AMRITA VISHWA VIDYAPEETHAM

BACHELOR THESIS

An Integrated Approach to Network Intrusion Detection & Prevention

Author:
Bhanu Prakash
AM.EN.U4ECE15020
Kaki Yeswanth
AM.EN.U4ECE15028
Sai Srinivas
AM.EN.U4ECE15033
S Balaji
AM.EN.U4ECE15052
Chandra Sekhar
AM.EN.U4ECE15061

Supervisor:
Aswathy K Nair
Assistant Professor
ECE

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Technology in Electronics and Communication Engineering

in the

Department of Electronics and Communication Engineering
Amrita School of Engineering
Amritapuri Campus

AMRITA VISHWA VIDYAPEETHAM

Contents

BONAFIDE CERTIFICATE

This is to certify that the project entitled Intrusion Detection in Computer Networks is the bonafide work carried out by **Bhanu Prakash (AM.EN.U4ECE15020)**, **Kaki Yeswanth(AM.EN.U4ECE15028),Srinivas(AM.EN.U4ECE15033),Balaji(AM.EN.U4ECE15052), Chandra Sekhar (AM.EN.U4ECE15061)** students of B Tech in Electronics and Communication Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Amritapuri Campus, in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in **Electronics and Communication Engineering** and that the project has not formed the basis for the award previously of any degree, diploma, associate-ship, fellowship or any other similar title.

Signature of the Supervisor:
Name of the Supervisor:
Place:
Date:

Signature of the Chairperson of the Evaluation Panel:
Name of the Chairperson of the Evaluation Panel:
Place:
Date:

Signature of the Department Chairperson:
Name of the Department Chairperson:
Place:
Date:

Declaration of Authorship

We, Bhanu Prakash (AM.EN.U4ECE15020),Kaki Yeswanth(AM.EN.U4ECE15028),Sai Srinivas (AM.EN.U4ECE15020),S Balaji (AM.EN.U4ECE15052),Chandra Sekhar (AM.EN.U4ECE15061) declare that this thesis titled, “Intrusion Detection in Computer Networks” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a bachelor’s degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where We have consulted the published work of others, this is always clearly attributed.
- Where We have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- We have acknowledged all main sources of help.
- Where the thesis is based on work done by ourself jointly with others, We have made clear exactly what was done by others and what we have contributed our-self.

Signed: _____

Date: _____

Acknowledgements

I would also like to express my gratitude for the immeasurable motivation and guidance provided by Sri. Mata Amritanandamayi Devi (AMMA), Chancellor of Amrita University.

We sincerely thank our college Amrita school of Engineering, Amritapuri for giving us an opportunity to undertake a minor part of final year project.

We are grateful to our project guide Ms. Aswathy K Nair for giving us the idea of Project and guiding us by giving valuable information throughout the mini project.

We are highly grateful to Dr. Sundararaman Gopalan (Head of the Department, ECE) for his commendable initiative and encouragment throughout the project. Our sincere thanks to Prof. Rajesh Khannan for giving us valuable suggestion during project review and all the members of project review panel.

We are indebted to Project Coordinators for their valuable support. At last we must express our sincere heartfelt gratitude to all the staff members of Electronics and Communication Engineering Department who helped me directly or indirectly during this course of work. Above all we are always thanks to God Almighty for giving us good health and mental strength for completing project.

Abbrevations

KNN : K Nearest Neighbors

SVM : Support Vector Machine

DOS : Denial of Service

POD : Ping of Death

UDP : User Datagram Protocol

AMRITA VISHWA VIDYAPEETHAM

Abstract

Faculty of Engineering
Amrita School of Engineering
Amritapuri Campus

Bachelor of Technology in Electronics and Communication Engineering

An Integrated Approach to Network Intrusion Detection & Prevention

by Bhanu Prakash (AM.EN.U4ECE15020),Kaki Yeswanth(AM.EN.U4ECE15028),Sai Srinivas (AM.EN.U4ECE15020),S Balaji (AM.EN.U4ECE15052),Chandra Sekhar (AM.EN.U4ECE15061)

In present generation, with the expansion of size of internet, security plays a crucial role in computer networks. Also with the advancement of Internet of things, earlier technology like firewall, authentication and encryption are not effective in ensuring complete security. This has lead to the development of IDS which is a process for monitoring events in a computer networks which are threats to computer security. With the help of various machine learning algorithms we have carried out the implementation of IDS. Machine learning technique increases the accuracy of anomaly detection in real time scenario. This work focuses on KNN classifier and SVM , which classify the program behavior as intrusive or not. After analyzing the data we have identified DOS occurs most number of times. So we have solved two types of DOS attacks Ping of Death (POD) and UDP. The algorithms have been implemented in Python language and implementation has been carried out in Wireshark and Cisco packet tracer.

Keywords: Computer Security, Intrusion Detection, Support Vector Machine, K Nearest Neighbors.

CHAPTER 1

1. Introduction

Computer security vulnerabilities prevail as long as we have flawed security measures, insecure software programs and weak authentication mechanisms and network protocols. Security products and Intrusion Detection System (IDS) are the existing methods to tackle the threats in computer networks. The latter is more intelligent as it gathers all the information and study the behavior and based on that it could characterize a programs' normal behavior. IDS detects network intrusions and protects the personal computer network from unauthorized users. It helps to distinguish between bad connections and good connections.

IDS use efficient machine learning algorithms that could learn the usual patterns to classify or distinguish the next coming behavior as anomaly or not. There are mainly four attacks that affect the Computer Networks. They are categorized into:

1) **DOS :**
   It is the cyber-threat in which intruder seeks to make a network resource unavailable to the user who is connected to the Internet temporarily by flooding a number of requests in attempt to overload the systems[9].

2) **R2L :**
   In this type of attack the user who is sending packets of information from the Internet does not have any access and it may have the possibility of being attacked or damaged. [9]

3) **U2R:**
   In this type of attack, the Intruder hacks the system with normal user account details and attacks the system and gain the super control over the system just like the local user. [9]

4) **Probing:**
   In this attack the Intruder scans the networking device and tries to find the weakness of the device which can later be damaged. It is used in data mining. [9]

The two types of Intrusion Detection are anomaly and misuse detection. The first type of detection deals with the behavior of user. Any change observed in the normal user behavior is called as anomaly detection. The other type of detection operates with desired type of Patterns by matching the original Patterns (Eg : Signatures). In Detail explanation about the two types of Detection is given below.

1) **Anomaly Detection:** Every Computer or a Person is capable of performing certain specific functions. Any change observed in the functionalities of a Person or a Computer is considered as anomaly behavior. For Example Consider a Teacher, The normal

functionalities of teacher is to teach students, give assignments and clarify doubts of a student. The anomalous functionalities of teacher can be allowing students to copy or helping them in exams. So, Here the change in behavior is observed and is called as anomaly detection.

Advantages:
1. No Knowledge needed for possible attacks.
2. Universal validity

Disadvantages:
1. High number of false Positives.
2. Difficulty in defining the normal behavior

2) Misuse Detection

The Basic Principle of this detection is describing an intrusion by indications and signs. These signatures should be given to IDS (Intrusion Detection System) in some abstract form. Later these signatures are used in matching the new signatures to identify an Intrusion.

Advantages:

1. Small number of false positives.
2. Small performance overhead

Disadvantages:

1. Not possible to detect new attacks.
2. Correctness of used patterns.

With Machine learning techniques like SVM (Support Vector machines) and kNN (k Nearest Neighbors) we could identify the intrusions and thereby take remedial security measures. Before getting an understanding on SVM and kNN, we should know what Machine Learning is about.

i. Machine Learning:

The term machine learning is a branch of Computer Science which is closely related to data mining. It is a study of algorithms and statistical models to improve the performance of a specific task. These algorithms are used in applications of Intrusion Detection in Computer Networks, spam emails and some other areas where it is impracticable to develop an algorithm for performing the tasks. There are many ways in which the machine learns, but the three important ways are

1. Supervised Learning:

It is the branch of machine learning that maps an input variable(x) to an output variable (Y) based on the labeled training Data. The algorithm is intelligent enough to map the function with a new input data, x and we could predict the output, Y. As the name implies the supervised learning algorithm acts as a supervisor which teaches the machine using well labeled data and when a new set of data is given as input it could predict the correct output from labeled data.
For Example you have one million coins with different currencies like rupee, euro and dollar weighing 1 gram, 2 gram and 3 respectively. The machine will predict the currency of the coin. When we feed this data to machine it identifies any coin of weight 1 gram is rupee and 2 gram is euro etc.

2. **Unsupervised Learning:**
   It is the branch of machine leaning that tests the data which has not been labeled or categorized. Here the role of algorithm is to classify or group the unsorted data *itself* based on patterns, similarities or differences. For Example Consider a list of cricket players with data set with runs scored and wickets taken. When we feed this data to machine it identifies and plot the graph between count of runs and count of wickets. The graph will contain two clusters with more number of runs less number of wickets and less number of runs and more number of wickets. The players with highest no. of runs and less no. of wickets are considered as batsmen and the players who has less no. of runs and more no. of wickets are considered as bowlers.

3. **Reinforcement Learning:**
   It is the branch of Machine learning that works on the principle of feedback of the data given. Suppose you have given the picture of dog to the machine and ask to identify it, The machine answers incorrectly and says cat or any other animals name.
   Then we give negative feedback to the machine saying that it id Dog. So the next time when you give the picture of Dog, it identifies as dog.

ii. Contribution of this work:

- Using the extracted data set we worked on DoS attack. We identified the type of attack
- Using kNN we were able to separate anomaly data set from actual data set. identified the anomaly data set from actual one.
- From the test results, it was shown that the system accuracy is high using this method.

iii. Overview

As Technology advances number of security threats are increasing exponentially. In this project we have discussed important basic network threats and how to overcome those network issues. Firstly we are going to take data set of networks attributes which gives the information about the IP address is malicious or not. Machine learning is most popular to classify the data. In

machine learning we have chosen most popular supervised algorithm KNN.KNN algorithm classified data based upon K nearest neighbors. So when we classify data using KNN algorithm, we will get result of IP address which are malicious. This project mainly focus to solve DOS attack.DOS is most powerful attack which destroys server by sending continuous request to server. To analyze this issue we need a network and server. To solve DOS attack we will setup network and check every IP address whether it is malicious or not. If IP address found as normal it request will send to server. In this project we are implementing Metropolitan Area Network using Packet Tracer. It is a tool which helps to create virtual networks and analyzing network real time. The classified data using KNN is given as input data manually to the Metropolitan Area Network. In that network every system tries to access server placed at remote network. The Router which routes to server is programmed with IP address to block. In routers Extended Access List feature helps to filter IP address to reach server.

