



**Intelligent Hybrid Approach for Classification Accuracy of  
Intrusion Detection System**

**Prepared By:**

**Mustafa Nihad Abbas**

**Supervisor**

**Prof.Dr.Mohammad Ahmad Alfayoumi**

**This Thesis Submitted in Partial Fulfilment of the Requirements for  
The Master Degree in Software Engineering**

**Isra University**

**Amman, Jordan**

**2019/2020**

## AUTHORIZATION STATEMENT

I am Mustafa Nihad Abbas, authorize Isra University to provide hard copies or soft copies of my thesis to libraries, institution or individuals upon their request.

Name: Mustafa Nihad Abbas

Signature:



Date:

17/02/2019

## اقرار تفويض

اني مصطفى نهاد عباس، افوض جامعة الإسراء للدراسات العليا بتزويد نسخ من رسالتي ورقياً و إلكترونياً للمكتبات او المنظمات او الهيئات و المؤسسات المعنية بالأبحاث و الدراسات العليا عند طلبها.

الإسم: مصطفى نهاد عباس

التوقيع: 

التاريخ: 19 / 10 / 14

The undersigned have examined the thesis entitled (**Intelligent Hybrid Approach for Classification Accuracy of Intrusion Detection System**) presented by (**Mustafa Nihad Abbas**) a candidate for the degree of Master of information technology in software engineering and hereby certify that it is worthy of acceptance.

17/12/2019

Date



Prof. Dr. Mohammad Ahmad Alfayoumi

15.12.2019

Date



Dr. Mudhafar Al Jarrah

16.12.2019

Date



Dr. Venus W. Samawi

## LIST OF CONTENTS

LIST OF CONTENTS .....	i
LIST OF TABLES .....	7
LIST OF FIGURES .....	8
ABSTRACT.....	10
CHAPTER ONE INTRODUCTION.....	1
1.1 OVERVIEW.....	1
1.2 PROBLEM STATEMENT .....	3
1.3 RESEARCH QUESTIONS.....	4
1.4 THE AIM OF THE STUDY .....	4
1.5 THE OBJECTIVES.....	4
1.6 THE SCOPE OF THE STUDY .....	5
1.7 RESEARCH PROCESSES .....	5
1.8 THESIS OUTLINE.....	7
CHAPTER TWO BACKGROUND AND LITERATURE REVIEW .....	8
2.1 INTRODUCTION.....	8
2.2 INTRUSION DETECTION SYSTEM (IDS).....	8
2.2.1 A brief history of IDS.....	10
2.2.2 Classification of IDS .....	11
2.2.3 Approaches to IDS .....	13
2.2.4 Challenges in IDS.....	16
2.3 Feature Selection.....	18
2.4 Feature Selection Evaluation Measures .....	20

2.5	FEATURE SELECTION PROBLEM IN IDS.....	23
	CHAPTER THREE DESIGN AND IMPLEMENTATION.....	26
3.1	INTRODUCTION.....	26
3.2	INTRUSION DETECTION SYSTEM (NSL-KDD).....	26
3.3	THE PROPOSED INTRUSION DETECTION SYSTEM.....	29
3.4	THE PROPOSED FEATURE SELECTION ALGORITHM.....	32
3.4.1	Firefly Algorithm .....	32
3.4.2	The Proposed FA.....	35
3.5	SUMMARY .....	39
	CHAPTER FOUR RESULTS AND DISCUSSION.....	40
4.1	Introduction .....	40
4.2	Experimental Settings .....	40
4.3	Results and Discussion.....	41
4.3.1	The Effect of Swarm Size and Number of Iterations .....	41
4.3.2	The Effect of <i>Swap</i> Variable .....	47
4.4	RESULTS COMPARTISON.....	49
	CHATER FIVE CONCLUSION.....	51
5.1	INTRODUCTION.....	51
	REFERENCES .....	55

