Intrusion Detection System with Multi Layer using Bayesian Networks

Jasreena Kaur Bains
Lovely Professional University, Punjab, India

Kiran Kumar Kaki
Lovely Professional University, Punjab, India

Kapil Sharma
Lovely Professional University, Punjab, India

ABSTRACT
In the era of network security, intrusion detection system plays a vital to detect real – time intrusions, and to execute work to stop the attack. Being everything shifting to internet, security became the foremost preference. In real world, the minority attacks R2L (Remote-To-User) and U2R (User-To-Root) are more hazardous than Probe and DoS (Denial-Of-Service) majority attacks. Present IDS are not much efficient to detect these low level attacks. Therefore, it is extremely important to improve the detection performance for the R2L and U2R attacks with the majority attacks. In this paper hierarchical layered approach for improving detection rate of minority attacks as well as majority attacks is propound. The propound model used Naive bayes classifier with K2 learning process on reduced NSL KDD dataset for each attack class. In this method every layer is individually trained to detect a single type of attack category and the outcome of one layer is passed into another layer to increase the detection rate and for better categorization of both the majority and minority attacks.

Keywords:
Intrusion detection system (IDS), Network security, Feature selection, naive bayes classifier, R2U, U2R, DoS

1. INTRODUCTION
The Network Security Can be defined as "It is the set of guidelines adopted by a Network administrator to prevent the unauthorized access, modification and misuse of the network traffic by attackers.ID (Intrusion Detection) techniques are used to strength security and increase resistance to internal and external attacks.

With the Increase in Technology the chance of malware and Intrusions get increase. The firewalls do not provide much security. In order to inspect network traffic many corporation are using proactive and reactive both types of IDS system.

The goal of the intrusion detection system is to monitor and control the network traffic to prevent the numerous types of intrusion attempt. When an intrusion detection system is deployed, it provides warning to the users indicating that system is under attack. These warnings can help users to prevent the attack by increasing resistance to attack.

There are two techniques which are used in the intrusion detection. Anomaly detection technique which is intrusive but not the anomalous, will results in very dangerous situation, because the system will not detect these intrusive activities. To avoid these problems the important thing is choosing the threshold level. Anomaly detection also requires the possible updating and keeping track of profile meters of the system which makes this technique an expensive one. Misuse Detection technique is based on a concept which considers that any attack can be represent in the form of a pattern which can detect even a small variation in the system. This means that these are the systems which can detect any known attack pattern but these systems give the problem in the case of unknown attacks because these systems can only detect the known attacks.

Network based vs. Host based system -Network based IDS are used to check the network traffic for searching any intrusion. It is placed at a point in the network where it check each data packet that coming in to the Network and try to detect any intrusion in these packets. One problem relating to network based IDS is that these are the easy targets of DOS attack as it servicing the entire segment alone. Host based IDS is useful in exploiting the vulnerabilities in the specific operating system. This type of IDS can only examine the log files of their own host; this type of IDS is unable to detect the attacks on the other hosts.

A network attack falls into one of four categories:

- Denial- of- service – A DoS attack tries to make resources unavailable to its users by flooding the network with more requests that the server can handle.
- Probe – probe attack are aimed to gain the information about the target host.
- User –To- Root- U2R attack involve the semantic details that are content based and has the local access to the user machines , also tries to gain the super user privileges.
- Remote- To-Local –R2L attacks are most difficult to detect because they involve both network level and host level features. In R2L attackers does not have an account on user machines, therefore tries to gain the access.

In [7] Author wrote a well-known paper that uses the Bayesian rule of conditional probability to point out that implication of the base-rate fallacy for intrusion detection. In [8] a system using a layered approach for intrusion detection was developed. Result shows that layered CRFs has very high attack detection rate 98.6% for Probe and 97.40% for DoS. However, they were outperformed by a significant percent for the R2L and U2R attacks. In [9] a system with three levels of attack depending on four main categories was developed. In the whole experimentation, the performance of naive bayes is compared with decision trees and found the performance of naïve bayes better than the decision trees.
In [10] An IDS was developed based on layered approach using naive bayes classifier which significantly increases the minority detection rate without hurting the performance of majority attacks. But there were two main issues observed

during the analysis phase. First, number of intrusions on the network is very small amount of the total traffic. Second, the attack groups are different.

