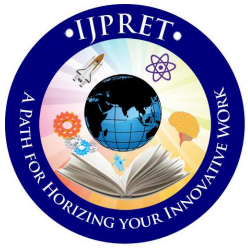
Research Article *ISSN: 2319-507X*
Rahul Dhulekar, IJPRET, 2013; Volume 1(8): 438-446 *IJPRET*



INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

INTRUSION DETECTION USING CONDITIONAL RANDOM FIELD AND LAYERED APPROACH

RAHUL R. DHULEKAR, PORNIMA NIRANJANE

1. ME 1st Year (Computer Science and Engineering), Babasaheb Naik College of Engineering, Pusad, District-Yavatmal, Maharashtra, India.
2. M. Tech. (Computer Science and Engineering), Asst. Professor (Dept. of CSE & IT), Babasaheb Naik College of Engineering, Pusad, District-Yavatmal, Maharashtra, India.

Accepted Date:

27/02/2013

Publish Date:

01/04/2013

Keywords

Intrusion detection system,
Layered approach,
Conditional random field,
Network security

Corresponding Author

Mr. Rahul R. Dhulekar

Abstract

Intrusion detection system is a collection of methods & various ways to detect malicious activities at the network as well as at the host level. Basic categories of intrusion detection systems are signature based detection and anomaly based detection. In this paper we aim at addressing the issues of accuracy and efficiency. This can be done using Conditional Random Field and Layered approach. We have also aimed at detection accuracy for User to Root (U2R) attacks and Remote to Local (R2L) attacks. We have overcome the shortcomings found in Data Clustering methods, Naïve Bayes, Hidden Markov models and neural networks. The Conditional Random Field demonstrates high attack detection accuracy while efficiency is demonstrated by Layered approach.

INTRODUCTION

The biggest issue for all the networks in today's enterprise environment is security. In order to secure the network infrastructure and communication various methods have been developed and an attempt to develop more methods is still going on. One such technique is intrusion detection. In this method we collect information from known types of attacks and use this information to check if any user is trying to attack our network or host computer. Thus, we can say that intrusion detection systems are a security mechanism which can detect, prevent & even at times repair attacks. Thus, we can define intrusion system as a combination of hardware or software to detect intruder activity.

Intrusion detection system can be classified in 2 ways-depending on how they are deployed and the data they use for analysis. Based on the mode of deployment they can be classified as host-based, network-based or application – based. Depending on the attack detection method they can be classified as signature-based or behavior-based.

1. RELATED WORK

There are number of methods that have been developed to detect intrusions or any kind of malicious activities. These include Naïve Bayes classifier, clustering, association rules, neural networks, genetic algorithms etc. Association rules are used in data-mining approaches which build classifiers by discovering relevant patterns of program and user behavior. But mining of features has a limitation to the entry level of packets. Data clustering methods such as fuzzy c-means are also deployed for this purpose. But the limitation here is that as the numeric distance between observations must be calculated and it is not always possible to represent observations in numeric form. Naive Bayes classifiers assume independence between the features in an observation due to which it is not frequently used. Another approach called neural networks require large amount of data for training which constraints its use. In order to overcome these shortcomings we have proposed an intrusion detection system using conditional random field and layered approach. CRF's are used for building robust

intrusion detection systems. CRF's are help to reduce the number of false alarm rates. By deploying layered approach we have achieved high accuracy.

2. PROPOSED SYSTEM

- **CRF (Conditional Random Field)**

Conditional models are probabilistic systems that are used to model the conditional distribution over a set of random variables. It provides the better framework as they do not make any unwarranted assumption on the observation and can be used to model rich overlapping feature among the visible observations.

CRF are discriminative models and directly model the conditions distribution, which is the distribution of interest for the task of classification and sequence labeling. These features are analyzed together they can provide meaningful information, which can be helpful for the classification of task. CRF's do not consider one feature to be independent and hence so perform better. The data sheet used in our experiments represent features of every session in relational form with only one label for me entire record.

In CRF, we represent the data in the form of a sequence and assign a label to every feature in the sequence.CRF increases the complexity and also increase the attack detection accuracy.

A. Important aim of CRF- To model the relationships among features of individual connections using a CRF.

B. Advantages of CRF

1. Higher attack detection accuracy.
2. Every element in the sequence is labeled such that the probability of the entire labeling is maximized.

