Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-6, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Real Time Intrusion Detection System Using Hybrid Approach

Nainesh V Chaudhari[1], Tejal Chaskar[2] , Renuka Amilkanthwar[3], Rekha Arjun[4] & Kalpana Kadam[5]

[1,2,3,4]B.E. Computer Engineering student, SKN-Sinhgad Institute Of Technology & Science, Lonavala , Pune University, India

[5]Asst. Professor, SKN-Sinhgad Institute Of Technology & Science, Lonavala , Pune University, India)

Abstract: *In our current society, the threat of cyber intrusion is increasingly high and harmful. With the rise of us age in computers, criminal activity has also shifted from physical intrusion into cyber intrusion. Intrusion detection system provide the ability to identify security breaches in a system. A security breach will be any action the owner of the system deems unauthorized. Existing system include anomaly detection or signature database. In the existing system they have implemented a clustering algorithm which matches connections that appear multiple times. In proposed system of our tool we have implemented a clustering algorithm as well as naive Bayes algorithm. In proposed system of our tool we are using hybrid approach. We are creating a Real Time Intrusion Detection System which will actively detect intrusion while the machine is running.*

Keywords: *Network Security, Intrusion Detection, Data Mining, Naive Bayes, Cyber security, Hybrid IDS, Bayes classifier*

Introduction

Internet has changed the life of human being completely. Applications of computer using internet are unlimited. Unfortunately, due to large scale use of internet, the risks and chances of attacks are also increased. So, it is essential to protect our system from different attacks. The word "Security" signifies the quality or state of being secure, that is to be free from any danger. Computer security is known as the protection of computing systems against threats to confidentiality, integrity, and availability. Protection ensures that the information is accessible only to those individuals who are authorize. Morality refers to the validity of data or integrity means that assets can be modified only by authorized parties and only in authorized way. Availability is the degree to which a system, subsystem or equipment is accessible and usable upon demand by an authorized party. Security threats are derived from various sources such as natural calamities, failure of services and people known as intruders. There are two types of intruders: the external intruders who are unauthorized users of the machines they attack, and internal intruders, who have permission to access the system with some restrictions. The traditional prevention techniques such as user authentication, data encryption, cryptography, avoiding programming errors and firewalls are used for computer security. If a password is inefficient and compromised, user authentication cannot prevent unauthorized use. They are generally impossible to protect against malicious attacks, internal attacks and weak modems. Intrusion detection is therefore required as an additional wall for protecting systems. The process of identifying the attacks in a system or network is called as intrusion detection. Intrusion detection is useful not only in ensuring successful intrusions, but also provides vital information for timely defending measures [4].

Intrusion Detection

The process of monitoring the events occurring in a computer system or network and analyzing them for sign of intrusions is known as Intrusion detection. Intrusion detection is classified into two types: misuse intrusion detection and anomaly intrusion detection. The misuse-based Intrusion Detection System uses a database of previous attack patterns and vulnerabilities as a reference. Each intrusion has some specific pattern. This pattern is called as signature. This pattern or signature are used to identify the attacks on the computer system or on the network. So, this is also called as signature –based Intrusion Detection System. An anomaly intrusion detection is any unusual event that occurs in an environment. This method is used to detect attacks that have not been defined yet [3].

The intrusion detection system can be divided into following two types depending on the architecture.

a) Network intrusion detection system

b) Host intrusion detection system

Network-based systems monitor the packet that flow over the network and collect data from network traffic (e.g., packets from network interfaces in promiscuous mode) while host-based systems collect events at the operating system level, such as system calls, or at the application level [5].

Literature survey

With the increasing growth of network-based service and sensitive information on networks, network security is getting more and more important than always. In [1] this paper naïve Bayes method is used. They have proposed a framework of NIDS [Network Intrusion Detection System] based on Naïve Bayes algorithm. The framework builds the patterns of the network services across data sets labelled by the services. With the built patterns, the framework recognizes attacks in the datasets applying the naïve Bayes Classifier algorithm. Compared to the Neural network based approach, we proceed towards achieving higher detection rate, less time consuming and has low cost factor. There is some drawback in this system which generates false alarm. As a naïve Bayesian network is a restricted network that has only two layers and assumes complete autonomy into the information nodes. This poses a limitation to this research work.

