Project Proposal

For

Network Intrusion Detection System

Submitted by

Fraz Ahmad, SP17-BSE-003

Department of Computer Science, CU, Islamabad.

25/07/2020

**Table of Contents**

Abstract 1

1 Introduction 1

2 Proposed System 1

3 Advantages/Benefits of Proposed System 2

4 Scope 2

5 Modules 2

5.1 Module 1: Neural Network Model: 2

5.2 Module 2: Desktop Application: 2

5.2 Module 2: Report 2

6 Dataset 2

7 System Limitations/Constraints 2

8 Tools and Technologies 3

9 Implementation 3

10 Conclusion 9

11 References 9

Project Category:

* Desktop Application/Information System

# Abstract

Machine and Deep learning-based Network Intrusion Detection System, it detects malicious activity, it is a multi-class classification of different web and network attacks. Model will be trained on UNB Intrusion Detection Evaluation dataset. Project will be desktop application which take a csv file of features extracted from PCAP files to predict malicious activity.

# Introduction

As project name explains it’s a system that detect malicious activity in the network, traditional IDS works with defined rules but it’s working is limited to old or known activities and unable to detect zero day attacks or new malicious activity. To detect unknown attacks or activity in the network the best solution is the use of Deep Learning/ Machine Learning to learn on old data and then predict the activity so we can protect our network from malicious activities. so I will use [UNB IDS dataset](https://www.unb.ca/cic/datasets/ids-2017.html) to train deep learning model to predict malicious behavior.so it solves the problem of detecting unknown and known malicious activities so its more powerful technique to defend networks.

# Proposed System

The proposed system uses machine learning algorithms and deep learning neural networks as model, then will choose highest accurate model for deployment, hopefully neural networks works well because problem is complex as it contains 80 features and its multiclass classifier. First dataset needs to be in a shape because it contains 8 files with resected to network traffic monitored and saved in a whole dataset creation process, first I will clean data and make it into a shape to load , then I will create neural network model with Tensorflow and Keras and save the best trained model for future use. Then for prediction purpose python build system application will take a csv file created form PCAP files to predict the network activity in perspective class. Dataset contains Brute force, DOS, BOT, infiltration, SQL injection and XSS attacks network activity. So, this dataset contains most of the attacks that’s how our system is robust enough to detect and malicious activity easily.

# Advantages/Benefits of Proposed System

* Unknown malicious activity detection
* Network protection from attacks and breaches
* Zero-day attack detection
* Dynamic system instead of static rule based traditional system

# Scope

The main functionality of a system is to take csv file or network activity flow and predict activity class belongs to, creation of csv file from PCAP files is not the part of this project it can be created with network flow meter created by UNB, and this project proposed system is not automated. So, user have to upload csv file to get classes of each activity. The main objective is to reduce false positives by getting higher accuracy on test and validation set. The second functionality is to visualize with charts what data contains, means what type of attacks are there and how much it’s like a report.

# Modules

There are three modules in this project.

## 5.1 Module 1: Neural Network Model:

Creation, training and saving of neural network model build with Tensorflow and Keras.

## 5.2 Module 2: Desktop Application:

Desktop Application GUI built with python tkinter library which holds saved model and read uploaded csv file and then get predictions from model.

## 5.2 Module 2: Report

This module provides functionality of reporting the analysis and visualization of uploaded csv file, means which and how much attacks are there in network data, number of attacks with perspective attack in chart.

# Dataset

The dataset is publicly available on internet and it is provided by Canadian institute of cyber security, user have to fill a form to get access to the dataset, which contains 8 files with respected attacks and weeks in given it will be collected. So, I will clean data and make it in a proper way to train my model. Dataset is available in this site <https://www.unb.ca/cic/datasets/ids-2017.html>.

# System Limitations/Constraints

The limitation and constraint of this system is that it is not automated mean it is not configured to automatically analyze network behavior user have to upload csv file to get prediction, second constraint is that I will not generate csv file from PCAP files.

# Tools and Technologies

Tools and technologies used in the project are:

Table 1: Tools and Technologies

|  |  |  |  |
| --- | --- | --- | --- |
| **Tools**  **And**  **Technologies** | **Tools** | **Version** | **Rationale** |
| Anaconda Individual Edition | 2020 | IDE (data science & machine learning) |
| Jupyter Notebook | 2020 | Web Based application for interactive data science and machine learning |
| Google Colab | - | GPU usage to train model |
| **Technology** | **Version** | **Rationale** |
| Python | 3.8 | Programming language |
| Tensorflow & Keras | 2.2.0 | Library |
| Pandas | 1.1.0 | Library |
| Matplotlib | 3.3.0 | Library |
| CSV | 13.1 | Library |
| Tkinter | 3.8.5 | Library |
| Pandas Table | 0.12.2 | Library |

# Implementation

To implement above solution, I choose Neural Network to train model, the reason to choose is the dataset is too complex and neural networks are far better than machine learning algorithms if dataset is big. So, my dataset is large enough to perform deep learning techniques and I also have expertise in deep learning.

First of all I have applied data preprocessing on my dataset, merging all files into one and make them binary as originally there are 8 files covering different attacks in categorical so I have merged all files and make them binary by converting benign label into 0 and all other attack labels to 1. Further I cleaned the data which contain NaN values, inappropriate characters, and further cleansing.

