Real Time DDOS Attack Detection Using Time Series Algorithms

Submitted By
Dhruvi Patel
18MCEI06

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INSTITUTE OF TECHNOLOGY
NIRMA UNIVERSITY

AHMEDABAD-382481
May 2019

Real Time DDOS Attack Detection Using Time Series Algorithms

Major Project

Submitted in fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering
(Information and Network Security)

Submitted By
Dhruvi Patel
(18MCEI06)

Guided By
Dr. K. P. Agrawal

तमसो मा ज्योतिर्गमय
NIRMA
UNIVERSITY
INSTITUTE OF TECHNOLOGY
NAAC ACCREDITED ‘A’ GRADE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INSTITUTE OF TECHNOLOGY
NIRMA UNIVERSITY
AHMEDABAD-382481

December 2019

Certificate

This is to certify that the major project entitled **"Real Time DDOS Attack Detection using Time Series Algorithms"** submitted by **Dhruvi Patel (18MCEI06)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering(Information and Network Security) of Nirma University, Ahmedabad, is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr. K. P. Agrawal
Guide & Associate Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sharada Valiveti
Associate Professor,
Coordinator M.Tech - CSE (INS)
Institute of Technology,
Nirma University, Ahmedabad

Dr. Madhuri Bhavsar
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr R. N. Patel
I/C Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Dhruvi Patel**, **18MCEI06**, give undertaking that the Major Project entitled ”**Real Time DDOS Attack Detection Using Time Series Algorithms**” submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering(Information and Network Security)** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student
Date:
Place:

Endorsed by
Dr. K. P. Agrawal
(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. K. P. Agrawal**, Associate Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Madhuri Bhavsar**, Hon'ble Head of Computer Science And Engineering Department, Institute of Technology, Nirma University, Ahmedabad for her kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. R. N. Patel**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Dhruvi Patel
18MCEI06

Abstract

DDOS attack is malicious attack in which attacker try to overwhelm the network traffic by flooding the target network. Attacker uses botnet to send large number of requests to target. DDOS attack can be in the form of TCP flood, SYN-ACK flood, HTTP flood, smurf attack. To Identify any type of DDOS attack it is necessary to differentiate between normal traffic flow and attack traffic flow. Time series forecasting and analysis helps to determine the network traffic pattern with the reference of time. Basically this patterns are learnt from historical data like data of last 30 days, data of one week, one hour, etc. This helps time series model to train on specific patterns.

Here in this application two types of time series models are used to detect DDOS attack, one is stochastic time series model and another is ANN time series model respectively ARIMA model and LSTM model. Training this model with historic time series data of network traffic, this machine learning models are capable of forecasting future network traffic.If traffic of current time lapse is affected due to DDOS attack then it can be detected by trained model.By evaluating and measuring performance of models, best model for system will be deployed for early detection of attack in the system.

Based on Performance of both models, best model for system is being deployed to detect DDOS attack in real time scenario.

Abbreviations

DDOS	Distributed Denial of service.
ARIMA	Auto regressive integrated moving average.
SARIMA	Seasonal Autoregressive Integrated Moving Average
LSTM	Long short-term memory .
RNN	Recurrent neural network.
AR	Auto regression.
MA	Moving average.
RMSE	Root Mean Square Error
ACF	Auto-correlation function
PACF	Partial Auto-correlation function

Contents

List of Figures

Chapter 1

Introduction

1.1 Problem statement

Detect DDOS attack in regular traffic flowing through network by time series forecasting of current traffic.Stochastic time series model like ARIMA and ANN time series model like LSTM is used as machine learning models for this purpose.Best model will be deployed to system after comparing the performance of models to recognize the attack situation.

1.2 Objective

- Goal is to detect DDOS attack in regular traffic flowing through network.
- Real time analysis of network traffic flow using machine learning
- Forecasting of future network traffic using ARIMA model and LSTM neural network
- After collecting the data, data preprocessing, data transformation, data reduction will be done in order to clean the data.
- In model training algorithm is selected and based on feature selection model is being trained.
- Trained model will forecast the future traffic of network, which gives early detection of attack.
- Comparing forecasted data and normal traffic flowing through system, application will detect DDOS attack.
- Also, Analyzing performance stochastic and ANN time series models and concluding best suitable model for detecting DDOS attack in system.

1.3 Scope of Project

Project will consist of creating a DDOS attack detection application based on time series algorithms. The project will be completed by May 2020.Main modules of application will include data collection, data preprocessing, EDA(Exploratory data analysis),Feature extractions, Model training, model evaluation and model deployment. Application will generate best suitable model to detect DDOS attack.

Chapter 2

Literature survey

1. **A Novel DoS and DDoS Attacks Detection Algorithm Using ARIMA Time Series Model and Chaotic System in Computer Networks(April 2016)** [1]
   Paper mainly focuses on detecting ddos attack using two features extracted from network traffic which is Source IP address and Number of packets. Researchers has used **ARIMA** model to predict the number of packets flowing in network for every one minute. Local lyapunov exponent computed to detect DDOS attack. Features are calculated by dividing numbers of packets to number of source IP address in every minute.
   BOX-COX transformation is used for data transformation. Positive value of lyapunov exponent shows abnormal behavior of network traffic and it shows ddos attack situation in network. Researchers has used **Darknet and CAIDA,DARPA1998** dataset to perform experiment and compare the prediction results. [1]

2. **Statistical Measures: Promising Features for Time Series Based DDoS Attack Detection(January 2018)** [2]
   By overwhelming normal network traffic attacker tries to disturbed regular network traffic and it results to ddos attack. Researchers has decomposed the time series and analyzed various components to compare behavior of internet traffic in attack situation. Researchers has extracted mainly four Statistical measures of time series which is **Hurst exponent, skewness, kurtosis, periodicity** from network traffic. **CAIDA** dataset is used for capturing time series data of every 1ms.Researchers has calculated all four parameters in normal traffic flow and attack traffic flow. As a result they have compared parameters from which kurtosis and Hurst exponents are best features to detect dos attack.. [2]

3. **DDoS Detection Algorithm Based on Preprocessing Network Traffic Predicted Method and Chaos Theory(2013)** [3]
   Researchers has purposed a novel approach in which first network traffic is pre processed by averaging model **simple linear AR model** and predicted network traffic. To detect anomalous behavior of network **neural network** is trained based on chaotic errors of prediction.

DARPA 1998 and 1990 datasets are used for experiment purpose. Also to improve performance of neural network back propagation is used for detection of attack, Result of experiments shows 93.75% successful attack detection. [3]

4. **Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data** [4]
Researchers has used **LSTM-RNN** to classify the data for intrusion detection in network. **KDD cup** data set is used as it has various categories like DDOS, network probes, remote to local attacks and user-to-root.
LSTM network with 43 input neurons, five target neurons, two memory block each containing two cells and peephole connections are used to build network. Target neurons are capable of discriminate between the normal traffic and attack for ddos attacks. [4]

5. **A Comparison of ARIMA and LSTM in Forecasting Time Series(2018)** [5]
Time series forecasting is very useful in finance ,economics, business and cyber security. There are some traditional methods like Autoregressive and auto regressive moving average, ARIMA are used for prediction and decomposition time series. Also advance machine learning and deep learning algorithms are used for forecasting time series data.
Researchers has compared **LSTM**(long short term memory) with traditional **ARIMA model** for comparison of error rates. LSTM is used is used to handle time series data as it can preserve the data for longer period also holds the features that are used for training.
Researchers has used dataset of **Yahoo finance** of one month duration. Dataset was being split in 70-30% for training and testing respectively. To check accuracy of prediction model root mean square mean error is used. After training, RMSE is calculated on test data for LSTM and ARIMA model from which LSTM performs better with 84-87% reduction in error rate. [5]

6. **A Long Short-Term Memory Enabled Framework for DDoS Detection** [6]
Researchers has proposed a LSTM based framework for detection of attack in network. Main idea is to give entire packet information to LSTM without doing feature engineering and evaluate the performance of model.
Researchers has used **CICIDS 2017** datasets for comparison of traditional machine learning approaches using feature engineering and LSTM with proposed methods. CICIDS 2017 dataset contains different DDOS attacks ‘hulk, TCP,UDP floods data with more than 80 features extracted from network traffic labeled as benign and malicious. Traditional approach of choosing best features to train machine learning algorithm performance batter than the proposed approach. [6]

Chapter 3

Basics of DDOS attack

DDOS attack is recognized as distributed denial of services attack where more than one attackers launches a attack and prevents a legitimate user to access the services of network, internet as well as intranet resources.
DDOS attack can be launched using various protocols like TCP,UDP,HTTP, ICMP on different layers such as network layer, transport layer, application layer. [7]
In ddos attacks like SYN flood, UDP flood, UDP Lag attacker send large number of packets to victim machine by using third party host or remote host using multiple ports to disturb the regular traffic on victim machine.

