# Real Time Intrusion Detection System using Machine Learning

Ajit Kalekar, Niranjan Kshatriya, Sneha Chakranarayan, Snehal Wadekar
Department of Computer Science,
AISSMS College of Engineering
University of Pune, Pune

***Abstract*--Today, world has come closer due to rapid increase internet. As technology has been developed many threats are emerged for the data security which is not at all good for sensitive data transactions, but as we know that the network security also posses equal importance in the computer infrastructure. Because of the intruders the security of the network has become serious problem. Thus to overcome this we are proposing this paper which is based on machine learning algorithm for intrusion detection using Naïve Bayesian Classifier, which is based on probabilistic model. This algorithm performs balance detections and keeps false positive rate at acceptable level for different types of real time networking attacks. In this, the system is trained by arranging the data attributes in a characterised format which eliminates the redundancy resulting in the reduction of complexity.**

***Keyword-: Real Time Intrusion Detection, Machine Learning, Naïve Bayes classifier*.**

## 1. INTRODUCTION

With the tremendous growth of network-based services and sensitive information on networks, network security is becoming more and more important than ever before. Intrusion detection techniques are the last line of defences against computer attacks behind secure network architecture design, firewalls, passwords, encryptions and personal screening. Despite number of intrusion prevention techniques available, attacks against computer systems are still successful. Thus, intrusion detection systems (IDSs) play a vital role in real time network security. An intrusion is defined as any set of actions that attempt to harm or damage the data which includes a deliberate unauthorised access to the information, manipulate information, or make a system unreliable. Intrusion detection system is a model designed to detect attacks among the various type of packets. It is the process of examining the events which occurs in a computer system or network and analysing them for presence of intrusions.

## 2. NAÏVE BAYES

The Naïve Bayes method is based on the work of Thomas Bayes (1702-1761). Bayes networks consider a very strong features independence assumption. The Naive Bayes classifier is a supervised learning algorithm based largely off of Bayes Theorem [14]:

$$P\,(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The Naïve Bayes algorithm[11] makes the assumption that all the input attributes are independent, such as one attribute doesn't affect the other in deciding whether or not a condition in the database. In Bayesian classification, we have an assumption that the given data belongs to a particular class, and then the probability for the assumption to be true is calculated. The approach [14] requires only one scan of the whole data. Also, if at some stage there are additional training data, then each training example can incrementally increase/decrease the probability that a hypothesis is correct. Thus, a Bayesian network is used to model a domain containing uncertainty

## 3. RELATED WORK

We have reviewed various papers of researchers. The contribution of researchers has been discussed:
Jonathan Palmer [7] have proposed Intrusion detection is an important part of network infrastructures but rule based IDSs can sometimes be difficult and time consuming to maintain. Many machine learning algorithms for intrusion detection have been researched over the past few years to help solve this problem. These techniques require a training data set and many researchers use the KDD '99 data set. In order for these techniques to perform well in real world environments, they must be trained with realistic data sets. This paper set out to determine if the KDD '99 data set was indeed suitable for this application. After attempting to design and perfrom an experiment to test the validity of the KDD '99 dataset, the direct results were inconclusive but lead to a further investigation of the data. Upon reviewing the data set itself and comparing its structure to the structure of some simulated attacks, it would appear that the KDD '99 data set is not the best choice for training machine learning algorithms that would be used in real world applications. Salem Benferhat, Abdelhamid Boudjelida, Habiba Drias [8] presented.The most adapted Bayesian classification model for intrusion detection is naive Bayes. They present several advantages due to their simple structure. Bayesian naïve networks construction is very simple; it is always easy to consider new scenarios (updates facility). Inference is polynomial, while inference in Bayesian networks with general structures is known to be a hard problem. Amjad Hussain Bhat, Sabyasachi Patra, Dr. Debasish Jena [9] have proposed Tree algorithm is used to classify network connections into intrusion and normal data based on a labelled training dataset that helps it in building classification patterns. In the second, anomaly detection part, we use hybrid approach of NB Tree and Random forest algorithm is used based on the

