



**HYBRID WEIGHT DEEP BELIEF NETWORK ALGORITHM FOR
ANOMALY-BASED INTRUSION DETECTION SYSTEM**

ZIADOON KAMIL MASEER

DOCTOR OF PHILOSOPHY

2022



Faculty of Information and Communication Technology

**HYBRID WEIGHT DEEP BELIEF NETWORK ALGORITHM FOR
ANOMALY-BASED INTRUSION DETECTION SYSTEM**

ZIADOON KAMIL MASEER

Doctor of Philosophy

2022

**HYBRID WEIGHT DEEP BELIEF NETWORK ALGORITHM FOR
ANOMALY-BASED INTRUSION DETECTION SYSTEM**

ZIADOON KAMIL MASEER

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

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022

DECLARATION

I declare that this thesis entitled “Hybrid Weight Deep Belief Network Algorithm for Anomaly-Based Intrusion Detection System “ 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 :

Supervisor Name : Ziadoon Kamil Maseer

Date :/12/2022.....

APPROVAL

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

Signature :

Supervisor Name : Assoc. Prof. Ts Dr Robiah Yusof

Date :/12/2022.....

DEDICATION

To my beloved mother and father

ABSTRACT

With an increasing number of recent services connected to the Internet, including cloud computing and Internet of Things systems, cyber-attacks have become more challenging. The deep learning approach plays a pertinent role in tracing new attacks in cybersecurity. Recently, researchers suggested a deep belief network (DBN) algorithm to construct and build a network intrusion detection system (NIDS) for detecting attacks that have not been seen before. However, the current DBN.NIDS model is still ineffective for large-scale real-world data due to some issues: 1) the pre-training of the DBN algorithm includes simple feature learning which does not work very well to extract important features from the attack data, 2) the classification task of the DBN algorithm is a poor detection for imbalanced class dataset and 3) the design of the DBN model could be weak and need to be continuously updated by modern definitions of abnormal to detect recent attacks. In this study, the Deep Belief Network algorithm was optimized and constructed to design an effective NIDS anomaly model. The optimized DBN algorithm, known as the HW-DBN algorithm, integrated through feature learning based on a Gaussian–Bernoulli Restricted Boltzmann Machine as well as classification task through a weight neuron network. The effectiveness of HW-DBN.NIDS was validated with real-world datasets that contained multiple attack types, complex data patterns, noise values, and imbalanced classes. A comparative analysis presented an HW-DBN.NIDS which was able to extract important features and detect the low frequency of modern attacks undetectable by other models. The results showed the proposed anomaly IDS model that outperformed the three models by achieving a higher recognition accuracy of 99.38%, 99.99%, and 1.00 for the Web, bot, and bot-IoT attacks in CICIDS2017 and CSE-CIC-IDS2018 dataset, respectively. In future, the HW-DBN algorithm can be proposed as an integrated deep Learning for the classification performance of attack detection models.