Machine learning algorithms are widely used for detection ddos attacks.Benign network traffics and traffic at the time attack are collected using network monitoring systems, these collected raw data is processed and useful features like source IP, destination IP, source port, flow duration, number of packets, timestamp, labels, etc are collected and based on that a cleaned dataset is created.
These cleaned dataset is used for training machine learning algorithms like decision tree, Logistic regression, SVM, Neural networks, ARIMA models to classify the network traffic. Based on the classified labels and traffic alerts are generated for benign and attack situation of the network.

Chapter 4

Classes of Time series models

Time series is class of problem where data collected from different interval of time are being analyzed, decomposed and forecasted by different time series models to get future predictions. Time series problems are of different types like weather forecasting, stock prize prediction, network traffic predictions etc.

There is correlation between different features of dataset and to check correlations between time series features autocorrelation functions are used as time series data can be stationary and non stationary.

Time series data can be handed by different machine learning algorithms like stochastic time series models, Deep learning time series models, Support vector time series models. [8]

4.1 Stochastic time series models

Based on the data, stochastic processes are carried out on mathematical objects by using random variables. There are different use cases of stochastic data on availability of historic data. [8] These models are autoregressive, moving average and integrated models, after combination of these models there is some algorithms like ARIME, ARFIMA, ARMA,SARIMA.

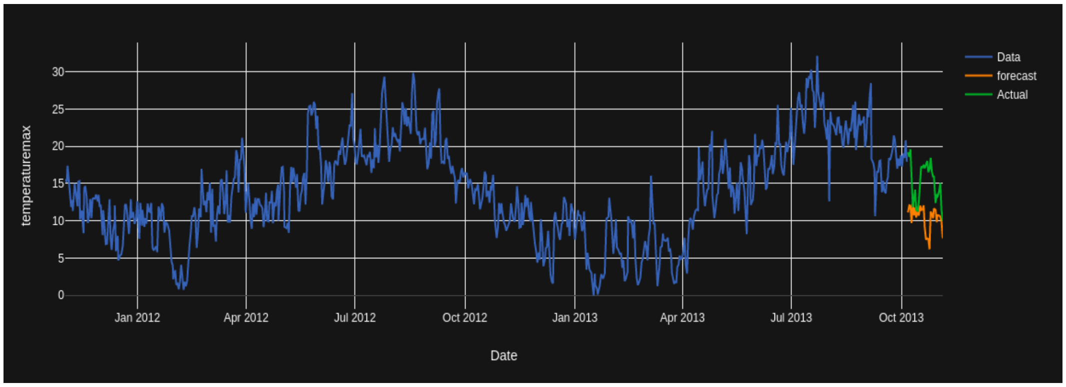


Figure 4.1: ARIMA model

4.1.1 ARIMA model

ARIMA is achromous of auto regressive integrated moving average model. ARIMA model is generalized model of ARMA,AR and MA to build composite time series model. [5]

AR- AR stands for auto regression, which basically uses dependencies of lagged observation(p) and observation and forms a regression model.
I-I stands for integrated and which basically uses measurements of differences of observations to make time series stationary.
MA- MA stands for moving average, which basically uses error rates while moving average model is used for lagged observation(q). [5]

ARIMA model is capable of handling non seasonal time series data as it has “integrate”. Partial and auto correlation functions are used for error differences and identifying frequency of difference. As ARIMA has moving average functionality it is helpful in forecasting time series data.

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}$$

Figure 4.2: ARIMA Equation

The equation for ARIMA model is containing calculation of AR and MA terms.Here xt is stationary variable,θ is autocorrelation calculated over lagged values of P,ϵ is residual errors. Another term is for calculation of MA with respect to q. This forms ARIMA model of order(p,q). [5]
Non stationary series can be converted to stationary by differentiating the present and past values of timeseries that makes ARIMA model of order(p,d,q).

ARIMA model is not able to handle the seasonality of timeseries data, Additional variant of ARIMA called SARIMA is able to handle the seasonality of timeseries data. SARIMA is denoted as (p,d,q) * (P,D,Q)S where p,d,q represents non seasonal order of AR,MA and differencing. And P,D,Q represents seasonal orders of AR,MA and differencing with S as seasonality component to ARIMA model. [5]

4.2 Deep learning time series models

Deep learning models are alternate to stochastic time series models, these models learns different patterns and regular detection to forecasting time series. [8] With help of intelligence of deep learning model seasonality are detected and data are generalized. Multilayer perceptron, RNN, LSTM are used for handling time series data.

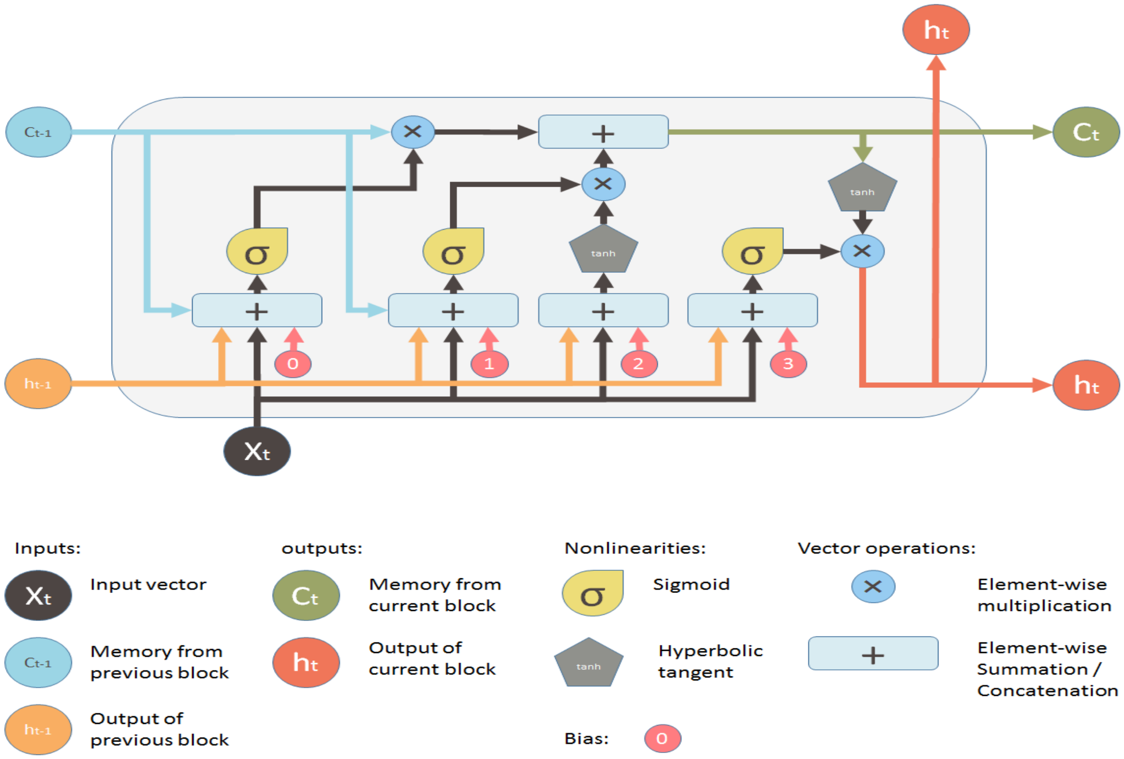


Figure 4.3: LSTM
[9]