## LIST OF TABLES

Table 3-1	The features of the NSL-KDD data set.....	28
Table 3-2	Distribution of attack records per NSL-KDD attack category .....	29
Table 4-1	Parameter Settings .....	41
Table 4 $\Psi$ -	The results of 15 run times for scenario 1 .....	42
Table 4 $\Upsilon$ -	The results of 15 run times for scenario 2 .....	42
Table 4 $\Xi$ -	The results of 15 run times for scenario 3 .....	43
Table 4 $\Theta$ -	The results of 15 run times for scenario 4 .....	43
Table 4 $\Gamma$ -	The results of 15 run times for scenario 5.....	44
Table 4 $\nu$ -	The results of 15 run times for scenario 6 .....	44
Table 4 $\wedge$ -	The results of 15 run times for scenario 7 .....	45
Table 4 $\rho$ -	The results of 15 run times for scenario 8 .....	45
Table 4-10	The summarized results for the proposed algorithm .....	46
Table 4-11	Results of the proposed algorithm for Swap = 5 .....	47
Table 4-12	Results of the proposed algorithm for Swap = 10 .....	47
Table 4-13	Results of the proposed algorithm for Swap = 15 .....	48
Table 4-14	Results of the proposed algorithm for Swap = 20 .....	48
Table 4-15	The results of all the algorithms .....	50

## **LIST OF FIGURES**

Fig 1-1	Research Process .....	6
Fig 2-1	The structure of IDS .....	9
Fig 2-2	Classification of IDS based on data collection and storage .....	11
Fig 2-3	Data analysis and process-based classification of IDS.....	12
Fig 2-4	Feature Selection Process .....	20
Fig 2-5	Types of feature selection evaluation measure.....	21
Fig 2-6	Filter-based feature selection.....	22
Fig 2-7	Wrapper-based feature selection .....	23
Fig 2-8	Process of Knowledge discovery.....	24
Fig 3-1	Block diagram of the proposed system .....	31
Fig 3-2	Flowchart of standard FA.....	34
Fig 3-3	Flowchart of GA-FA .....	36
Fig 3-4	The structure for each firefly.....	37
Fig 3-5	A graphical illustration for Crossover operator.....	38
Fig 4-1	The effect of swarm size and iteration number on the accuracy.....	46



## LIST OF ABBREVIATION

IDS	Intrusion Detection System
DDOS	
R2L	
GA	Genetic Algorithm
GWO	Grey Wolf Optimizer
PSO	Particle Swarm Optimization
FFA	Firefly Algorithm
ACO	Ant Colony Optimization
NFL	No-Free Lunch Theorem
GD	Gradient Decent
KNN	K Nearest
SVM	Support Vector Machine
NSL	Network Socket Layer
BP	Backpropagation
PCA	Principle Components Analysis
RF	Random Forrest
SOM	Self-Organization Maps
SFLA	Shuffled Frog Leaping Algorithm
BBO	Biogeography-Based Optimization
PSO	Particle Swarm Optimization
ABC	Artificial Bee Colony
CS	Cuckoo Search
BA	Bat Algorithm

## ABSTRACT

Intrusion detection system (I.D.S) is an essential component, which enhances the security of computer systems by actively detecting all forms of attack at the early stages. The main use of IDS is the monitoring of the network traffics and analyzing the behavior of the users in searching for any abnormal activity or attack signature for real-time intrusion detection. The main weakness in any IDS is their inability to offer adequate sensitivity and accuracy; coupled with their inability to process enormous data. To address these issues (such as the increasing traffic, huge behavior profiles, large signature databases, and the inability of differentiating normal behaviors from the suspicious ones), several algorithms have been developed. Hence, the main aim of this work is to choose the differentiating features for the development of an optimal machine learning algorithm which can offer high detection rates, fast training, and testing processes offline. The proposed machine learning model contains a feature selection algorithm (wrapper type) which is based on the integration of the Binary Firefly algorithm enhanced for feature selection by crossover operator taking from the genetic algorithm, called (GA-FA) with the Naïve Bayesian Classifier (NBC). The performance of the proposed model was tested on NSL\_KDD data sets prepared by the MIT Lincoln Laboratory. The model testing was based on several experiments and different scenarios (the effect of swarm size, number of iterations, and the *Swap*). For evaluating the ability to select the minimum number of features with the higher value of classification accuracy, the algorithm worked perfectly and selected a comparable number of features. The model achieved the best average accuracy of 97.011%. In conclusion, the proposed feature selection algorithm has the ability to select the most relevant features which enhance the classification accuracy of the network intrusion detection system.