DOS has many categories, in that Ping Of Death is the most powerful DOS attack. In our project we are going to develop Ping Of Death attack. If we develop an attack, analyze the attack lot of possibilities and underlying principles for forming the attack can be learnt. This attack implementation gives fundamentals and ideas for creating intelligent software's which solves Ping of Death attack in future. Firstly we have chosen Wireshark, a software which helps to monitor and analyze the incoming requests and outgoing requests from the PC. In Ping of Death we are specifically focusing on ICMP Protocol. ICMP Protocol helps for communication of systems in same network. For humans communication lot of software's like facebook, whatsapp are developed. For System communication ICMP protocol will be so helpful. For example if system-A want to communicate or to get status of System-B. It can use Command PING followed by respective IP address of System. This Ping command send a request to respective IP address system using data as packets. If respective system receives it will send some data packets as acknowledgement to the sender. If we send requests thousands and thousands in seconds, systems can't handle. If we apply same concept and logic to servers than we have developed Ping Of Death attack. Internet is so vast, so we have found a software (UDP Unicorn) which automatically sends requested no of packets to the particular IP address within given time. We have extracted the data from Wireshark after applying the Protocol filter-UDP. The data has been programmed to notify an UDP attack based on the number of Packets.

Classification of data using KNN algorithm is gives us less efficiency. we would like to improve the accuracy of data classification. So we have chosen another most famous algorithm called Support Vector Machine(SVM).It classifies the data with good accuracy and time for computation also decreases compared to KNN algorithm. In SVM data can be classified using basic concept hyper plane. Hyper plane which divided the data into groups. Accuracy is based on hyper plane and support vectors. support vectors are data points near to hyper plane. Hyper plane can be linear, circle, ellipse etc. It all depends upon data points.

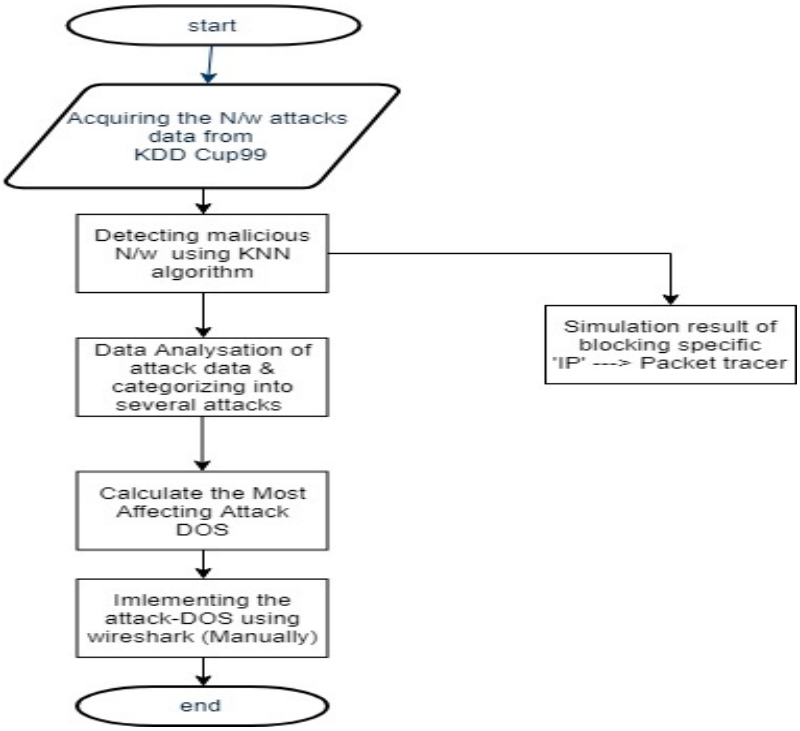


Fig. Overview

CHAPTER 2

1. Literature review:

[1]

In this paper, the Author refers to the incidents of DDoS attacks on online sites such as Yahoo, Gmail, and many other resources are related to the web which cause DDoS attacks. These attacks disconnect the network or ruin the devices by overwhelming. But these threats are temporary and are a threat to the systems with very few resources and are considered as a major problem, and should be prevented by further researches. The Author in this research presented the TRA to give statistics of the attacks. TRA that verifies how many times a same kind of packet occurs in the analysis which is composed of a

TCP flag and Protocol. After finding results the authors says that knowing network traffic using Traffic Rate Analysis declare that many variations between normal traffic and DDoS attack traffic rate, which was examined on the model based on single SVMs.

[2]

In this paper, the author points out the benefits of the Internet and the increasing network threats and the methods to detect the threats. The Internet security corporation, Symantec, which is well known to all reveals from the annual ISTR (Internet Security Threat Report) that cyber criminals continue to damage over consumers and all other web businesses. Cenzic, a web security company in 2014 reports that 96% of the tested internet applications are in a state of being exposed to the possibility of being attacked or getting damaged by cyber criminals. These reports suggests that safety measures should be taken for the effective internet security. Anderson was the first person who introduced Intrusion detection. Later, many researches have been carried out. The 2 general methods of the Intrusion Detection are anomaly detection and misuse detection. The first type of detection deals with the behavior of user. Any change observed in the normal user behavior is called as anomaly detection. The other type of detection operates with desired type of Patterns by matching the original Patterns (Eg : Signatures)

So There are many techniques evolved to conduct intrusion detection, out of which the SVM is the most popular and efficient techniques. Corinna Cortes and Vapnik in 1993 proposed the SVM algorithm and published the algorithm in 1995. It is a supervised learning model that analyse the data which is used for regression analysis and classification. It has a bunch of training examples that come under one or the other categories which builds a system that results new input conditions to the categories. This model consists a representation of categories as points in space which are divided by a wide range. New points are matched in to the same space and identifies to which category they fall.

[3]

In this paper the author states that the two algorithms KNN and SVM has high false alarm rate and low detection efficiency, they proposed two new algorithms improved K-means and multi-level SVM to overcome these disadvantages.

Improved K-Means:

As we know that previous KNN algorithm is easy to implement but it has a drawback of sensitiveness to the evaluation of clustering centers. The improved K-means algorithms chose each axis of the cluster attributes between extreme values of data for KNN algorithm.

Their motto is to mention the initial point of the cluster manually and compute the extreme values of data. Then multiple iterations are carried out to find out the optimal cluster.

Multi-Level SVM:

The basic idea of the Multi-Level SVM is they initially use the improvised K-means algorithm in dividing data into several clusters such as normal and abnormal. Then this algorithm classifies these clusters which are highlighted as abnormal in distinguishing kind of attack and reveals the detection of the network threats.

[4]

In this paper the author mentions about the threats in cyber-security which is a growing concern in the present environment. A unsupervised anomaly based Intrusion Detection system which uses statistical techniques in the process of detection. Although this anomaly based algorithm has many advantages but it has high number of false alarms. So in this work they have developed an SVM algorithm which is a machine learning technique that provides a second line that reduces the false alarms. They are going to assess the performance of Intrusion Detection System against 1 class SVM's and 2 class SVM's.

[5]

The Research on this paper gave many ideologies in anomaly detection. intrusion detection systems (IDSs) and network intrusion detection systems (NIDS) recognizes the attacks with accuracy in identifying such attacks. It mainly gives a anomaly detection schemes for knowing novel network intrusion detections with results on KDDCup'99 data set.

In Bayesian classification, under this method, Initially an estimation is made to the data in order to specify the classes or not , and calculation of probability is marked as true considering a practical approach for such kind of problems and gives only one scan of the complete data. If in any case of additional training data, then every sample follows process of incrementing increase/decrease of the probability that or assumption or estimation is correct.

The usage of BPN algorithms are also used to verify with the observed results.

[6]

This paper says about phase based models of defense and counter measure strategy and how CND incorporates threat specific intrusion analysis and defensive mitigations.

The Improvised Explosive Device delivery chain models all data through adversary funding to take an attack whereas Co-ordinated intelligence focuses on every step in IED threat chain in order to overcome these threats. This method also provides a way for developing model in antiterrorism planning to support the commanders achieve the best ways to protect from others. Explanation of vast Intrusion kill chain which actually engages an adversary in order to build a new desired effects.

This paper mainly deals with the intrusions for which the aggressor need to develop a payload to breach a certain trusted boundary. It also scopes on the definitions of Computer network attacks (CNA) and computer network Espionage (CNE) for the better intrusion through kill chains.

[7]

In this paper, An IDS is considered to a burglar alarm. For an instance, a simple lock system in the house which guards the house from robbery. But somehow anyone tries to break the lock system and gets to enter into house, it will be burglar alarm, indicating the lock was broken and gives an indication of an alert to the owner by raising a attentive alarm. Other Non-local users when connected to network by dialing via modem gets installed somehow in any private network and shall not be detected by the firewall. IPS which is a most effective prevention technology that makes network traffic to detect and stops vulnerability exploits.There are only 2 types of prevention to the system which are categorized as Network (NIPS) and the Host (HIPS). These systems take actions to protect networks and systems. The attacks that are mainly focused in this paper are as follows

Intrusion attacks or User to Root Attack (U2R) [7.1]:

Logon Attacks [7.2]:

A Logon attack can ignore the process of authentication and grants the user with double benefits.

Denial-of-Service (DOS) Attacks [7.3]:

SYN Attack [7.3.1]:

SYN attack means Synchronization attack. Here the attacker continuously sends the flood of SYN requests to the assigned target in order to get the advantage of the resources of the server and causes to the unresponsiveness to the system.

