Journal of Theoretical and Applied Information Technology
15th September 2017. Vol.95. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EFFECTIVE APPROACH FOR INTRUSION DETECTION USING KSVM AND R

[1]M. NAGA SURYA LAKSHMI, [2]DR Y RADHIKA

[1]Research Scholar, Dept. of CSE, GIT, GITAM University, Visakhapatnam, Andhra Pradesh, India
[2]Professor, Dept. of CSE, GIT, GITAM University, Visakhapatnam, Andhra Pradesh, India
[1]mnslakshmi.gitam@gmail.com, [2]radhika@gitam.edu

ABSTRACT

Nowadays there is an incredible escalation of the usage of computers over various networks and application domains, which in turn increases the security threats in terms of intrusions. An intrusion may happen either internally or externally and the traditional approaches used in intrusion detection are unable to meet the requirements of preventing and detecting an intrusion. For the detection of different attacks, intrusion detection occupied important work for the maintaining of privacy and reliability in network resource. In this paper, the methodologies of Data Mining has been used for increasing the performance in the IDS, and to handle Some of the problems like data Preparation, pre-processing of the data, data classification and Intrusion detection are being solved using different techniques like Dynamic Data Preparation (DDP), Hybrid Rule-based Pre-processing, Efficient Kernel Based Support Vector Machine (EKBSVM) and Decisive Neural net using R (DNR) respectively. The proposed techniques have produced better accuracy, specification and minimized the false alarm rate (FAR).

Keywords: *Anomaly Detection, Classification, DDoS, Intrusion Detection, Misuse Detection*

I. INTRODUCTION

Despite the wide development of data innovation, security has stayed one testing territory for PC and systems. The quantities of hacking and interruption episodes are expanding year on year as innovation takes off. Security danger comes from outer gatecrashers as well as from inner clients as abuse. The firewall will be able to break the system and it can open the framework into the system, and is unable to differentiate between good or bad activity. Consequently, if there is a requirement to permit an opening to a system, then a firewall which is a static rule based, unable to protect from intrusion attempts. In contrast, Intrusion Detection Systems (IDS) can examine the hostile action on these openings. Conversely, Intrusion Detection Systems (IDS) can screen for threatening movement on these openings. The generic aspect of the IDS is represented in figure 1.1.

1.1. The Systems of Intrusion Detection

In the present computing world, the necessary and important elemental of IDS is the network security architecture. In order for the characterization of IDS, here the importance is to know/ it is important to know that they have to be aware of an *intrusion*".

The intrusion can be categorized in terms of *integrity, confidentiality, and accessibility*. An action or event causes a violation of confidentiality of the system. An action or event causes a violation of integrity if it permits shifting the circumstances of locating the resources, in a machine in an unlawful aspect. Likewise, the action or the event may cause a violation of the accessibility Sometimes the real users may be prohibited for the accessing of the services or its resources which are there in a computer. IDS have the options to track what actions are being performed in the system or on the web and observing it and to analyze the cipher of attacks. For monitoring or analyzing the attacks, IDS will act like a software or hardware which automatically processes its events.

Due to the fast escalation of attacks, numerous intrusion detection systems anticipated in research. A few fundamental components are similar to the existing system and the rest vary from the proposed system. Figure 1.1 exhibits the generic design of IDS. Figure1.1 shows some of *the* detected like *misuse and anomaly units,* etc. *Audit Trail Dataset* collects the data to find events and processes the data to convert in the proper format. The *Feature Extraction unit* is the key aspect of IDS. For detecting several intrusive behaviors, an alarm is set to detect.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1.1.1 Different Types of IDS

IDS has categorized as no of classes upon the elements are shown in figure 1.1.

IDS Based on Host and Network: With the help of Audit *Trail Dataset*, It is divided into two types of IDS. *Network-based* IDS (NIDS) gathers the text in the form of packets from the network system that is being monitored. Basically, the NIDS is a *sniffer system*. It is easy to deploy individual OS. They offer improved security against *DDoS* attacks. When network traffic is encrypted, this type of Intrusion Detection system is unable to scan content or protocol and also detection becomes hard on advanced switching networks, as the data packets are not reachable to NIDS. Above the same standard, the second one is the *host based* IDS (HIDS). It gathers the text in the form of operating system log files, utilization of CPU, System Calls, and the network event logs from the host, which is being protected. These systems are unproductive by switching networks or encrypted traffic whereas HIDS are operating system dependent and thus it requires several prior forecasts before functioning. These systems are very capable of detecting attacks like a *buffer overflow*.

