INTRUSION DETECTION SYSTEM USING ENHANCED ADABOOST ALGORITHM

N.Nandhakumar[1], *S.Vasanthi*[2]

ABSTRACT

Intrusion detection system is the process of detecting computers or networks for unauthorized activities or file modification. It can monitor network traffic and detecting the network attacks such as Denial of service, Probe, R2L, U2R Attacks. More specifically, Intrusion detection system tools to detect specific computer attacks and unauthorized use of system, and to alert the proper users usually Network Administrators about detection of attacks in system. These techniques help in them determining what qualifies as an intrusion (attacks) versus normal. The Intrusion detection can be done through a Cascading classifier approach which is used to increase the detection rate of the rare network Classification Algorithm. At First, the Training set network data are sent to both algorithm for learning purpose. Bayesian network detects rare attacks in a better way. So those data records which are classified as a "NORMAL" by J48 Model network attack categories like R2L, U2R Attacks. Here J48 Classification algorithm was cascaded with Bayesian are sent to Bayesian Network Model and Classified. This result in a better detection rate of rare network attacks. This paper explains the splitting of KDD training sets into two sets and we use Particle Swarm Optimization for selecting important features from two training sets for learning and then Enhanced Adaboost (Ensemble) classifier was cascaded with Bayesian Network Classifier. This algorithm repeatedly calls weak learner (classifier) to increase classification accuracy rate. This result in increasing the detection rate of both rare and non-rare network attack categories.

Keywords - Intrusion Detection, Ensemble Classifier, Adaboost, R2L, DoS, PSO and etc...

I. INTRODUCTION

With increased usage of computers and access of internet, net work attack of device or computer also increased. An intrusion detection system (IDS) is a software application or device which monitors network attack or system activities for malicious processing or policy violations and produces reports to a management.

Network security is a protection of a computer networks and its services from unauthorized modification, destruction, or disclosure by the unauthorized persons called "INTRUDERS".

Protecting computer networks requires proper maintenance and implementation of many different types of network security. Hackers and dissatisfied employees are not the only threats to our network systems, data and device. Inappropriate physical access and Lack of security

[1]M.Phil Scholar, Dept. of Computer Science, PSGCAS, Coimbatore. E-mail : nandhamphil@gmail.com
[2]Asst. Professor, Dept. of Computer Science, PSGCAS, Coimbatore E-mail : sn_vasanthi@yahoo.com

awareness to systems increase the risks on data and devices. Intrusion detection is an effective and efficient security mechanism to detect and or protect the computer network from various attacks in the Network[1].

Intrusion Detection System

Arun Hodigere (2001) suggested intrusion means attempting to break into or misuse the system. The Intruder may be from outside the network (unauthorized users) or legitimate users of the network (authorized users). It can be physical, system or remote intrusion. There are different ways to intrude such as unexpected combinations, buffer overflows and unhandled input and race conditions. Intrusion Detection is defined as the act of detecting actions that detect to compromise the confidentiality, availability or integrity of a resource (generally server resource). It provides security management for computers and networks. It analyzes and gathers information from various areas within a system or network to identify intrusions and misuse.

An Intrusion Detection System (IDS) detects an intruder that is unexpected, unauthorized or unwanted people or programs on the computer network. The goal of intrusion detection is to prevent and detect unauthorized use of services. In order to achieve this, the intrusion Detection System detects all incoming and outgoing data (traffic).[1]

Classification of IDS

The intrusion detection system classification is Shown in figure1. It gives better understanding of their methods and limitations. The methods are analyzed based on the kind of input information.

