**A Machine Learning-Based Classification andPrediction Technique for DDoS Attacks**

**Alternative title:**

Predictionof DDoS attacks using machine learning based approach.

**Aim:**

We proposed a complete systematic approachto detect DDOS attack using machine learning algorithm.

**Abstract:**

 Distributed network attacks are referred to as Distributed Denial of Service (DDoS)attacks. These attacks take advantage of specific limitations that apply to any arrangement asset, such asthe framework of the authorized organization's site. In the existing research study. It is necessary to work with the latest dataset to identify the current state of DDoSattacks. In this presented work, used a machine learning approach to predict DDoS attack types.For this purpose, used Random Forest and XGBoost classification algorithms. To access the researchproposed a complete framework for DDoS attacks prediction. To meet the proposed objective, we used UNWS-np-15 dataset and Python was used as a simulator. After applying themachine learning models, we generated a confusion matrix for identification of the model performance.In the first classification, the results showed that both Precision (PR) and Recall (RE) are 88% for theRandom Forest algorithm. In the second classification, the results showed that both precision(PR) and Recall(RE) are approximately 90% for the XGBoost algorithm

**Synopsis:**

Distributed network attacks are referred to, usually,as Distributed Denial of Service (DDoS) attack. A DDoS attack sends different requests (with IP spoofing) to the target web assets to exceedthe site's ability to handle various requests, at a given time,and make the site unable to operate effectively and efficiently\_ even for the legitimate users of the network. Typically,the target of various DDoS attacks are web applicationsand business websites; and the attacker may have differentgoals.

**Existing System:**

 CNN and RNN both are two different algorithmsthat can be used for different purposes. For example, CNN isused for feature extraction and RNN is used for regression intime series data utilization. Though both CNN and RNN based model producing accurate results, it is very long and time consuming process.

**Problem Definition:**

 The authors used the CNN andRNN model for intrusion detection.This is avery long and time-consuming process. Therefore, it is very important to perform advanced machine learning techniquesto model optimization that train the best model for highlyaccurate work.

**Proposed System:**

Among the machine learning techniques, random forestand XGBoost both are powerful supervised learning models.Both are applicable and used for classification problems. Therandom forest algorithm is approximately 100 times fasterthan other algorithms and best working for classificationproblems.

**Advantage:**

 It is approximately100 times faster than the random forest and best for forbiddata analysis. Both are simple and faster than other algorithmin terms of execution times.

**Algorithm:**

 After preprocessing dataset, that data will be given to the machine learning algorithm. Machine learning algorithm analyzes the data and predict types of DDOSs attack .

**Random Forest Classifier**

 A random forest algorithm is a combination of the decisiontree. It is very fast compared to other classification. Now afterfeature scaling the next step is the machine learning classification model. In our proposed work we used a random forest classification algorithm. The random forest, which is one ofthe most popular and powerful machine learning classification algorithms, is used for reaching a lot of decisions in theproposed model.In the first classificationwe observed that both the Random Forest Precision (PR)

**XG Boost**

In the era of machine learning and artificial intelligence, theXGBoost algorithm is known as the queen by scientific andacademic researchers. Most of the researchers considering asa weapon for big data utilization. This model also working on tree but 100 times faster than other models. The XGBoostlearning model have very fast speed, scalability, efficiencyand simplicity. This model is more reliable for big data.This model is working on probability. The confusion matrixand outcomes of the classification, are given below, for theXGBoost method.XGBoostPrecision (PR) and Recall (RE) are approximately.

**Modules:**

* Dataset Collection
* Data Pre-process
* Detection

**Dataset Collection:**

CollectedUNSW-nb15 dataset from GitHub1 that containsfeatures' data about the DDoS attacks. This datasetis provided by the Australian Centre for Cyber Security (ACCS). The dataset consists of different featuresabout the DDoS attacks including an ID number, Protowhich presents medium of the network, label of the attacks,and attacks' cat which presents the severity of the DDoSattacks.

**Data Preprocessing:**

 Data preprocessing it is very important and time-consuming part of data analysis. here we are going to separate relevant data from irrelevant data and convert it to quality information. For this step we are using statistical techniques to clean data and replace those values which are not important in our experimental analysis. This is essential of every data analysis for the initial phase examination. After that, we will be able to convert informationinto reliable form. After analyzing data in the data pre-processing phase, we also observed and identified that our datasets are almost clean.

**Detection of DDoS:**

 To design and develop an approach using supervised machine learning classifiers for DDoS attack detectionbased on different techniques.

 We have studied various methodologies whichare used for detection of Distributed Denial-of-Service(DDoS) Attacks on Software Defined Network, basedon the findings andresults we have concluded that theAttribute based Double of TransductiveConfidenceMachines for Random forest classifier method gives moreefficient way to find out anomalous flow in Software DefinedNetwork

**Hardware Requirements:**

* Hard Disk : 500GB and Above
* RAM : 4GB and Above
* Processor : I3 and Above

**Software Requirements:**

* Operating System : Windows 10 (64 bit)
* Software : Python
* Tools : Anaconda

**Conclusion:**

In this paper, we proposed a complete systematic approachfor detection of the DDOS attack. First, we selected theUNSW-nb15 dataset from the GitHub repository that containsinformation about the DDoS attacks. This datasetwas provided by the Australian Centre for Cyber Security. Then, Python and jyupter notebook areused to work on data wrangling. Secondly, we divided thedataset into two classes i.e. the dependent class and theindependent class. Moreover, we normalized the dataset for the algorithm. After data normalization, we applied the proposed,supervised, machine learning approach. The model generated prediction and classification outcomes from thesupervised algorithm.

**Future Work :**

Looking to the future, for functional applications, it is important to provide a more user-friendly, faster alternative to deep learning calculations, and produce better results with a shorter burning time. It is important to work on unsupervised learning toward supervised learning for unlabeled and labeled datasets. Moreover, we will investigate how non-supervised learning algorithms will affect the DDoS attacks detection, in particular, we non-labeled datasets are taken into account

 Input data

Machine Learning

Model Creation

Test data

Prediction result

Preprocessing

**Architecture Diagram:**