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 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 pengecaman 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

LIST OF TABLES

LIST OF FIGURES

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
$\sum$	-	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 bais)
ε	-	Epsilon (samll value)
g	-	Gradient or slop 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 Transifer 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).

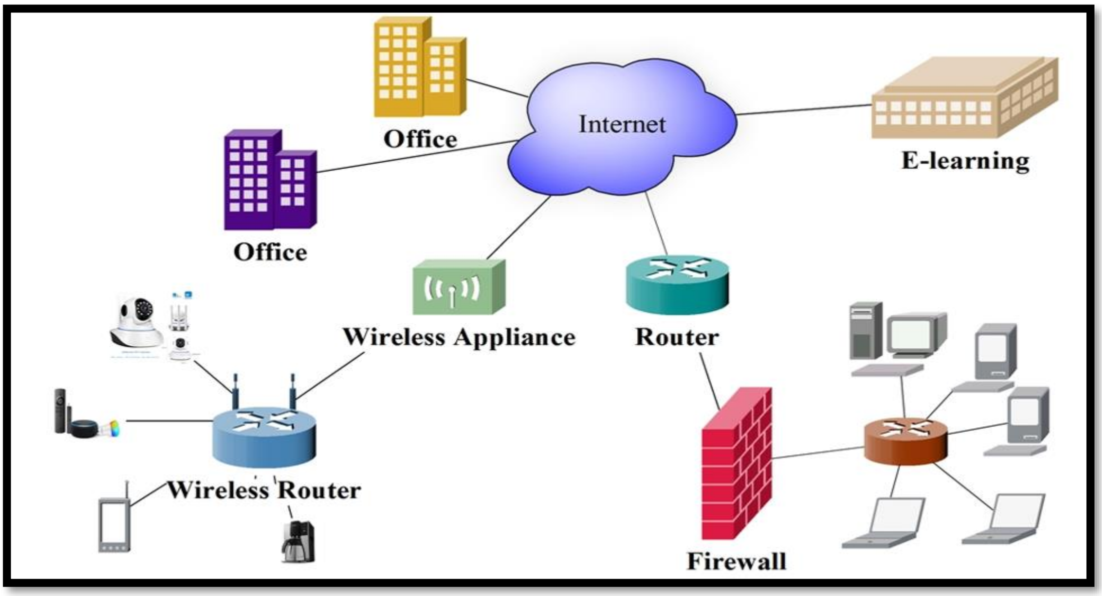
5. Mustafa Hasan Al-Bowarab, Nurul Azma Zakaria, Zaheera Zainal Abidin, **Ziadoon Kamil Maseer**. "Review on device-to-device communication in cellular based network systems". *International Journal of Engineering & Technology*, 7 (3.20) (2018) 435-440. (ISI indexed, Q4 (2020).

CHAPTER 1

INTRODUCTION

1.1 Overview

Over time, more and more distinct services are being explored online via the Internet. The Internet have millions of automotive serv1ices 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 ground-breaking 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