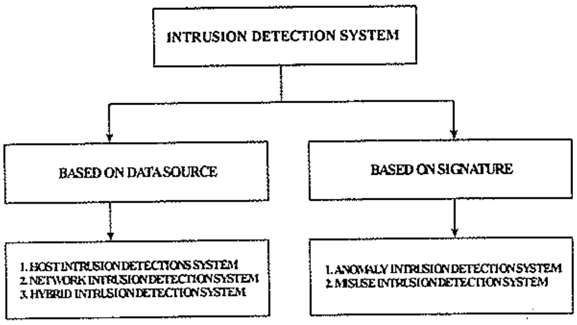

INTRUSION DETECTION SYSTEM
BASED ON DATASOURCE
BASED ON SIGNATURE
1. HOST INTRUSION DETECTIONS SYSTEM
2. NETWORK INTRUSION DETECTION SYSTEM
3. HYBRID INTRUSION DETECTION SYSTEM
1. ANOMALY INTRUSION DETECTION SYSTEM
2. MISUSE INTRUSION DETECTION SYSTEM

Figure 1 : Classification diagram of Intrusion Detection System (IDS)

Anomaly- Intrusion Detection

The anomaly intrusion detection refers to intrusions that can be detected based on behavioral change of user. Anomaly detection may be further divided into static detection and dynamic anomaly detection.[11][9]

Misuse -Intrusion Detection

Misuse or Signature intrusion detection refers to intrusion that follows well-defined patterns of attacks that exploit weaknesses in application software and system.[2]

Host- Intrusion Detection System

Host intrusion detection systems runs on individual host or devices on the network (Each system in a network is a Host). A HIDS monitors the incoming and outgoing data only from our device and will alert the user or administrator, if any threats happen in the system (host).

Network- Intrusion Detection System

Network intrusion detection is an Attempt to gain unauthorized access to network resources. NIDS systems are placed at a strategic point within the network to monitor system traffic and from all devices in the network. A NIDS monitors all incoming and outgoing packets.[8]

Hybrid- Intrusion Detection System

This detection system combines multiple levels different IDS technologies in together a single system so it can theoretically produce a much stronger IDS.[5][6]

II. Data Set Description

Since 1999, KDD Cup 99 has been the most widely used data set in the evaluation of anomaly detection method. This data set is prepared by Stollfo et al. and is built based on the data captured in DARPA 98 IDS evaluation program. The intrusion attacks fall in one of the following four categories:

1) *Denial of Service Attack (*DoS)

2) *User to Root Attack* (U2R)

3) *Remote to Local Attack* (R2L)

4) *Probing Attack*

KDD „99 features can be grouped into three groups:

1) Basic features: Basic features category encapsulates all the attributes which can be extracted from a TCP/IP connection.

2) Traffic features: Traffic features group includes features that are computed with respect to a window interval and is divided into two groups:

a) "Same host" features:

b) "Same service" features:

3) Content features: this is unlike most of the Probing and DoS attacks, the U2R and R2L attacks don t have any intrusion frequent sequential patterns. *Content features* are because the probing attacks and DoS involve many connections to some host(s) in a short period of time. This includes number of failed login attempts. [7]

In the KDD training set, total of 26,198 records are present. Out of this records Normal and Dos contains 12,400 and 8,250 records respectively. And Probe category contains only 2,600 record and R2L and U2R category contains only 215 and 22 records respectively. So this result in Normal and Dos dominates R2L and U2R attack categories. This makes the poor detection rate of this R2Land U2R attacks. So we need to minimize this dominant effect of non-rare categories over rare category attacks. Table shows distribution of 5 categories in KDD training set.[4]

Table 1 : FULL TRAINING SET

Attack Types	Number of Records
Normal	12,400
Dos	8,250
Probe	2,600
R2L	215
.U2R	22

So we need to train rare attacks separately with one classifier and non-rare attacks separately with other classifier. Then these two classifiers that are trained separately evaluated using KDD test set. So we separate KDD training set into two data training sets. First set contains 11,800 Normal records and all Dos and Probe records. The Training set2 contains both rare attacks i.e. R2L and U2R records and 400 randomly selected Normal records. The reason for picking 400 Normal records in the training set2 was classifier2 needs Normal records for training, because we sent those records that are detected as Normal by the first classifier to second classifier when testing the cascaded IDS model.