ALGORITMA RANGKAIAN PEMBERAT HIBRID KEPERCAYAAN MENDALAM UNTUK SISTEM PENGESANAN PENCEROBOHAN BERASASKAN ANOMALI

ABSTRAK

Serangan siber telah menjadi lebih mencabar dengan peningkatan bilangan perkhidmatan yang disambungkan ke internet, termasuk pengkomputeran awan dan sistem internet pelbagai benda. Pendekatan pembelajaran mendalam memainkan peranan penting dalam mengesan serangan baharu dalam keselamatan siber. Baru-baru ini, penyelidik mencadangkan algoritma rangkaian kepercayaan yang mendalam (DBN) untuk membina sistem pengesanan pencerobohan berasaskan rangkaian (NIDS) untuk mengesan serangan yang belum pernah dilihat sebelum ini. Walau bagaimanapun, model DBN.NIDS semasa masih tidak berkesan untuk data dunia sebenar yang berskala besar disebabkan beberapa isu: 1) pra-latihan algoritma DBN termasuk pembelajaran ciri mudah yang tidak berfungsi dengan baik untuk mengekstrak ciri penting daripada data serangan, 2) tugas pengelasan algoritma DBN adalah pengesanan yang lemah untuk set data kelas yang tidak seimbang, dan 3) reka bentuk model DBN mungkin lemah dan perlu dikemas kini secara berterusan oleh definisi moden yang tidak normal untuk mengesan serangan terkini. Dalam kajian ini, algoritma Deep Belief Network telah dioptimumkan dan dibina untuk merekabentuk model anomali NIDS yang berkesan. Algoritma DBN yang dioptimumkan, dikenali sebagai algoritma HW-DBN, disepadukan melalui pembelajaran ciri berdasarkan Mesin Boltzmann Terhad Gaussian–Bernoulli serta tugas pengelasan melalui rangkaian neuron berat. Keberkesanan HW-DBN.NIDS telah disahkan dengan set data dunia sebenar yang mengandungi berbilang jenis serangan, corak data kompleks, nilai hingar dan kelas tidak seimbang. Analisis perbandingan mempersembahkan HW-DBN.NIDS yang dapat mengekstrak ciri penting dan mengesan serangan moden berkekerapan rendah yang tidak dapat dikesan oleh model lain. Keputusan menunjukkan model IDS anomali yang dicadangkan telah mengatasi ketiga-tiga model dengan mencapai ketepatan pengesanan yang lebih tinggi iaitu 99.38%, 99.99% dan 1.00 untuk serangan Web, bot dan bot-IoT dalam set data CICIDS2017 dan CSE-CIC-IDS2018. Pada masa hadapan, algoritma HW-DBN boleh dicadangkan sebagai pembelajaran mendalam bersepadu untuk prestasi klasifikasi bagi model pengesanan serangan.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious, the Most Merciful

Thanks to Allah SWT for everything, I was able to achieve and for everything I tried, but I was not able to achieve.

To my supervisor Associate Professor Ts. Dr. Robiah binti Yusof you are truly a supervisor. I am greatly appreciative for her support and guidance, most importantly, for providing me the freedom to pursue my ideas and find my own path in research.

In addition, I have gained a wealth of experience and knowledge working under her supervision, which will always be my delight to share along my life's journey. Furthermore, I would like to thank Ts. Dr. Nazrulazhar bin Bahaman for his support and guidance as a co-supervisor.

TABLE OF CONTENTS

	PAGE
DECLARATION	
APPROVAL	
DEDICATIONS	
ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF SYMBOLS	xii
LIST OF ABBREVIATIONS	xiv
LIST OF PUBLICATIONS	xvii
CHAPTER	
1 INTRODUCTION	1
1.1 Overview	1
1.2 Related work	4
1.3 Research Problem	7
1.4 Research Question (RQ)	12
1.4.1 Sub Research Questions	13
1.5 Research Objectives (ROs)	14
1.6 Research Methods	15
1.7 Research Contribution	16
1.8 Research Scope	18
1.9 Research Organisation	18
1.10 Summary	21
2 LITERATURE REVIEW	22
2.1 Introduction	22
2.2 Chapter Outline	23
2.3 Chapter Objectives	23
2.4 Overview of an Intrusion Detection System	24

