Classifying Attacks in NIDS Using Naïve- Bayes and MLP

Ruchira Gurav, Prof. A.A. Junnarkar

`

Abstract— **With the rapid growth in telecommunication and network security has become a crucial and important issue, so there must be a system which can monitor ongoing traffic and identify whether the data is of normal or attack type with better accuracy, speed, and less false alarm rate. There are several algorithms or conventional techniques with certain advantages and disadvantages .The proposed system is a multilevel, hybrid approach which classifies data not only in attack or normal type but also into sub-attack categories with better accuracy and speed. It makes use of two techniques naive-Bayes which is a probabilistic method for classifying data with less time complexity and Multilayer Perceptron which performs efficiently for large data sets and for unseen data. This approach makes use of different machine learning techniques at each level and, reduced feature set so as to reduce redundancy and dimensionality of input layer. So the proposed model of Intrusion Detection System is designed in such a way that achieves accuracy more than 90% by providing high detection and classification rate with less false alarm rates and speed.**

Index Terms— **Intrusion Detection System, Promiscuous mode, dimensionality, alerts, legitimate user**

I. Introduction

The highly connected computing world has also equipped the intruders and hackers with new facilities for their destructive purposes. The costs of temporary or permanent damages caused by unauthorized access of the intruders to networks and computer systems have urged different organizations to, increasingly; implement various systems to monitor data flow in their networks. These systems are generally referred to as Intrusion Detection Systems (IDSs)

1.1 There are basically four types of attacks:

- **Denial of Service (DoS):**
  A DoS attacks is a type of attack in which the hacker makes a memory resources too busy to serve legitimate networking requests and hence denying users access to a machine e.g. apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm, etc.

- **Remote to User attacks (R2L):**
  A remote to user attack is an attack in which a user sends packets to a machine over the internet, and the user does not have access to in order to expose the machines vulnerabilities and exploit privileges which a local user would have on the computer, e.g. x lock, guest, send mail dictionary etc.

- **User to Root Attacks (U2R):**
  These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges, e.g. Perl, x term.

- **Probing:**
  Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be exploited so as to compromise the system. This technique is commonly used in data mining, e.g. satan, saint, ports weep, m scan, n map etc.

1.2 Types of IDS:

***1.* Signature based (Misuse) IDS:**

This type of intrusion detection system contains a database of know vulnerabilities. It monitors traffic and seeks a pattern or a signature match. This means, it operates in much the same way as a virus scanner, by searching for a known identity or signature for each specific intrusion event. It can be placed on a network to watch the network vulnerabilities or can be placed on a host. Signature-based IDS examine ongoing traffic, activity, transaction, or behavior for matches with known patterns of even specific to known attacks and it raises alarm only when the so called match is found.

***2.* Anomaly based IDS:**

Also known as Heuristic or Behavior based, Anomaly based IDS analyzes the traffic patterns and determine normal activities. After that, it applies statistical or heuristic measures to event to determine if they match with this normal behavior. Events which do not match with the accepted normal behavior patterns are considered as attacks . By creating patterns of normal behavior, anomaly based IDS systems can observe when current behavior deviates statistically from the normal. This capability theoretically gives anomaly-based IDSs abilities to detect new attacks that haven't been seen before or close variants to previously known attacks. It

Manuscript received July, 2015.

Ruchira Gurav, *Computer Engineering, Pune University Pune, India, 9552956455*

Prof. Aparna A. Junnarkar, *Computer Engineering , Pune University, Pune, India.*

means, these types of IDS may identify any possible attacks

3. Host based IDS (HIDS):
Host-based IDS can analyze activities on the host it monitors at a high level of detail. It can often determine which processes and/or users are involved in malicious activities. It can monitor events that are local to a host and can detect successful or failure of attacks that cannot be seen by a network-based IDS.

4. Network based IDS (NIDS):
Network based IDS monitors the traffic on its entire network segment. This is generally accomplished by placing the network interface card in promiscuous mode to capture all network traffic segments. These network traffic packets are checked network by the IDS to find the attacks. Network based IDS can reassemble packets, look at headers, determine if there are any predefined patterns or signature match.

1.3 Features:

There are in all 41 features available with the help of which attacks can be detected. Instead of describing all the features, here we divide them into three groups and provide descriptions and examples for each group.