Table training set-1and set-2 shows the two training sets which contains full of non-rare attacks.[11][12][13]

Table 2 : FEATURES SELECTED BY PSO FOR TRAINING SET1

S.NO	FEATURES SELECTED
1	Protocol type
2	Flag
3	Src_bytes
4	Dst_bytes
5	Wrong_fragment
6	Urgent
7	Hot
8	Logged_in
9	Root_shell
10	Is_host_login
11	Srv_count
12	Serror_rate
13	Same_srv_rate
14	Diff_srv_rate
15	Dst_host_rate
16	Dst_host_diff_srv_rate
17	dest_host_same_src_port_rate
18	dst_host_serror_rate

Table 3 : FEATURES SELECTED BY PSO FOR TRAINING SET 2

S.NO	FEATURES SELECTED
1	Protocol_type
2	Service
3	Flag
4	Dst_bytes
5	Urgent
6	Logged_in
7	Num_compromised
8	Root_shell
9	Num_outbound_cmds
10	Is_host_login
11	Count
12	Diff_srve_rate
13	Srv_diff_host_rate
14	Dst_host_count
15	Dst_host_diff_srv_rate
16	Dst_host_samesrc_port_rate
17	Dst_host_rerror_rate
18	Dst_host_srve_error_rate

III. Proposed Work of Cascaded Classification For IDS

Data mining ("Knowledge Discovery in Databases"(KDD)), is the process of employing one or more machine learning techniques to analyze and discover patterns in large data sets. Data mining is one of the popular techniques for detecting intrusions. It is helpful to detecting new vulnerabilities and intrusions, type of attacker behaviors and provides decision support for intrusion.

The anomaly based intrusion detection system improves the accuracy by making necessary changes in the classification process. In the existing system, Cascaded classifier approach was used. J48 and Bayesian network classification was cascaded and detection rate of rare network category attacks was increased. But detection rate of rare attacks was low, for example R2L detection rate was just 45% and U2R attacks detection rate was 17.5%. So, in the proposed system Enhanced Adaboost ensemble classification algorithm was cascaded with Bayesian Network classification in order to further increase the detection rate of rare network attacks and also non-rare network attack Probe. Also to improve the detection rate further, we use Particle swarm optimization for selecting the important features from the training dataset.[11]

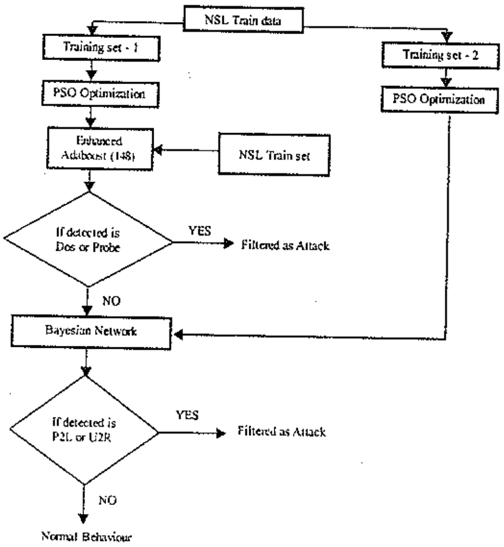


Figure 2 : PROPOSED WORK

J48 DECISION TREE CLASSIFICATION

J48 is a Weka implementation of a C4.5 decision tree classification algorithm. ID3 decision tree, where the main advantage of C4.5 over ID3 was it handles both continuous and discrete data in a dataset. And also it handles missing values in a dataset very effectively then ID3.

In the case of IDS, C4.5 generally classifies non rare attacks better than any other classifiers. That is its classification accuracy of dos and probe attacks was generally high. But its detection rates for rare attacks like R2L and U2R are generally not good up to the expected level. So C4.5 algorithm was used as first classifier for the classification of datasets where non rare attacks classification accuracy was nearly 100%. Those records which are classified as normal by C4.5 classifier was once again sent for classification to the next classifier in the cascaded approach.