2.4.1	IDS Definition	24
2.4.2	Network IDS (NIDS)	25
2.4.3	Network Intrusion Detection System Approaches	26
2.5	Anomaly-NIDS Framework	31
2.6	Input Data Traffic Phase	32
2.6.1	Benchmark dataset	34
2.6.2	Real-life Dataset	37
2.6.3	Realistic Dataset	37
2.6.4	Discussion of NIDS Dataset Issues	39
2.7	Pre-processing Data Stage	49
2.7.1	Sampling Methods	49
2.7.2	Feature Extraction/Reduction (FR)	53
2.7.3	Feature Selection (FS)	55
2.8	Detection Phase	59
2.8.1	Shallow Supervised Learning Methods	59
2.8.2	Shallow Unsupervised Learning Methods	65
2.8.3	Deep Learning Technique	69
2.9	Reviewed and Analysis of Anomaly-NIDS	78
2.9.1	Effectiveness of Anomaly-NIDS	79
2.9.2	Used Network Dataset	82
2.9.3	Algorithm for Anomaly-NIDS	84
2.10	Anomaly-NIDS Gap Analysis	86
2.11	Summary	90
3	RESEARCH METHODOLOGY	92
3.1	Introduction	92
3.2	Research Methodology	92
3.3	Research Approach	93
3.4	Research Framework	93
3.4.1	Stage 1: Investigation on Research Gap	95
3.4.2	Stage 2: Optimization Methods	95
3.4.3	Stage 3: Design of Anomaly-NIDS	97
3.4.4	Stage 4: Implementation and Evaluation of anomaly-NIDSs	106
3.4.5	Stage 5: Documentation	109
3.5	Summary	109
4	ANOMALY-NIDS DESIGN	111
4.1	Introduction	111
4.2	Proposed Model Design	113
4.2.1	Phase 1: Pre-processing Raw Data	114
4.2.2	Phase 2: Building Anomaly-NIDS Model	114
4.2.3	Phase 3: Evaluation Model	115
4.3	Dataset Traffic Description	116
4.3.1	CICIDS2017 Dataset Description	116
4.3.2	Traffic Scenarios Description	117

4.3.3	CICIDS2017 Realistic Features	119
4.3.4	CSE-CIC-IDS2018 dataset	121
4.4	Summary	121
5	IMPLEMENTATION AND EVALUATION	123
5.1	Introduction	123
5.2	DL-NIDSs Implementation	124
5.2.1	Introduction of Deep Learning Model	125
5.2.2	Instruments and Tools	126
5.2.3	Implementation of Anomaly-NIDSs	128
5.3	DL-NIDSs Evaluation	138
5.3.1	RBM.NIDS Evaluation	138
5.3.2	DBN.NIDS Evaluation	140
5.3.3	AE.NIDS Evaluation	142
5.4	Comparison Result	151
5.4.1	Web Attack Scenarios	151
5.4.2	Bot Attack Scenarios	152
5.4.3	CICIDS2018 Dataset	153
5.5	Summary	154
CHAPTER 6	CONCLUSION AND FUTURE WORK	155
6.1	Introduction	155
6.2	Research Summary	155
6.3	Research Contribution	159
6.4	Research Limitations	162
6.5	Discussion	163
6.6	Future Trend	164
6.6.1	Generative Model	164
6.6.2	Effectiveness of Anomaly-NIDS	165
6.6.3	Low Computational of Anomaly-NIDS	165
6.6.4	Anomaly-NIDS Update	166
6.7	Summary	166
REFERENCES		168
APPENDICES		184

LIST OF TABLES

TABLE	TITLE	PAGE
1.1	Research Gap	12
1.2	Main Research Question	12
1.3	Sub Research Questions	14
1.4	Research Objectives Summary	15
1.5	Summary of (RP), (RQs), (ROs), and (RMs)	16
1.6	Research Contributions (RCs)	17
1.7	Outline of (RP), (RQs), (ROs), and (RCs)	17
2.1	Comparison of Signature, Anomaly and Hybrid NIDS	30
2.2	Evaluation parameters of NIDS Dataset	33
2.3	Comparison of NIDS Datasets	47
2.4	Comparison of Sampling Methods	52
2.5	Advantages and Disadvantages of Feature Selection Methods	58
2.6	High Precision, Recall And F-Score From Appendix A	82
2.7	NIDS DLs Evaluation	85
4.1	List of Attack Types in CICIDS2017	117
4.2	CICIDS2017 Attack Distribution and Description	118
4.3	Imbalance Classes Distribution	120
5.1	Hardware and Software Specification	127
5.2	Deep Learning Parameters	135
5.3	Proposed Anomaly-NIDS Configuration	136