The hybrid IDS [2] is efficient to detect known and unknown intrusion. In this application of the data mining algorithm to original connection records how to effectively get the corresponding frequent patterns in the key to study. But there is some drawback in this system to select appropriate and representative original data and to filter precisely useless rules. In this research [3] they used both anomaly detection and a signature database using data mining techniques. They test the ability of data mining using a clustering technique to discover DOS attacks. [6] Clustering refers to the grouping of similar data. This grouping allows users to see pattern of reoccurring activities or popular trends. The description of reoccurring activities is highly compatible with description of denial of service attacks. Our solution provides a tool that would run data mining tools against a log file to detect patterns that may be considered an unauthorized activity. The tool gains additional patterns as time goes by and grows more effective. There is some drawback in this system, they use only clustering algorithm and only clustering is not that much efficient for detecting the attacks.

Data Mining Techniques for Intrusion Detection

In proposed system of our tool we are using hybrid approach. In this hybrid approach we have implemented a clustering as well as naive Bayes algorithm. In clustering technique, we use k-means algorithm. In this system we are using serialization concept of java for storage purpose.

K-Means Algorithm

K-means is one of the simplest unsupervised learning algorithms that solve. well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriority. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{C} \sum_{j=1}^{Ci} (||xi + vj||)^2$$

where,
”xi – vj” is the Euclidean distance between xi and vj.
’ci’ is the number of data points in i^{th} cluster.
’c’ is the number of cluster centers.

Naive Bayes Algorithm

The Naive Bayes algorithm is a heavily simplified Bayesian probability model [7]. In this model, consider the probability of an end

result given several related evidence variables. The probability of end result is encoded in the model along with the probability of the evidence variables occurring given that the end result occurs. The probability of an evidence variable given that the end result occurs is assumed to be independent of the probability of other evidence variables given that end results occur [1].

Bayes theorem:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Let, X be a data tuple. In Bayesian terms, X is considered as "evidence". H be some hypothesis such as that the data tuple X belongs to a specified class. For classification problems, the probability that the hypothesis H holds given the "evidence" or observed data tuple X. P(H|X) is the posterior probability of H condition on X. P(X) is the prior probability of X.

Naive Bayesian Classifiers with an example

The following example is a simple demonstration of applying the Naïve Bayes Classifier. This example shows how to calculate the probability using Naïve Bayes classification algorithm.

Table 1. Naive Bayes Classifiers example

RID	Age	Income	Student	Credit rating	ClassBuys_ Computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle	Medium	No	Excellent	Yes
13	Middle	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

.

Predicting a class label using naïve Bayesian classification, we wish to predict the class label of a tuple using naive Bayesian classification from the training data as in the above table.

The data tuples are described by the attributes age, income, student and credit rating. The class label attribute, buys_computer, has two distinct values (namely, {yes, no}). Let C1 correspond to the class buys_computer=yes and C2 correspond to buys_computer=no. The tuple we wish to classify is

X = (age=youth, income=medium, student=yes, credit_rating=fair)

Calculate Initial Probability

P (buys_computer = yes) = 9/14 =0.643
P (buys_computer =no) = 5/14 =0.357

Calculate Individual Probability

P (age=youth | buys_computer=yes) =2/9=0.222
P (age=youth | buys_computer=no) =3/5=0.600

P(income=medium|buys_computer=yes)=4/9 =0.444
P(income=medium|buys_computer=no)=2/5 =0.400

P(student=yes|buys_computer=yes) =6/9=0.667
P(student=yes|buys_computer=no) =1/5=0.200

P(credit_rating=fair|buys_computer=yes) =6/9=0.667
P(credit_rating=fair|buys_computer=no) =2/5=0.400

Using the above probabilities, we obtain

P (X | buys_computer=yes)=
P (age=youth | buys_computer=yes)*
P (income=medium | buys_computer=yes)*
P (student=yes | buys_computer=yes) *
P (credit rating=fair | buys_computer=yes)

=0.222 * 0.444 * 0.667 * 0.667 =0.044

Similarly, P(X | buys_computer=no)
= 0.600 * 0.400 * 0.200 * 0.400 = 0.019.