Group 1 includes features describing the commands used in the connection (instead of the commands themselves). These features describe the aspects of the commands that have a key role in defining the attack scenarios. Examples of this group are number of file creations, number of operations on access control files, number of root accesses, etc.

Group 2 includes features describing the connection specifications. This group includes a set of features that present the technical aspects of the connection. Examples of this group include: protocol type, flags, duration, service types, number of data bytes from source to destination, etc.

Group 3 includes features describing the connections to the same host in last 2 seconds. Examples of this group are: number of connections having the same destination host and using the same service, % of connections to the current host that have a rejection error, % of different services on the current host, etc.. During inspection of the data it turned out that the values.

In the intelligent Multi Level system proposed by Sherif M. Badr [1], there are three levels for classifying attacks and for each level it makes use of different machine learning techniques. The results have shown that, Decision trees can work efficiently for generalizing problem and for unseen data likewise MLP also performs well but the main drawback is use of C5 decision tree leads to high false alarm rate and Exhaustive Algorithm shows low classification rate for U2R attack.

S. Revathi , Dr. A. Malathi [2] , they have checked on various classification techniques like MLP, Random Forest Naïve Bayes, JRIP, how they perform for classifying and detecting attacks. Results of this paper show that MLP performs efficiently with better accuracy for detecting U2R attacks. R. Naja, Mohsen Afsharchi [3], in this paper experimental result shows that Naive- Bayes is faster at making decisions regarding classification with better accuracy.

Mohammad Reza Norouzian, Sobhan Merat [4], in this paper it is experimentally proved that how Artificial Neural Networks (MLP) perform efficiently for large data sets with better accuracy, whereas P. Ganesh Kumar, D. Devaraj [5], have proved experimentally that if reduced feature set is used then gives better results and saves on time and computation overhead and leads to reduced redundancy or dimensionality reduction.

3.1 NIDS Architecture General Flow:

Description of the Architecture:
-Packet Monitor: This module monitors network stream real time and capture packets to serve for the data source of the NIDS.
-Preprocessor: In preprocessing phase, network traffic collected and processed for use as input to the system.
Classifying attacks in NIDS Using MLP and Naive Bayes
- Feature Extractor: This module extracts feature vector from the network packets (connection records) and submits the feature vector to the classifier module.
-Classifier: The function of this module is to analyze the network stream and to draw a conclusion whether intrusion happens or not.
-Decision: When detecting that intrusion happens, this module will send a warning message to the user.
-Knowledgebase: This module serves for the training samples of the classifier phase. As you know, the artificial neural networks can work effectively only when it has been trained correctly and sufficiently. The intrusion samples can be perfected under user participation, so the capability of the detection can improve continually. All of these modules together make the NIDS architecture system based on the artificial neural networks.

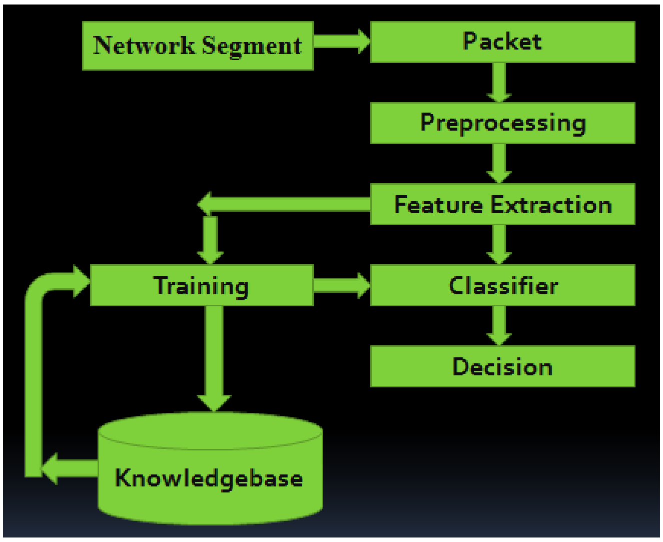

Fig 1. NIDS Architecture

II. Proposed Method

The following figure shows the flow of Proposed Approach:

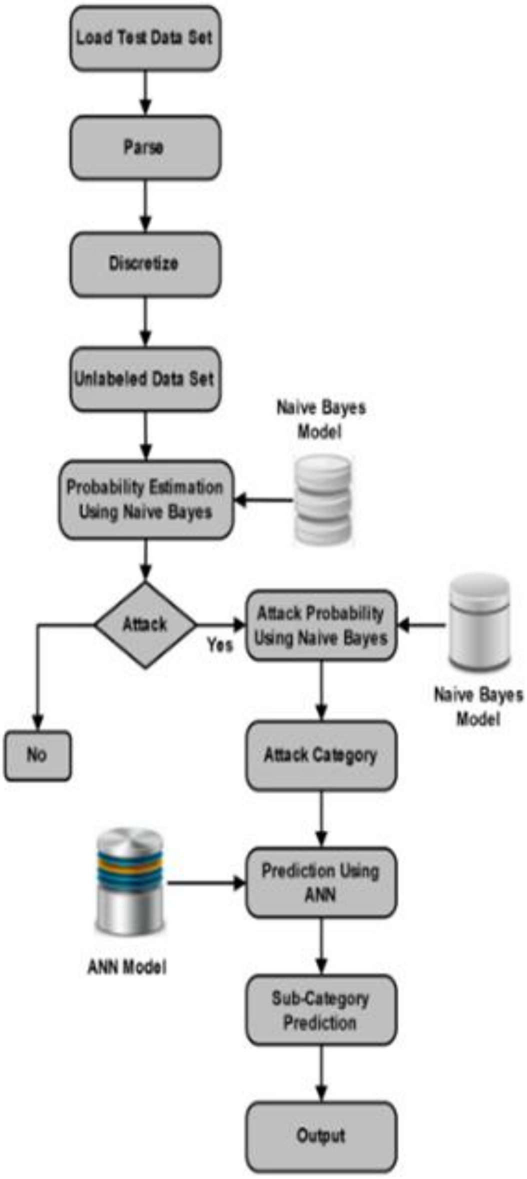

Fig 2. Proposed Approach

The proposed system is a multilevel - hybrid approach, has the capability of classifying network intruders into a set of different levels. Where Level 1, classifies the network records to either normal or attack. The second level can identify four categories/classes like DoS (Denial of Service), Probe, U2R (User to Root), R2L (Root to Local) . The data is input in the first level which identifies if this record is a normal record or attack. If the record is identified as an attack then the module would raise a flag to the administrator that the coming record is an attack then the module inputs this record to the second level which identifies the class of the coming attack. Level 2 modules pass each attack record according to its class type to level 3 modules. Level 3 consists of 4 modules one for each class type (DOS, Probe, R2L, U2R). Each module is responsible for identifying the attack type of coming record. Modules of each level can use different machine learning techniques. Hybrid model will improve the performance to detect intrusions. The proposed approach makes use of Naive- Bayes method for Level 1 and Level 2 whereas Artificial Neural Networks that is Multi-Layer Perceptron for Level 3.

III . SYSTEM COMPONENTS

Input to the system:

The input to the system is Feature vector , there are in all 41 features like protocol, service used, destination bytes, source bytes etc. ,but some features are redundant so to avoid curse of dimensionality and to reduce redundancy 13 features are to be considered. Some features have binary values and some have numeric integer values so to make it continuous feature vector is normalized in range of [0,1].

Techniques Used:

1] Naive-Bayes:

Naive Bayes follows the probabilistic approach and predicts chances of current data moving on are of type normal or attack. This technique is efficient in terms of time complexity and generates results in less time with better accuracy.

-Let X be a data sample

-H be a hypothesis that X belongs to class C

-Classification is to determine P(H/X), the probability that the hypothesis holds given the observed data sample X

$$P(H \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid H)P(H)}{P(\mathbf{X})}$$

2] Artificial Neural Networks- Multi Layer Perceptron:

Neural Networks is biologically inspired by a nervous system.MLP is a multilevel feed forward neural network wherein output of first layer is fed as a input to the next layer, followed by a activation function. The learning of neurons is carried out with the help of back propagation rule.

1) Present the neural network with a number of inputs

2) Check how closely the actual output generated for a specific input matches the desired output.

3) Change the neural network parameters to better approximate the outputs.

There are two different learning methods for the neural networks: supervised and unsupervised. In supervised learning method, the network learns the desired output for a given input or pattern. The well known architecture of supervised neural network is the Multi-Layer Perceptron (MLP); the MLP is employed for Pattern Recognition problems. MLP performs well for unseen data and also for large data sets.