Misuse/Anomaly Based IDS: One more standard for Classifying IDS is from processing or detection viewpoint. In the *detection method*, it is divided into 2 types of IDS. Misuse-based can be called as *signature-based*, it preserves a large collection of signatures of known attacks in the database. Ahead of the reception of data from the *dataset*, here the data will be compared with the data in the database. Then an alarm will be triggered, if some match occurs. It is a demanding task in the misuse-based method for indicating the signatures.

This research focuses mainly on this issue whereas the attacks are not capable of detecting zero-day attacks because these attacks are not specified in the database. The good thing in this type of IDS is that the false alarm rate is too small in IDS. The anomaly based IDS is present in this class and it can also be called as behavior-based systems. These systems study the normal behavior instead of loading the known signatures based attacks, and these can be analyzed and observed. Any divergence in the original behavior is measured as suspected. An alarm is set to find attacks. So these works from the hypothesis, that if anomalous behavior or action is different than the usual behavior immediately it can be detected. By description, they are competent of removing *zero-day* violations in the system whereas they undergo many false alarms if they deviated the usual activity and it can be identified that an attack has occurred. Through this hypothesis it is clear that the anomalous behavior or action can be easily detected.

The Feature Extraction console collects the relevant features of each attack from the data base console using an effective pre-processing method using Support Vector Machine technique based on the intrusion detection system used in the machine. When new attacks are observed then the refinement is made on feature dataset. With updated feature training and testing datasets are executed by using efficient classifier techniques for classifying patterns of normal and attacks.

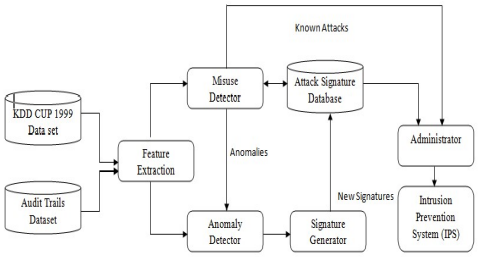

Figure 1.1- Generic view of Intrusion Detection System

1.2. The 1999 Dataset of KDD CUP

The dataset selected in the fifth international Conference on KDD Process of knowledge Discovery and Data Mining tools. The aim of the contest task was to frame an intrusion detector for network security, a foretelling IDS model proficient of differentiating among intrusion or attacks, called as bad connections, and normal connections, called as good connections. This standard database consists of audited data, designed using large range of attacks which have been simulated in the environment of military network. The datasets are obtained from DARPA-98 network data. Every connection in the network is described using 41 features, which provide information regarding BF-Basic Features, CF-Content Features, TTF-Time-based Traffic Features and HTF-Host-based Traffic Features. The attack classification is done by using class label considered as a 42nd feature, and it is used to distinguish the connection as normal or attack (the type of attack). About five million records are used

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for designing the training dataset and more than half million records are used for creating testing dataset. Four categories of attacks are used for both testing and training datasets; they are Denial of Service, Remote-2-Local, User-2-Root and Probe. The Majority of Pattern reorganization and classification techniques tested and trained on KDD IDS datasets are unable to identify major U2R and R2L attacks. These observations are taken to investigate further to identify the limitations and shortcomings of the KDD-99 dataset to dispute that these datasets should not be used in pattern reorganization or classification techniques used for detecting misuse activities of these two U2R and R2L attack categories.

It is analytically observed that, there are dissimilar hypothetical results for U2R and R2L. These techniques are analyzed by cross switching of the roles of both training and testing datasets, and relative and subjective analytical rules are generated separately on testing and training datasets through the decision tree approaches in data mining classification.