5.4	RBM.NIDS Classification Report	138
5.5	DBN.NIDS Classification Report	140
5.6	AE.NIDS Classification Report	145
5.7	HW-DBN.NIDS Classification Report	148
5.8	Web Attack Results	152
5.9	Bot Attack Result	153
5.10	CICIDS2018 Results	153

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Modern Infrastructure of Network	1
1.2	Unit 42 IoT Threat Report of 2020	2
1.3	DBN.NIDS Challenges	9
1.4	Bot Attacks Dataset Distribution	10
1.5	Web Attacks Dataset	11
1.6	Thesis Outline	20
2.1	Outline of Chapter Two	24
2.2	Network-based IDS	26
2.3	Scheme of Anomaly-NIDS	28
2.4	General Intrusion Detection System Framework	31
2.5	IDS Framework	32
2.6	Briefly Taxonomy of Network Intrusion Dataset	34
2.7	Oversampling Technique	50
2.8	AIDS methods	59
2.9	ANN Architecture	60
2.10	DT Architecture	61
2.11	RF Architecture	63
2.12	SOM Architecture	68
2.13	Deep Learning and Machine Learning	70
2.14	Deep Learning Architecture	71

2.15	Frequency of ML and DL Techniques	79
2.16	Frequency of DL Algorithms	80
2.17	Conventional ML Algorithms for NIDS	81
2.18	Percentage of Simulation and Realistic Dataset	83
2.19	Frequency of Used Dataset	84
2.20	Derived Gap and Research Problem	87
3.1	Research Methodology	93
3.2	Proposed Research Framework	94
3.3	DBN Algorithm	98
3.4	HW-DBN Algorithm	99
3.5	Conventional GB-RBM Algorithm	103
3.6	Improved GB-RBM	104
3.7	Weighted Neuron Network	105
3.8	Benchmarking Evaluation	106
4.1	Chapter Outlines	112
4.2	Proposed Anomaly-NIDS Model	113
4.3	HW-DBN Algorithm	115
4.4	Distribution of Web And Bot Attacks Dataset	121
5.1	Outline of Chapter 5	124
5.2	Implementation Structure	125
5.3	Deep Learning Phases	126
5.4	Algorithm for Benchmarking Implementation	129
5.5	Benchmark Implementation	129

5.6	RBM.NIDS Confusion Matrix	140
5.7	DBN.NIDS Confusion Matrix	142
5.8	AE Training Metric for Web Attacks	143
5.9	AE Training Metric for Infiltration Data	144
5.10	AE Training Metric for Infiltration Data	144
5.11	AE.NIDS Confusion Matrix	146
5.12	HW-DBN Training Metric for Web Attacks Data	147
5.13	HW-DBN Training Metric for Infiltration Data	147
5.14	HW-DBN Training Metric for CICIDS2018 Data	148
5.15	HW-DBN.NIDS Confusion Matrix	150

LIST OF SYMBOLS

a, b	-	Variables Value
Min	-	Minimum Value
Max	-	Maximum Value
m	-	Mean Values
n, k	-	Constant Value
p	-	Instance
Z	-	Array of elements
h	-	Hidden units
v	-	Input units
j	-	Hidden unit number
i	-	Input unit number
W	-	Weight Matrices
w	-	Element of weight
Σ	-	Summation data
E, e	-	Energy function
P	-	Probability of data
σ	-	Sigmoid function
log	-	Logarithmic value
∂	-	Derivative function
t	-	Step number

N	-	Gaussian Distribution Form
b	-	Bias of hidden layer
e	-	Exponential value
L	-	Loss function
θ	-	Theta (weights and bias)
ε	-	Epsilon (small value)
g	-	Gradient or slope function
η	-	Constant learning rate
β	-	Beta learning rate
Y	-	Label or y-axis
x	-	Element of row