Ping of Death [7.3.2]:

In this category, The attacker will send a ping request to the targeted system which is greater than 65K bytes in size which leades the system to be crashed on its own The General or default regular size should be in the range of 56-84 bytes in case of assuming Internet protocol header.

[8]

In this paper, The problem solving of having a general taxonomy for network threats is classified in which a pre-defined taxonomy which was expressed for categorizing the network attacks based on following factors such as source, threat is active or passive in OSI model.So according to the survey, There are only three algorithms that helps in problem solving with some outstanding results which are ANN, k-means and SVM.

THREATS TAXONOMY

It mainly deals with constructing a general taxonomy for security threats of high priority in helping researchers. Getting tools build up with a capability of detection of several attacks ranging from specific to Zero-day attacks.

Threat Sources :
Identifying network threats and providing a classification which is as following criteria
(1) The Threat sources, (2) The Affected layer on OSI model
(3) Both Active and Passive threats. The Attackers actually know that if one of the attacks is targeted to any single layer of the OSI model, then the other layers are obviously affected. The taxonomy presented is mainly focused on the Target layer of attack. The much known attacks like adware, spyware and gathering information are actually considered as passive attacks whereas DoS, Impersonation and Virus are under category of active attacks. However, few attacks neither come under active nor of passive ones until their use is known.
Packet forging :

It is the another form of networking attack. This is also known as injection process or doping actually.The attacker will generate packets that are similar to the network so as to steal the information to confuse with the duplicate packets which had been inserted by the attacker.

Example :
Whenever the attacker intercepts the process of communication between two or more items or block of datasets, When It starts it Either have to listen to or control the communication between them and alter(swap) between them. this attack is considered to be a 'Man in the Middle' attack

CHAPTER 3
3. Methods Adopted:

1.K-Means Clustering

It comes under unsupervised machine learning deals with unlabeled data (data which don't have specific categories). The main motive of the algorithm is to find clusters in data. The no. of groups is mentioned by variable K. This algorithm is iterative and groups data points to one of the groups based upon the features that are provided in the Dataset. Data points are grouped based upon the similarities in features. Centroids of K clusters helps to label new data points. Every data point in dataset is always assigned to a any of the K cluster. Centroids is formed based upon the minimum distance of data points. These algorithm is used to find patterns in complex data sets and forms groups. After executing algorithms groups are formed, then any test data can be grouped to the correct groups.
.

1. Data assignment for clusters

Each centroid indicates as one of the clusters. In step 1 Euclidean distance is calculated between centroids and data points and finally each data point of dataset is mapped to its nearest centroid.

$$d(\mathbf{p},\mathbf{q}) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}.$$

d(p , q)=Euclidean distance between p(p_1…..q_n) and q(p_1….q_n)

2. Updating Centroid

In this step, the centroids are get updated based upon mean of all data points.This algorithm iterates between these two steps until maximum number of iterations are reached or no data points changes clusters or groups or the sum of distances are minimized.

$$\mu_k := \frac{1}{|C_k|} \sum_{i \in C_k} x^{(i)}$$

μ_k=Centroid

C_k=set of examples which are assigned to Centroid K.
For example $x^{(1)}$ and $x^{(2)}$ are data points ,then C_k becomes 2.

i. Intrusion detection algorithm based on KNN

KNN is one of the simplest algorithm used for classification and it is also most used learning algorithm. KNN uses the database consists of datapoints which are separated into two several classes to predict the classification of new datapoint. The model structure is formed from the data. We are providing data set along with training data. Based on nearest neighboring (Euclidean distance) it allocates new unclassified data point to one of the cluster.

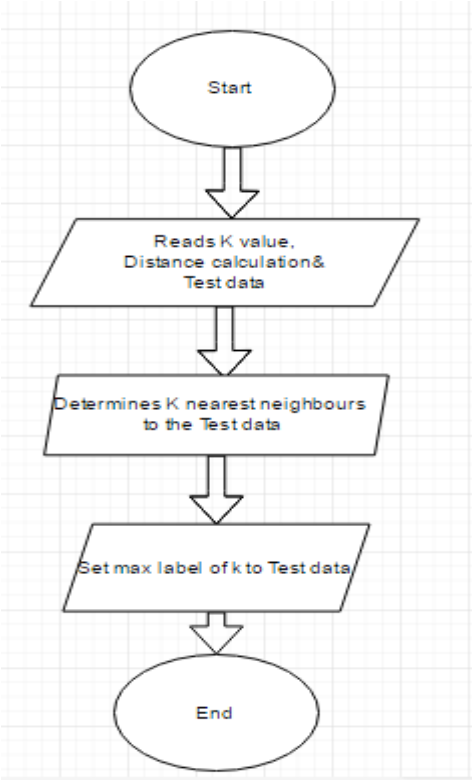
Application of KNN in DoS attack

The intruder broadcasts lot of data packets to stop the communication and to terminate the services. So due to DoS attack services will be terminated between client and server. Every data packet has IP addresses, TCP flag, Time duration, Checksum, and other fields. Normal data packet and flooded data packet won't have all same fields. Based on these fields K-means clusters into two groups (Normal and anomaly). The K-means clusters the data and analysis the data packets. So when a new data packet arrives based on the fields it goes to either of clusters.

Algorithm

Firstly KNN reads the K value and finds the Euclidean distance and Tests data. KNN reads k nearest neighbors to Test data. Finally sets maximum label class of k to Test data. There will be iteration between these steps

Flow Chart of KNN

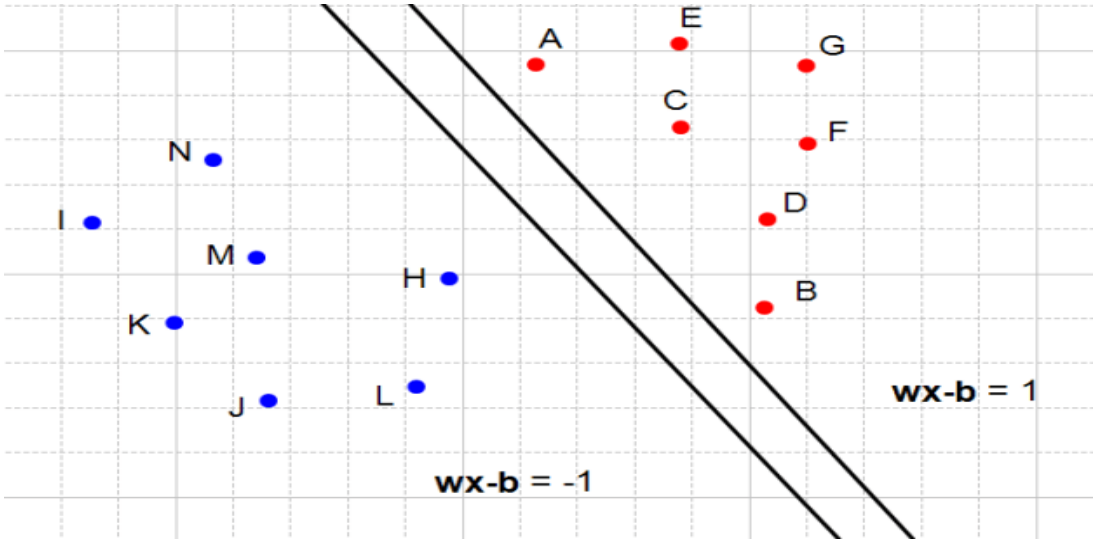


ii. Intrusion detection algorithm based on SVM

SVM algorithm comes under supervised machine learning. It is further classified into both classification and regression. SVM mostly used in solving classification problems. In this algorithm we represent each data point as a point in n- dimensional space. Here ‘n’ is no of features we have and with the values corresponding to each feature are plotted as values at particular coordinate. After plotting we will find hyperplane that differentiates two classes. Then we will go through classification. SVM is best technique in separating two different classes.

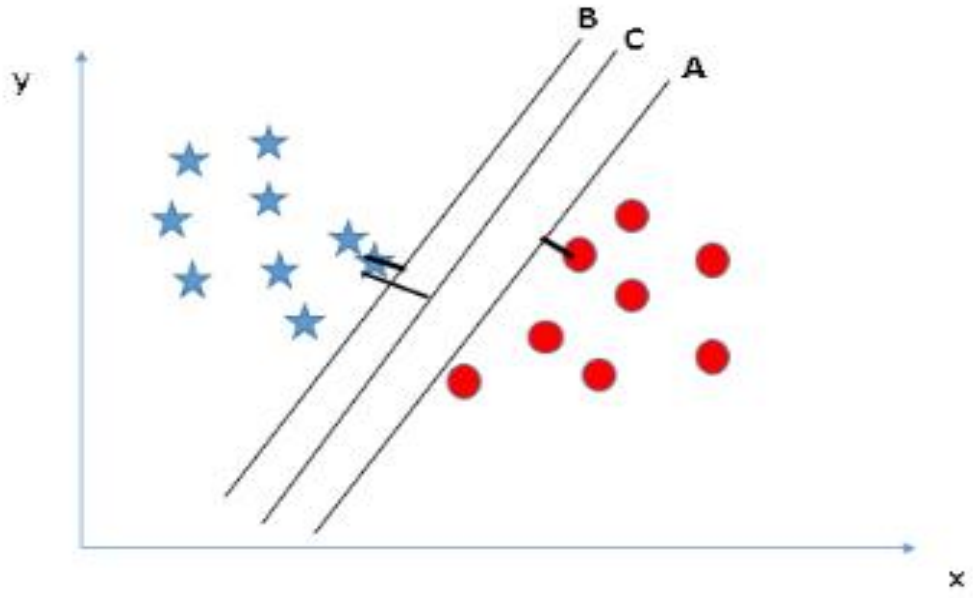
How to find right hyperplane

Rule 1:
We need to select the hyperplane which separates the two classes best.