In this paper, machine learning algorithm naive bayes classifier with K2 learning has been evaluated on the NSL KDD dataset to detect the attacks form the above four categories with hierarchical layered approach for improving detection rate of minority attacks as well as majority attacks. This method constructs the network based intrusion detection system for anomaly detection.

Rest of the paper is organized as follows: section 2 gives the definition of proposed work, the proposed methodology described in section 3; in section 4 finally the paper is concluded with their conclusion.

2. PROPOSED WORK

In this paper hierarchical layered approach for improving detection rate of minority attacks (R2L and U2R) as well as majority attacks (DoS and Probe) is propound. The propound model used Naive Bayes classifier with K2 learning on reduced dataset NSL KDD for each attack class. In this method every layer is individually trained to detect a single type of attack category and the outcome of one layer is passed into another layer to increase the detection rate and for better categorization of both the majority and minority attacks.

3. PROPOUND METHODOLOGY

The methodology of proposed system consists of different phases. Each step is arranged in pipelined manner.

3.1 K2 learning process

To implement the proposed system first we need to perform the K2 learning process after the creation of deployment area in the Netbeans by integrating with the WEKA tool.

3.2 Naive bayes classifier

Naive bayes classifier is supervised machine learning algorithm as well as statistical method for classification. After K2 learning process naive bayes classifier is used to classify the different types of attacks. It works on strong independence relation assumption [9], that is, features are Independent in the context of a session class and the probability of one attribute does not affect the probability of the other. It is defined as follows:

$$P(c|X) = \frac{P(X|c)\,P(c)}{P(X)}$$

i.e.

$$(c|X) = P(x1|c)\,P(x2|c)\ldots P(xn|c)\,P(c)$$

Where,

- P (c|x) is the posterior probability of class (target) given predictor (attribute).
- P(c) is the prior probability of class.
- P (x|c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

The time complexity for learning a naive bayes classifier is O (Np), where N is the number of training examples and p is the number of features.

The complete work is defined under the Bayesian Network classifier. There are three main steps to construct a Bayesian network by domain knowledge:

1) Determine the number and the meanings of the variables in the interested domain.

2) Determine the direct influence relationships among variables in the domain.

3) Determine the conditional probabilities given the structure of the Bayesian networks from step 2.

3.3 Feature selection

Feature selection is an essential data processing step prior to applying a learning algorithm. In other words, the feature selection process is used to remove the different features from the original dataset.

3.4 Working of the hierarchical layered model

The hierarchical layered approach is implemented with the WEKA tool integrating with Netbeans by selecting set of features for every layer rather than using all the 42 features of the dataset. It will categories the intrusion into one of four classes; Probe, DoS, R2L and U2R. We will train and test each layer differently to detect an attack. For example, first layer of our proposed model is trained to detect Probe attack only and outcome of first layer is passed into second layer for better categorization .The second layer is trained to detect DoS attack and outcome of this layer is passed to next one. Figure 1 represents the hierarchical layered mode