4.2.1 LSTM

LSTM is advanced version of RNN which mainly used to predict time series related data as it can remember and hold the data for longer number of observation. LSTM consist of different dates for each of its cell and data in the cell are filtered, disposed and will be added to next coming layer. Sigmoid function is applied to neural network.
[5] There are three types of gate in LSTM:
Forget gate: It gives output signifying to keep result or ignore the result.
Memory gate: It stores new data to cell and apply tanh function to generate output vector for output gate.
Output gate: Output gate is responsible for output based on new and filtered data.
LSTM model takes training data and it is to be fitted for number of epochs, for given memory units. Mean squared error and other functions are used as optimizer. Then it forecast the result on test data. [5]

Chapter 5

Flow of Implementation

5.1 Block diagram of implementation

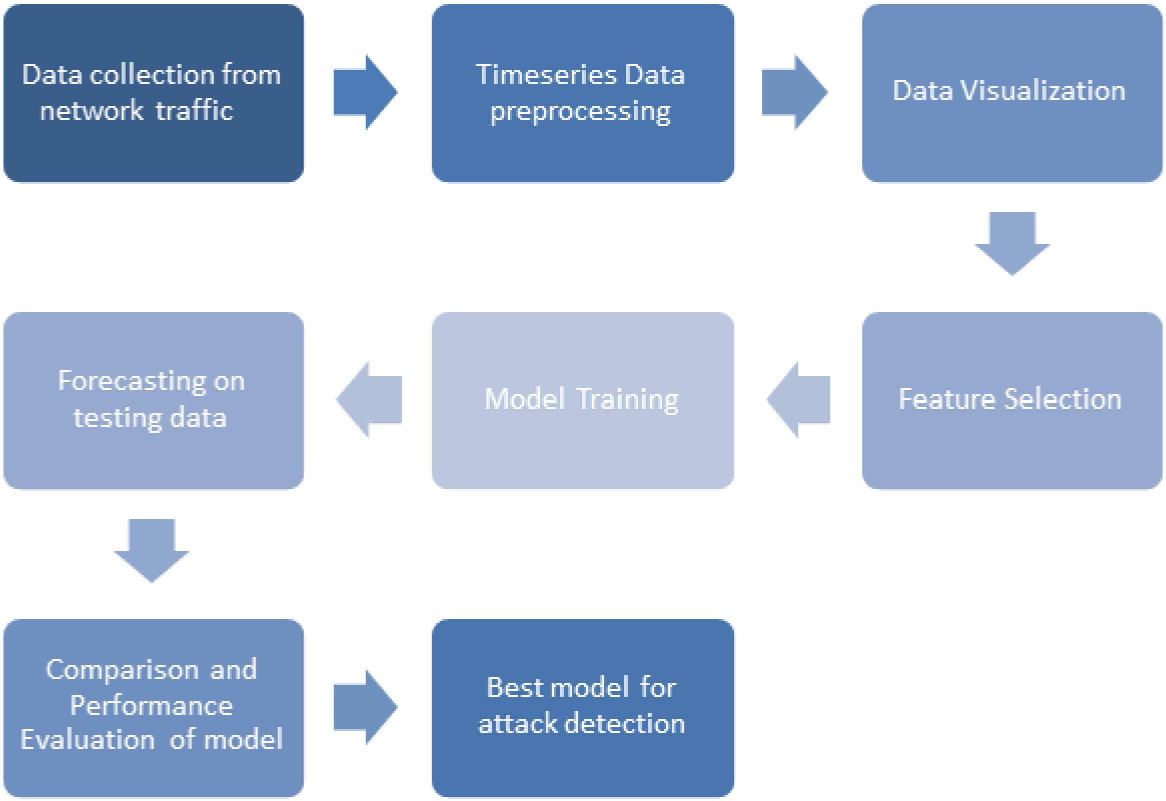

Data collection from network traffic
Timeseries Data preprocessing
Data Visualization
Feature Selection
Model Training
Forecasting on testing data
Comparison and Performance Evaluation of model
Best model for attack detection

Figure 5.1: Flow chart of Implementation

The diagram shows the flow of implementation for ddos attack detection using time series algorithms.
First raw data are collected from network where ddos attack launched by attacker. Collected raw data are transformed in machine learning features for further processing. In data preprocessing all the statistic of data are calculated and collected data are resampled according to time.

After that in Data visualization time series data are visualized to check Stationary and noise in data. Based on that suitable feature are selected and Trained on SARIMA and LSTM. On testing data forecasting has done to check get future predictions. And after that based on the performance and comparison of models best suitable model is concluded to detect ddos attack.

5.2 Algorithm For DDOS detection

Algorithm 1 DDOS attack detection using ARIMA,SARIMA, LSTM

```
Input: Series
Output : RMSE for forecasted data

Resample the data for each second
visulization of timeseries data to select feature
Tranformation of data
split data : 70 % for Train and 30% for test
Prepare data for benign traffic
Forecating :
for each t in range(test) do
    $model \leftarrow ARIMA, SARIMA, LSTM(Train)$
    $predictions \leftarrow model.forecast()$
end for
$MSE = MeanSquareError(test, Predictions)$
$RMSE = sqrt(MSE)$
return RMSE
```

Chapter 6

Experiment Evaluation

6.1 Dataset Description

In this project CICDDOS 2019 dataset is used which is developed by researchers of Canadian institute of cyber security.
The data set contains mainly two types of ddos attacks like realistic ddos attacks like DNS, MSSQL, PORTMAP, NetBios and exploitation ddos attack like syn flood, UDP flood and UDP lag. [7]

For performing DDOS attack researchers has used following architecture. Third party performs ddos attack on victim network which is protected by firewall and contains on web server and two switches to operate the 4 PCs. [7]

The data is collected from real time PCAPs of network and machine learning related more than 80 features are extracted by using CICFlowMeter-V3.
Here in this experiment model are trained on exploitation attacks for which datasets are divided into training dataset and testing dataset.