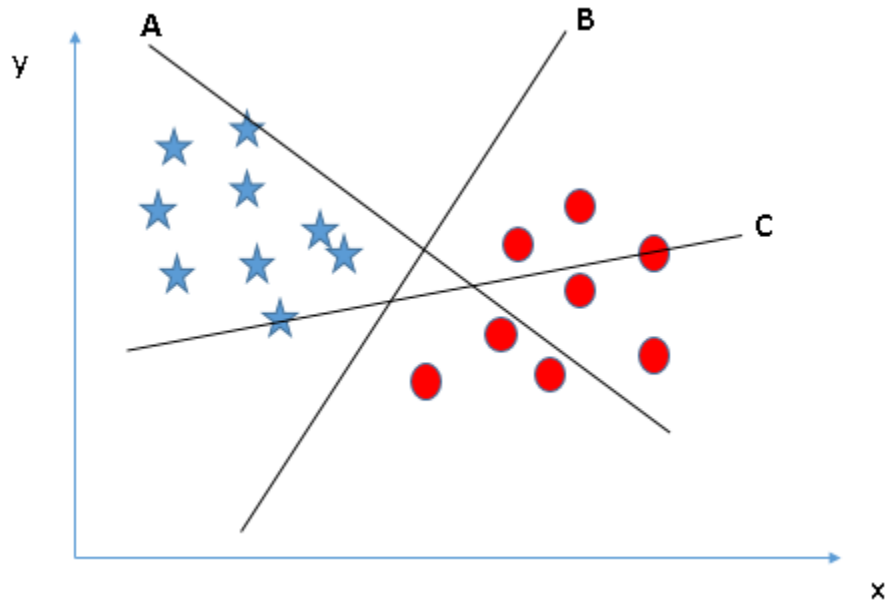


Rule 2:

If we find three hyperplanes which separates two classes best(Three hyperplanes are in parallel). Among those hyperplanes we need to choose the hyperplane based upon **Margin**(distance). Margin of data point and hyperplane should be maximum. We need to choose hyperplane with higher margin. If we select hyperplane with least margin , there will be miss classification.

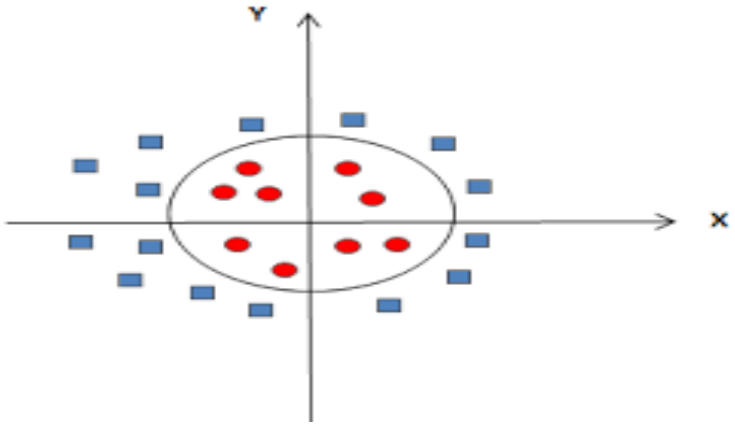


Rule 3:
If we find a hyperplane which has high margin and doesn't classify two groups and we find a hyperplane which has less margin and classify two groups. In these two hyperplanes hyperplane with less margin case should be considered.



Rule 4:
Suppose if we have two kinds of classes, if we have data point of one class presented in territory another class. We need to ignore about data point and draw hyperplane with maximum margin.

Rule 5:
Hyperplane won't be always a straight line, sometime it may be circle. When data points of one class are surrounded by data points of another class. We need to choose circle as hyperplane. After plotting hyperplane, it is easy classify dataset.



Advantages

- SVM works well, if we have clear margin for separation
- If we have high dimensional spaces, It will effective
- If we have dimensions more than no of samples then it will be so effective.

Disadvantages

- If we have complex data set, it does not perform well because the training time requires is larger.
- Classes will be overlapping, if data set has more noise.

CHAPTER 4

Implementation

Intrusion attack and detection is simulated in two environments: Cisco Packet Tracer, Wireshark and UDP Unicorn.

i. Cisco Packet Tracer

Metropolitan Area Network and Malicious IP addressing is implemented in Cisco Packet Tracer. It is a Simulation software that makes users to create their own network topologies. This software also allows user to simulate the configuration of router and switches using user interface and command line interface.

Blocking IP Address using Cisco Packet Tracer

Establishing MAN in Cisco Packet Tracer

Step1

For proper communication in network, every PC is assigned with Subnet Mask, Default gateway, IP address. PC of same network ID are connected through Switch. Switches help to forward the data between PC's and it also forwards and receives the data packets of Router.

Step2

Router helps to communicate between Remote Networks (PC of different Network). PC of one network tries to access web server located at another network , it should communicate through Routers. Every router forms its routing table. It provides path for data packets to reach destination. Each router has knowledge of only networks that are connected to it. So, in this step we had connected PC's of different networks are connected to routers. Default gateway of PC helps to reach the router It has connected. Hops of router are like bridges between routers. In this step We are going to configure the router hops

Step3

Step1 and step2 makes successful connection establishment between PC's of different networks. In this step we are going to check whether receiving data packets of one PC to another either in same network or remote network through PING command. In cisco packet tracer every PC is provided with terminal. In that terminal **ping IP address** helps us to know whether proper network is established between them.

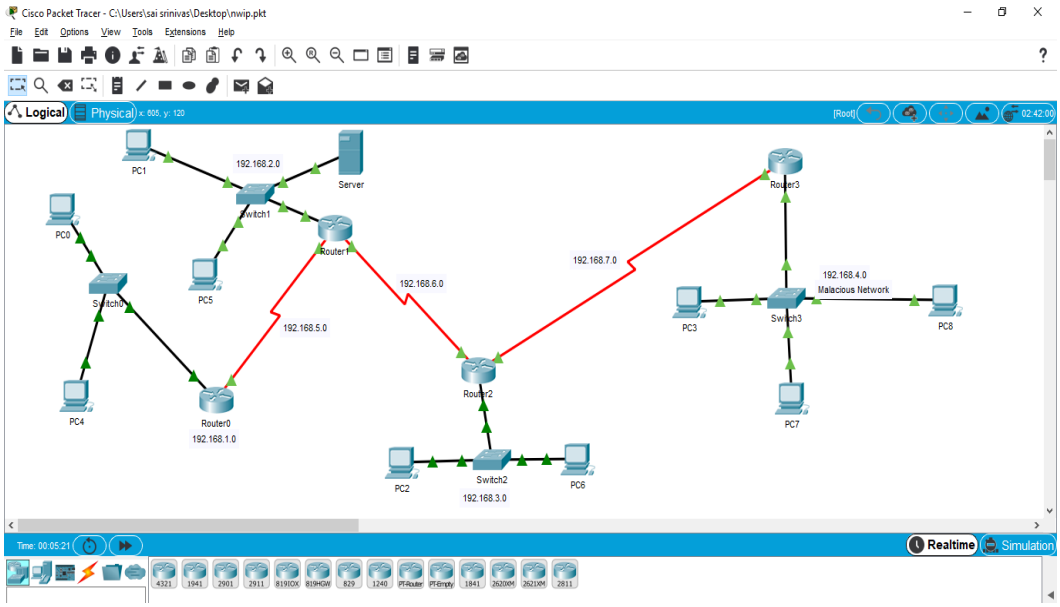


Fig . Metropolitan Area Network in Cisco Packet Tracer

Malicious IP Address blocking using Standard Access list

In this step PC with malicious IP address are blocked in the router of sever located. Access Control List (ACL) works like filters and enables to control packets permitted or denied in or out of a network. Every Router in Cisco packet tracer is provided with CLI (Command Line Interface). Commands for denying the ip address in router through standard access list are **en** helps to enable the router, **configure terminal** helps us to configure terminal, **sh access-lists** gives list of access-list created, **access-list no deny IP address wildcard subnet** blocks the specified IP address the router, **access-list no permit any** permits all the IP address in network except the blocked IP address.

ii. WireShark

It is a software for analyzing the messages between the executing protocols entities is called a packet sniffer. **Sniffs** represents the messages sent or received from the computer. Packet Sniffer is an application which can capture and analyze a Network. It collects the details of various protocol fields and displays back. It is packet analyzer which uses packet capture library in the personal computer. It identifies messages which were sent and received by apps and protocols running background on our personal computer but never tries to send packets by itself.

PACKET SNIFFER- Analogous to TASK MANAGER:

- It shows all the information regarding the IP address, protocols, Length etc.,
- It displays all the Protocols such as TCP, HTTP, FTP.
- Manual Detection of malicious files when followed.
- It Predicts DOS attack based on Highest count among all the Packets in Wireshark.

iii. Attacks

a) POD

It is a type of attack which crashes, freezes, or destabilizes the targeted device by sending oversized packets using a simple ping command.

ICMP Protocol:

- The Internet Control Message Protocol(ICMP) is the protocol in the Internet Protocol Suite.
- It is used in all Network devices, such as routers, to send error logs and packets information.

Implementation:

- After classifying the type of attack, we are implementing PoD using Wireshark Packet sniffer analyzer manually.
- Using command prompt we are sending number of requests to the destination IP address through ‘PING’ command from source which are connected in the same LAN.
- Then we are analyzing the data using Wireshark.
- By using Python Programming language, we have counted the number of times we got requests from the source.
- If the requests exceeds the threshold value, then we have classified it as Ping of Death.

