ISSN 2395-1621

Network Intrusion Detection using Host Based IDS

[#1]Babar Namrata S., [#2]Kharare Vaishali S., [#3]Patil monali S.,
[#4]Prof. Kavita S. Sawant

[#123]Department of Computer Engineering
[#4]Prof. Department of Computer Engineering

Savitribai Phule Pune University, Pune-411043 , India

ABSTRACT

Due to increasing incidents of Cyber attacks, building effective intrusion detection systems are essential for protecting information system security,As attacks on network security are increasing day by day so we are introducing network intrusion detection system using host based IDS. In computerized world such as E-commerce , Artificial Intelligence , web services etc. are having online applications and "security" is the major parameter. In this paper we are detecting malicious attacks occurring to the system by maintaining a log. In this paper we are using two algorithms namely, Naive Bayes and K-means where, probability estimation has been done by naive Bayes and K-means is used for clustering. An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious. Packet does have many features by analyzing features of packets we are achieving final result.

ARTICLE INFO

Article History

Received :3rd February, 2016

Received in revised form :

5th February, 2016

Accepted : 8th February, 2016

Published online :

16th February, 2016

Keywords: **Network Security, Intrusion detection system, Feature selection, Data Mining, Naive Bayes classifier.**

I. INTRODUCTION

In Network Intrusion Detection System we analyze and monitor the behavior of a system and all the malicious packets coming to system. Many unauthorized activities are being observed. Internal attackers such as disgruntled employees are also harmful and instead of SVM i.e. Support vector machine and ANN i.e. Artificial Neural Network we are using Naive Bayes algorithm. SVM gives us the result in binary format it doesn't specify type of attack and "ANN" demands for large database, so we prefer "Naïve Bayes Algorithm". Intrusion detection system (IDS) is a tool that is being used to protect organization from attacks from different sources. IDS can handle large amount of data without affecting performance and without dropping data. Intrusion Detection is the problem of identifying unauthorized use, misuse, and abuse of computer systems by both system insiders and external intruders. As we are aware of "malicious ports". malicious ports are something which can be attacked very easily, Here features of packets like SYC ,hop limit ,length etc. are analyzed "SYC" the value of this flag should be"0"the system may detect this as "a Malicious attack", range of packet length is also feed to the database to detect the attack. IDS which can detect a wide variety of attacks reliably and efficiently when compared to the traditional network IDS.

II. RELATED WORK

ADAM (Audit Data Analysis and Mining) is an intrusion detector built to detect intrusions using data mining techniques. It first absorbs training data known to be free of attacks. Next, it uses an algorithm to group attacks, unknown behaviorr.
ADAM has several useful capabilities, namely;

Classifying an item as a known attack
Classifying an item as a normal event.
Classifying an item as an unknown attack.
Match audit trial data to the rules it gives rise.
First of all, the system is being trained; system has given real-time packets and malicious packets. we apply K-means for clustering and normalization has been done on the packets.On the basis of features of packets. "Transformation" converts the packet features into binary

format .i.e. 1 or 0. Again probability estimation has been done by using Naive Bayes algorithm.

Estimation calculation includes-
1)Individual probability.
2)Initial probability.
3)Final probability.

The probability of Yes i. e. packet is normal.
And probability of No i.e. packet is malicious. Both are calculated. And the greater one is considered as the final result.

III.PROPOSED SYSTEM

Here we proposed a model "Network Intrusion Detection System using host based IDS". it is host based because we have to install software on all machines . We are capturing packets through network by using JAVA libraries i.e. jpcap / winpcap. As there are administrative functions also , system does have proper user_id and password. It allows to access the system only for legitimate users.

LOAD DATA SET:

For Training Purpose Of dataset, First System needs to introduced with some data set which contains the malicious as well as normal packets. Real time packet and attacking modules are come under this phase.

PARSE:

Parsing is nothing but the divination of packets .it gives the idea about the no of normal packets and malicious packets.

LABELLED DATASET:

As the packets coming to system must be labelled very efficiently. Packets does have many features like length, hop limit, SYN ,source port ,destination port etc. Eg. flag has two values such as true or false but system fails to understand this parameters so binary representation of these values is must such as 1 and 0.Lenth of packet is also labelled into two types such as high and low.

DISCRETIZATION:

It is nothing but the packets are discretised into particular interval values. Which makes the classification easy. It makes the classification simple for k means algorithm.
Eg. Suppose Length of a packet is less than 500 then it is classified into one group i.e low and length more than 500 is discretized into another group i.e. high.

DISCRETE LABELLED DATASET:

Labelled data typically takes set of unlabelled data. It provides data with some sort of meaningful “tag”,”label”.or “class” i.e somehow informative or desirable to know.

PROBABILITY ESTIMATION:

Estimation of transition probabilities offer one way to characterize the past changes in credit quality of obligors and are cardinal inputs to many risk management applications. And initial probability and Individual probability calculation comes under the training phase of system. And final probability calculated in detection phase of system.

PREDICTION:

Final probability calculation has been done under this block.

Final probability = initial probability * individual probability

MODEL:

Even if logout has been done by authorized person, then also packets coming to system will be stored in model.

OUTPUT:

System gives the detection of malicious packets.

Introduction Of Modules: It consist of two phases namely

1)Training Phase :-
2)Detection Phase:-

Training Phase:-

- Capture packets.
- Add Manage Train Dataset.
- Normalize Dataset.
- Train Dataset

Capture packets:-

As, very first the system captures packets arriving to system from network. Jpcap / winpcap libraries are being used for capturing packets.

Add Manage Train Dataset:-

Here dataset is managed it consist of attacker module and non-attacking dataset.