similarity of connections features. There are challenges in anomaly detection, one challenge is the imbalance between intrusion types in real network connections datasets which are used as training data to our detection system. Some types of intrusions like denial of service (DoS) have many network connections than other intrusion types like user to root (U2R) and normal have more than connection among all; so, any data mining approach will be interested in decreasing the overall error rate of the system regardless of intrusion types, which causes increasing the error rate of the minority attacks like U2R, although these attacks are very dangerous than majority attacks. Upendra [10] have showed C4.5 selected 7 attribute technique for intrusion detection and performed feature set reduction and evaluated their performance. From the result, it is observed that after applying the feature selection from 41 attributes to 11 and 7 attributes. The overall performance of C4.5 has increased their performance than NB. Mohan Banerjee, Roopali Soni [11] research on two learning algorithms of data mining i.e. K-means and Naive Bayes classifier. K-means is a clustering algorithm, which work to provide grouping to data sample on the basis of their similarities and dissimilarities. Naive Bayes classifier is classification algorithm which correctly classifies the intrusion/attack. The combination of these two algorithms used in order to improve accuracy, precision rate and reduce the false positive rate. In this paper, the apply one of the efficient data mining algorithms called k-means clustering via naïve bayes classification for anomaly based network intrusion detection. Experimental results on the KDD cup'99 data set show the novelty of our approach in detecting network intrusion. It is observed that the proposed technique performs better in terms of Detection rate when applied to KDD'99 data sets compared to a naïve bayes based approach. Neethu B [12] have proposed IDSs based on human experts, intrusion detection techniques using machine learning have attracted more and more interests in recent years. Machine learning is a field of study which provides the computers with the ability of learning from previous experience. Machine learning is based heavily on statistical analysis of data and some algorithms can use patterns found in previous data to make decisions about new data. Dewan Md. Farid, Nouria Harbi, and Mohammad Zahidur Rahman [13] have paper introducing a new hybrid learning algorithm for adaptive network intrusion detection using naive Bayesian classifier and ID3 algorithm, which analyzes the large volume of network data and considers the complex properties of attack behaviours to improve the performance of detection speed and detection accuracy. In this paper we have concentrated on the development of the performance of naïve Bayesian classifier and ID3 algorithm. It has been successfully tested that this hybrid algorithm minimized false positives, as well as maximize balance detection rates on the 5 classes of KDD99 benchmark dataset. The attacks of KDD99 dataset detected with 99% accuracy using proposed algorithm. Mrutyunjaya Panda and Manas Ranjan Patra [14] have proposed a framework of NIDS based on Naïve Bayes algorithm. The framework builds the patterns of the network services over data sets labelled by the services. With the built patterns, the framework detects attacks in the datasets using the naïve Bayes Classifier algorithm. Compared to the neural network based approach, our approach achieve higher detection rate, less time consuming and has low cost factor. However, it generates somewhat more false positives. As a naïve Bayesian network is a restricted network that has only two layers and assumes complete independence between the information nodes. This poses a limitation to this research work. In order to alleviate this problem so as to reduce the false positives, active platform or event based classification may be thought of using Bayesian network. We continue our work in this direction in order to build an efficient intrusion detection model. Prof. D.P.Gaikwad and Dr.R.C.Thool [1] have done survey on architecture taxonomy and product of IDS. They mention limitation of various IDS available in market that complete attack prevention is not realistically attainable due to the configuration and administration, system complexity, and abuse by user. They have discussed some aspects of IDS such as role of IDS, categories of IDS, modes of IDS. They proposed the general architecture, network parameter and architectural taxonomy. General architecture consist of three components namely Sensor (agent), Analyzer, User Interface. Sensor collects information and sends it to analyzer. Analyzer determines which intrusion occurs and user interface is used for interaction between system and users. This architecture improves performance of system. They have discussed various features of different IDS's such as Snort, MacAfee and Tripwire. As a future work author are going to develop wireless network IDS system.

## 1. PROPOSAL OF SYSTEM

We are proposing the Real Time Intrusion Detection System using Naïve Bayes algorithm which mainly implies the detection of abnormal packets using past experience of the system. Here the incoming packets are analysed and categorised according to values of the attributes to produce dataset. Using this data set the next arriving packets are detected as normal or abnormal packets. If abnormal packets are detected reporting can be done.

The various modules of the system are as follows:

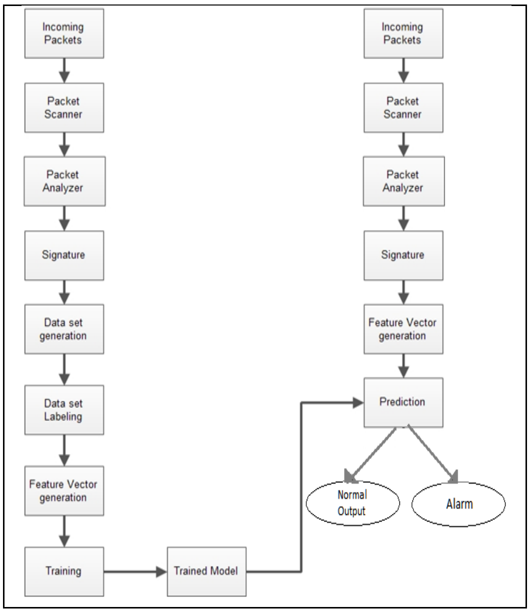

Fig1: Basic system function

*Incoming packet:*

In our project directly connected to network (we capture online packet) and transfer to packet scanner.

*Packet Capture:*

In Real time networking many packets are transferred from source to destination. In this module live packets from the network are captured and passed to the pre-processor module. According to the concept of machine learning the data is grouped or clustered based on its features or characteristics for e.g. protocols used by the packets.

*Packet scanner:*

After catching the packet we use packet scanner for the purpose of scanning the packet. Packet scanning is the important part of our system.

