www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 3 March 2015, Page No. 10842-10845

Improving efficiency of hybrid Intrusion Detection System using k-means and Naïve Bayes

Pushpak singha, Rahul lakkadwala, Anup sheth, akshay gaikwad, Megha V. kadam

Student (UG), Department of Computer Engineering,
A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India
pushpaksingh93@yahoo.in

Student (UG), Department of Computer Engineering,
A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India
anupscool@gmail.com

Student (UG), Department of Computer Engineering,
A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India
rlakkadwala@gmail.com

Student (UG), Department of Computer Engineering,
A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India
akshayakshay147@gmail.com

Assistant Professor, Department of Computer Engineering,
A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India
megha.desai1@gmail.com

Abstract: ***Today, world has come closer due to rapid increase of internet. So computer security is of big concern now. As technology has been developed many threats are emerged for the data security which is not at all good for sensitive data transactions. So it is necessary to build high level security to provide safe communication between various networks. Intrusion detection systems are built to detect the attacks. Because of the intruders, the security of the network has become serious problem. Thus to overcome this we are proposing this paper for intrusion detection using k-means and Naïve Bayesian Classifier, which is depend on probabilistic model. This algorithm performs attack detection and keeps false positive rate at low level for different types of networking attacks.***

Keywords: Real Time Intrusion Detection, k-means, Naïve Bayes classifier, serialization.

.

1. Introduction

We all know computer networks are gaining a lot of importance these days. From many years threat to data security have emerged as demon for internet. Data needs to be secured for meeting is the first need of people. Attackers, intruders have caused havoc in data security and a wide threat which is usually unending. Intrusion means an unauthorized entry to possibly violently exploit an organization. Thus to overcome intrusions we are going to develop an intrusion detection system using two algorithms which are k means and naive bayes. Intrusion Detection Systems are like a burglar alarm for your computer network. It detects unauthorized access attempts. They are the first stage of defense for your computer system. An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system [5].Our system will monitor packets coming from different network and compares them with known data bases. For the data storage we are using serialization concept. If mismatch it’s a possible new threat, this new threat it again stores in a databases, finally manages and trains packets according to various fields such as hop limit, lengths, flags etc. This it displays list of attacks which are malicious. The main motivation of using data mining methods in intrusion detection is automation. Data mining techniques, such as decision tree (DT), naïve Bayesian classifier (NB), neural network (NN), support vector machine (SVM)[4], k-nearest neighbors (KNN),and genetic algorithm have been widely used to analyze network logs to gain intrusion related knowledge to improve the performance of IDS in last decades. So to prevent the system from attack it is necessary to make the system attack detector [6]. To apply data mining techniques in intrusion detection, the collected logs or audit data needs to be preprocessed and converted to the format that suitable for mining. Next, the reformatted data will be used to develop a clustering or classification model. Data mining provide decision support for intrusion management, and also help IDS for detecting new vulnerabilities and intrusions by discovering unknown patterns of attacks or intrusions. Intrusion detection is a process of gathering intrusion [7] related knowledge that occurred in the computer networks or systems and analyzing them for detecting future intrusions.

2. Naive bayes

The Naïve Bayes method is based on the work of Thomas Bayes (1702-1761). Bayes theorem considers a very strong feature which works on independence assumption. The Naive Bayes classifier is a supervised learning algorithm which is based largely off of Bayes [8].

Theorem: A conditional probability is the likelihood of some conclusion, C, given some evidence/observation, E, where a dependence relationship exists between C and E.

This probability is denoted as **P(C |E)** where

$$P(C\mid E)=\frac{P(E\mid C)P(C)}{P(E)}$$

The Naïve Bayes algorithm takes all the input attributes as independent attribute, so that one attribute doesn't affect the other in deciding whether or not a condition is in the database. In Bayesian classification, we have an assumption that the given data belongs to a particular data set, and then the probability for the assumption to be true is calculated. The approach [9] requires only one scan to go through all data. Also, if at later stage there are additional training data, then each training data can incrementally increase/decrease the probability that a hypothesis is acceptable.

3. Serialization

Serialization is the process of converting some in-memory object to another format that could be used to either store in a file share over the network. After a serialized object has been written, it can be easily read from the file and deserialized, that is the information and bytes that represent the object and its data can be used to recreate the object in memory. This is very useful when communicating between various systems. The serialization format could be either interoperable or non-interoperable Serialization is the process of turning data (e.g. variables) into a representation such as a string, that can be written and read back from for example a file or the database. Serialization is used when large amounts of data have to be stored in flat files and retrieved at a later stage. But to achieve this without serialization, it becomes too repetitious, error-prone and complicated as the data structure is complex.