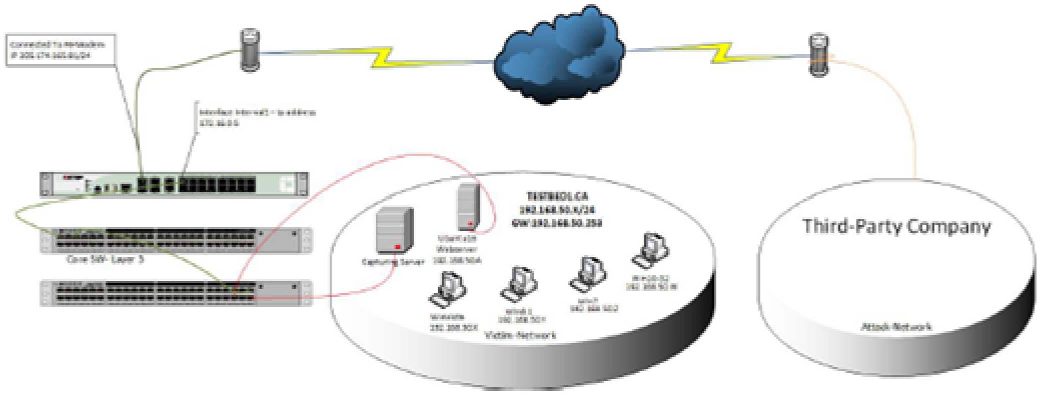


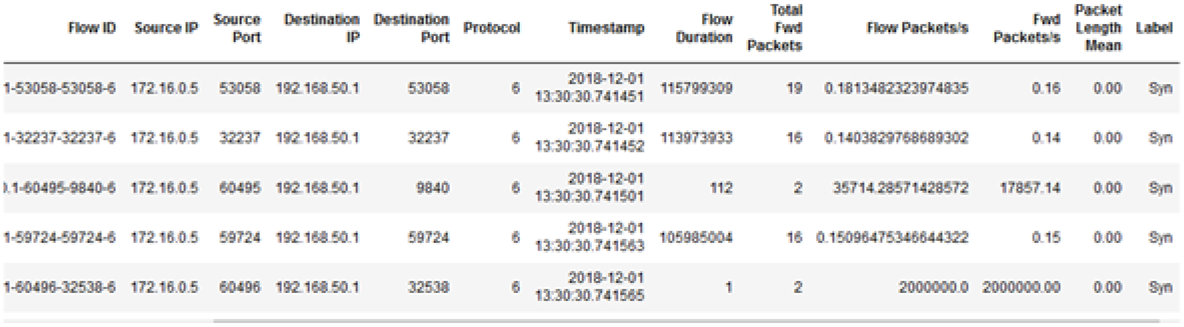
Figure 6.1: DDOS Attack Network

6.2 Libraries Used

- Libraries: statsmodel,sklearn,Keras,Tensorflow,plotly,Pandas,Numpy,MatplotLib
- Environment: Pycharm 2019
- Documentation: Latex
- Presentation: Latex

6.3 Implementation

Initial records of dataset showing major features from dataset like flowID, Source IP, source port, destination IP, destination port, flow duration , total packtes in forward direction, Mean length of packets and label.



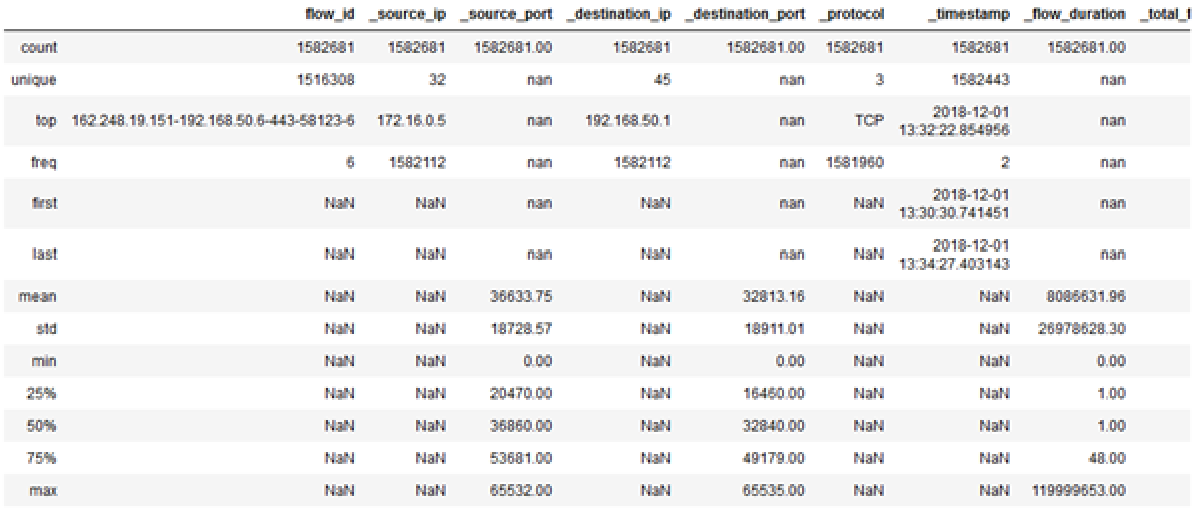
Flow ID	Source IP	Source Port	Destination IP	Destination Port	Protocol	Timestamp	Flow Duration	Total Fwd Packets	Flow Packets/s	Fwd Packets/s	Packet Length Mean	Label
1-53058-53058-6	172.16.0.5	53058	192.168.50.1	53058	6	2018-12-01 13:30:30.741451	115799309	19	0.1813482323974835	0.16	0.00	Syn
1-32237-32237-6	172.16.0.5	32237	192.168.50.1	32237	6	2018-12-01 13:30:30.741452	113973933	16	0.1403829768689302	0.14	0.00	Syn
).1-60495-9840-6	172.16.0.5	60495	192.168.50.1	9840	6	2018-12-01 13:30:30.741501	112	2	35714.28571428572	17857.14	0.00	Syn
1-59724-59724-6	172.16.0.5	59724	192.168.50.1	59724	6	2018-12-01 13:30:30.741563	105985004	16	0.15096475346644322	0.15	0.00	Syn
1-60496-32538-6	172.16.0.5	60496	192.168.50.1	32538	6	2018-12-01 13:30:30.741565	1	2	2000000.0	2000000.00	0.00	Syn

Figure 6.2: CICDDOS 2019 Dataset

6.3.1 Data Processing

In Data processing benign and attack data are segregated based on the labels. After that datatype of each feature is calculate, missing values, infinite values are handled.

Also the data is type casted based on the datatype and detailed description of data is calculated which gives insights like count, unique values , frequency, top values, mean, std and distribution of data.

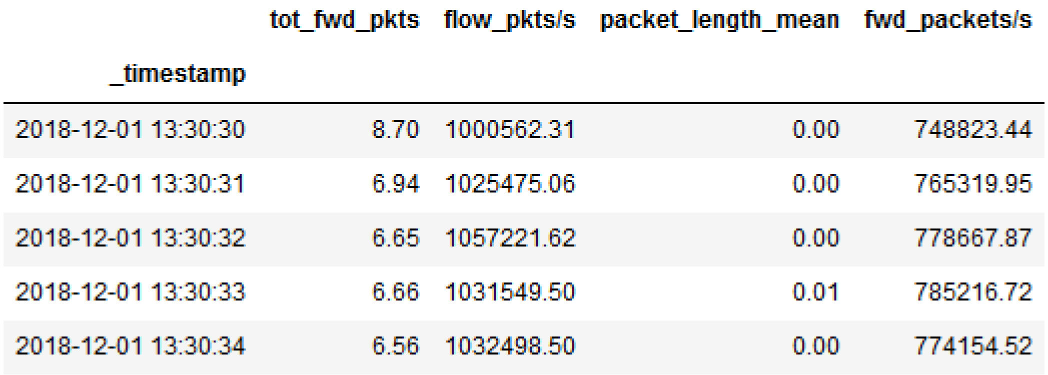


	flow_id	_source_ip	_source_port	_destination_ip	_destination_port	_protocol	_timestamp	_flow_duration	_total_f
count	1582681	1582681	1582681.00	1582681	1582681.00	1582681	1582681	1582681.00	
unique	1516308	32	nan	45	nan	3	1582443	nan	
top	162.248.19.151-192.168.50.6-443-58123-6	172.16.0.5	nan	192.168.50.1	nan	TCP	2018-12-01 13:32:22.854956	nan	
freq	6	1582112	nan	1582112	nan	1581960	2	nan	
first	NaN	NaN	nan	NaN	nan	NaN	2018-12-01 13:30:30.741451	nan	
last	NaN	NaN	nan	NaN	nan	NaN	2018-12-01 13:34:27.403143	nan	
mean	NaN	NaN	36633.75	NaN	32813.16	NaN	NaN	8086631.96	
std	NaN	NaN	18728.57	NaN	18911.01	NaN	NaN	26978628.30	
min	NaN	NaN	0.00	NaN	0.00	NaN	NaN	0.00	
25%	NaN	NaN	20470.00	NaN	16460.00	NaN	NaN	1.00	
50%	NaN	NaN	36860.00	NaN	32840.00	NaN	NaN	1.00	
75%	NaN	NaN	53681.00	NaN	49179.00	NaN	NaN	48.00	
max	NaN	NaN	65532.00	NaN	65535.00	NaN	NaN	119999653.00	

Figure 6.3: Dataset Description

As the data is collected on each millisecond of network traffic, to process the data effectively the data is resampled per second and timestamp is used as index to process timeseries. Below are the meaning of each of the important features that can be used to identify the

DDOS situation in network.
Tot fwd pkts – total number of packets in forward direction
Flow pkts/s - flow of packets/second
Packet length mean - mean value of packet length
Fwd packets/s - total number of packets in forward direction



	tot_fwd_pkts	flow_pkts/s	packet_length_mean	fwd_packets/s
_timestamp				
2018-12-01 13:30:30	8.70	1000562.31	0.00	748823.44
2018-12-01 13:30:31	6.94	1025475.06	0.00	765319.95
2018-12-01 13:30:32	6.65	1057221.62	0.00	778667.87
2018-12-01 13:30:33	6.66	1031549.50	0.01	785216.72
2018-12-01 13:30:34	6.56	1032498.50	0.00	774154.52

Figure 6.4: Resampling of data

6.3.2 Data visulization

In data visualization data, the features are plotted according to the timestamp and based on that pattern of timeseries is analyzed.

