Machine Learning based Network Security

Tapasya Bari[1], Vaidehi Borwankar[2], Nipoon Donta[3], Prateek Jha[4], Nishta Verma[5]

[1,2,3,4]BE EXTC Student, Rajiv Gandhi Institute of Technology, Mumba , India.
[5]Assistant Professor, Rajiv Gandhi Institute of Technology, Mumbai, India.

---***---

Abstract - *Machine Learning Algorithms are special Category of Algorithms which allows software applications to become accurate in predicting outcomes without being explicitly programmed, Large Amount of data can be analyzed to give results and work is automated using machine learning. Machine learning algorithm are used to build classifiers which classify the packets as malicious or non- malicious.*

The main aim is to perform traffic monitoring, analyse it and govern the intruders. On basis of features of the captured packets on the network, traffic classifier system is developed by using dataset of botnet traffic. The classifier is designed to detect DDOS (Distributed Denial of Service) virus detection system based on data-mining algorithms to detect the infected computers quickly using Bayes Classifier

Keywords: Machine learning, Malicious packet, Distributed Denial of service, Bayes Classifier, Attacker, packet capturing

1. INTRODUCTION

With the rapid expansion of Internet during recent years, security has become an essential issue for computer networks and computer systems. In current modern network size of captured data is increasing exponentially. One can see on television or from newspapers about the news of hackers stealing confidential data for illegal usage, and they may use a variety of methods such as Distributed Denial of Service (DDoS), Spam and Trojan. These methods require the cooperation of many computers, so hackers often spread out malicious software to achieve the goal of attacks. Therefore, it is important to enhance network security while developing and applying new information technologies to prevent from illegal access to confidential information.

1.1 Algorithm

The present paper relates to an algorithm which helps in classification of malacious and normal traffic. This type of classification is important for network security and privacy. Traffic classification classifies network on basis of different parameter. Later well assigned port number were used to classify network. In fast growing network many application are using dynamic changed port number which is making port based classification monotonous job. After this payload came into picture. This classification canachieve good accuracy and they can be accessed and inspected properly.

1.2 Classifier

In spite of good accuracy and proper inspection payload has its own limitation. It is slow and consume lots of resources. Many of author in research community has proposed automatic mechanism for derivation of payload features. Some promising result were proved, but still their approaches have their own limitation. The methodology discussed by them requires lots of processing time and large memory. Since size of the network data is increasing day by day with the advance technology researcher have been using machine learning technique based on the features to classify data Machine learning based algorithms create the classification model by using the large data set and calculated features. Moreover, the statistical properties based features of the network traffic is also becoming important for machine learning based classifications such as packet length statistics for a network traffic flow, for example the minimum, mean, maximum, standard deviation of the packet sizes., a good traffic classifier can be developed with the consideration of the Machine learning(ML) based techniques and based on these calculated features statistics. While ML classifiers have shown good efficiency and promising accuracy, payload-based classifiers has more accuracy

2. METHODOLOGY

PHASE1

In this network this system, two modules system and user are included. Flowchart of the website and data packet capturing are given below

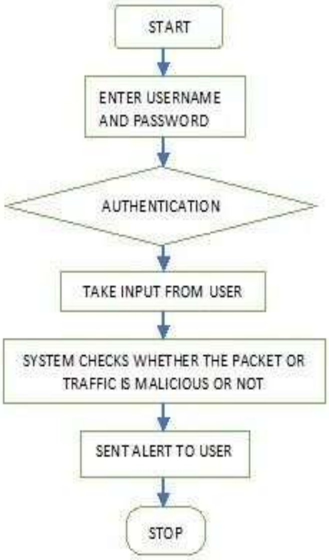

START
ENTER USERNAME AND PASSWORD
AUTHENTICATION
TAKE INPUT FROM USER
SYSTEM CHECKS WHETHER THE PACKET OR TRAFFIC IS MALICIOUS OR NOT
SENT ALERT TO USER
STOP

Fig 1. Flow chart of dummy website

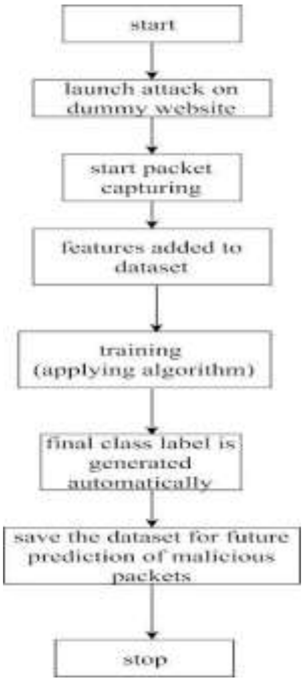

start
launch attack on dummy website
start packet capturing
features added to dataset
training (applying algorithm)
final class label is generated automatically
save the dataset for future prediction of malicious packets
stop

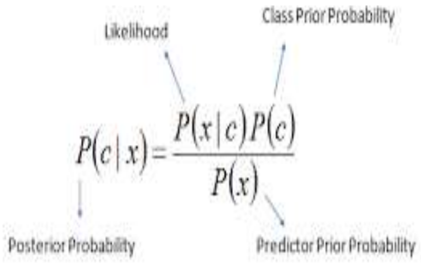
Fig 2. Flowchart Of ML algorithmn

The data being uploaded on dummy website which is further attacked by a system operator. After clicking on packet capturing option, the packets that are being transferred to any server accessing same site within the same network are captured. The captured packets include all the network parameters which are then classified by machine learning algorithm.

PHASE 2

COMPARITIVE ANALYSIS NAÏVE BAYES

According to the analysis, Naïve Bayes classifier are simple probability classifiers which are based on the Bayes theorem. It performs classification accurately and prediction as compared to the others. It requires several parameters for the prediction. The basic theory of Bayes classifier is based on the Bayesian theorem and the formula is given below.



$$P(c|X) = P(x_1|c)\times P(x_2|c)\times\cdots\times P(x_n|c)\times P(c)$$

Where X represents unknown data and C represents known class. This study used UDP packets, total IP number and port number for outward connections. The important part of Bayes classifier is to compute probability that X appears in known C class.

SVM

SUPPORT VECTOR MACHINE

It is a supervised learning model with associated learning algorithm which analyze data by using regression analysis.

A model is built by using the SVM training algorithm which assigns examples to new category or other. It makes non-probabilistic binary linear classifier

We have designed machine learning algorithm namely naïve based algorithm and support vector machine algorithm (SVM) in order to differentiate the network traffic. They have a very high accuracy rate for spam detection being as high as 98%. The data set is created by capturing the packets from applying

DDOS (Distributed Denial Of Service) attack. The data processing, feature extraction and data labelling was done on the captured packets. Each module has its own significance. We propose to make a real time system with the real time classifier for the classification of internet traffic.

3. CONCLUSION

We have designed machine learning algorithm namely naïve based algorithm and support vector machine algorithm (SVM) in order to differentiate the network traffic. They have a very high accuracy rate for spam detection being as high as 98%. The data set is created by capturing the packets from applying DDOS (Distributed Denial Of Service) attack. The data processing, feature extraction and data labelling was done on the captured packets. Each module has its own significance. We propose to make a real time system with the real time classifier for the classification of internet traffic.

REFERENCES

1) A P2P Botnet Virus Detection System Based On Data- Mining Algorithms Wernhuar Tarng, Cheng-Kang Chou and Kuo-Liang Ou October 2012

2) Using Machine Learning Techniques to Identify Botnet Traffic Carl Livadas, Bob Walsh, David Lapsley, Tim Strayer Internetwork Research Department.