3.4	PROJECT FLOW CHART.....	40

3.5	PROJECT MILESTONES.....	40
3.6	CONCLUSION.....	43
<b>CHAPTER 4: ANALYSIS AND DESIGN.....</b>		<b>44</b>
4.1	INTRODUCTION .....	44
4.2	PROBLEM ANALYSIS .....	44
4.3	REQUIREMENT ANALYSIS .....	44
4.3.1	SOFTWARE REQUIREMENT .....	44
4.3.2	HARDWARE REQUIREMENT.....	45
4.4	PROJECT DESIGN .....	46
4.4.1	DATASET .....	46
4.4.2	DATA PREPROCESSING .....	48
4.4.3	FEATURE SELECTION.....	49
4.4.4	DATA SPLITTING .....	50
4.4.5	CLASSIFICATION.....	51
4.4.6	CONCLUSION.....	53
<b>CHAPTER 5: IMPLEMENTATION.....</b>		<b>54</b>
5.1	INTRODUCTION .....	54
5.2	SOFTWARE DEVELOPMENT ENVIRONMENT SETUP.....	54
5.3	PROCESS MODULE .....	55
5.3.1	COLLECTION OF DATASET .....	55
5.3.2	DATA PREPROCESSING .....	55
5.3.2.1	REMOVE IRRELEVANT FEATURES .....	56
5.3.2.2	CATEGORICAL ENCODING .....	58
5.3.3	FEATURE SELECTION.....	62

5.3.4	TRAIN AND TEST DATA.....	65
5.3.4.1	SPLITTING DATA .....	65
5.3.4.2	MODEL FILE.....	67
5.3.4.3	PREDICT FILE .....	68
5.4	CLASSIFICATION .....	68
5.4.1	CODE REVIEW .....	69
5.4.2	EXECUTE .....	72
5.5	RESULT .....	72
5.5.1	RESULT OF TESTING MODEL .....	72
5.5.2	ACCURACY TABLE .....	74
5.6	CONCLUSION.....	74
	<b>CHAPTER 6: DISCUSSION .....</b>	<b>75</b>
6.1	INTRODUCTION .....	75
6.2	DISCUSSION OF THE PROJECT .....	75
6.3	DISCUSSION ON THE PROPOSED METHOD.....	76
6.4	CONCLUSION.....	77
	<b>CHAPTER 7: PROJECT CONCLUSION .....</b>	<b>78</b>
7.1	INTRODUCTION .....	78
7.2	PROJECT SUMMARY .....	78
7.3	PROJECT CONSTRAINT .....	79
7.4	PROJECT CONTRIBUTION.....	79
7.5	PROJECT LIMITATION .....	79
7.6	FUTURE WORK.....	79

7.7	CONCLUSION.....	79
	<b>REFERENCES.....</b>	<b>80</b>
	<b>APPENDIX A – PROJECT GANN CHART .....</b>	<b>82</b>



## LIST OF TABLES

	PAGE
<b>Table 1 – Problem statement.....</b>	<b>3</b>
<b>Table 2 – Project Question .....</b>	<b>4</b>
<b>Table 3 – Project Objective .....</b>	<b>4</b>
<b>Table 4 – Project contribution .....</b>	<b>5</b>
<b>Table 5 – Tools to launch DoS attack.....</b>	<b>15</b>
<b>Table 6 – Shows the list for both type algorithm.....</b>	<b>20</b>
<b>Table 7 – List files in the dataset gained[11]. .....</b>	<b>21</b>
<b>Table 8 – Comparison of all critical review.....</b>	<b>34</b>
<b>Table 9 – Project milestone .....</b>	<b>43</b>
<b>Table 10 – Software requirement .....</b>	<b>45</b>
<b>Table 11 – Hardware requirement .....</b>	<b>45</b>
<b>Table 12 - file include during download the dataset.....</b>	<b>47</b>
<b>Table 13 – brief for NSL-KDD Dataset.....</b>	<b>47</b>
<b>Table 14 – the feature description for the dataset[11].....</b>	<b>48</b>
<b>Table 15 - equation of every kernel in SVM.....</b>	<b>52</b>
<b>Table 16- Shows the summarize of dataset .....</b>	<b>66</b>
<b>Table 17 – Parameter description.....</b>	<b>69</b>
<b>Table 18 – Training model parameter description .....</b>	<b>70</b>
<b>Table 19 – y_predict description.....</b>	<b>71</b>
<b>Table 20 – Accuracy table .....</b>	<b>74</b>
<b>Table 21 – Shows the description of the proposed model.....</b>	<b>77</b>
<b>Table 22 – Project Gann Chart.....</b>	<b>82</b>