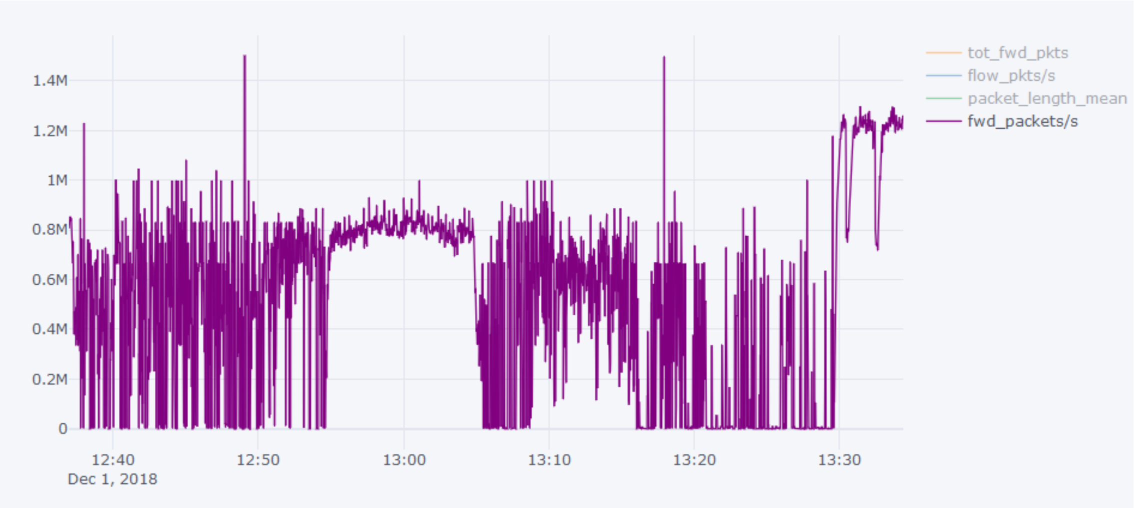


Figure 6.5: Forward packets per second vs Timestamp

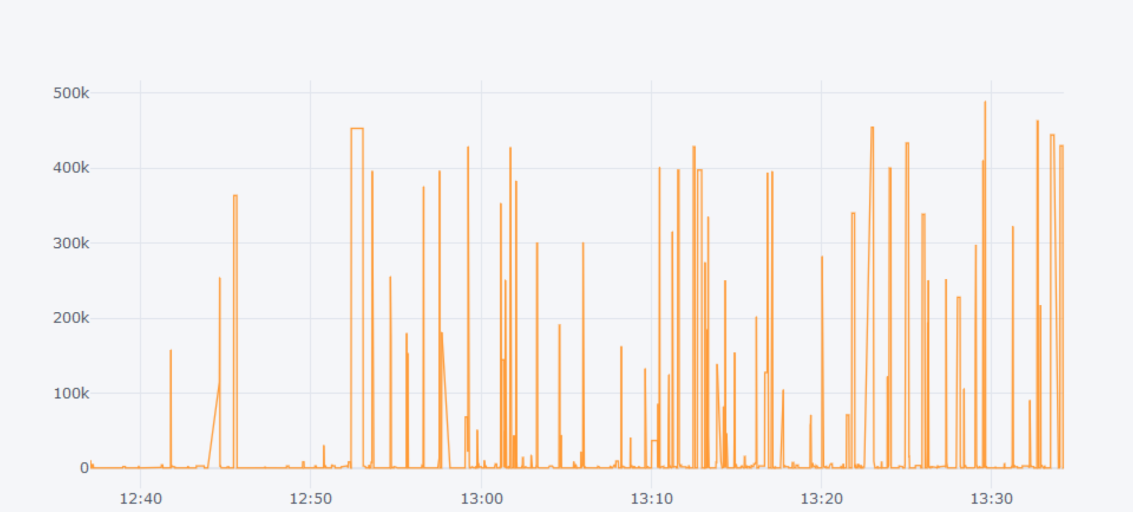


Figure 6.6: Forward packets per second vs Timestamp - Benign Traffic

6.3.3 Feature Selection

For detection of ddos attack most correlated feature is total number of packets from sources(Attackers) to destination(victim machine).

As attacker tries to flood the network, there is huge difference observed in number flowing from source to destination. Here when the traffic was benign the max number of observed packets are 490k while maximum number observed during Attack situation is 1.5M .
After selection of feature the dataset is splitted into training and validation sets to measure the performance of model.
Also dataset for benign traffic is created to evaluate the performance of model during normal situation of network traffic.

Timeseries Decomposition

The timeseries data is decomposed into trend, seasonality and noise using "additive' or "multiplicative" model to get details of each component.
Where Trend is value of timeseries in increasing or decreasing manner.
Seasonality is repetition of pattern in time series data.
Residual errors are random white noise in series. [10]

ACF and PACF Plots

Auto correlation and partial auto correlation plots are used to identify the values of p and q, d from ACF and PACF plots respectively. With the help of ACF the linearity of data is identified which determines the value of lag p.
On the other hand PACF used for identifying MA and differentiating term by differentiating timeseries and observing the lag values at different interval. [5]