The 1999 KDD Dataset is utilized to accept the adequacy of the Hybrid IDS. The originators of interruption discovery dataset mainly depend on the 1998 DARPA activity for to assess of frameworks in distinctive philosophies. The Military system consists of three machines with different frameworks and administrators. For the parody distinctive IP locations are used to produce activity. To record all the movement activities for the TCP dump position we use a sniffer. The re-enacted period for the system is given as seven weeks. And now the attacks in the system are categorized into four types which are as below:

Denial of Service (DoS): The intruder tries to prohibit genuine customers from using network services.

Remote to Local (R2L): The intruder does not record in the machine, which results to get entrance.

Sender to Root (U2R): The intruders have a neighborhood for casualty machines and to increase master client benefits.

Probe: Intruder trying to get data from the objective host. The following parameters categorize the attributes of IDS.

The 41 attack feature set is categorized based on the above-mentioned parameters.

BF-Basic Features- 9
CF-Content Features- 13
TTF-Time-based Traffic Features-9
HTF-Host-based Traffic Features-10

Generally, computers either single or connected through a network are mainly uncovered to probably damaging admittance by hackers in unauthorized manner. In this aspect we require an effective mechanism of intrusion detection. Normally, an Intrusion detection system detects unauthorized administration of computer machines mainly using the internet. The basic components of an Intrusion detection system are sensor unit for generating security attacks, an event monitoring console, alerting mechanism to detect an intrusion and central database console to record logged information of intrusions and uses a standard system for generating rules based on the alerts from security information received. There are a number of ways to categorize an intrusion detection system based on the sensor location and type of attacks and techniques used by the central engine console for the alert generation. The Intrusion detection attacks are stated as follows

Back Attack: It is the DDoS attack across the web server, an intruder sends the request through the URL's containing hidden information. This attack will make server to slow down when the server is trying to process and server is unable to process other requests.

Ftp-write attack: This attack is R2L user attack and advances due to the mis configuration of FTP.

Imap Attack: It causes a buffer overflow in the imap server, which allows attackers to access remotely and to execute random instruction with high level root rights.

Ipsweep attack: It is generated in the network to sweep the host's information, which are on a network and this vital information is useful to an intruder in searching for machines

Land attack: This DDoS attack is well suitable for traditional TCP/IP protocol implemented machines. This can be created by sending the spoofed Packet of SYN with the same source and destination address.

The Load module attack: This U2R attack is generated against Sun 4 Operating systems, which uses Xnews windowing system. An unidentified intruder can gain access to the root level. This attack loads two loadable kernels (dynamical) into the current system which is in running state.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Neptune attack or SYN flood attack: In general, SYN flood attack may occur during the connection establishment process of TCP/IP. The intruder sends IP spoofed packets continuously, which will cause exhaust of memory due to the heavy flooding of SYN spoofed packets.

The Nmap attack: Nmap is a network scan tool, and it supports different types of one flag bits of SYN, FIN and ACK values of both TCP and UDP. It allows the user to specify the port numbers to scan, and waiting time between each port access.

Perl attack: This attack can create a malfunction during the perl implementation

Phf attack: This attack corrupts an unidentified CGI scripts, which are used to execute the http server with high privileged values.

Ping of Death attack: This attack may affect traditional operating systems, and it may cause to respond a system in unpredictable due to overwhelmed packets of IP. It may cause crashing and repeated rebooting of the machine, and freezes the resources.

The Satan attack: This attack is the predecessor of the older SAINT scanning tool

Smurf attack: The intruders will use echo request packets of ICMP protocol and IP packets are broadcasted from remote login locations, which causes DDoS attack.

Teardrop attack: It exhibits a DDoS attack that causes a flaw in the TCP/IP implementation in older operating systems.

The layout of this paper is organized as follows. Section 2 refers to literature review. Section 3 describes the motivation. Section 4 defines the problem statement of the proposed work. Section 5 explains the implementation of the proposed work. Section 6 includes experimental study. Section 7 refers to conclusion, acknowledgement and future enhancement.

2. LITERATURE REVIEW

Intrusion Detection System was primarily proposed by J.Anderson in the year 1980 [1].

W.R. Cheswick, has classified existing firewalls into three types based on the gateways they are application gateway, packet filtering and circuit filtering and these types can be more than one at a time [2].