## LIST OF FIGURES

	PAGE
<b>Figure 1 – Statistics cybercrime in Malaysia.....</b>	<b>2</b>
<b>Figure 2 – Shows the outline for the next discuss topic. ....</b>	<b>8</b>
<b>Figure 3 – Shows the DoS vs DDoS attack.....</b>	<b>12</b>
<b>Figure 4 – TCP 3-Way Handshake.....</b>	<b>13</b>
<b>Figure 5 – Shows the stages of the ICMP.....</b>	<b>14</b>
<b>Figure 6 – Illustrate the ICMP flood attack. ....</b>	<b>14</b>
<b>Figure 7 – Shows the snort detecting DoS attack.....</b>	<b>16</b>
<b>Figure 8 - Shows the machine learning train and test flow.....</b>	<b>17</b>
<b>Figure 9 – Shows the concept of Bag of Words in Feature Extraction .....</b>	<b>19</b>
<b>Figure 10 – Shows the type of attack in the dataset.[11] .....</b>	<b>22</b>
<b>Figure 11 – Shows the accuracy gain from different algorithm[11].....</b>	<b>22</b>
<b>Figure 12 – Shows the algorithm for training in this project[12].....</b>	<b>25</b>
<b>Figure 13 – Shows the Receiver Operating Characteristics (ROC) curves that illustrated the performance of the proposed model in detecting application-layer DDoS attack in real situation[12].....</b>	<b>25</b>
<b>Figure 14 – Shows the architecture while generating the data[13]. ....</b>	<b>26</b>
<b>Figure 15 – Shows the performance of the SMO algorithm in predicting the attack[13]. ....</b>	<b>27</b>
<b>Figure 16 – Shows the result for test data 1[13]. ....</b>	<b>27</b>
<b>Figure 17 - Shows the dataset used in testing the proposed model[14].....</b>	<b>28</b>
<b>Figure 18 - shows the accuracy result obtain by few popular ml algorithm by using balanced dataset[14]. ....</b>	<b>29</b>
<b>Figure 19 – Indicate the accuracy of predicting SYN flood by using SVM [15]</b>	<b>30</b>



**Figure 20 - Machine Learning approach by using Multiple Linear Regression for detecting DDoS attack in the network. [16]..... 31**

**Figure 21 - IG result for all the features [16]..... 32**

**Figure 22 – Flow stages on chosen methodology..... 37**

**Figure 23 – Project Flowchart ..... 40**

**Figure 24 – Project design in this research..... 46**

**Figure 25 - black line that separate class blue and class red. The black line is called as hyperplane..... 51**

**Figure 26 - shows the support vector line to separate the class blue and red. The support line will find the nearest class. The distance between support vector line and hyperplane is called as margin. .... 51**

**Figure 27 – NLSVM example. This type of SVM cannot be done with simple match the hyperplane. So, kernel is needed to class the non-linear type of SVM ..... 52**