C] Dataset:

NSL-KDD dataset which has been used as effective benchmark dataset to compare various machine learning techniques. The main advantages of NSL KDD dataset are:

1) No redundant records in the train set

2) No duplicate record in the test set

3) The selected records is inversely proportional to the Percentage of original records in KDD data set.

The training dataset consists of 21 different attacks out of the 37 present in the test dataset. Most novel attacks are present in test dataset which are not present in training data. The 4 major attack categories: DoS, Probe, U2R and R2L.

IV. RESULTS

Accuracy measures are:

- *True positives (tp)* : The number of items correctly labeled as belonging to the positive class
- *True negative (tn)* : The number of items incorrectly labeled
- *false positives (fp)*: The number of items incorrectly labeled as belonging to the class
- *false negatives (fn)* : The number of items not labeled as belonging to the positive class but should have been

Accuracy= (tp+tn)/(tp+tn+fp+fn)

Features /Attributes which are fed as input to the system : Attack Type, Services, Flag, Src_Byte, Dest_Byte, Count, Srv Count, Dest_host_count, Dest_host_srv_count.

Learning Techniques	False Alarm Rate	Mean Square Error	Accuracy
Multilayer Perceptron	5%	2.33%	97.81%
Exhaustive	10.3%	8.33%	54.4%

V. CONCLUSION

After studying results of previous systems we have analyzed that exhaustive gives only 54.4% of classification rate where Multilayer Perceptron performs efficiently with accuracy more than 90% and Naive Bayes generates results more quickly than decision Tree classifiers .So our system makes use of Naive Bayes and MLP technique which is a Multi-Level Hybrid approach. Naive-Bayes is efficient in terms of time complexity than any other algorithm It Results into less false alarm rate. Multi-Layer Perceptron is efficient in terms of accuracy. ANN works better for incomplete and unseen data with no requirement of updating records if properly trained and even works efficiently for large data sets. The proposed method Classifies attacks not only in normal or threat but in sub- attack categories so that accordingly preventing actions can be taken. If we use reduced feature set it helps in reducing redundancy and curse of dimensionality while selecting features so saves on time with better accuracy. So the proposed model of Intrusion Detection System is designed in such a way that achieves efficiency by providing high detection and classification rate with less false alarm rates and speed.

REFERENCES

[1] Sherif M. Badr, Implementation of Intelligent Multi-Layer Intrusion Detection Systems (IMLIDS), International Journal of Computer Applications (0975 8887) Volume 61 No.4, January 2013

[2] S. Revathi, Dr. A. Malathi, Detecting User-To-Root (U2R) Attacks Based on Various Machine Learning Techniques, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 4, April 2014

[3] R. Naja, Mohsen Afsharchi, Network Intrusion Detection Using Tree Augmented Naive-Bayes, CICIS'12, IASBS, Zanjan, Iran, May 29-31, 2012

[4] Mohammad Reza Norouzian, Sobhan Merati, Classifying Attacks in a Network Intrusion Detection System Based on Artificial Neural Network s , ICACT 2011

[5] P. Ganesh Kumar and D. Devaraj , Intrusion Detection Using Artificial Neural Network with Reduced Input Features, ICTACT JOURNAL ON SOFT COMPUTING,JULY 2010, ISSUE: 01

[6] Dewan Md. Farid, Nouria Harbi, Emna Bahri, Mohammad Zahidur Rahman, Chowdhury Mofizur Rahaman. "Attacks Classification in Adaptive Intrusion Detection using Decision Tree". World Academy of Science, Engineering and Technology 63, 2010

[7] Giovanni Vigna, Steve Eckmann, and Richard A. Kemmerer. The STAT Tool Suite. In Proceedings of DISCEX 2000, Hilton Head, South Carolina, January 2000. IEEE Computer Society Press

[8] J Cheng and R Greiner, From Feature Selection To Building Of Bayesian Classifiers: A Network Intrusion Detection Perspective, Proc. 14th Canadian Conference On AI, 2001

[9] M Pater, H Kim, and A Pamnam, State Of The Art In Intrusion Detection System.

[10] M Panda and M.R Patra, Network Intrusion Detection Using Nave Bayes, International Journal Of Computer Science And Network Security 7 (2007), no. 12, 258263.

Ruchira Gurav received the B.E. degree in Information Technology Engineering from BVCOEW, Pune University in 2013. She is now pursuing M.E. degree in Computer Engineering from P.E.S.'s Modern College of Engineering and her area of interest is Networking and Security