ENSEMBLE CLASSIFIERS

Ensemble classifiers are the classification algorithms, which were not a new technique, but it iteratively calls the weak classifiers like ID3, J48 or Bayesian classification until classification accuracy was 100%.These Ensemble classifiers are based on the weighting scheme i.e. assigning weights to each record (instances) in the data set. Basically there are two types of ensemble classifiers. They are Boosting and Bagging. Boosting was based on prediction by majority voting for a particular class. Boosting provides better detection rate than Bagging. AdaBoost i.e. Adaptive Boosting was one of the popular Boosting classification approach used for better classification and so good detection rate in the case of IDS.[10][16].

CASCADED CLASSIFIERS

Cascaded classifiers are the new technique in that instead of using single classification scheme for classification, two classifiers was used. The main advantage of using this cascaded classifier approach was to increase the detection rate of attacks (classification accuracy). This approach mainly used to increase the detection rate of rare network attack categories. There are some classification algorithm which classifies non rare attacks better and some other classification algorithm like Bayesian network which classifies rare attacks better compared to other classification algorithm. Here the dichotomization of the training data set was done. In this step rare and non rare attacks are separated and model was trained by two data sets separately. This results in the reduction of dominant effects of non rare attack categories to rare attack categories. So if two classifiers are cascaded overall detection rate of network attacks was better compared to use of the single classification technique. Especially the classification accuracy of rare attack categories was increased compared to both single and multiple classifiers.[10][14]

BAYESIAN NETWORK

Bayesian Network is a classification algorithm which was based on the probabilistic model. Bayesian network is a graphical model which allows representation of dependencies among subset of attributes. It is also known as belief networks in probabilistic network. These networks specify joint conditional probability distributions. It provides casual relationship in which learning can be performed.

Naive baye s classifier assumes that the effect of given class attribute value is independent of other attributes values. This assumption is known as class conditional independents. This simplifies computation. When this was true Naïve bayes was more accurate model than all the other classifiers. But in reality dependencies between attributes exist in a dataset. So naïve baiyes fails when dependencies exist.[3]

Advantages of Bayesian networks include explicit uncertainty characterization, efficient and fast computation, and quick train. But Bayesian network provides better classification accuracy in classifying rare network attacks like U2R and R2L attacks. So those data records which are classified as normal by C4.5 (First classifier) are allowed once again to the Bayesian classifier and it resulted in reduction of misclassification as normal for rare attack categories.[14][16]

IV. Performance Comparison of Detection Rate With Other Approaches

The Main performance measure for evaluating the intrusion detection system model was detection rate.ie how accurately IDS (classifier) classifies the test data set. Here we take Cascading of Enhanced Adaboost with J48 Classifier as a base learner and Bayesian network model for classification. There are 19 attributes (including class label attribute). First KDD test data set (22,544 records) are sent to Enhanced Adaboost with J48 as base learner model and only those records that are predicted as "Normal" by Enhanced Adaboost Model was sent to Bayesian Network Model and the overall detection rate by this cascading model was only 85.9163% (17340 records was correctly classified).

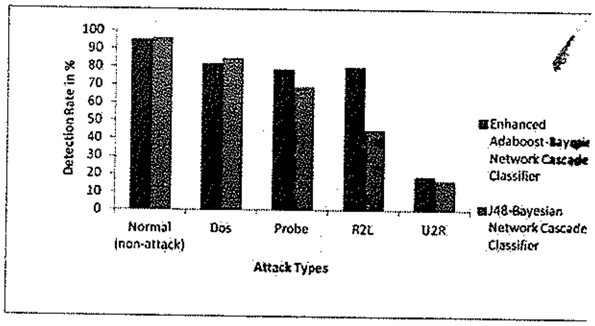


Figure 3 : PERFORMANCE COMPARISON

Table 4 : Comparative Analysis of Enhanced Adaboost-Bayesian Network cascaded classifier with J48-Bayesian Network cascaded classifier

Attack Category	Using Enhanced Adaboost & Bayesian Network Cascaded Classifier	J48-Bayesian Network Cascaded Classifier
Normal(non-attack)	95	96
Dos	82	81
Probe	79	69
R2L	80	49
U2R	19	17

As a result, metrics are computed in dataset in the validating phase and the obtained good result for all attacks and normal data.