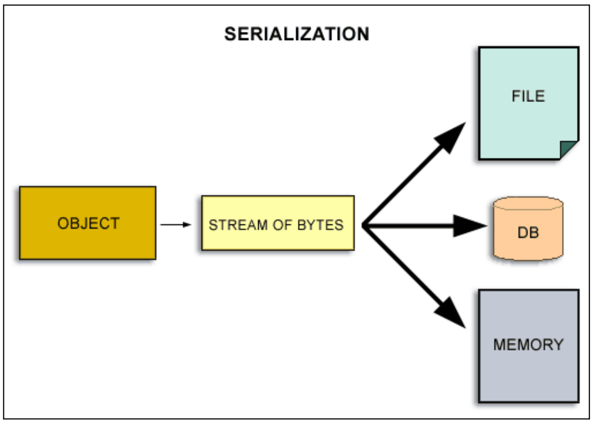

Fig.1 serialization

Fig.1 translates data structures, objects into a format that can be stored in file or buffer and reconstructed if needed. The object in memory are serialized into a binary form that can be stored Serialization is objects encoding into other language.

4. Problem analysis

Traditional Intrusion Detection System (IDS) focus only low level attacks and only generate isolated attacks to achieve greater security in detecting malicious activities. Current intrusion detection is often associated with high false alarm with moderate accuracy and detection rates, for all types of attacks. Naive bayes is very fast to distinguish between normal and abnormal as it processes training set only once to store statistics and use it to predict the unforeseen record. With the help of k-means for clustering and naïve bayes for classification, we can overcome from this problem.

5. Proposed technique

Here we are presenting general idea on a new hybrid model for intrusion detection system which will increase efficiency as compare existing intrusion detection system. In proposed model we are using serialization for data mining concept. Data mining techniques is very useful in many different fields. Over the past ten years, a growing number of research techniques have applied data mining to various problems in intrusion detection. In this we will apply data mining for detection of intrusions. With the help of k-means and naïve bayes, we will detect attacks with good accuracy good detection rates [10]. We propose a model which uses two different techniques K-Mean clustering, Naive Bayes. For the first stage in the proposed hybrid IDS model, we are grouping similar data instances based on their behaviors by using a K-Means clustering as a pre-classification. In second stage, we are using Naïve Bayes classifier. For classifying the resulting clusters into classes like normal and abnormal. Surely it minimizes the false alarm rate.

6. Proposed solution

We are proposing the Real Time Intrusion Detection System using k-means and Naïve Bayes algorithm which detects abnormal packets using past experience of the system. Here the incoming packets are captured, analyzed and categorized according to values of the attributes to produce dataset. Using this data set store in our database, the next arriving packets are detected as normal or abnormal packets. If abnormal packets are detected necessary steps can be taken. Here we are reporting to admin through email. In our project directly connected to network (we capture online packet) and transfer to packet scanner.

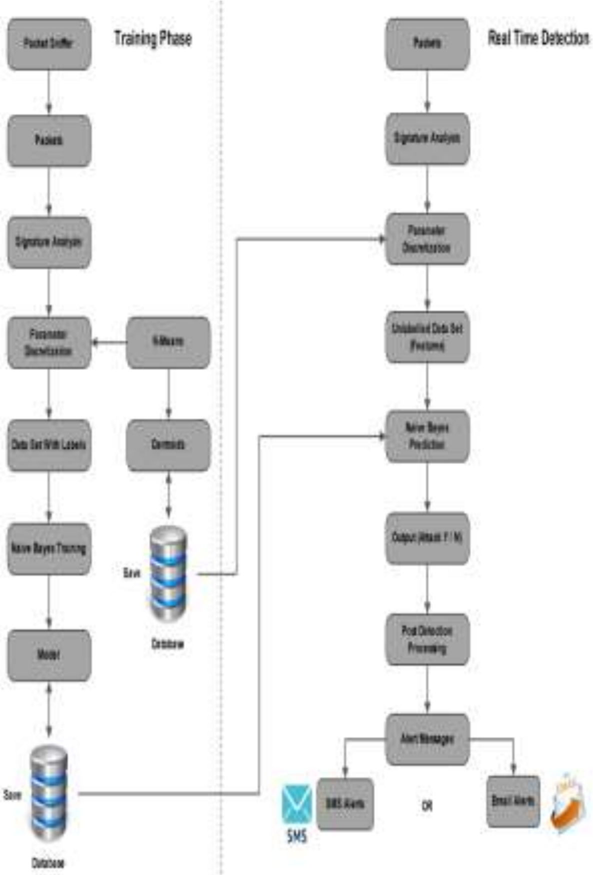

Fig2: Basic system function

Following are the modules of the system:

Capturing packets*:*
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 next stage. 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.

***Scanning**:*
After catching the packets we use packet scanner to scan the packet. Packet scanning is the necessary part of our system.

Analyzing*:*
Packet analyzer is also called as Packet sniffer. As data flow across the network, packets which are coming to our system are captured with the help of packet sniffer. Then it decodes the packet, raw data showing the attributes of various fields in the packet and analyzes its content according to specification.

Parameter:
By taking attribute value we can define the signature for each packet to store in the database.