- **Layered Approach**

Layered approach represents a sequential Layered approach and it is based on ensuring:-

1. Availability
2. Confidentiality
3. Integrity of data and
4. Services over network

Goal of Layered approach:-

It reduces the computation and overall time required to detect anomalous events. Every layer in Layered approach framework is trained separately and then deployed it

sequentially. Here basically layers act as filters that block any anomalous connection, thereby eliminate the need of further processing at subsequent layers enabling the quick response to intrusion. Aim of Layered approach:-it can be implemented to improve the overall system efficiency. We select the features for each layer based upon the type of attacks that the layer is trained to detect.

- **Integrating Layered Approach With Conditional Random Field**

For our Intrusion detection system we integrate the Layered approach with CRF to build a single system that is accurate in detecting attacks and efficient in operation [3] [4].

- **Attacks**

An attack can be defined as any method, process or means to illegally or maliciously aim to compromise network security [1].

A. Probe Attacks

These attacks are also called as scan attacks as they automatically scan a network of computers or a DNS server with the aim of finding the active ports, valid IP addresses, operating system types etc. It can also be considered as an attempt to gain access to a computer and its files by tracing its weak points in the computer system. Features such as “duration of connection”, ”source bytes”, ”protocol type”, ”service” etc are important to detect probe attacks.

B. DoS Attacks

A Denial-Of-Service can be considered as an attempt to make network or computer resource unavailable to its users. It basically consists of efforts of a single individual or a group of people to make some service unavailable temporarily.DOS attacks are usually implemented by enforcing the target computers to reset or by consuming all its resources so that it can no longer provide the intended services to the users. Features such as “src_bytes”, “protocol type”,” dst_host_same_srv_rate” etc are important to detect DOS attacks.

C. User-to-Root (U2R) Attacks

In this type of attacks a local user on a machine is able to obtain privileges normally reserved for the UNIX super user or the Windows NT administrator.

There are various types of User-To-Root attacks. Buffer overflow is the most common one in this category. Features such as "number of file creations", "number of shell prompt attempts" etc are focused to detect User-To-Root attacks.

D. Remote-to-Local (R2L) Attacks

In this an attacker does not have an account on the victim machine gains local access to the machine, furtively escapes files from the machine or even modifies data in transit to the machine. Features such as "duration of connection" , "service requested ","number of failed login attempts" etc are used for detecting R2L attacks.

- **Algorithm**

Considering the network behavior and the increasing development in the attack fashion makes it necessary to develop a fast intrusion detection algorithm having fast detection rates and low false alarm rates. Thus we implement Adaboost algorithm to the Intrusion Detection System [5].

Experimental results show that the Adaboost algorithm has low computational complexity and error rates compared to the existing algorithms used for Intrusion Detection System. The reasons for applying Adaboost algorithm are as follows:

1) It is the one of most important machine learning algorithm, easy at implementation. It has been implemented in many pattern recognition problems such as face recognition. But this algorithm hasn't been implemented so far in the field of Intrusion Detection.

2) It recognizes the misclassification made by weak classifiers in recognizing the type of input.

3) Adaboost applied with weak classifiers is faster compared to when applied with strong classifiers.

The framework of Adaboost algorithm consists of four main parts namely:

1. Feature extraction

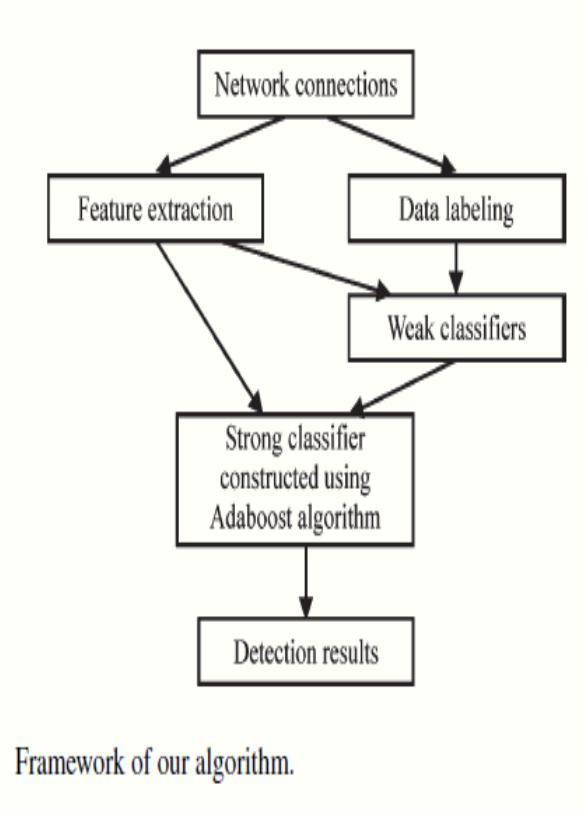
2. Data Labeling

3. Weak Classifier

4. Strong Classifier

A. Feature Extraction

For network connections there are present a number of features for various types of attacks. We aim at categorizing the important features of the existing 41 features using



Framework of our algorithm.

the feature extraction mechanism of the four major types of Intrusions (Probe Attack, Dos Attack, User to Root and Remote to Local).