Both SVM and C4.5 are compared by Ektefa the classifier performance does not suit for real time complex problems. The performance of C4.5 is better compared with other techniques [3].

To improve intrusion detection using unlabeled data, Ching-Hao et al. proposed Co-training framework. The proposed method shown less error rate than existing methods, the proposed method has shown enhanced accuracy [4].

Denning, D.E has proposed Detecting and monitoring mechanism on abnormal patterns of audit data to prevent security violation. The Proposed method uses profiles for behavior representation in terms of statistical models and metrics [5].

To deal with multidimensional dataset, hybrid feature selection is proposed by Sethuramalingam. S. The proposed method has removed an inconsistent and redundant feature that decreases the performance of classification. For selecting significant features of the dataset genetic technique has combined with information gain. The proposed method has shown better accuracy when features are combined [6].

John Mchugh has proposed a mechanism of intrusion detection with the combination of the brute force method which is used to evaluate the intrusions and the proposed method deals with misuse detection based on signature and anomaly detection [7].

Prof. Ujwala Ravale et al. has proposed intrusion detection mechanism using k-means clustering and RBF Kernal functions of SVM used in the classification model design. The proposed system has produced decreased number of attributes related to each data point [8].

Gao Xiang, Wang Min has proposed unsupervised method; it uses a large dataset as training data and has recorded less accuracy. To conquer this problem, a semi supervised approach has been proposed [10].

The combination of J48 and RBF is proposed by Panda, the proposed method classifies data into separate classes like Attack or Normal. Both proposed methods show more error prone and Root Mean Squared Error [11].

Lane T has proposed Markov decision process, which is based on the combination of detecting both anomaly and misuse attack. The Semi supervised method is applied in building the classifiers [12].

Clustering using fuzzy logic has been proposed by Qiang Wang, Vasileios Megalooikonomou, the statistical properties of a cluster and Euclidean distances are used to evaluate the proposed approach [13].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Hybrid PSO technique has been proposed by Holden, which can deal with attributes of type nominal. Using simple rule set the proposed method shows enhanced accuracy [14].

Zhang Fu, Marina Papatriantafilou, Philippas Tsigas has proposed two different methods to reduce Denial of Service Attacks. Primarily, separation of malicious traffic from legitimate traffic and secondly, performance degrading strategies applied on legitimate traffic [17].

SVM and PSO are combined by Ardjani, to optimize the performance of SVM. To measure the accuracy, cross validation of 10-fold is done. The proposed method shows better accuracy with more execution time [19].

Sink tree model has been proposed by Zhang Fu. Marina Papatriantafilou, Philippas Tsigas, Wei Wei, in order to prevent denial of service attacks, DDoS attacks are not only in the target machine they compromises the whole network [23].

Zhang et al. proposed the concept of dividing the network into clusters using proactive techniques. Each packet requires permission to move into other clusters [25].

Chien-Yi Chiu et al. has proposed semi supervised approach in order to reduce the rate of false alarm by developing an alert filter with a high detection rate. In the proposed system, features of both semi supervised and supervised learning approaches are same [26].

Vincenzo Gulisano et al. has proposed aggregation method to monitor high speed traffic by prefixing IP address to detect anomalies related to Distributed denial of service attacks in packet streaming fashion [29].

Acknowledgement based synchronous communication is proposed by Zhang Fu, Marina Papatrianta Filou, Philippas Tsigas using static clock drift. The proposed technique detects direct attack based on the speed of clock drift [30].

Li Jimin, Zhang Wei, KunLun Li has proposed SVM technique using Tri-state training to increase the speed and rate of accuracy [31].

To find clusters without using labeled data, Monowar H. Bhuyan has proposed a tree based approach. The CLUS technique is used to label the dataset using clustering technique. The proposed system has proposed better results for mixed type network data and for the numeric data [32].

In the proposed approach done by Catania Carlos A, Garino Carlos, has shown more time consumption in the preprocessing of network data and in turn it increase the load on the administrator [33].