**Figure 28 – Process module task..... 55**

**Figure 29 – Data Preprocessing phase ..... 55**

**Figure 30 – Step 1 Phase 1..... 56**

**Figure 31 – Step 2 Phase 1..... 57**

**Figure 32 – Step 3 Phase 1..... 58**

**Figure 33 – Step 1 Phase 2..... 59**

**Figure 34 – Complete categorical encoding..... 60**

**Figure 35 – Step 2 Phase 2..... 61**

**Figure 36 – Step 1 Feature Selection ..... 62**

**Figure 37 – Step 2 Feature Selection ..... 62**

**Figure 38 – Step 1 of splitting data..... 65**

**Figure 39 – Step 2 of splitting data..... 66**

**Figure 40 – Shows the summarize of data after split..... 67**

**Figure 41 – Shows the process of generate the model..... 67**

**Figure 42 – Shows the process of predict the generated train model..... 68**

**Figure 43 – Model code review ..... 69**

**Figure 44 – Train result code review..... 70**

**Figure 45 – Generate the y\_predict for confusion matrix ..... 71**

**Figure 46 – Shows the code for print the report ..... 71**

**Figure 47 – Shows the code for print the confusion matrix ..... 71**

<b>Figure 48 – Shows the process of train model .....</b>	<b>72</b>
<b>Figure 50 – Show the prompt save file name and the average of the training accuracy .....</b>	<b>72</b>
<b>Figure 51 – Shows the report and confusion matrix from test process.....</b>	<b>73</b>
<b>Figure 52 – 10-Fold Cross Validation score.....</b>	<b>76</b>



## LIST OF ABBREVIATIONS

<b>FYP</b>	-	<b>Final Year Project</b>
DDoS	-	Distributed Denial of Service
DoS	-	Denial of Service
HTTP	-	Hypertext Transfer Protocol
UDP	-	User Datagram Protocol
ICMP	-	Internet Control Message Protocol
TCP	-	Transmission Control Protocol
SYN	-	Synchronize
IDS	-	Intrusion Detection System
SVM	-	Support Vector Machine
HIDS	-	Host-Based Intrusion Detection System
NIDS	-	Network Intrusion Detection System
PPS	-	Packets Per Seconds
RPS	-	Request Per Seconds
SYN-ACK	-	Synchronize-Acknowledge
CPU	-	Central Processing Unit
CMD	-	Command Prompt
OS	-	Operating System
RBF	-	Radial Basis Function
CIC	-	Canadian Institute for Cybersecurity

## LIST OF ATTACHMENTS

	<b>PAGE</b>
<b>Appendix A</b>	
<b>Project Gann Chart</b>	<b>82</b>



## CHAPTER 1: INTRODUCTION

### 1.1 INTRODUCTION

Distributed Denial of Service (DDoS) are the extended for Denial of Service (DoS). As knowledge, DoS is the type of attack that require and launched by the single attacker machine. While DDoS are the combination of multiple machine that are set to be an attacker machine that will be used to perform network hack like DDoS. DDoS attack will be the high impact attack for the network because the attack is launch by many devices or machine in one single time.

The problem for DDoS attack is, sometimes owner for the attacker machine does not know that they involved in the attack because the script for DDoS attack is inserted on the malicious software or program that downloaded on the internet. There are few methods to launch the Dos or DDoS attack and the common is by launching the flood attack. Flooding attack can be categorized into several version depending on what protocol packet that attacker used for example HTTP flood, UDP flood, ICMP flood or commonly known as Ping flood or the other name is Ping of Death. The type of flood that will be discuss on this project is TCP flood. TCP flood on DDoS or DoS attack is using handshake in 3-way TCP handshake process[1].

The attack is starting on the first handshake step which is synchronize (SYN) the common name for this type of attack is called TCP SYN flood attack. In general, flow of this attack is the attacker send many requests to the server by using TCP protocol. Server will busy to respond the huge amount of request launched by attacker and it will be denied other real request due to focusing on the SYN attack request. Effect from this attack will be the serious problem because victim might having lost due to their business cannot be operated while the attacked were launched.

## 1.2 PROBLEM STATEMENT (PS)

Hacking nowadays become a trend as method to bringing down someone else. According to (Basyir, 2021) on his article in New Straits Time Malaysia, Malaysia have faced a huge amount of loss since 2017 due to cybercrime frauds and attack. The amount stated is RM 2.23B. The rate statistic of attack is indicated on pie chart below.

