PES
UNIVERSITY

HIGH LEVEL DESIGN DOCUMENT

Deep Learning based Network Intrusion Detection System hosted on Cloud

UE17CS490A – Capstone Project Phase – 1

Submitted by:

Archana C	**PES1201701384**
Chaitra H P	**PES1201701370**
Khushi M	**PES1201701416**
T P Nandini	**PES1201700064**

Under the guidance of

Dr. Sivaraman Eswaran
Associate professor
PES University

August - December 2020

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

TABLE OF CONTENTS

1. Introduction

An intrusion detection system (IDS) is a device or software application that monitors a network for malicious activity or policy violations. Any malicious activity or violation is typically reported or collected centrally using a security information and event management system. Some IDS's are capable of responding to detected intrusion upon discovery. Intrusion detection systems can be classified into the following categories:

a) Network Intrusion Detection system(NIDS): This IDS monitors the network traffic flowing in and out of the system to detect traces of malicious activity.
b) Host Intrusion Detection System(HIDS): This IDS monitors a particular host's/system's important OS files for malicious activities.

Intrusions can be detected by classifying access to the systems into predefined malicious activities, but this won't take care of the novel ways devised by attackers to breach the security of a system.So we propose a Anomaly based Hybrid Intrusion Detection System that uses Deep learning algorithms to detect malicious activities.

The proposed model trains a model using DL algorithms that learns about the characteristics of a malicious activity using a pre-existing dataset and with each detection improves and optimizes the model.

The model consists of 2 stages post building the model:

1. Extraction of minimal features that best represent the data set.
2. Classification of the system access into a class of attack if it is an attack.

The proposed model uses UNSW-NB15 dataset to train the DL model using Convolution Neural Network(CNN).

[ADDITIONAL]

A proactive module is an addition to the model to predict an intrusion that could probably occur in the near future.

This model has 3 components:

a) IDS module: This is a NIDS
b) Data Processing Module: It interacts with the data repository and control model building.

c) Proactive Forecasting module: Produces DL model using ML methodologies in the development phase.

Finally this model offers real time detection of intrusions and a feature that forecasts the intrusions in near future.

2. Current System [if applicable]

Over the past few years, a number of models and approaches based on traditional machine learning have been proposed for network intrusion detection. Examples include SVM, KNN, ANN,Random forest etc. Yanxia Sun *et al [1]* proposed an IDS using deep learning with feed forward deep neural networks (FFDNNs) coupled with a filter-based feature selection algorithm on NSL-KDD dataset. Mohammad Mehedi Hassan *et al* [2] have used Hybrid deep learning model to efficiently detect network intrusions based on CNN and a weight-dropped LSTM network on UNSW-NB15 dataset. Farruhk Aslam Khan *et al* [3] proposes a novel two-stage deep learning (TSDL) model, based on a stacked auto-encoder with a soft-max classifier, for efficient network intrusion detection evaluated on two datasets KDD99 dataset and UNSW-NB15 dataset. Vinaykumar *et al.* [4] presents a DNN model with 1 input layer, 5 hidden layers and 1 output layered architecture with ReLU as activating function, cross entropy as loss function and stochastic gradient descent as optimization method on the KDD99 dataset as the hybrid IDS approach. Giang Nguyen *et al.* [5] proposes a proactive Deep Learning Model to predict future possible attacks using big data technologies on real data collected from an existing network IDS(ZEEK/Bro).

3. Design Details

3.1 Innovativeness

- Using Doc2vec and language models like HMM and N-gram for efficient text analysis.
- Using modern optimization methods like Basin Hopping, Adam methods to train the neural net which are more efficient than back propagation which has the vanishing gradient problem.
- Experimenting with different activation functions like leaky ReLU which solves the zero gradient issue from ReLU on different hidden layers.
- Testing on recent datasets and on realistic networks using testbeds.

3.2 Dataset

- The UNSW-NB15 dataset comprises more complicated types of attacks to evaluate deep learning models and is more recent- 2015.

3.3 Optimization

- Hyper parameter tuning like learning rate, number and depth of hidden layers, number of iterations can be varied and tested.

3.4 Legal Implications

- privacy might be an issue because this analyses some headers during feature extraction.

3.5 Usage limitation

- For Industrial usage the model must classify the given packet into attack or benign in near real time provided it follows the constraint on input to the model.
- Model must not produce a High False Positive Rate and should not classify an attack as a benign packet.
- Models must be trained periodically on up-to-date network traffic features.

3.6 Specific Requirements

- High computing GPUs

3.6.1 Hardware Requirements

- GPUs for high computation.
- 4GB RAM (minimum)
- Intel i5 processor

3.6.2 Software Requirements

- Python
- Jupyter Notebook
- Cloud instances eg: Amazon EC2

- Tensorflow (1.15)
- keras (2.2.5)

3.7 Assumptions

- The dataset covers all types of records equally so as to reduce bias against any one type of attack.
- The model is able to train periodically on up-to-date network traffic features in an offline mode and it is able to detect intrusion attacks in an online mode.

3.8 Interoperability

Can be integrated with any system to detect intrusions to the system.

3.9 Performance

- Classify the given network packet with low latency.
- The model should achieve high accuracy with low False Positive Rate(FPR)

3.10 Security

- The model must be cryptic to avoid misuse by users.

3.11 Reliability

- The goal is to reduce false positives and make the IDS reliable.

3.12 Maintainability

- Constant upgrades to train dataset.

3.13 Scalability

- Handling any workload depends on hardware and software optimization.

3.14 Usability

- Easy to use and must be portable.

Appendix A: Definitions, Acronyms and Abbreviations

SVM - Support Vector Machine
KNN - K- Nearest Neighbour

ANN - Artificial Neural network

DL - Deep Learning

Appendix B: References

[1] Sydney Mambwe Kasongo And Yanxia Sun, “A deep learning method with filter based feature extraction for wireless intrusion detection system”,IEEE, 2020.

[2] Mohammad Mehedi Hassan, Abdu Gumaei, Ahmed Alsanad, Majed Alrubaian, Giancarlo Fortino, “A hybrid deep learning model for efficient network intrusion detection in big data environment “,ScienceJournal 2020.

[3] Farrukh Aslam Khan (Senior Member, IEEE), Abdu Gumaei , Abdelouahid Derhab, and Amir Hussain , “A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection”, IEEE 2019.

[4] R. Vinayakumar, Mamoun Alazab, K. P. Soman , Prabaharan Poornachandran, Ameer Al-nemrat, And Sitalakshmi Venkatraman,”Deep Learning Approach for Intelligent Intrusion Detection System”, IEEE 2019

[5] Giang Nguyen,Stefan Dlugolinsky, Viet Tran And Alvaro Lopez Garcia, ” Deep Learning for Proactive Network Monitoring & Security Protection” , IEEE 2020.