3. MOTIVATION

The research concentrates on providing solutions to the issues in intrusion detection communities that help administrators in performing pre-processing, classification, labeling of data and to mitigate the outcome of Distributed Denial of Service Attacks. Due to the great enlargement of attacks, which makes the task rigid, attacks can be identified only after it happens. To overcome this situation, recurrent updating of profiles is necessary. The Reduced workload of administrator increases the detection of attacks. Data mining includes many different techniques to accomplish the desired tasks. All of these techniques aim to fit a model for an approved data and even analyzes the data and replicate a model which is neighboring to the data being analyzed.

4. PROBLEM STATEMENT

The main aim of this research is to handle problems like data Preparation, pre-processing of the data, data classification and Intrusion detection are being solved using different techniques like Dynamic Data Preparation (DDP), Hybrid Rule-based Pre-processing, Efficient Kernel Based Support Vector Machine (EKBSVM) and Decisive Neural net using R (DNR) respectively.

For the detection of the problems, this research has been implemented using some of the methods of Data mining. Nowadays the network administrators are mainly using this pattern signature. The reality is that the existing task deal with the problem that it is required for achieving intrusion recognition and not others. To solve the above issues the following solutions were made,

- To solve the Classification problem, an Efficient Kernel Based Support Vector Machine approach is used; it rightly classifies the data without any misclassification.
- To reduce network administrator's workload, Decisive Neural net technique is applied in this research paper

Journal of Theoretical and Applied Information Technology
15th September 2017. Vol.95. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. IMPLEMENTATION

5.1. Architecture:

The Proposed architecture IIDSS (Intelligent Intrusion Detection Security System has the following key modules. They are,

Preparation- Pre-processed data is available readily for the execution, Attack related data is collected from KDD Cup dataset. Information about 23 different attacks is used for intrusion detection.

Pre-processing- Data pre-processing is done using WEKA tool by using RemoveUseless() function to eliminate redundant attributes, out of 41 attributes few attributes can be eliminated for the performance improvement in the system.

Classification-Using KSVM (Kernal Based Support Vector Machine), the type of attack can be classified effectively.

Intrusion detection Module: This module is combined with R and neural net package for better detection of intrusions. The modules deployed as shown in figure 1.2

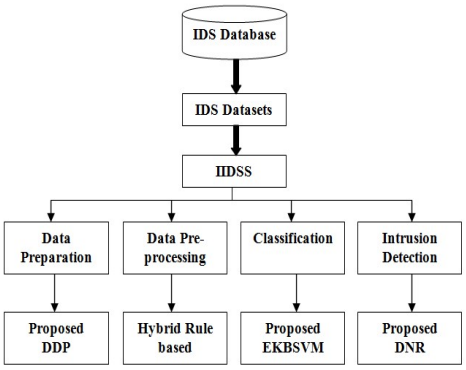

Figure-1.2 Proposed Architecture

5.2. Data Preparation

Data preparation can be done by using proposed DDP (Dynamic Data Preparation) technique, and steps required for the data preparation of KDD cup data is depicted as follows in the given technique.

Dynamic Data Preparation Technique

Input: KDDCUP dataset

Output: Attack wise csv files

Step 1: Create an Array //Array is used to store sample
Training records form the input data
Static ArrayList[],

Step 2: Select sample records for data preparation,
Int noOfattacks, noOfnormal // Number of
Maximum instances of each attack

Step 3: Create file Dataset_Anomaly.csv and Dataset_Misuse.csv // File path for built Datasets

Step 4: Create file with Optimized_Attack type //File
Path for Input files

Step 5: Create String array Attacks // an array that
Stores the types of attacks, *String Attacks[]*

Step 6: Read inputs data records and create a list for
Anomaly, *for i=0 to attacks.length,*
Repeat Step 6

Step 7: Read inputs data records and create a list for
Misuse, *for i=0 to attacks.length,*
Repeat Step 7

Step 8: Generate output files based on the Attributes
List

The Knowledge Data Discovery cup 99 data with the total of 42 attributes are used in this study. The KDD has different types of data versions, In this 10% of training data has been used in the data preparation phase and has a total of 6212 instances. Each instance has been derived with featured attributes of 42. In this paper, only 41 attributes are divided into 4 classes as mentioned below.

- Here Basic (B) is the feature for TCP connections which has individuality.
- Content (C) denotes domain knowledge.
- Traffic (T) are computed features of a two-second time window
- Host (H) features are used to assess the attacks from the last two seconds.