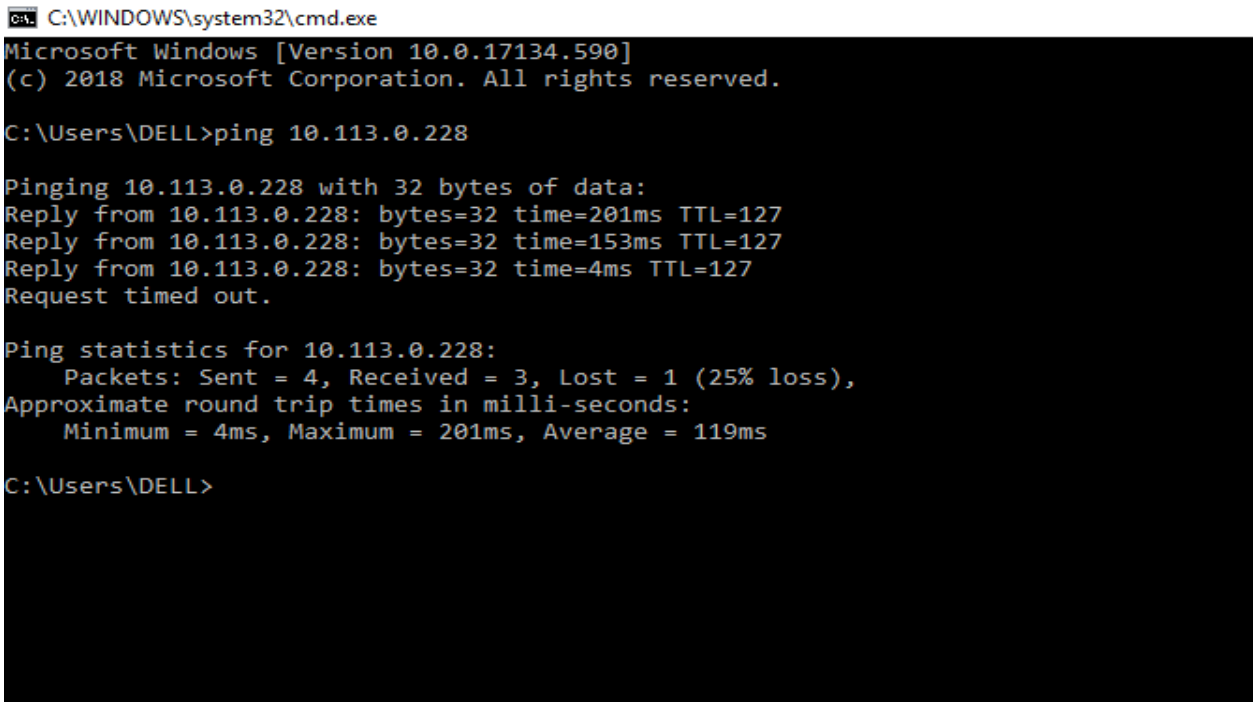


Fig. Sending Ping Requests using Command Prompt

In [5]:
```
import csv
def loadDataset(filename):
    count=0
    with open(filename, 'r') as csvfile:
        lines=csv.reader(csvfile)
        dataset=list(lines)
        for x in range(len(dataset)-1):
            if(dataset[x][3] == "192.168.212.5" and dataset[x][4] == "ICMP"):
                count=count+1
    if(count>200):
        print("It is DOS attack")

def main():
    loadDataset(r'C:\Users\DELL\idsdata.csv')

main()
```

```
It is DOS attack
```

Fig. Code Snippet of Detecting DOS Attack

b) UDP Flooding

- This attack comes under DOS in which a high no. of UDP packets being sent to targets with the idea of Flooding or overwhelming.
- At the receiver end, it filters out for applications associated to
- Datagrams and if it is found none then there is a send-back “Destination Unreachable” packet.
- System is unresponsive and overwhelmed to the clients with the increment of UDP packets received and answered.
- In the UDP flood attack, Intruder anonymizes the attack by tricking the IP address with the packets to make identify that it does not return ICMP packets to host.
- As a result of UDP Flooding, the server becomes exhausted when the firewall of a system tries to protect the targeted server eventually leading to legitimate traffic.
- To implement a UDP flood attack manually, there are a more no. of commercially-available software packages like UDP Unicorn and other.

ATTACK DESCRIPTION:

- User Datagram Protocol (UDP) is a session-less and connectionless networking protocol.
- As it depends on lower overhead, there is no need of Three-way handshake like TCP.
- A High amount of “best effort” traffic is sent over these channels to any destination in the absence of an Initial Handshake.
- UDP cannot filter out specific packet formats, and thus attackers advances to create the large packets, fill it up with junk numbers or text and send it to the host while being attacked.

Implementation:

UDP Unicorn:
This is an open-source DoS attack software that destroys a computer's network connection by continously sending UDP packets with garbage data.

Working:

UDP primarily knows making use of steps in which server runs responds to a UDP packet which is sent to one of the ports. It goes through two steps in response, server receives a packet of UDP in the particular port

1. Checking whether there are programs executing that are currently analysing for requests in the specific port.
2. A ICMP (ping) packet is known for alerts with the tag “destination unreachable” whenever there are no programs at the receiver end.

For instance, consider the routing calls of hotel receptionist with UDP Flooding. Initially the receptionist gets the phone call where the caller attempts for a connection to ask out for a specific room. The receptionist then looks up through the list to make sure it is available or not and makes sure to take that particular call or not.Once the receptionist realizes that there is no attempt of taking the calls, they need to pick up the phone and reply that guest might not be taking the call. If suddenly all the lines try to come in connection at once with similar requests then they will quickly become overwhelmed.

Fig. Bots Request

When the transmission of UDP packets is supposed to happen, each packet include the IP address of the wanted source device. During such DDos attack, SPOOFING UDP PACKETS is transmitted instead from their own IP address for which attackers are not exposed to anyone. As a result to this, checking and responding to each received UDP packet, the target’s resourcs becomes exhausted when a immense flooding of UDP attacks takes place at receiver end, resulting in DOS to normal traffic.

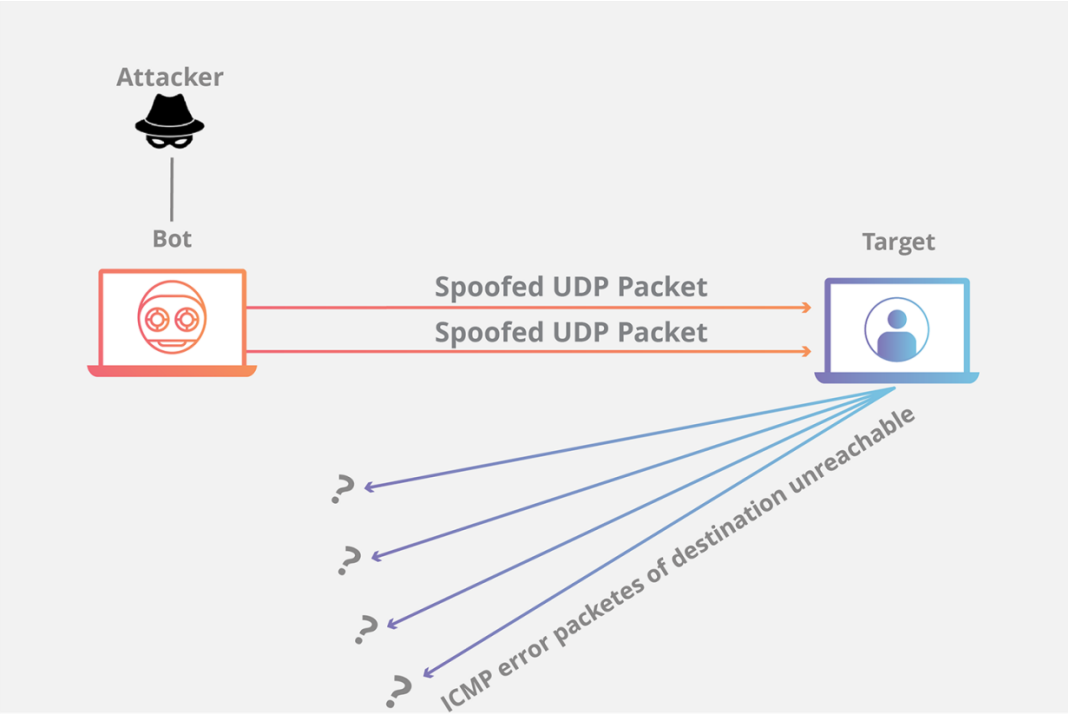


Fig. UDP Attack Architecture Overview

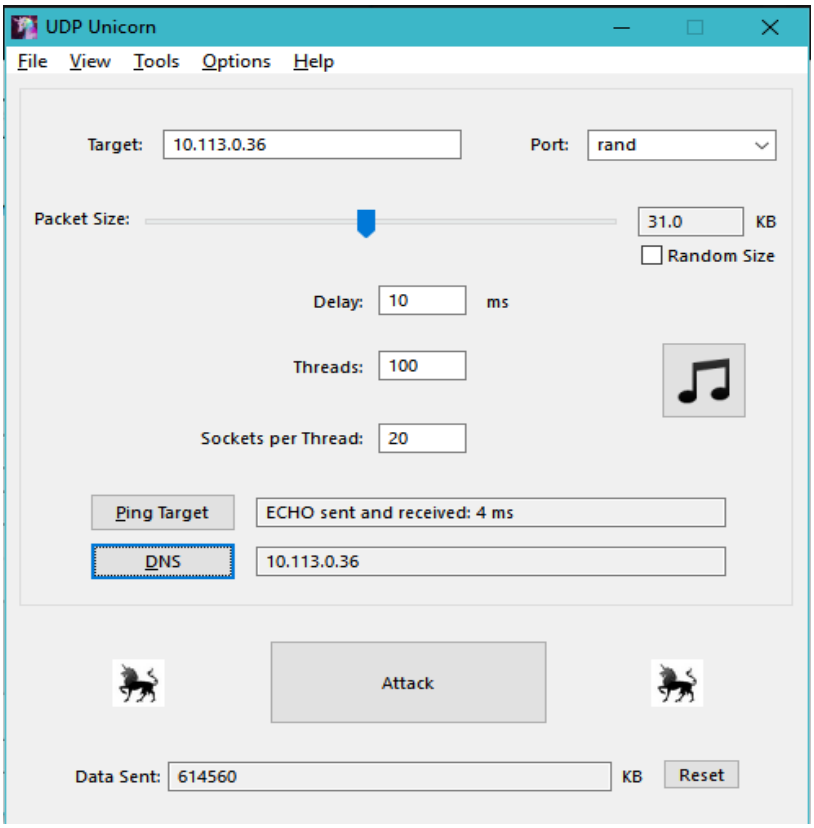


Fig. UDP Unicorn Dialog Box

iv. Alternative Methods:

a) IP address Identification

Whenever a data packet needs to send between users, IP address is one of the most important detail to be addressed. Every data packet is assigned with IP address. When a data packet is received, we are going to check whether the incoming data is authentic or not.
User will have blacklisted IP addresses and we are going to compare incoming data packet IP address with blacklisted IP address database.