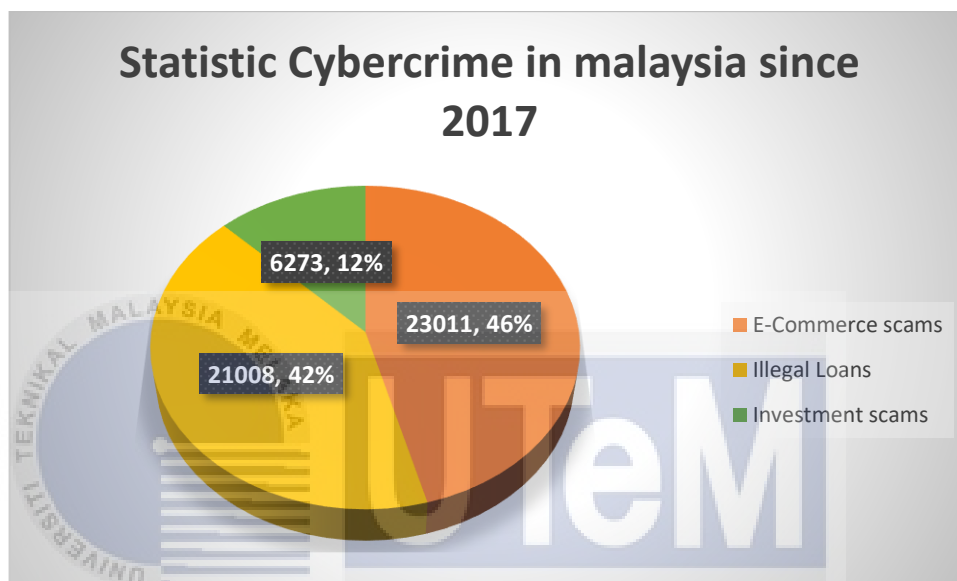


Figure 1 – Statistics cybercrime in Malaysia

This means, cyber-attack is the serious case that need a specialist expertise to solve this problem. According to figure 1. The rate shows the statistic for the scams or cybercrime fraud. But, if we deep inside the cases. In order to launch that attack, attacker also will be involved in cybersecurity threat attack such as data breach, and others network type of attack. This means, they also will have to launch the attack to the networks that relate with the Intrusion Detection System IDS. The IDS is the general terms of network attack.

As mentioned, IDS is only the term. Inside of IDS topic, there are generous of attack that widely use example phishing, deface, denial of service or Distributed denial of Service and others. If someone or some organization get stuck in this attack, they might cause a huge lost especially for a business owner or profit organization.

In general, if the e-commerce website on attack, customer might not reach the website or have been generate to another website that are set by attacker just to dropping the opponent business <sup>(1)</sup>. Besides, the network or traffic of the attacked network might congest and having overflow because there are several attack techniques that generate or transmit huge fake traffic towards attacked network just to prevent the opponent from succeed<sup>(2)</sup>. Regarding to the situation, it is recommended for a server or system have an effective technique to detect and prevent the flow of traffic in the network. Based on the problem, NDDoS – TCP SYN FLOOD DETECTION USING SVM is the best solution to prevent this problem. The problem statement (PS) for this plan is shown in Table 1.1.

PS	PROBLEM STATEMENT
PS1	Business drop cause from bad hacker launching DDOS attack to the network
PS2	Network getting slower due to congested traffic by flooding attack

**Table 1 – Problem statement**

اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

### 1.3 PROJECT QUESTION (PQ)

There are four project questions came based on problem statement before this study. The mentioned question is summarized into table 1.2 below:

PS	PQ	PROBLEM STATEMENT
PS1	PQ1	How to classify the traffic towards network?
	PQ2	How to prevent attacker sending huge transmission on the network?
PS2	PQ1	How to detect normal transmission and fake transmission?
	PQ2	Is the machine learning algorithm can be used to measure the accuracy in detecting the spam or attack traffic towards network?

**Table 2 – Project Question**

### 1.4 PROJECT OBJECTIVE (PO)

Project objectives (PO) is the goals for this research. Table 1.3 below are the objectives on this project:

PS	PQ	PO	PROJECT QUESTION
PS1	PQ1	PO1	To study taxonomy of DoS and DDoS attack focuses on TCP SYN Flood.
	PQ2	PO2	To develop a classification system that can detect the attack towards network
PS2	PQ1		PO3
	PQ2		

**Table 3 – Project Objective**