A screenshot of a computer

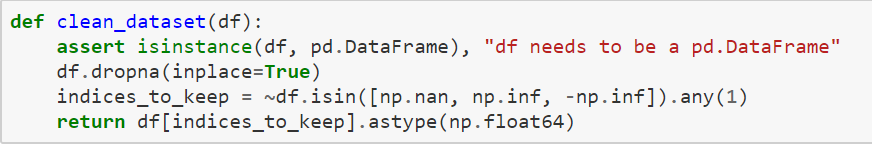
Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a cell phone

Description automatically generated



After cleaning I applied normalization and convert data into X and Y. Then after preprocessing my neural network model is:

A screenshot of text

Description automatically generated

It’s a 7-layer model and 2 of them are batch normalization and dropout which are used to drop the over fitting problem. To remove the over fitting problem and make my model generalize I used l1 regularization, dropout layer, and batch normalization. As dataset is big enough and total features are 78 so its complex that’s why I created deep network having many layers. Further I used ADAM optimizer and binary cross-entropy as a loss function. After lot of training and too many iterations the results are:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Training Accuracy is 98% and Testing Accuracy is also 96% so here u can see that there is no overfitting nor underfitting. Model training accuracy and loss in graph is.

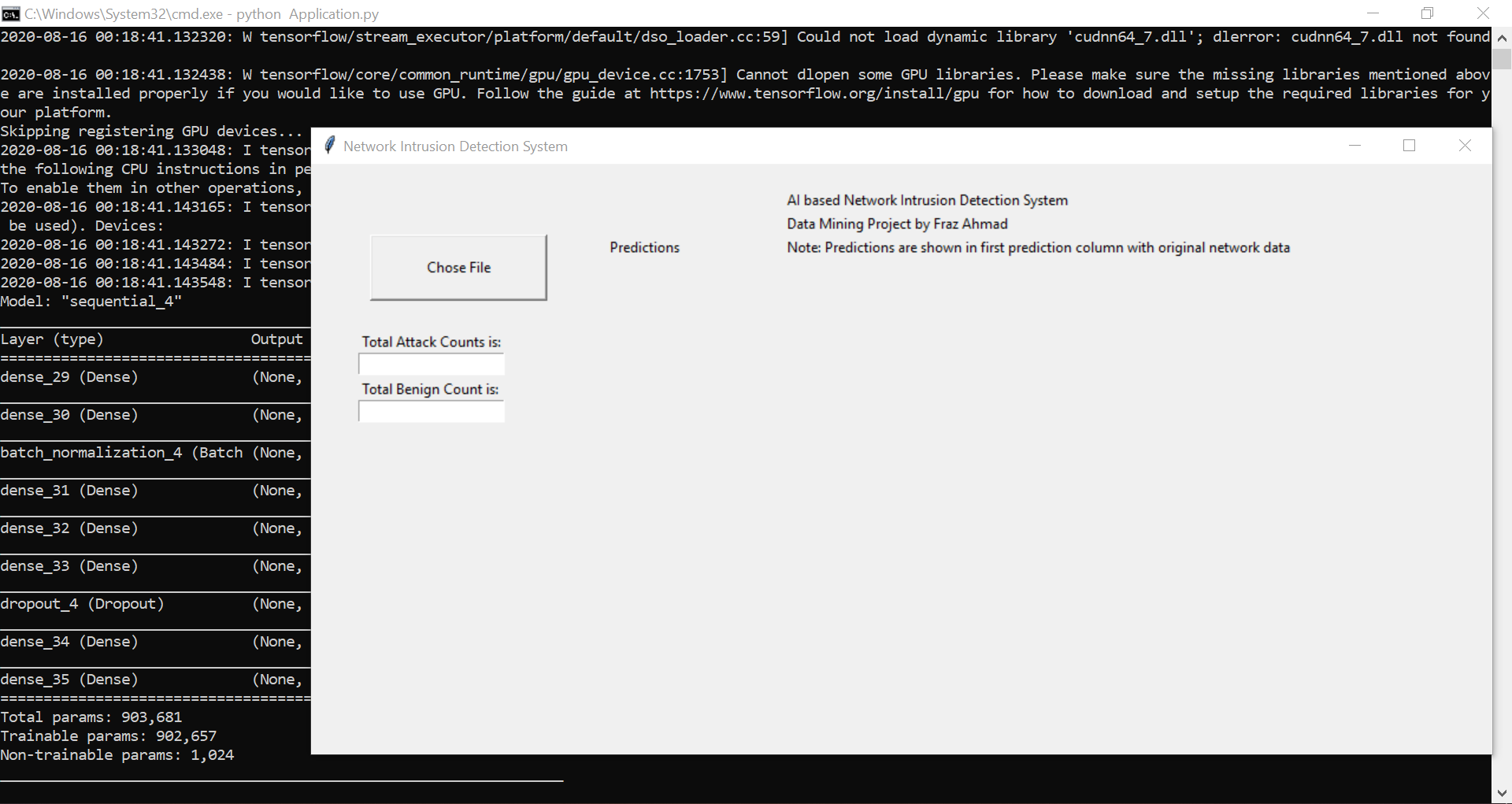
A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Now Trained model is saved and can be deployed in any server or network but I have created a desktop application and load trained model in it and get predictions on user uploaded csv file.



A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

# Conclusion

The project is related to cyber security for network security to detect intrusions in network, the IDS is based on Deep Learning technology to detect modern attacks in a modern way. So, I will train a neural model which predicts a network activity and classify them in different attacks and Benin, it’s a multi-class classification model.

# References

* Project Video link: <https://pern-my.sharepoint.com/:v:/g/personal/sp17-bse-003_isbstudent_comsats_edu_pk/EVX-mf8RnzxBghaAJVExNHwBLQByDujlhHQoBlcoy_K_6Q?e=kSTiQ7>
* github link: <https://github.com/Frazad/Network-Intrusion-Detection-System-Deep-Learning.git>
* Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Purtogal, January 2018
* <https://www.unb.ca/cic/datasets/ids-2017.html>