LIST OF ABBREVIATIONS

Acronym	Term
AD	- Anomaly-based Detection
AE	- Auto-Encoders
AIDS	- Anomaly Intrusion Detection System
ANN	- Artificial Neural Networks
CAIDA	- Center of Applied Internet Data Analysis
CAN	- Controller Area Network
CD	- Contrastive Divergence
CICIDS2017	- Canadian Institute for Cybersecurity 2017
CNN	- Convolutional Neural Network
CSM	- Cost-Sensitive Method
DL	- Deep learning
DBN	- Deep Belief Network
DNN	- Deep Neuron Network
DoS	- Denial of Service
D-RBM	- Discriminative RBM
DT	- Decision Tree
EM	- Expectation–Maximization
FAR	- False Alarm Rate
FFNN	- Feedforward neural network
FN	- False Negative
TN	- True Negative
TP	- True Positive
FP	- False Positive
FS	- Feature Selection
FTP	- File Transfer Protocol
G-mean	- Geometric Mean
GR	- Gain Ratio
GR-RBM	- Gaussian–Bernoulli RBM

HDL	-	Hybrid Deep Learning
HIDS	-	Host IDS
HTTP	-	Hypertext Transfer Protocol
HW-DBN	-	Hybrid Weighted Deep Belief Network
IDS	-	Intrusion Detection System
IG	-	Information Gain Feature-Feature evaluator
IMAP	-	Internet Message Access Protocol
Info gain	-	Information Gain
IoT	-	Internet of Thing
IR	-	Imbalance ratio
ISCX	-	Information Security Center of Excellence
KDD cup 99	-	Knowledge Discovery and Data Mining
K-NN	-	K-nearest neighbors
LAN	-	Local area network
LDA	-	Linear Discriminant Analysis
LLE	-	Locally Linear Embedding
LR	-	Logistic Regression
ML	-	Machine Learning
ML	-	Machine Learning
NB	-	Naive Bais
NIDS	-	Network Intrusion Detection System
NSL-KDD	-	Network Security Laboratory-KDD
PCA	-	Principal Component Analysis
PCAPs	-	Packet Capture Application Programming
POP3	-	Post Office Protocol 3
R2L	-	Remote to Local
RBM	-	Restricted Boltzmann Machines
Recon	-	Reconnaissance Attacks
ReLU	-	Rectified Linear Activation
RF	-	Random Forest
RNN	-	Recurrent Neural Network

ROC	-	Relative Operating Characteristic
ROS	-	Random Oversampling
RUS	-	Random Under sampling
SNIDS	-	Signature Network Intrusion Detection System
SDL	-	Supervised Deep Learning
SDN	-	Software-Defined Networking
SMOTE	-	Synthetic Minority Oversampling Technique
SMTP	-	Simple Mail Transfer Protocol
SPA	-	Stateful Protocol Analysis
SSH	-	Secure Shell Protocol
SVM	-	Support Vector Machine
T-SNE	-	T-distributed Stochastic Neighbor Embedding
U2R	-	User to Root
UDL	-	Unsupervised Deep Learning
UNB	-	University of New Brunswick
UNIBS	-	University in Brescia
WCEL	-	Weight Cross-Entropy Loss
W-NN	-	Weighted Neural Network
HW- DBN.NIDS	-	Network Intrusion Detection System using Hybrid Weight Deep Belief Network
DBN.NIDS	-	Network Intrusion Detection System using Deep Belief Network