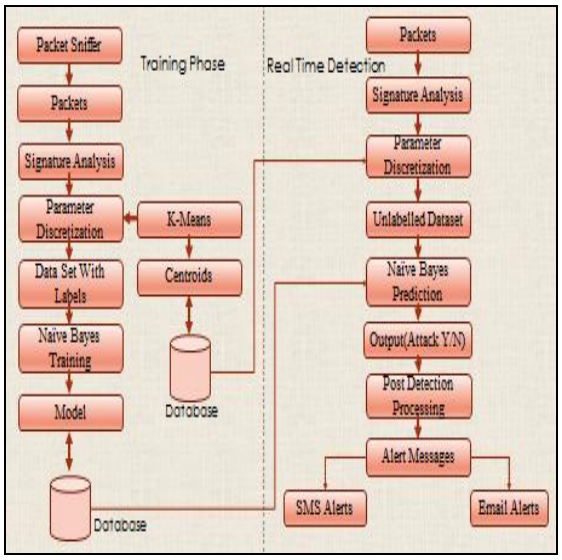
Calculate Final Probability

To find the class, Ci, that maximizes P(X|Ci)P(Ci), we compute

P(X|buys_computer=yes)
P(buys_computer=yes)
= 0.044 x 0.643 = 0.028
P(X|buys_computer=no) P(buys_computer=no)
= 0.019 x 0.357 = 0.007

Therefore, the Naive Bayesian classifier predicts buys_computer = yes for tuple X

1. Experimentation Setup and Results Analysis



Ошибка! Источник ссылки не найден.

Above figure shows detail flow diagram of experimentation setup. Above diagram contain two phases:

i. Training Phase

In this phase we can train the dataset. In the training phase the system constructs a model using the training data to give maximum generalization accuracy. In training phase, we can give input as well as output.

ii. Real Time Detection Phase

This is the real time detection phase. In this phase we can give input then output is generated automatically using naive Bayes algorithm.

Finally, Figure 1 shows the IDS using Naive Bayes consists of incoming packets and detected packets. Incoming packets are the real time packets and detected packets are the intrusions detected from real time packets (data). Table 2 shows the experimental results of detecting attacks and the normal behavior of packets.

Table 2. Experimental results.

Types	Result in (%)
SYN flood	93.46
TCP data flood	92
UDP flood	90.68
Normal	96

2. Conclusion

This paper gives the solution for the threat of cyber intrusion is increasingly high and harmful. Intrusion detection system provide ability to identify security breach in a system. In this paper we use the data mining algorithms. In this paper we create a real time intrusion detection system, which will actively detect intrusion while the machine is running. After running through our tool we were able to successfully detect the attacks. We will also like to create some prevention techniques for detected attacks.

Acknowledgements

We are thankful to Asst. professor. V. D. Thombre & Asst. professor Bhagyashree Patle for providing various resources such as laboratory with all needed software platforms, continues Internet connection, for our project.

References

[1] Mrutyunjaya Panda and Manas Ranjan Patra, "NETWORK INTRUSION DETECTION USING NAÏVE BAYES", IJCSNS International Journal of Computer Science Technology and Network Security, VOL.7 No.12, December 2007.

[2] Duanyang Zhao, Qingxiang Xu,"Analysis as well as Designs for Intrusion Detection System Based on Data Mining" ,2010 Second Inerntional Workshop on Education Technology and Computer Science.

[3] Jonathon Ng, Deepti Joshi,"Applying Data Mining Techniques to Intrusion Detection", 2015 12th International Conference on Information Technology.

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

[5] Manish Kumar, Dr. T. V. Suresh Kumar, Dr. M. Hanumanthappa," Intrusion Detection System Using Decision Tree Algorithm" 978-1-4673-2101-3/12/$31.00 ©2012 IEEE

[6] S.J.Russel, and Norvig, "Artificial Intelligence: A modern approach (International edition), Pearson US imports & PHIPES, Nov 2002.

[7] https://en.wikipedia.org/wiki/Clustering

[8] Mrs. G. Subbalakshmi, Mr. K. Ramesh and Mr. M. ChinnaRao, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), ISSN : 0976-5166, vol. 2 no. 2 Apr-May 2011.