B. Data Labeling

The layer performs the objective of labeling the incoming data as +1 when it is found normal or -1 when it is found malicious. This layer categorizes the data as normal or attack.

C. Weak Classifier

In Adaboost algorithm weak classifiers are nothing but decision stumps. Its structure is analogous to that of a decision tree with a root node and two child nodes. For each feature data a decision stump is constructed. The decision stump contains only one comparison operation for testing the sample data. Thus the testing time for decision stump is relatively low.

D. Strong Classifier

Strong classifier is constructed using a number of weak classifiers. These are trained using sample data for categorizing the sample data into normal or attack.

Let H = {h f } is the set of constructed weak Classifiers.

Let the set of sample data be {(x1,y1),…,(x i, y i) ,…,(x n, y n)

Where x i is ith feature vector, y I (+1,-1) is the label

of i^{th} feature vector denoting $\in$ whether the feature vector is normal or not and n is the size of the input data set. Let {w …… w n} be the set of sample weights.

Step 1) Training

Initialize weights Wi (1) (i=1,……, n)

Satisfies $\sum_{i=1}^{n} Wi(1) = 1$

In Adaboost algorithm initial weights of the sample data have a strong influence on the mean classification errors. This is found unfavorable for intrusion detection because it is necessary to reduce false alarm rates than the mean errors.

Step 2) Let $_j$ be the sum of the weighted classification errors of weak classifiers hj [y i h i (xi)]

Where I[j] = $\sum_{i=c}^{n}$ Wi (8) 1

$\begin{cases} 1 & r - true \\ 0 & r - false \end{cases}$

When 'True' indicates that the sample data is falsely classified giving rise to false alarm and when False indicates that the sample data is correctly classified, hence avoiding false alarm.

Step 3) Testing

Every decision stump has only on comparison operation for the purpose of testing thus leading to reducing the testing time and improvising the efficiency of the algorithm [1] [4].

- **Computational Complexity**

The computational complexity is derived from the construction of decision stumps and strong classifiers. For every decision stump, all sample data should be searched for each feature. Thus the complexity for the construction of decision stump is (nM) where n is the total number of samples and M is the total number of features. There are in total T iterations for the construction of strong classifier.

Thus the computational complexity is (nTM) there is present only one comparison operation in each decision stump for testing a sample data. Hence leading the test time to be low and thus improving the efficiency of the algorithm and reducing the computational complexity.

4. CONCLUSION AND FUTURE WORK

By implementing intrusion detection system using conditional random field & layered approach we have achieved accuracy & efficiency. Attack detection rate are improved by implementing conditional random field. The time required to train & test the model can be reduced by using layered approach. Most of intrusion detection systems fail to detect U2R and R2L attacks, whereas our IDS can effectively detect such attacks. The biggest advantage of our IDS is that the number of layers can be increased or decreased dynamically. This in turn provides great flexibility to the network administrators. In future our method can be used for extracting features that help in the development of signature for signature-based systems.

- **Comparison of results**

When compared our IDS with other systems, we make the following conclusions. The main drawback in data clustering method is that it requires calculating the numeric distance between observations. This has a prerequisite that the observations must be numeric which is not always possible. If we consider Naïve Bayes [2] then it assumes strict independence between the features. This in turn reduces the attack detection accuracy .Hidden Markov model are unable to model long range dependencies, which thus makes it unsuitable for use.

Decision trees are also extensively used for intrusion detection because of their high speed of operation. They also possess high attack detection accuracy. Neural networks require a high amount of data for training which becomes a major drawback.

5. ACKNOWLEDGMENT

The authors would like to thank the anonymous viewers for their valuable remarks and comments.

6. REFERENCES

1. http://www.symantec.com/connect/articles/ preventing-and-detecting-insider-attacks-using-ids

2. http://danielowen.com/NIDS

3. www.giac.org/paper/g sec/3323/ids-features- ips/105477

4. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.7500&rep=rep1&type=pdf

5. R. Bace and P. Mell, Intrusion Detection Systems, Computer Security Division, Information Technology Laboratory, Nat'l Inst. of Standards and Technology, 2001.