This phase of data preparation is designed and developed to detect an anomaly and misuse based intrusion detection system using neural networks. Prepare datasets to select the number of records randomly for each and every type of attack. Data preprocessing is done by using “removeuseless()” function. Data Preparation steps are as follows

1. In the initial step, KDD cup 10% data has been portioned attack wise and stored each file with a naming convention

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Optimized_"name_of_the_attack".

2. Java program named DataPreparation.java is used to extract records randomly from each of the files named as Optimized_'name_of_attack' created in step1.

3. The CSV files Dataset_Anomaly and Dataset_Misuse are generated after executing the "DataPreparation.java" code, which is later used for classification.

5.3. Data Pre-processing

Hybrid Rule-based Pre-processing approach is proposed to rearrange the information that will be prepared. Disentangling would mean evacuating characteristics that did not bode well. Pre-processing results of Anomaly and Misuse data and attribute information are displayed in fig 1.3.

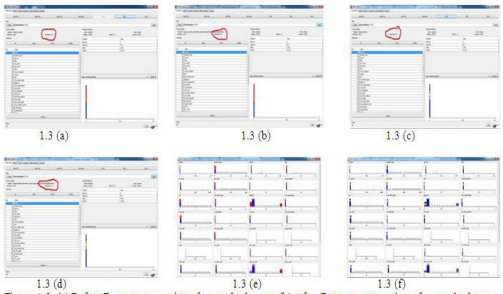

Figure 1.3- (a) Before Data pre-processing of anomaly dataset, (b) After Data pre-processing of anomaly dataset, (c) Before Data pre-processing of misuse dataset, (d) After Data pre-processing of misuse dataset, (e) Attribute feature set of anomaly dataset, (f) Attribute feature set of misuse dataset

The preferred standpoint is that expelling characteristics would decrease the measure of information getting handled, which would build the execution of the neural system. The drawback to this can be if imperative properties are accidently expelled then the precision of distinguishing an interruption will endure. The exit from this is utilizing different cycles of evacuating certain qualities and after that utilizing certain ascribes to make sense of what suits best. Weka's RemoveUseless() function is used. This function evaluates characteristics that as a rule don't differ much. Every single steady characteristic is expelled and traits that surpass the most extreme rate of a change parameter are likewise evacuated. The generated "R" scripts likewise check the nature of a quality in light of the change in property tests. Trait lessening was completed for qualities that don't contribute even 1 percent of the total variety of the information set.

The Hybrid rules of Proposed technique will remove attributes do not change at all, or that change too much. All variable attributes are deleted automatically. The proposed technique has shown better performance in dealing with missing class values and string class values. The performance of the technique is measured by using the percentage of variance value as a parameter. The formula for calculating Variance percentage is

$$P_v = (N_d / N_t * 100)$$

Where N_d is the total number of distinct values and N_t is the total number of values and P_v is the variance percentage. If any attribute value is greater than P_v, then the attribute will be deleted during the pre processing phase.

5.4. Data Classification

Efficient Kernel Based Support Vector Machine (EKBSVM) is proposed for data classification. The Radial Basis function (RBF) kernel is combined with our proposed technique for attacks classification. The proposed kernel function is derived as follows

$$K_{rbf}(p,p') = \exp[-||p-p'||^2 /$$

$$K_{rbf}(p,p') = \exp[-\left|\frac{(p-m)}{s} - \frac{(p'-m)}{s}\right| 2/\gamma]$$

$$Krbf\ (p,p') = \exp\ [-||(p-m)/s||2/\gamma\ \exp\ [-||(p'-m)/s||2/\gamma$$

$$X\ \exp\ [((p-m)/s).((p'-m)/s)/\gamma$$

$$F(p) = \exp\ [-||(p-m)/s||2/\gamma.h\,(p,p')$$

Where kernel spread parameter level, m is the mean and s is the mean square deviation and RBF kernel projection for the maximum values (infinite dimensions) and the form of the proposed function is as follows.

$$K_{rbf}(p,p') =< \varphi(p), \varphi(p') >$$

Where is used as parameter for the projection of vectors p into space.