Example

John is tennis player, who play tennis based upon different conditions like outlook, humidity , wind etc Outlook ,humidity , wind are different categories for a data point. When we take one category like humidity, we have two subsets like high and normal. For subset like high and normal we find probabilities based upon the training data. So, if we get new data we will give priority to high probability and makes decision according to it.

b)Time Stamping

Every data packet assigned time stamp, when it sends data packet from sender-A to sender-B. we will be tracking data packet sent timing and arrival timing. Suspicious data may have different timings, based on that we are going to detect intrusion.

CHAPTER 5

Results and Analysis:

We have implemented the K-Nearest Neighbour Algorithm with the combination of defining several methods.

Method 1: Importing Data

We have defined a method in Python language to import the data.For the Implementation we have imported a file in csv format.Imported csv, so that we read the data from csv format file.

Method 2: Classifying the data into Training Set and Testing Set

We have used the split ratio as 0.66:0.33; to split the data.Made use of an Inbuilt-function random.random() which returns the next floating point number in the range[0.0,1.0).We have classified the data-set into Training Set if the value of random.random is greater than the split value.

Method 3: Calculating the Euclidean-Distance

Calculated the Euclidean-Distance by Implementing the Distance formula

Method 4: Extracting the Neighbours based on the 'K' value

Calculated distances are sorted and the 'K' nearest neighbours are appended in a list. The 'K' parameter plays an important role in this method.

Method 5: Extracting the Response from the K-Nearest Neighbours

Based on the majority votes from the neighbours, returning the most up-voted class as the predicted class.

Method 6: Calculating the Accuracy

Based on the ratio of the correct value and predicted value the Accuracy is measured.

neptune

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
tcp	private	REJ	0	0	255	10	0.04	0.06	0	0	0	0	1	1	neptune
tcp	private	REJ	0	0	255	1	0	0.06	0	0	0	0	1	1	neptune
tcp	ftp_data	SF	12983	0	134	86	0.61	0.04	0.61	0.02	0	0	0	0	normal
icmp	eco_i	SF	20	0	3	57	1	0	1	0.28	0	0	0	0	saint
tcp	telnet	RSTO	0	15	29	86	0.31	0.17	0.03	0.02	0	0	0.83	0.71	mscan
tcp	http	SF	267	14515	155	255	1	0	0.01	0.03	0.01	0	0	0	normal
tcp	smtp	SF	1022	387	255	28	0.11	0.72	0	0	0	0	0.72	0.04	normal
tcp	telnet	SF	129	174	255	255	1	0	0	0	0.01	0.01	0.02	0.02	guess_passwd
tcp	http	SF	327	467	151	255	1	0	0.01	0.03	0	0	0	0	normal
tcp	ftp	SF	26	157	52	26	0.5	0.08	0.02	0	0	0	0	0	guess_passwd
tcp	telnet	SF	0	0	255	128	0.5	0.01	0	0	0	0	0.66	0.32	mscan
tcp	smtp	SF	616	330	255	129	0.51	0.03	0	0	0	0	0.33	0	normal
tcp	private	REJ	0	0	255	2	0.01	0.07	0	0	0	0	1	1	neptune
tcp	telnet	S0	0	0	235	171	0.73	0.07	0	0	0.69	0.95	0.02	0	neptune
tcp	telnet	SF	773	364200	38	73	0.16	0.05	0.03	0.04	0	0.77	0	0.07	normal
tcp	http	SF	350	3610	71	255	1	0	0.01	0.04	0	0	0	0	normal
tcp	http	SF	213	659	255	255	1	0	0	0	0	0	0	0	normal
tcp	http	SF	246	2090	35	255	1	0	0.03	0.05	0	0	0	0	normal
udp	private	SF	45	44	255	255	1	0	1	0	0	0	0	0	normal
tcp	private	REJ	0	0	255	18	0.07	0.07	0	0	0	0	1	1	neptune
tcp	ldap	REJ	0	0	255	19	0.07	0.05	0	0	0	0	1	1	neptune
tcp	pop_3	S0	0	0	255	87	0.34	0.01	0.01	0	1	1	0	0	mscan
tcp	http	SF	196	1823	255	255	1	0	0	0	0	0	0	0	normal
tcp	http	SF	277	1816	36	255	1	0	0.03	0.02	0	0	0	0	normal

Fig. Dataset with Attributes

```
c=89.09
print('accuracy =',c)
```

```
accuracy = 89.09
```

Fig. Accuracy report of KNN

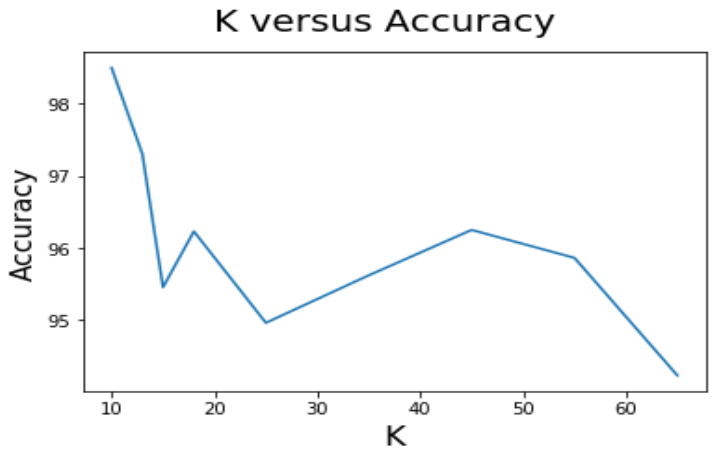


Fig. K vs Accuracy Graph

After classifying the IP Address using KNN Algorithm the IP Address was manually given using Cisco Packet Tracer.

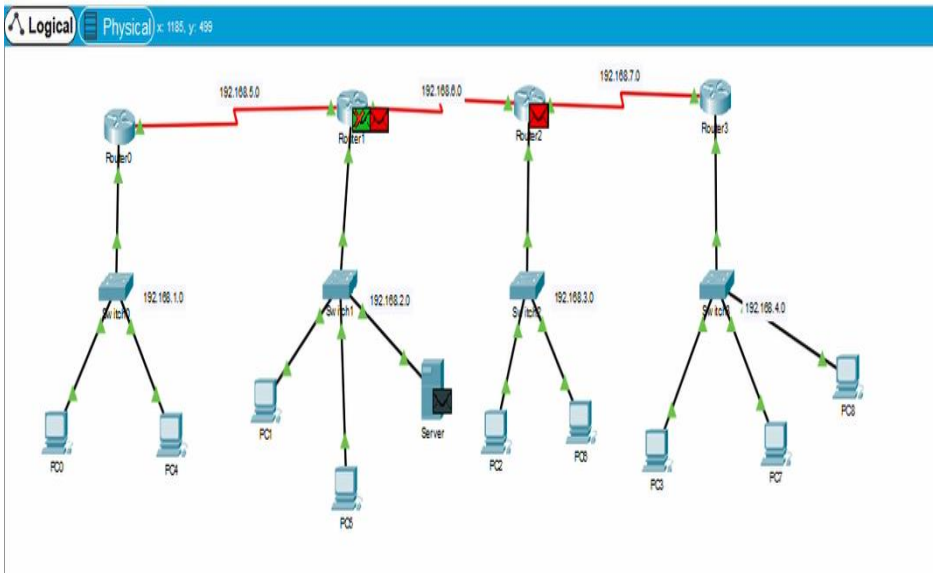
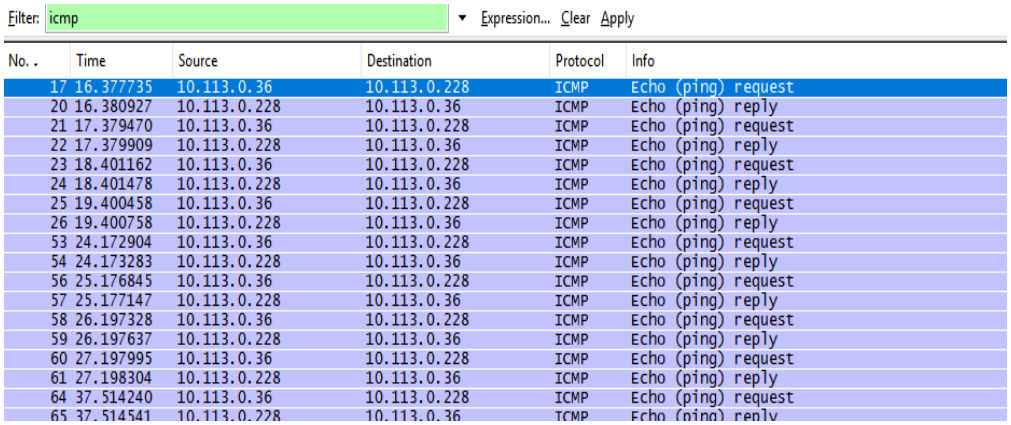