Normalize Dataset:-

Packets have features like length, hop limit, flag, SYN, destination port. Feature extraction has been done here by using k-means algorithm clusters are form to define a particular cluster normalization has been done. Particular number is assigned to cluster for normalization purpose.

Train Dataset:-
Dataset should be trained by wise person system for accurate result .Trained dataset accuracy of system has been examine at time of detection phase.

Detection Phase:-

- Capture Packet
- Normalization
- Detection
- Result
- Maintain Log

Capture Packet:-
In detection phase when system captures packets through network, the packet may be real time packet or attacking packets.

Normalization:-
Packets have features like length, hop limit , flag , SYN, destination port. Feature extraction has been done here by using k-means algorithm clusters are form to define a particular cluster normalization has been done. Particular number is assigned to cluster for normalization purpose.

Detection:-
In detection component the current packet coming to the system detected.

Naïve bayes algorithm is used for detection.

Result:- System gives the result in case of malicious packets are captured.
Maintain Log:- System maintain "log" after malicious packets are captured.
It is used for the recovery of the system.

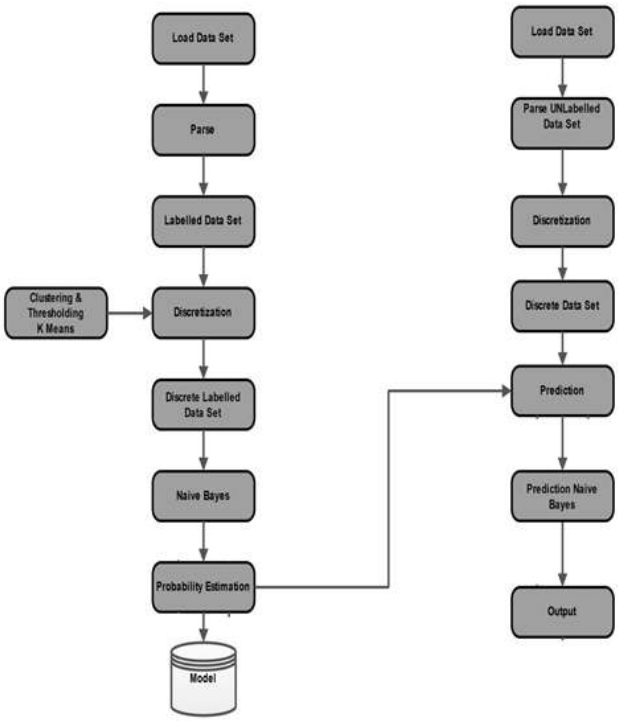

fig 1:system architecture

IV. ALGORITHM

A. K-means Clustering algorithm:-

Steps in K_MEANS algorithm:

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

1) Defined value of k.
2) Read length values.
3) Set initial cluster.

While true

1) Calculate distance between each centroid.
2) Calculate mean .
3) Get new cluster.
4) If (previous cluster and current cluster is not equal)

Then process again.
Else
Stop
End if
End while.

Advantages of K_MEANS:

1) Fast, robust and easier to understand.
2) Relatively efficient.
3) Best result.

B. Naïve Bayes Algorithm:-

The naive Bayesian classifier, or simple Bayesian classifier, works as follows:

This algorithm works on probability estimation probability estimation falls under three categories such as :

1) Initial probability
2) individual probability
3) Final probability

Steps of in naïve bayes algorithm :-

1) capture UDP , TCP and ICMP packets.
2) Generate dataset along with labeled dataset.
3) For all packet P to N total packet N.

Where ,
P = Read packet features.
N = Transform Packets.
Transform packet ( )
If length < threshold value
Assign 0
Else
Assign 1
Return value.

4) End.
5) For P to N ,
   - calculate initial probability
   - calculate individual probability
   - calculate final probability
6) End.

C. Mathematical Model

Let, S be the system divided as, S={s , In, If ,Ia ,F, e, Success,Failure}
Where,
s= Start state
e= End State

Success = Intrusion Attack Detected.
Failure = Intrusion Attack Not Detected if dataset is not trained properly.

In =no. of packet capture through network.

In = {In1, In2 , In3…….Inn}
where,
If=If be the features of packet.
If= {If1, If2 ,If3…….Ifn}
Where,
IA=Ia be the set of final attack.
Ia={Ia1,Ia2,Ia3……..Ian}
Where
F be the set of function of system
F = (F1, F2, F3, F4, F5,F6,F7,F8,F9)

1) Training Phase:

F1= trained dataset()
F2=add manage dataset()
F3= capturing packets( )
F4= feature of packets( )

2) Detection Phase:

F5= discretization ( )
F6= normalization( )
F7=detection using N.B( )
F8= intrusion attack( )
F9= Maintain log of Intrusion Attacks()

V. CONCLUSION

Sstem proposed a framework of NIDS based on Naïve Bayes algorithm. The framework builds the patterns of the network services over data sets labeled by the services. With the built patterns, the framework detects attacks in the datasets using the naïve Bayes Classifier algorithm.

REFRENCES

[1] "Symantec-Internet Security threat report Highlights (Symantec.com)",http://www.prdomain.com/companies/Symantec/new releases/Symantec_internet_205032.htm.

[2] R.Durst, T.champion, B.witten, E.Miller, and L.Spagnuolo, "Testing and evaluating computer Intrusion detection system" communications of ACM,Vol.42, no.7,pp 53-61, 1999.

[3] A.Sung & S.Mukkamala, "Identifying important Features for intrusion detection using SVM and neural networks," in symposium on application and the Internet, pp 209-216,2003.

[4] D.Barbara, J.Couto, S.Jajodia, and N.Wu, "ADAM:A test bed for exploring the use of data mining in Intrusion detection", SIGMOD, vol30, no.4, pp 15-24, 2001.