The improved version of Support vector machine technique is combined with derived kernel function

The improved version of Support vector machine technique is combined with a derived kernel function

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$f(x) = \sum_{x_j \in S} \alpha_j y_j K(x_j, x) + b$$

Where x_j denotes training patterns, $y_j \in \{+1, -$ represents the corresponding class label and S denotes set of vectors used in the proposed technique. The pseudo code for the proposed kernel is as follows.

```
Pseudo code for proposed Kernel function
function (sigma = 1)
{
  rval <- function(x, y = NULL)
   { if (!is(x, "svector"))  stop("x must be a svector")
     if (!is(y, "svector") && !is.null(y)) stop("y must a svector")
     if (is(x, "svector") && is.null(y))
     {
       return(1)
     }
     if (is(x, "svector") && is(y, "svector"))
     {
       if (!length(x) == length(y))
         stop("number of dimension must be the same on both data points")
       return(exp(sigma * (2 * crossproduct(x, y) - crossproduct(x) -
         crossproduct(y))))
     }
  }
  return(new("rbfkernel", .Data = rval, kpar = list(sigma = sigma)))
}
```

The technique for the proposed classification is as follows

EKBSVM Technique

Input: Training Dataset

Output: Classified data

Step 1: Select the training dataset.

Step 2: Proposed Kernel function is deployed

Step 3: The EKBSVM training is executed on training

Data

Step 4: The trained dataset is loaded for testing.

Step 5: The testing data, structured fields are given for

Classification of test data

Step 6: The EKBSVM classifier works based upon the

Proposed training structure

Step 7: The classification results are obtained.

Step 8: The classification result contains the detected

Attacks for the protocols

The proposed EKBSVM is an improved version of the traditional SVM approach which does classification using supervised learning approaches. Proposed technique maps linear vectors into non-linear space. Derived kernel function is used to construct hyper plane space by splitting features space. Semi supervised approach is used in the proposed EKBSVM technique in which prediction is done by setting target attribute values. The proposed technique is carried out in an iterative approach for generating decision function by using training dataset.

The training dataset is combination both target and predictor values. If the proposed technique is able to predict an attack values for the chosen target value, then it is called the function of classification.

5.5. Intrusion Detection

Decisive Neural net and R technique (DNR) ksvm uses SMO technique used to solve the SVM-QP problem and most SVM formulations. For multiclass classification with k classes, k is greater than 2, Ksvm uses the 1-to-1 method, in which k * (k-1)/2. The suitable class can get by voting the method; classification inconvenience can be solved by a single quadratic problem involving all the classes. If the analyst variables contain factors, the formula interface has to be used to get an accurate model matrix.

The plot job for binary classification ksvm objects reflects the outline of the resolution values with the consequent SV featured. The neural net predicts can be a return class for the probability errors of classification by giving the parameter meter type for probability. The crisis of problem selection is moderately given with experimental. By keeping RBF kernel in observation, we can gain values which are of range from 0.1 to 0.9 quintiles of the \|x- x'\| statistics and also with the help of this kernel function, we can estimate the quintiles and we can use median values.

1. Defining of the optimal hyper plane: maximizing the margin
2. For nonlinearly separable problems: For misclassification, we have a penalty term.
3. Mapping data for the high dimensional space to classify with linear decision surfaces: reformulate the problem so that the data is mapped implicitly for this space.

The vectors give about hyper planes that can support of type support vectors. For that, we use SVM to find the maximized margin between both the classes for classification.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Technique for Anomaly

1. Perform read operation using read.csv - function
2. Read the table
3. aRow = nrow (function (x), dim(x) [1L])
4. aCol = ncol (function (x) , dim(x) [2L])
5. Sub=Sampling of records
6. Generate anomalyTrainingSet
7. Generate anomalyTestSet
8. Ksvm

a. SV type: C-svc, parameter: cost C = 1

b. kernel function for Gaussian Radial Basis Hyper parameter: sigma = 0.015

9. Anomaly Prediction
10. Generate Confusion Matrix

Technique for Misuse

1. Perform read operation using read.csv - function
2. Read the table
3. mRow = nrow (function (x) ,dim(x) [1L])
4. mCol = ncol (function (x), dim(x) [2L])
5. Sub=Sampling of records
6. Generate misuseTrainingSet
7. Generate misuseTestSet
8. Ksvm