LIST OF PUBLICATIONS

1. **Z. K. Maseer**, R. Yusof, N. Bahaman, S. A. Mostafa and C. F. M. Foozy, "Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset," in *IEEE Access*, vol. 9, pp. 22351-22370, 2021, doi: 10.1109/ACCESS.2021.3056614. (ISI indexed, Q1, IF = 3.367 (2020))
2. **Z. K. Maseer**, R. Yusof, S. A. Mostafa, N. Bahaman, and C. F. M. Foozy, "DeepIoT.IDS: Hybrid Deep Learning for Enhancing IoT Network Intrusion Detection". *Computers, Materials and Continua (CMC)*. (ISI indexed, Q2, IF = 5.410 (2020)). Accepted.
3. Hassan, Ali Abdul-hussian and Shah, Wahidah Md and Othman, Mohd Fairuz Iskandar, Hassan, Hayder Abdul Hussien and **Ziadoon Kamil Maseer**," Unequal clustering routing algorithms in wireless sensor networks: A comparative study", *Jour of Adv Research in Dynamical & Control Systems*, Vol. 10, 02-Special Issue, 2018. (ISI indexed, Q4, IF = 0.31 (2020)).
4. Mustafa Hasan Kathim, Nurul Azma Zakaria, Z.Zainal Abidin, **Ziadoon Kamil Maseer**, Ali Hasan Alzamili, "Framework of Meta-Heuristic Based Computational Load Balancing in Wireless Sensor Network", *International Journal of Advanced Science and Technology* Vol. 29, No. 9s, (2020), pp. 1423-1431. (ISI indexed, Q4, IF = 0.48 (2020)).
5. Mustafa Hasan Al-Bowarab, Nurul Azma Zakaria, Zaheera Zainal Abidin, **Ziadoon Kamil Maseer**. "Review on device-to-device communication in cellular based network systems". *International Journal of Engineering & Technology*, 7 (3.20) (2018) 435-440. (ISI indexed, Q4 (2020)).

CHAPTER 1

INTRODUCTION

1.1 Overview

Over time, more and more distinct services are being explored online via the Internet. The Internet has millions of automotive services connecting with endpoints. This new technology of conventional networks is known as IoT networks, smart energy grids, industrial machines, building automation, and many personal assistance devices (Fredrik Jejdling, 2019; Sivanathan et al., 2019). IoT stands for an intricate and dynamic network that links device endpoints to deliver services as shown in Figure 1.1. Many security concerns are raised by the diversity and volume of data transformed via networks. These groundbreaking technologies have greatly increased the risk and threat of attacks.

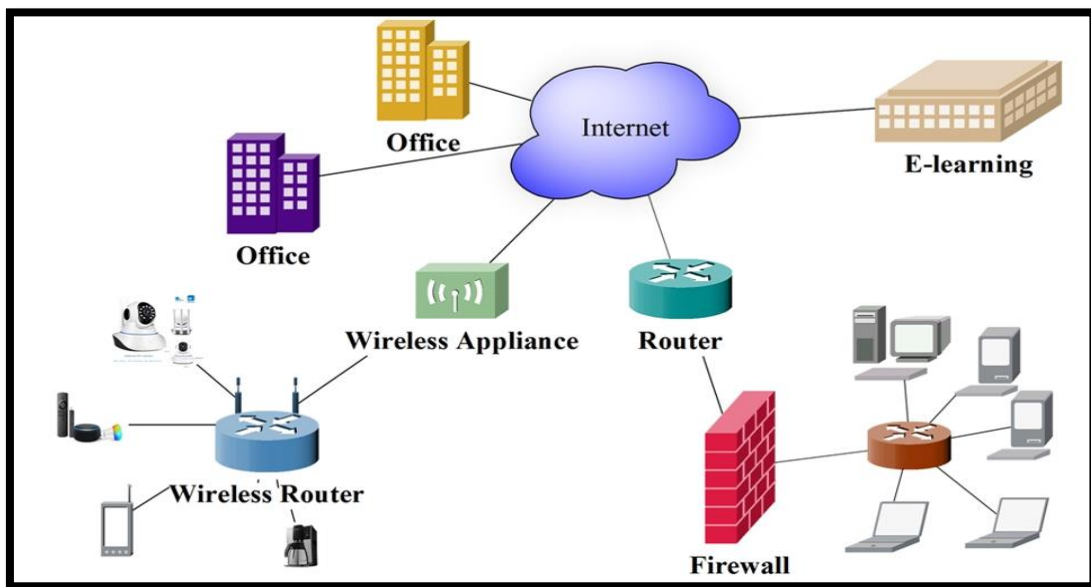


Figure 1.1 Modern Infrastructure of Network

A threat report from Unit 42 IoT was just released, and it surveyed more than 1.2 million IoT devices in the American healthcare and IT sectors. The report demonstrated how