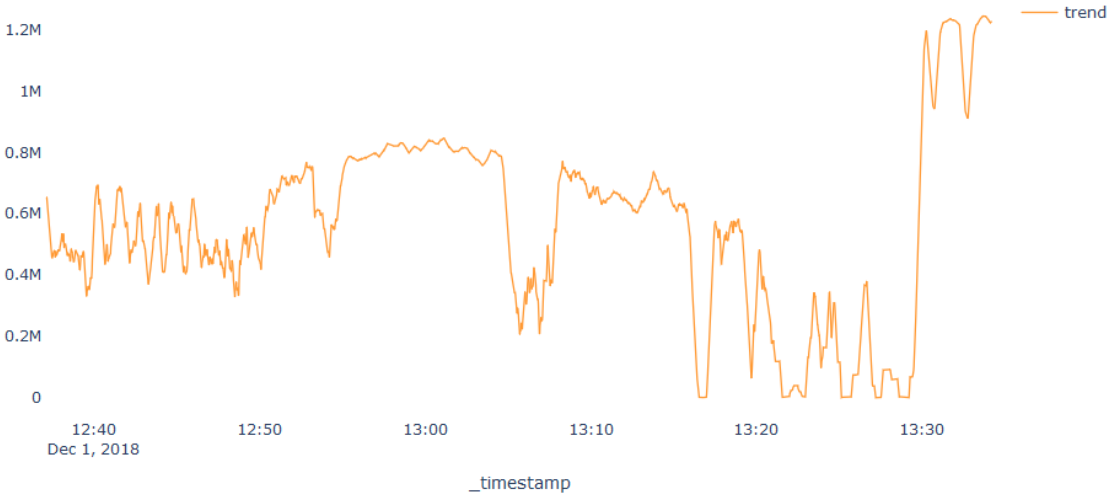


Figure 6.7: Observed trend in series

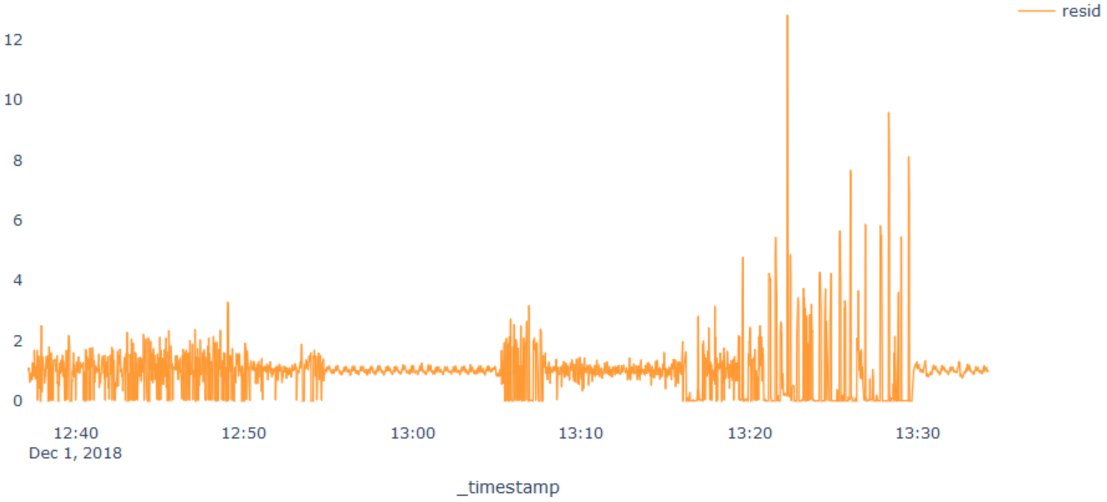


Figure 6.8: Residual errors in series

Here the ACF plots shows that present values are highly correlated with past values.

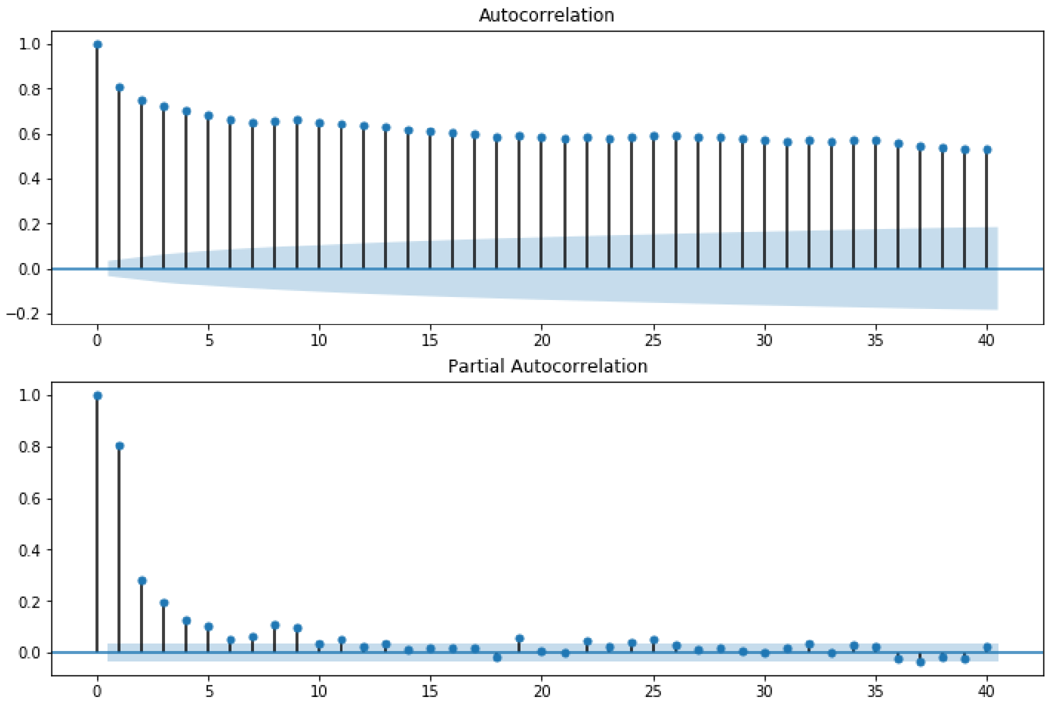
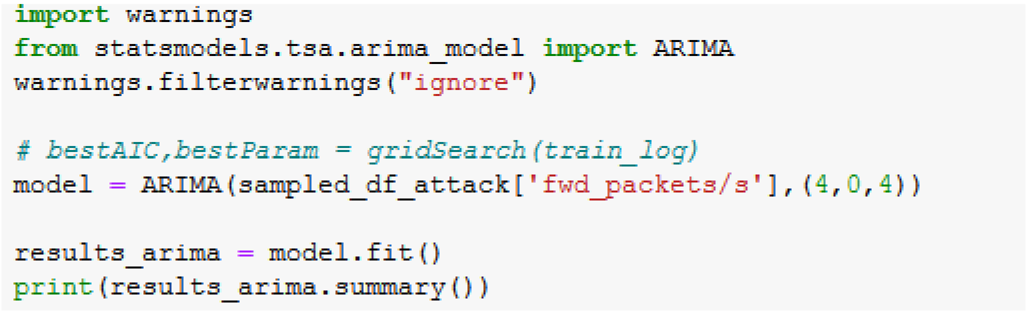


Figure 6.9: ACF - PACF plots

6.3.4 ARIMA model

ARIMA model is trained on FwD pkt with best params calculated through Auto ARIMA algorithm i.e (4,0,4) with AIC value of 93320.
After that data is forecasted on test data and plotted into graph which is almost same as the observed value of timeseries.
When data is forecasted on benign traffic there is huge difference in forecast as model is trained on DDOS dataset so it is able to identify benign traffic.



```
import warnings
from statsmodels.tsa.arima_model import ARIMA
warnings.filterwarnings("ignore")

# bestAIC,bestParam = gridSearch(train_log)
model = ARIMA(sampled_df_attack['fwd_packets/s'],(4,0,4))

results_arima = model.fit()
print(results_arima.summary())
```

