LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

Cloud based Network Intrusion Detection System using Deep Learning Algorithms

UE17CS490B – Capstone Project Phase – 2

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

January - May 2021

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

PES UNIVERSITY

LOW LEVEL DESIGN AND IMPLEMENTATION DOCUMENT

TABLE OF CONTENTS

1. Introduction

1.1 Overview

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 Cloud based Network Intrusion Detection System using 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. Since it is hosted on the cloud, it will be scalable.

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.

1.2 Purpose

Cyber attacks are always evolving in quality and both quantity. According to PurpleSec "In 2017 there were over 130 large-scale, targeted breaches in the U.S. per year, and that number is growing by 27% per year". With increase in the internet traffic world wide , the influx of network traffic into any system has also seen a major spike over the years. The attacker can take advantage of this and over flood the system with dummy traffic and make the systems unresponsive.

1.3 Scope

There are malicious users who have the means to evade signature based IDS systems. Hence the need for Neural networks is on the rise, as they are well suited to picking up new patterns of attacks readily and not dependent on signatures or rules. Our proposal is to use a Deep Neural Net model to build a hybrid IDS which can be employed in real time.

2. Design Considerations, Assumptions and Dependencies

2.1 Design Considerations

Using modern optimization methods like 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. Hosting the IDS on the cloud. The UNSW-NB15 dataset comprises more complicated types of attacks to evaluate deep learning models and is more recent- 2015. Hyper parameter tuning like learning rate, number and depth of hidden layers, number of iterations can be varied and tested. privacy might be an issue because this analyses some headers during feature extraction.

2.2 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.

2.3 Dependencies

- Python, keras and tensorflow (for DL models)
- AWS EC2 instances
- System with minimum 8GB RAM

- Jupyter Notebook
- Intel i5 processor

3. Proposed Methodology/Approach

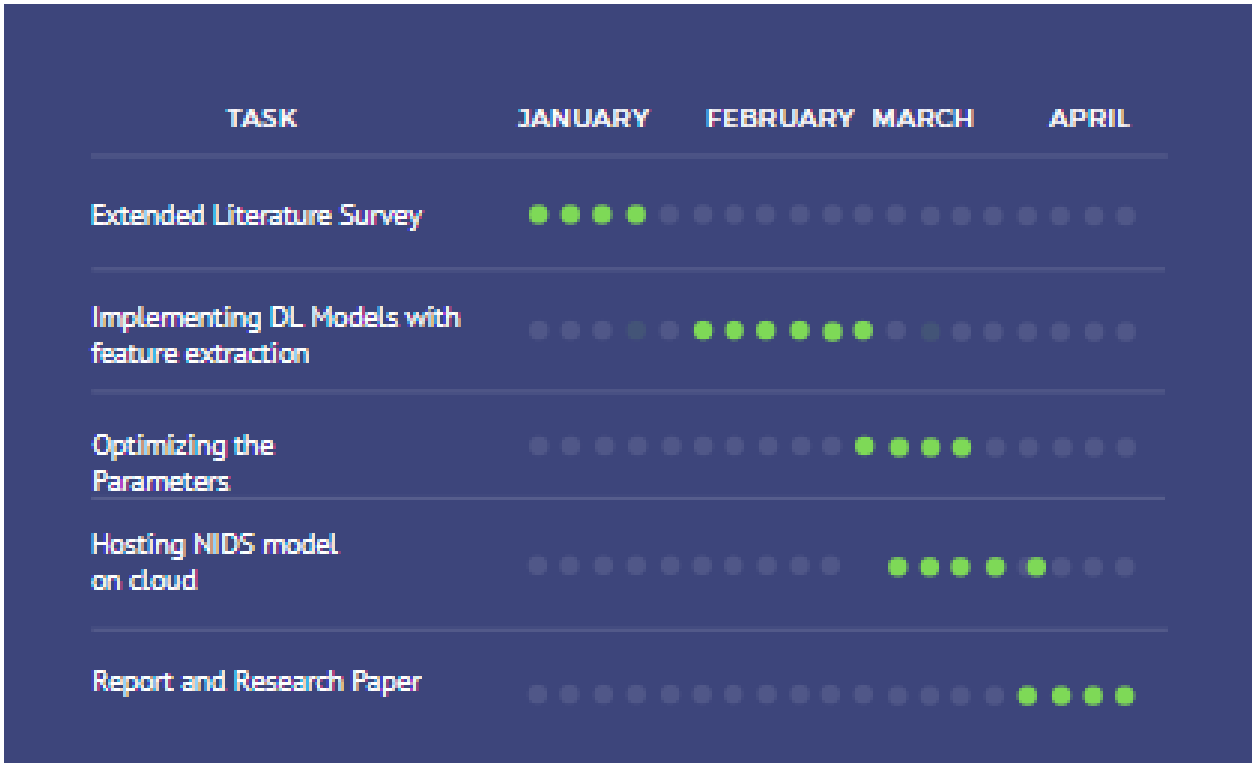
3.1 Algorithm and Pseudocode

```
# 1. define the network
model = Sequential()
model.add(Dense(4096,input_dim=41,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(2048,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(1024,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(512,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(1024,activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Dense(1))
model.add(Activation('sigmoid'))
```

3.2 Implementation and Results

Cloud Implementation is done by creating an AWS EC2 instance with two containers using flask framework for making API calls. And compared the results of different Machine Learning and Deep Learning Models on the UNSW-NB15 dataset. Among the different Machine Learning models implemented, Random Forest and Decision Tree achieved the highest accuracy of 86%. adding more number of hidden layers to LSTM and GRU did not improve the accuracy whereas that was not the case with DNN. The LSTM and GRU achieved around 88.78%, 88.79% accuracy respectively and Deep Neural Network with five hidden layers achieved the highest accuracy of 95.02%.

3.3 Further Exploration Plans and Timelines



Appendix A: Definitions, Acronyms and Abbreviations

ML - Machine Learning

DL - Deep Learning
IDS - Intrusion Detection System
LSTM - Long Short Term Memory
GRU - Gated Recurrent Unit
CNN - Convolutional Neural Network
DNN - Deep Neural Network