Data set labeling*:*
After generating the signature, we collect the signature for the corresponding packet and generate the data set and it is used for labeling for defining the corresponding packets

Training model*:*
It contains set of trained data set which used to detect attack in the packet.

Detection*:*
Based on the training model can detect the packet is normal packet or abnormal packet.

Output*:*
Based on the detection (i.e. normal or abnormal) we generate output in the form alarm for abnormal packet.

7. Proposed technique

We have reviewed various papers of researchers. The contribution of researchers has been discussed:

A.M Chandrasekhar and K.Raghuveer [1] have proposed Intrusion detection is an important part of network security but rule based IDSs can sometimes be difficult and time consuming to maintain. Different types of machine learning algorithms for intrusion detection have been researched over the past few years to solve this problem. These methods require a training data set and many researchers use the KDD ’99 data set. To perform these techniques well in real time, they must be trained with realistic data sets. This paper set out to determine is the KDD ’99 data set was indeed suitable for this application. After attempting to design and perform an experiment to test the validity of the KDD ’99 dataset, the results were inconclusive because it works in the static data set. After reviewing it proves KDD ’99 data set is not the good choice for training machine learning algorithms that would be used in real world applications for real time detection. Prof. D.P.Gaikwad and Dr.R.C.Thool [2] have done survey on architecture taxonomy and product of IDS. They proposed the general architecture, network parameter and architectural taxonomy. They discuss about various IDS available in market that it cannot detect all types of attack because of system complexity, configuration and administration. They presented some aspects of IDS which are role of IDS, modes of IDS, categories of IDS. Roshan Chitrakar, Chuanhe Huang [3] presented. The most adapted Bayesian classification is naive Bayes for intrusion detection. They have simple structure. Bayesian naïve networks construction is simple as compare to others; it is easy to consider new model or scenario. 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. Dewan Md. Farid, Nouria Harbi, and Mohammad Zahidur Rahman [11] discussed a new hybrid learning algorithm for network intrusion detection using naive Bayesian classifier and ID3 algorithm, which analyzes the large number of network data and its complex properties to improve the performance of detection accuracy. In this paper, proposed solution applies two data mining algorithms called k-means clustering via naïve bayes classification for intrusion detection. Mrutyunjaya Panda and Manas Ranjan Patra [12] have implemented a framework of NIDS based on Naïve Bayes algorithm. With the built patterns, the framework detects attacks using the naïve Bayes Classifier algorithm. As compare to the neural network based, our approach achieve higher detection rate, less time consuming. Mohan Banerjee, Roopali Soni [13] research on two algorithms of data mining which are K-means and Naive Bayes classifier. K-means is used for clustering algorithm, which is use for grouping of data to the data sample. Naive Bayes classifier is a classification algorithm which classifies the intrusion/attack. These two algorithms used to improve accuracy and reduce the false alarm rate. It has been successfully tested that this hybrid algorithm always minimized false positives. As a future work author are going to develop wireless network IDS system.

8. Conclusion

This paper will improve detecting speed and accuracy of the system, and proposing more efficient method to detect abnormal attacks in the system. In this paper, a hybrid model through combination of K-Means clustering, Naïve Bayes classifier is presented. In which it performs balance detections and keeps false positives at acceptable level for different types of network attacks. Due to the large volumes of security audit data and dynamic properties of intrusion behaviors’, different data mining based intrusion detection techniques have been applied to host-based data in the last decades. The main characteristic of misuse detection technique is to compare the incoming packets with the predefined knowledge in order to decide whether it is an attack or not. In anomaly detection technique it looks for any unexpected changes in behavior of a system against what is considered to be normal behavior. The proposed design of IDS, aims to be more accurate and detect malicious attack.

References

[1] A.M Chandrasekhar, K.Raghuveer,”Intrusion detection technique by using K-means, Fuzzy Neural Network and SVM classifiers”, proceedings of ICCCI, pp1-7, 2013.

[2] 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.

[3] Roshan Chitrakar, Chuanhe Huang, “Anomaly Based Intrusion Detection Using Hybrid Learning Approach of Combining k-Medoids Clustering and Naïve Bayes Classification”, 8th IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), 2012.

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

[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] 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.

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

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

[10] K Qazanfari, M S Mirpouryan, H. Gharaee, "Novel Hybrid Anomaly Based Intrusion Detection Method", 6th IEEE International Symposium on Telecommunications (IST), 2012.

[11] 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.

[12] 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

[13] 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.

Author Profile

Pushpak Singha pursuing Bachelor‘s degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India

Rahul Lakkadwala pursuing Bachelor‘s degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India

Anup Sheth pursuing Bachelor‘s degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India

Akshay Gaikwad pursuing Bachelor‘s degree from Savitribai Phule Pune University in A.I.S.S.M.S College of Engineering, Pune, Maharashtra, India

Megha V. Kadam, Assistant Professor in A.I.S.S.M.S College of engineering, completed Master's Degree from Pune University, Maharashtra. 6 papers have been published and presented in various international conferences as of now working in the area of Network Security