Fig. Blocked IP Address denied by router

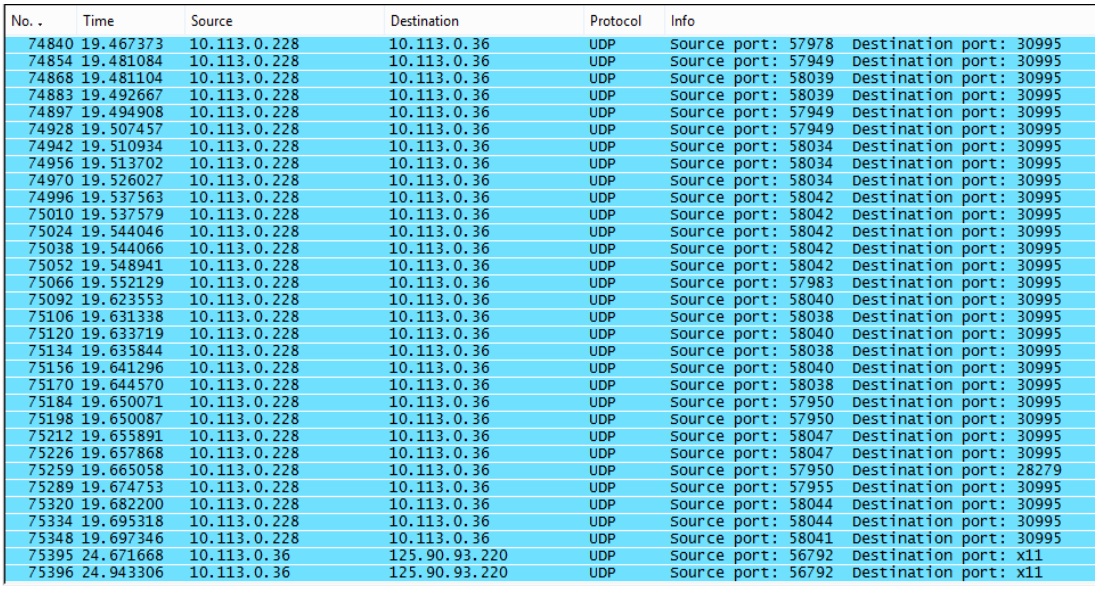
After blocking the IP Address we have analysed the data using Python Programming Language and classified DOS Attack occurs most number of times. So we have classified two types of DOS attacks Ping of Death (POD) and UDP.



Filter: icmp ▾ Expression... Clear Apply

No.	Time	Source	Destination	Protocol	Info
17	16.377735	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
20	16.380927	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
21	17.379470	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
22	17.379909	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
23	18.401162	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
24	18.401478	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
25	19.400458	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
26	19.400758	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
53	24.172904	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
54	24.173283	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
56	25.176845	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
57	25.177147	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
58	26.197328	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
59	26.197637	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
60	27.197995	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
61	27.198304	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply
64	37.514240	10.113.0.36	10.113.0.228	ICMP	Echo (ping) request
65	37.514541	10.113.0.228	10.113.0.36	ICMP	Echo (ping) reply

Fig. Wireshark Capture View after Continuous Pinging(ICMP).



No.	Time	Source	Destination	Protocol	Info
74840	19.467373	10.113.0.228	10.113.0.36	UDP	Source port: 57978 Destination port: 30995
74854	19.481084	10.113.0.228	10.113.0.36	UDP	Source port: 57949 Destination port: 30995
74868	19.481104	10.113.0.228	10.113.0.36	UDP	Source port: 58039 Destination port: 30995
74883	19.492667	10.113.0.228	10.113.0.36	UDP	Source port: 58039 Destination port: 30995
74897	19.494908	10.113.0.228	10.113.0.36	UDP	Source port: 57949 Destination port: 30995
74928	19.507457	10.113.0.228	10.113.0.36	UDP	Source port: 57949 Destination port: 30995
74942	19.510934	10.113.0.228	10.113.0.36	UDP	Source port: 58034 Destination port: 30995
74956	19.513702	10.113.0.228	10.113.0.36	UDP	Source port: 58034 Destination port: 30995
74970	19.526027	10.113.0.228	10.113.0.36	UDP	Source port: 58034 Destination port: 30995
74996	19.537563	10.113.0.228	10.113.0.36	UDP	Source port: 58042 Destination port: 30995
75010	19.537579	10.113.0.228	10.113.0.36	UDP	Source port: 58042 Destination port: 30995
75024	19.544046	10.113.0.228	10.113.0.36	UDP	Source port: 58042 Destination port: 30995
75038	19.544066	10.113.0.228	10.113.0.36	UDP	Source port: 58042 Destination port: 30995
75052	19.548941	10.113.0.228	10.113.0.36	UDP	Source port: 58042 Destination port: 30995
75066	19.552129	10.113.0.228	10.113.0.36	UDP	Source port: 57983 Destination port: 30995
75092	19.623553	10.113.0.228	10.113.0.36	UDP	Source port: 58040 Destination port: 30995
75106	19.631338	10.113.0.228	10.113.0.36	UDP	Source port: 58038 Destination port: 30995
75120	19.633719	10.113.0.228	10.113.0.36	UDP	Source port: 58040 Destination port: 30995
75134	19.635844	10.113.0.228	10.113.0.36	UDP	Source port: 58038 Destination port: 30995
75156	19.641296	10.113.0.228	10.113.0.36	UDP	Source port: 58040 Destination port: 30995
75170	19.644570	10.113.0.228	10.113.0.36	UDP	Source port: 58038 Destination port: 30995
75184	19.650071	10.113.0.228	10.113.0.36	UDP	Source port: 57950 Destination port: 30995
75198	19.650087	10.113.0.228	10.113.0.36	UDP	Source port: 57950 Destination port: 30995
75212	19.655891	10.113.0.228	10.113.0.36	UDP	Source port: 58047 Destination port: 30995
75226	19.657868	10.113.0.228	10.113.0.36	UDP	Source port: 58047 Destination port: 30995
75259	19.665058	10.113.0.228	10.113.0.36	UDP	Source port: 57950 Destination port: 28279
75289	19.674753	10.113.0.228	10.113.0.36	UDP	Source port: 57955 Destination port: 30995
75320	19.682200	10.113.0.228	10.113.0.36	UDP	Source port: 58044 Destination port: 30995
75334	19.695318	10.113.0.228	10.113.0.36	UDP	Source port: 58044 Destination port: 30995
75348	19.697346	10.113.0.228	10.113.0.36	UDP	Source port: 58041 Destination port: 30995
75395	24.671668	10.113.0.36	125.90.93.220	UDP	Source port: 56792 Destination port: x11
75396	24.943306	10.113.0.36	125.90.93.220	UDP	Source port: 56792 Destination port: x11

Fig. Wireshark Capture View after UDP Attack

UDP Conversations

Address A	Port A	Address B	Port B	Packets	Bytes	Packets A->B	Bytes A->B	Packets A<-B	Bytes A<-B	Rel Start	Duration	bps A->B	bps A<-B
10.113.0.228	58021	10.113.0.36	34665	1	802	1	802	0	0	0.002437000	0.0000	N/A	N/A
10.113.0.228	58032	10.113.0.36	34665	1	802	1	802	0	0	0.034198000	0.0000	N/A	N/A
10.113.0.228	58030	10.113.0.36	34665	1	802	1	802	0	0	0.051128000	0.0000	N/A	N/A
10.113.0.228	58029	10.113.0.36	34665	1	802	1	802	0	0	0.053944000	0.0000	N/A	N/A
10.113.0.228	58043	10.113.0.36	34665	1	802	1	802	0	0	0.088993000	0.0000	N/A	N/A
10.113.0.228	58047	10.113.0.36	14367	1	802	1	802	0	0	0.100469000	0.0000	N/A	N/A
10.113.0.228	58045	10.113.0.36	34665	1	802	1	802	0	0	0.103371000	0.0000	N/A	N/A
10.113.0.228	58044	10.113.0.36	14367	1	802	1	802	0	0	0.106966000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57958	1	797	0	0	1	797	0.139905000	0.0000	N/A	N/A
10.113.0.228	57959	10.113.0.36	15209	1	802	1	802	0	0	0.286821000	0.0000	N/A	N/A
10.113.0.228	57956	10.113.0.36	43342	1	797	1	797	0	0	0.356891000	0.0000	N/A	N/A
10.113.0.228	58001	10.113.0.36	15209	1	801	1	801	0	0	0.367868000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57965	1	797	0	0	1	797	0.413450000	0.0000	N/A	N/A
10.113.0.228	57975	10.113.0.36	15209	1	800	1	800	0	0	0.424311000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57975	1	797	0	0	1	797	0.441238000	0.0000	N/A	N/A
10.113.0.228	57996	10.113.0.36	15209	1	802	1	802	0	0	0.457584000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	58011	1	797	0	0	1	797	0.552807000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	58009	1	797	0	0	1	797	0.572671000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	58012	1	797	0	0	1	797	0.587872000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57970	1	797	0	0	1	797	0.689738000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57971	1	797	0	0	1	797	0.693886000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57976	1	797	0	0	1	797	0.888228000	0.0000	N/A	N/A
10.113.0.228	58043	10.113.0.36	15209	1	801	1	801	0	0	0.922665000	0.0000	N/A	N/A
10.113.0.228	57984	10.113.0.36	15209	1	797	1	797	0	0	1.007634000	0.0000	N/A	N/A
10.113.0.228	57989	10.113.0.36	15209	1	797	1	797	0	0	1.012981000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57988	1	797	0	0	1	797	1.059771000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57986	1	797	0	0	1	797	1.074643000	0.0000	N/A	N/A
10.113.0.36	60982	10.113.0.228	57966	1	802	0	0	1	802	1.092959000	0.0000	N/A	N/A

Fig. UDP Conversations

After classifying the data using KNN Algorithm, we have got less Accuracy. So we have implemented SVM (Support Vector Machine) for better Accuracy.