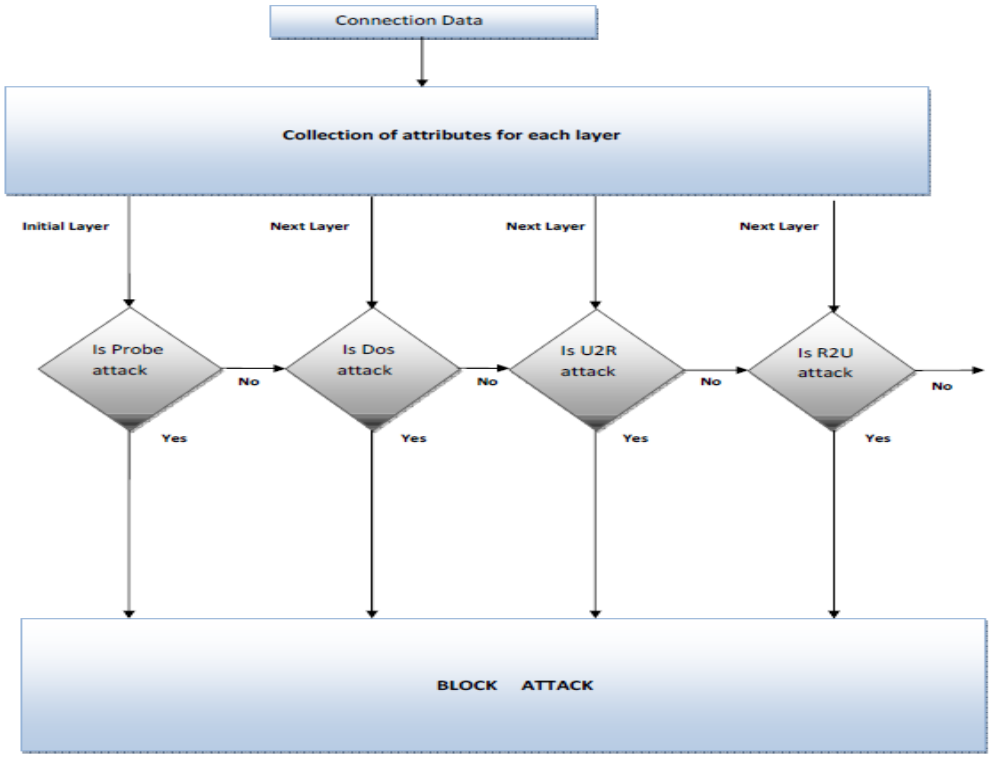

Figure1: Working of the propound model

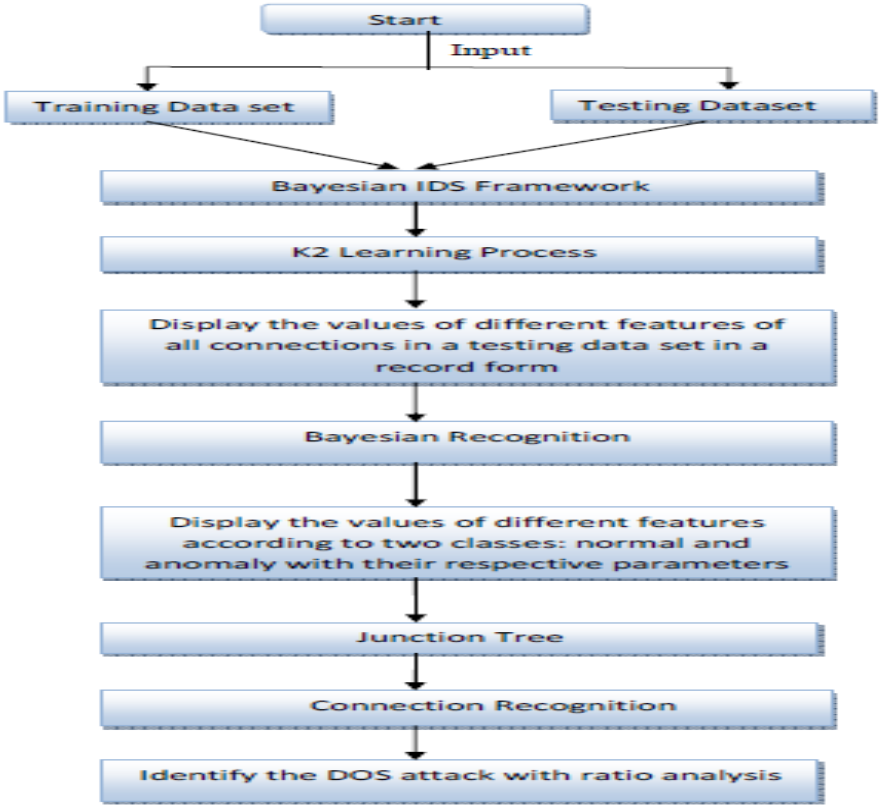

Figure2: Framework for proposed intrusion detection system

3.5 Dataset description

To process with proposed approach, NSL KDD dataset is used. Dataset represent the data as rows of TCP\IP dumps where each row consist of computer connection .Packet information in TCP dump file is summarized into connections. A connection is a sequence of TCP packet starting and ending at some defined time and data flows between source IP address and target IP address under well define protocol or normal connection. Each computer connection has 41 features and these features are grouped in to four categories before feeding the data to the Bayesian network, for learning or testing, raw network traffic has to be pre-processed and summarized into connections or high-level events.

9 of the 41 features are used, these features are

- Protocol type: type of the protocol, e.g. tcp, udp, etc.
- Service: network service on the destination, e.g., http, telnet.
- Land: 1 if connection is from/to the same host/port, 0 otherwise.
- Wrong Fragment: number of wrong" fragments.
- Num, failed logins: number of failed login attempts.
- Logged in: 1 if successfully logged in, 0 otherwise
- Root shell: 1 if root shell is obtained, 0 otherwise.
- Is guest login: 1 if the login is a "guest" login, 0 otherwise.

4. CONCLUSION

In this paper, machine learning algorithm naive bayes classifier with K2 learning has been evaluated on the NSL KDD dataset to detect the attacks form the four categories with hierarchical layered approach for improving detection rate of minority attacks as well as majority attacks. The outcome of one layer is passed into another layer to increase the detection rate and for better categorization of both the majority and minority attacks. The proposed hierarchical layered model with naive bayes classifier will result in better prediction of minority classes as well as majority classes.

5. REFERENCES