Figure 6.10: Code Snippet of ARIMA

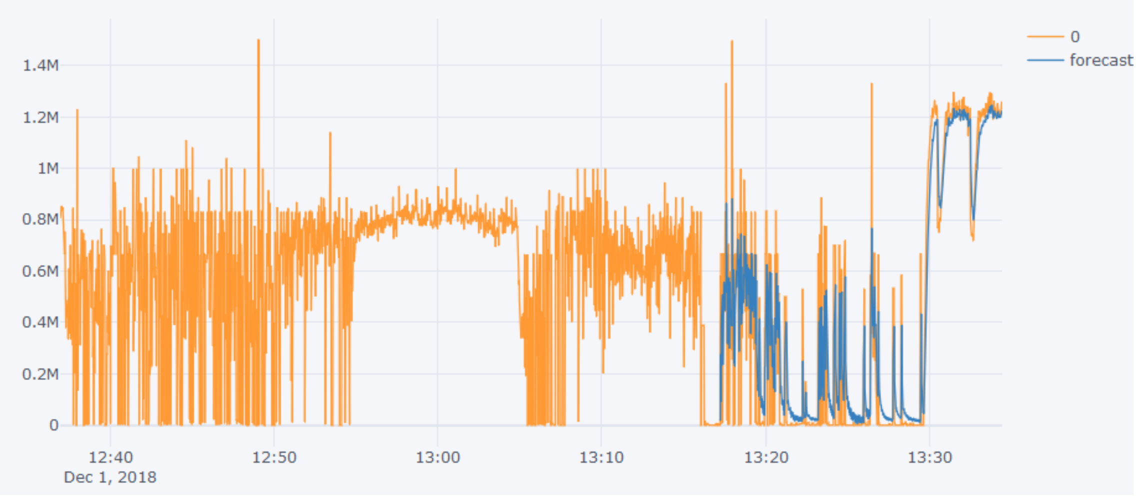


Figure 6.11: ARIMA forecast

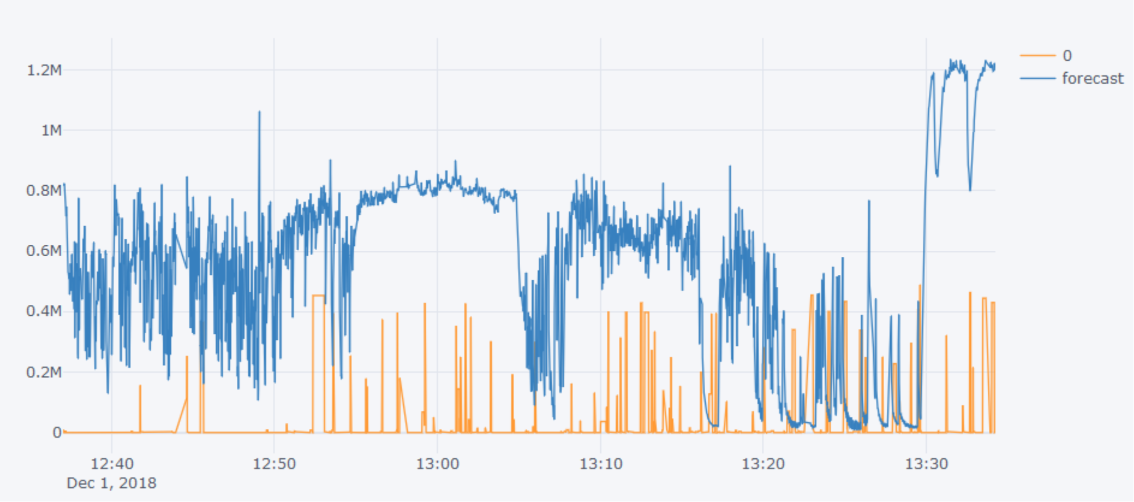


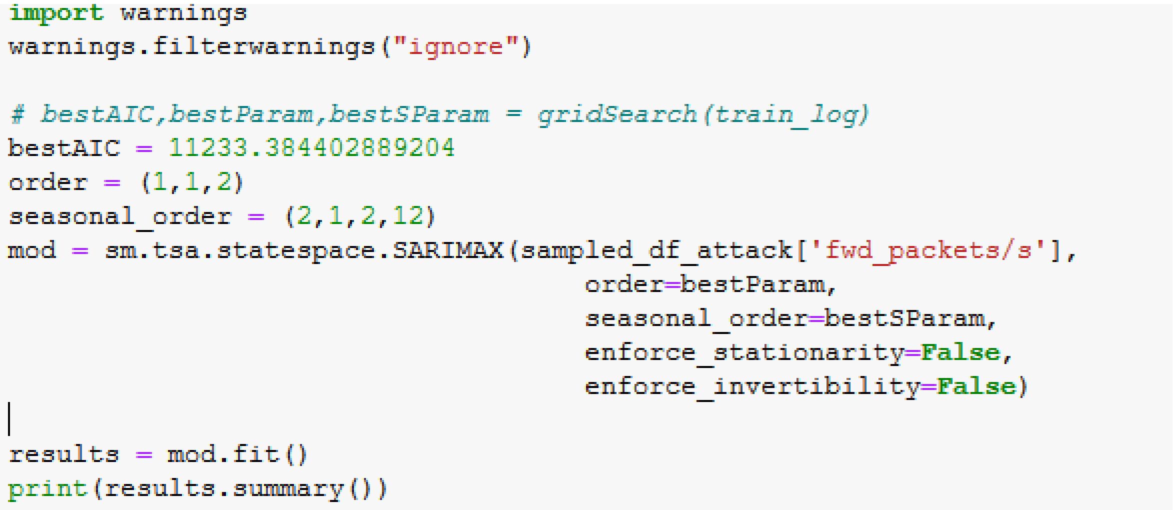
Figure 6.12: Forecast on benign traffic

6.3.5 SARIMA model

SARIMA model is trained on FwD pkt with best params calculated using grid search and from grid search based on the minimum value of AIC value of parameters are used to train model.

Here best param (p,d,q) * (P,D,Q)S are (1,1,2)*(2,1,2)12 with best AIC of 11233.

After that future traffic is forecasted for next 10 min, which shows there will be DDOS kind of situation.



```
import warnings
warnings.filterwarnings("ignore")

# bestAIC,bestParam,bestSParam = gridSearch(train_log)
bestAIC = 11233.384402889204
order = (1,1,2)
seasonal_order = (2,1,2,12)
mod = sm.tsa.statespace.SARIMAX(sampled_df_attack['fwd_packets/s'],
                                order=bestParam,
                                seasonal_order=bestSParam,
                                enforce_stationarity=False,
                                enforce_invertibility=False)

results = mod.fit()
print(results.summary())
```

Figure 6.13: Code Snippet of SARIMA

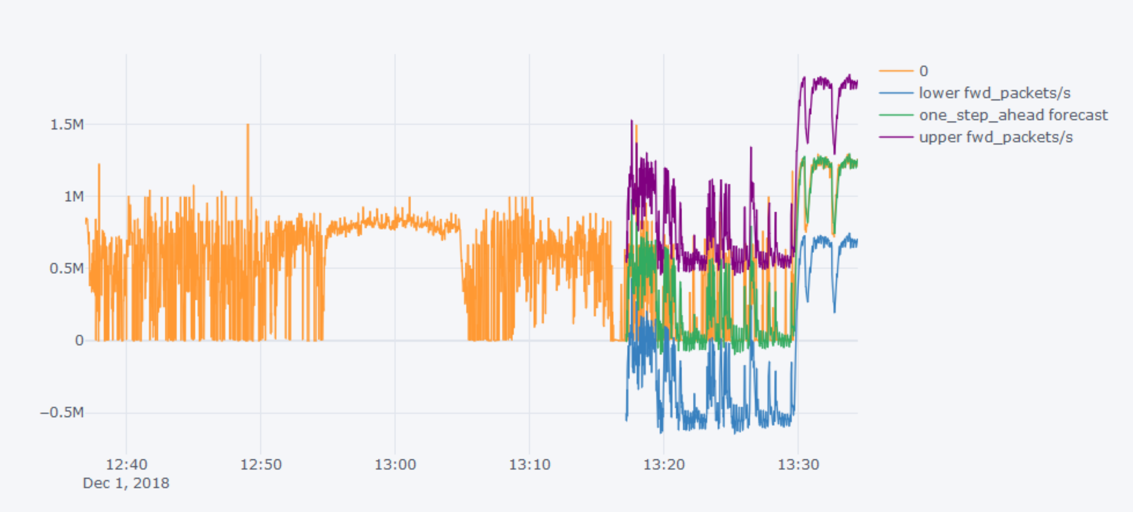


Figure 6.14: SARIMA forecast

6.3.6 LSTM Model

For LSTM first data is normalized using min max normalization. After that the target is generated by adding window component to data, so the problem is turned into supervised problem.

Architecture of model contains one input layers and 4 blocks of one LSTM layer with one dense layer for output.