*Packet analyzer:*

Packet analyzer is also known as Packet sniffer. As data streams flow across the network sniffer capture each packet and if needed decodes the packet, raw data showing the values of various fields in the packet and analyzes its content according to specification.

Packet analyzer specially used for:
- Analyzing the network problem.
-Detecting network misuse by internal or external user.

*Signature:*

Using the attribute value and parameter we can define the signature for each packet.

*Data set generation:*

After generating the signature, we collect the signature for the corresponding packet and generate the data set and it is used for labeling.

*Labeling:*

Labelling is used for defining the corresponding packet.

*Training model:*

Training model contain set of trained data set which used to detect attack in the packet.

*Prediction:*

Based on the training model we make the prediction that the packet is normal packet or abnormal packet.

*Output:*

Based on the prediction (i.e. normal or abnormal) we generate output in the form alarm for abnormal packet.

## 4. CONCLUSION

In this paper, learning algorithm for network intrusion detection using naive Bayesian classifier is presented, which performs balance detections and keeps false positives at acceptable level for different types of network attacks, and eliminates redundant attributes. The proposed algorithm also addresses some difficulties of data mining such as handling continuous attribute, dealing with missing attribute values, and reducing noise in training data. Due to the large volumes of security audit data as well as the complex and dynamic properties of intrusion behaviours, several data mining based intrusion detection techniques have been applied to network-based traffic data and host-based data in the last decades.
However, there remain various issues needed to be examined towards current intrusion detection systems (IDS).

## REFERENCES

[1] Prof. D.P.Gaikwad and Dr.R.C.Thool, Architecture Taxonomy and Product of IDS,International Conference on Computer Applications,Computer Application-II,doi:10.3850/978-981-08-7304-2_0382.

[2] Christine Dartigue, Hyun Ik Jang, and Wenjun Zeng Computer Science Department University of Missouri-Columbia Columbia, Missouri, USA , Services Research Conference on A New Data-Mining Based Approach for Network Intrusion Detection, 2009 IEEE Seventh Annual Communication Networks and Services Research Conference.

[3] Sandhya Peddabachigari, Ajith Abraham, Johnson Thomas,Department of Computer Science, Oklahoma State University, USA paper on Intrusion Detection Systems Using Decision Treesand Support Vector Machines.

[4] G.Prashanth, V.Prashanth, P.Jayashree, N.Srinivasan, IEEE-International Conference on Signal processing, Communications and Networking ,Using Random Forests for Network-based Anomaly detection at Active routers ,Madras Institute of Technology, Anna University Chennai India, Jan 4-6, 2008. Pp93-96.

[5] Mehdi MORADI and Mohammad ZULKERNINE, paper on A Neural Network Based System for Intrusion Detection and Classification of Attacks.

[6] Amira Sayed A. Aziz,Mostafa A. Salama,Aboul ella Hassanien,Sanaa El-Ola Hanafi paper on Artificial Immune System Inspired Intrusion Detection System Using Genetic
Algorithm.

[7] Jonathan Palmer, international paper on Naive Bayes Classification for Intrusion Detection Using Live Packet Capture.

[8] Salem Benferhat, Abdelhamid Boudjelida, Habiba Drias, research on An Intrusion Detection Approach Based on Tree Augmented Naive Bayes and Expert Knowledge.

[9] Amjad Hussain Bhat, Sabyasachi Patra, Dr. Debasish Jena, IEEE paper on Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines, june 2013.

[10] Upendra,Assistant Professor, CSE Department, NIT Raipur, C.G., India, International Journal of Emerging Trends & Technology in Computer Science on An Efficient Feature Reduction Comparison of Machine Learning Algorithms for Intrusion Detection System, Volume 2, Issue 1, January – February 2013.

[11] Mohan Banerjee, Roopali Soni, MTech (CSE) Scholar, Department of Computer Science & Engineering, Thakral College of Technology, Bhopal HOD & AP, Department of Computer Science & Engineering, Thakral College of Technology, Bhopal, International Journal of Science, Engineering and Technology Research on Design and Implementation of Network Intrusion Detection System by using K-means clustering and Naïve Bayes , Volume 2, Issue 3, March 2013.

[12] Neethu B,,Department of Computer Science, Amrita University,International Journal of Electronics and Computer Science Engineering on Classification of Intrusion Detection Dataset using machine learning Approaches.

[13] Dewan Md. Farid1, Nouria Harbi1, and Mohammad Zahidur Rahman2,Department of Computer Science and Engineering,Jahangirnagar University, International Journal of Network Security & Its Applications (IJNSA) on Combining Naïve Bayes and Decision Tree For Adaptive Intrusion Detection.

[14] Mrutyunjaya Panda and Manas Ranjan Patra, Department of E &TC Engineering, G.I.E.T., Gunupur, India, Department of Computer Science, Berhampur University, Berhampur, India, International Journal of Computer Science and Network Security on Network Intrusion Detection Naïve Bayes, VOL.7 No.12, December 2007258,Manuscript received December 5, 2007,Manuscript revised December 20, 2007