V. Conclusion

The work proposes the improving detection rate of a rare network attack categories. In the existing system, the J48 and Bayesian Network classifications are cascaded. The accuracy is not achieved up to the expectation, especially in the case of R2L attacks (only 45%) and true positive rate is increased and false positive rate is reduced in the system, compared to use of normal single classifier approach. In proposed system, Particle Swarm Optimization was used for selecting only important features needed for better learning and then Enhanced AdaBoost Ensemble classification algorithm was cascaded with Bayesian Network is proposed to improve the detection rate of both rare and non rare network attack categories and reduce the false positive rate of the system. This results in improvement of the classification accuracy i.e. detection rate was improved for both rare and non-rare attacks.

References

[1] Hodigere A.,et al (2001), *"Intrusion Detection System "*, shiraz university.

[2] Topallar M, Depren O, Ciliz MK, Anarim E, (2005) *"An intelligent intrusion detection system for misuse detection and anomaly in computer networks Expert System Appl 29(4):713–722.*

[3] Dewan Md. Farid, Mohammad Zahidur Rahman ,Chowdhury Mofizur Rahman (2010) *"An efficient Intrusion Detection based on Boosting and Naïve Bayesian Classifier".*

[4] Dokas D., Ertoz.L, Lazarevic.A Srivastava.J, and Tan.N (2002), *"Data Mining for Network Intrusion Detection,"* Proc. By NSF Workshop Next Generation Data Mining (NGDM 02), pp. 21-3

[5] Dasgupta.D and Gonzalez.F," Anomaly Detection using real-valued negative selection *"Genetic Programming and Evolvable Machines(2003), Volume 4, number 4, pp. 383 - 403.*

[6] Kristopher Kendall, *"A Database of Computer Attacks for the Evaluation of Intrusion Detection Systems (1999),"* Proceedings Darpa Information Survivability Conference And Exposition (Discex).

[7] Lee.L and Stolfo.S (1998), *"Data Mining Approaches for IntrusionDetection, Seventh USENIX Security Symp (Security 98),"* pp. 79-94.

[8] Marsland.S (2003), *"Novelty detection in learning systems ",* Neural computing surveys, volume 3 , pp.157-195.

[9] Shon T, Moon J *"A hybrid machine learning approach to network anomaly detection"* (2007).

[10] Zhenwei Yu, Jeffrey J.P. Tsai *"An efficient intrusion detection system using a boosting-based learning algorithm"* (2009).

[11] Journal of Software vol.6, No.12 *"An efficient hybrid clustering-PSO for Anomaly intrusion detection"* by . Zheng H (2011)

[12] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html,(2010), KDDCup 1999 Intrusion Detection Data , website.

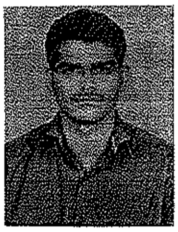
[13] http://iscx.ca/KDD/(2010), "KDD 99 NETWORK ATTACKS DATASETS " website.

[14] P.Rajesh & P.Natesan *"Cascaded classifier approach based on Adaboost to increase detection rate of rare network attack categories IEEE.ORG*

[15] http://www.swarmintelligence.org/ Particle Swarm Optimization for feature selection website http://en.wikipedia.org/wiki/AdaBoost website

Author's Biography

S.Vasanthi, has completed Master of Computer Science (MSc) from Hindustan College of arts and Science and Master of Philosophy(M.Phil) in PSG College of arts and science .Presently she is working as Assistance Professor at Computer Science Department at PSG College of Arts and Science . She has total 06 years of teaching experience in UG course. Her area of interests include Dataminig , and Networking.



N.Nandhakumar, has completed Master of Computer Science (MSc) Barathiyar University and Pursuing Master of Philosophy(M.Phil) in PSG College of arts and science(Under guidance of S. Vasanthi Asst. Professor Department of computer science Psgcas.) His area of interests include Dataminig, and Networking.