Per PREDICTED CLASS:

View percentages per predicted class including postivepredictive values (PPV) and False Discovery Rates (FDR)

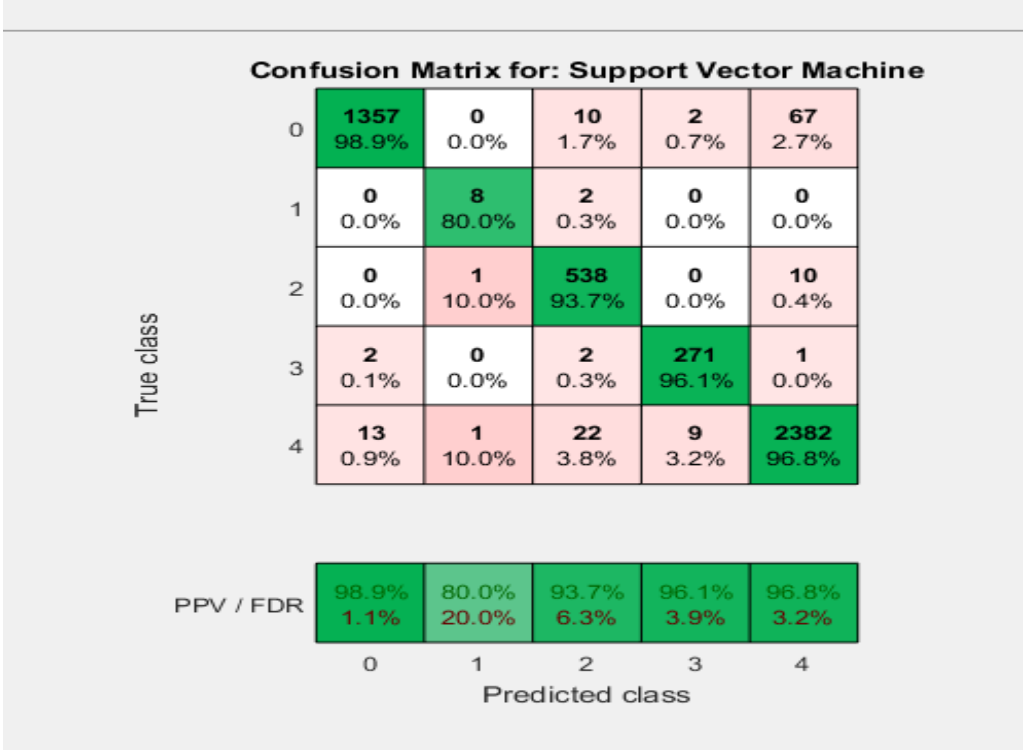


Fig. Confusion Matrix for 5-class SVM

Classes
0 : DOS
1 : U2R
2 : R2L
3 : Probing
4 : Normal

Per true class:

View percentages per true class including TPR
And FNR

.

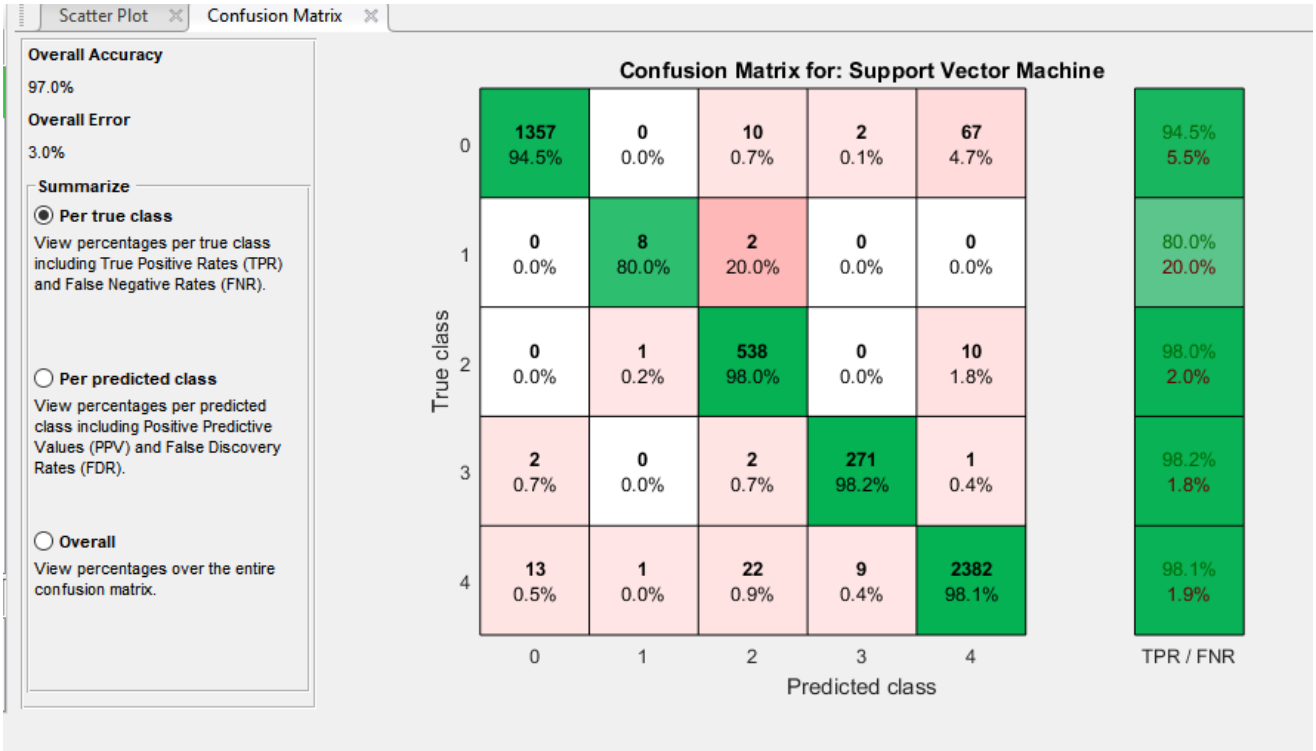


Fig. Confusion Matrix for 5-Class SVM

Classes
0 : DOS
1 : U2R
2 : R2L
3 : Probing
4 : Normal

Per PREDICTED CLASS:

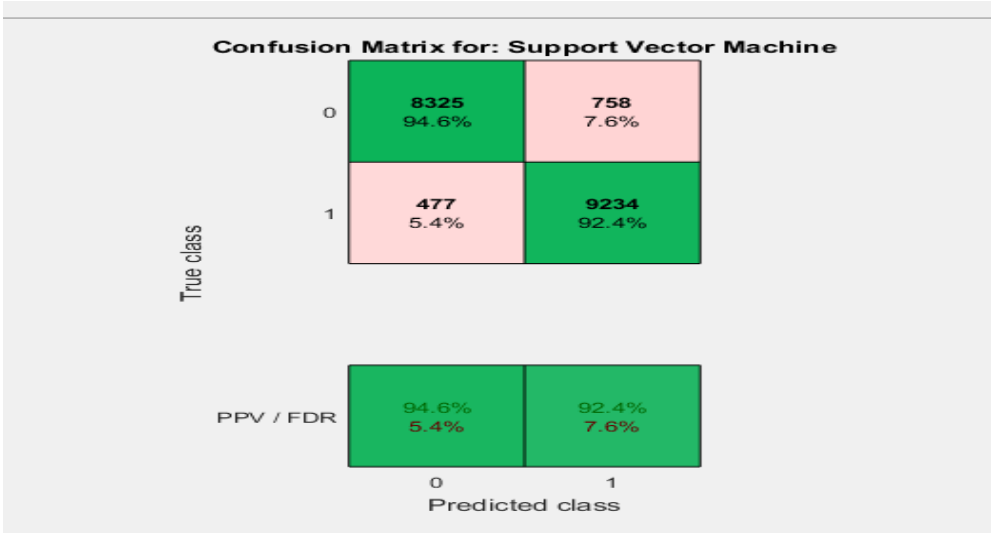


Fig. Confusion Matrix for 2-Class SVM
Classes
0 : Intrusive
1: Normal

Per true class:

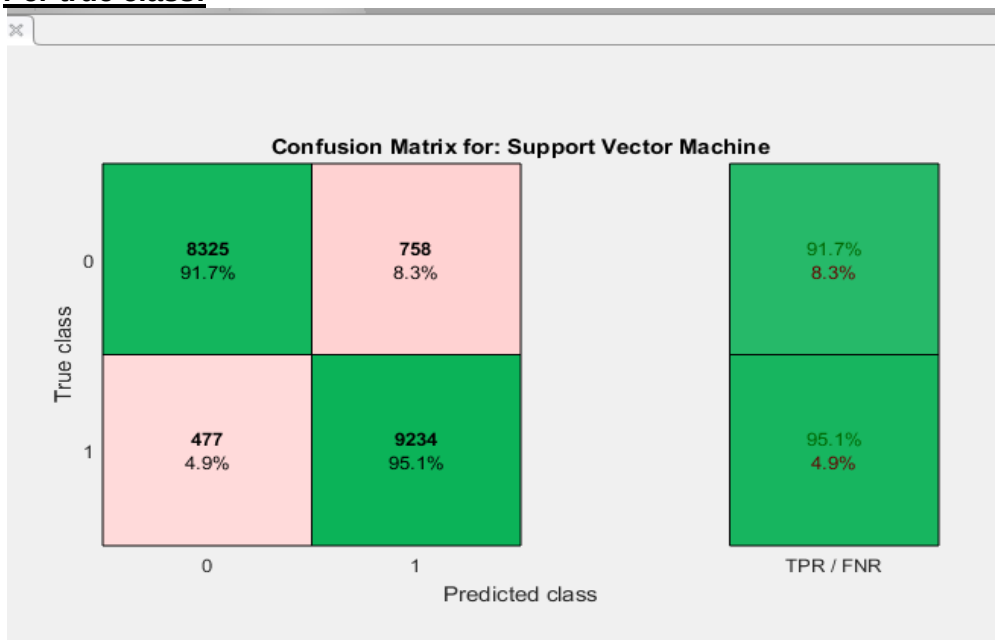


Fig. Confusion Matrix for 2-Class SVM
Class 0 : Intrusive
Class 1 : Normal

CHAPTER 6

Conclusion

Initial phase of the work was implemented with KNN algorithm for identifying program behaviour and label as intrusion or not. An integrated system that blocks malicious IP address along with intrusion detection is carried out in Cisco Packet Tracer and using Python language. This has effectively blocked blacklisted IP addresses. Second phase of the work has been carried out in real time environment in which DoS attack was created and pinged to a LAN network. Later the attack was detected and classified as DoS. We have worked on Denial of Service (DoS) attack and worked on extracted data set from KDDCUP96. Our results give a good accuracy for intrusion detection using kNN algorithm. As an extension of this work and to solid the accuracy rate, the same is implemented using SVM algorithm. The work can be extended to time stamping. Time stamping works well in DoS as it will examine the arrival time and rate of packets.