[1] T. Subbulakshmi, S. Mercy Shalinie, A. Ramamoorthi Detection and Classification of DDoS Attacks Using Machine Learning Algorithms, European Journal of Scientific Research ISSN 1450-216X Vol.47 No.3 (2010), pp.334-346

[2] Reyhaneh Karimazad , Ahmad Faraahi, An Anomaly-Based Method for DDoS Attacks Detection using RBF Neural Networks, International Conference on Network and Electronics Engineering IPCSIT vol.11 (2011) IACSIT Press, Singapore

[3] Subbulakshmi, T.Shalinie, S.M. GanapathiSubramanian, V.Balakrishnan, K. Anand Kumar. Kannathal, K. Detection of dDoS attacks using Enhanced Support Vector Machines with real time generated dataset, Advanced Computing (ICoAC), 2011 Third International Conference

[4] Saman M. Abdulla, Najla B. Al-Dabagh, Omar Zakaria, Identify Features and Parameters to Devise an Accurate Intrusion Detection System Using Artificial Neural Network, World Academy of Science, Engineering and Technology 2010.

[5] I Ahmad, A B Abdulah, A S Alghamdi, K Alnfajan, MHussain, Feature Subset Selection for Network intrusion Detection Mechanism Using Genetic Eigen Vectors, Proc .of CSIT vol.5 (2011)

[6] H Nguyen, K Franke, S Petrovic Improving Effectiveness of Intrusion Detection by Correlation Feature Selection, 2010 International Conference on Availability, Reliability and Security,IEEE Pages-17-24 .

[7] S.Axelsson, "The base rate fallacy and its implications for the difficulty of Intrusion detection", Proc. Of 6th. ACM conference on computer and communication security 1999.

[8] Kapil Kumar Gupta, BaikunthNath and Ramamohanarookotagiri, "A layered approach using conditional random fields for intrusion detection", IEEE Tranc.on Dependence and secure computing, Vol.7, 2010

[9] Nahla Ben Amor, Salem Benferhat, ZiedElouedi "Naive Bayes vs Decision Trees in Intrusion Detection Systems" AC'04, March 14-17, 2004,

[10] Neelam Sharma, Saurabh Mukherjee, Layered Approach for Intrusion Detection Using Naive Bayes Classifier, International Conference on Advances in Computing, Communications and Informatics (ICACCI-2012), August 3-5, 2012, Chennai, T Nadu, India.