0

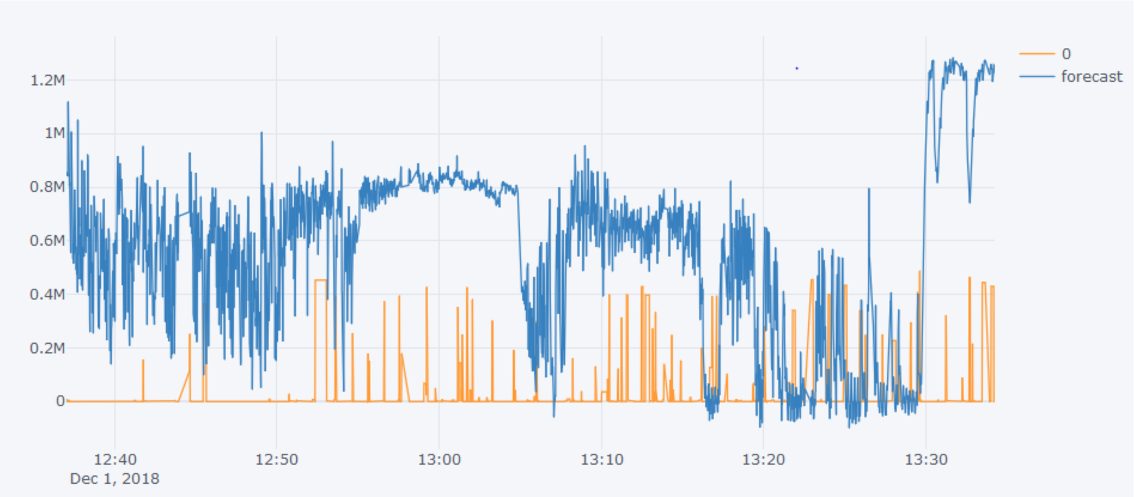


Figure 6.15: Forecast on benign traffic

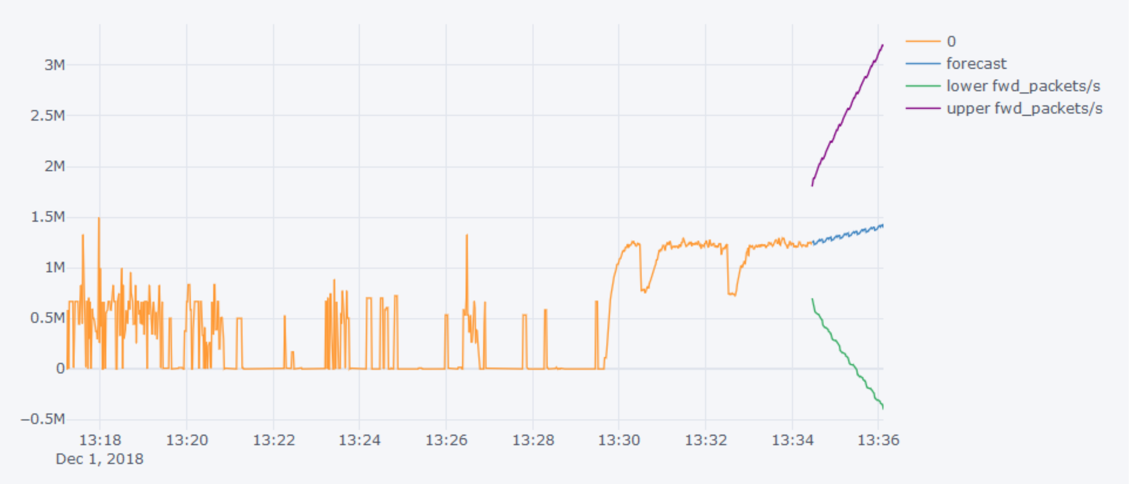


Figure 6.16: Forecast of Future traffic

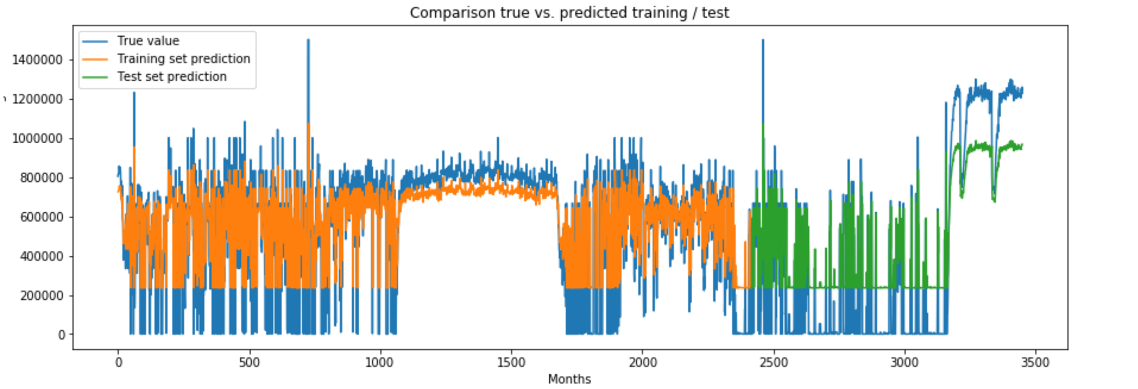
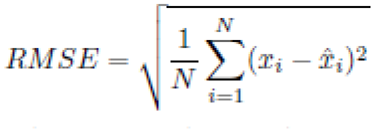


Figure 6.17: LSTM forecast

6.4 Results



$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}$$

Figure 6.18: RMSE
[5]

RMSE is used for measuring accuracy of prediction given by trained model.It calculates the errors between predicted and actual values . The formula for RMSE is as below.

Where the difference between actual values and predicted values are calculated for all N observations and divided by total number of observations. And after that squared root is taken.
With RMSE, Mean absolute error and mean absolute percentage errors is calcuated to check performace of models on the test data.

Below are the comparision of models on different errors.

Algorithm	RMSE	MAE	MAPE
ARIMA	12.18	11.78	11.33
SARIMA	13.0	13.0	24.0
LSTM	12.58	12.38	10.96

Table 6.1: Perfomance of Models

Chapter 7

Summary and conclusion

7.1 Summary

Real time ddos attack detection mainly focus on the detection of ddos attack on the given system by capturing the network traffic and extracting the features from CICDDOS 2019 dataset from Canadian institute of cyber security.
Machine learning model is trained by using ARIMA,SARIMA and LSTM to detect the ddos attack from the sequence of time series. The models are used to predict the network traffic on test data and able to predict the data.
Based on the predicted data and test data RMSE is calculated to check performace of trained models.
Also the predictions and forecast is generated for benign traffic of CICDDOS data, on which the value of RMSE is was high.
Based on the values of RMSE on test data and benign data the alert messages are generated for DDOS situation in the network.
Here from all the three models performance of ARIMA is noticeably good. All three models can be used to forecast the traffic.Also SARIMA can be used to make future forecast for next couple of minutes. Still the performance of LSTM can be improved by experimenting batch size, epochs, optimizer.

7.2 conclusion

DDOS attack is back door to viruses and worms to enter into any computer system and network.In order to secure system machine learning models are accurate then traditional signature based methods.ARIMA and LSTM models are trained to identify initial stage of ddos attack.

Bibliography

[1] S. Meysam, T. Nezhad, M. Nazari, and E. A. Gharavol, “A Novel DoS and DDoS Attacks Detection Algorithm Using ARIMA Time Series Model and Chaotic System in Computer Networks,” *IEEE Communications Letters*, vol. 20, no. 4, pp. 700–703, 2016.

[2] R. F. Fouladi, C. E. Kayatas, and E. Anarim, “Statistical Measures : Promising Features for Time Series Based DDoS Attack Detection †,” 2018.

[3] Y. Chen, X. Ma, and X. Wu, “DDoS Detection Algorithm Based on Preprocessing Network Traffic Predicted Method and Chaos Theory,” *IEEE Communications Letters*, vol. 17, no. 5, pp. 1052–1054, 2013.

[4] R. C. Staudemeyer and C. W. Omlin, “Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data,” pp. 218–224.

[5] S. Siami-namini and N. Tavakoli, “A Comparison of ARIMA and LSTM in Forecasting Time Series,” *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1394–1401, 2018.

[6] X. Liang and T. Znati, “A Long Short-Term Memory Enabled Framework for DDoS Detection,” *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.

[7] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, “Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy,” no. Cic, 2019.

[8] S. Halder and S. Ozdemir, *Hands-On Machine Learning for Cybersecurity: Safeguard your system by making your machines intelligent using the Python ecosystem.* Packt Publishing, 2018. [Online]. Available: https://books.google.co.in/books?id=LR2CDwAAQBAJ

[9] lstm image. [Online]. Available: https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714

[10] timeseries decomposition. [Online]. Available: https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/