a. SV type: C-svc (classification) , parameter : cost C = 1

b. Gaussian Radial Basis kernel function. Hyper parameter: sigma = 0.015

9. Misuse Prediction
10. Generate Confusion Matrix

6. RESULTS

6.1. Comparison table

The existing system was developed from the concept of Hybrid PSO and C4.5. In this study, The IDS system is resided in the concepts of support vector machines (SVM) implemented in R. In this work „neural net" package available in R is used for implementing the neural network. The results obtained show acceptable accuracies. The results are shown in Table 2.

Table 2: Obtained Results, compared with existing systems

Techniques	% of Sensitivity	% of Specificity	% of Accuracy	% of FAR
C4.5	86.56	82	93.22	1.55
SVM	83.81	64.28	87.17	3.21
C4.5+ACO	89.25	85.41	95.05	0.86
SVM+ACO	87.41	67.96	90.81	2.41
C4.5+PSO	92.50	88.38	95.36	0.73
SVM+PSO	90.05	70.80	91.56	1.93
EDADT (Efficient Data Adapted Decision Tree)	96.85	92.35	98.11	0.19
IIDSS Technique	99.81	99.90	99.62	0.01

6.2. Graphs- Anomaly, misuse, FAR

The existing system is developed from the concept of Hybrid PSO and C4.5. This approach is developed using support vector machines and R language. This model has used the neural net package available in R for implementing the neural network. The neural net package permits adaptable settings through custom-decision of mistake and actuation work. Moreover, the figuring of summed up weights is executed. This package computes a method for objects of class for neural networks, characteristically created by the neural net and this function computes the output of all neurons for precise uninformed covariate vectors given a trained neural network. The Results of Anomaly and Misuse attacks detection is presented in Figure 1.5 and the Comparison results with the existing system are shown in figure 1.6.

The classification results for anomaly detection are obtained based on the parameters like type of classification "C-svc" with cost C=1 and hyper parameter Sigma=0.015. The total numbers of support vectors generated are 3027 with a training error 0.032846.

The misuse detection results are obtained based on the parameters like classification type "C-svc" with cost C=1 and sigma = 0.015. The total number of support vectors generated is 3140 with a training error 0.107408.

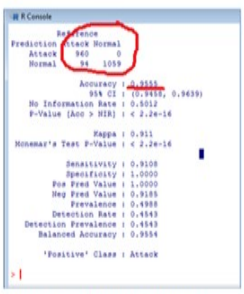

Figure 1.5 (a) – Results of Anomaly Detection

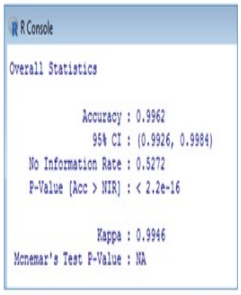

Figure 1.5 (b) – Results of Misuse Detection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

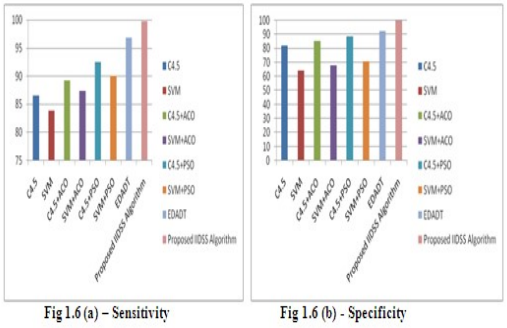

Fig 1.6 (a) – Sensitivity

Fig 1.6 (b) - Specificity

Fig 1.6 (c) – Accuracy

Fig 1.6 (d) – FAR

Figure 1.6: Comparison of proposed model with existing methods (a) Sensitivity, (b) Specificity, (c) Accuracy, (d) FAR

In this research paper, data mining methodologies have been used for intrusion detection. The Proposed method will distinguish the features of Known features and unknown attacks. This work of intrusion detection is carried out using data mining tools with a sample of 6212 records of KDD Cup 1999 dataset to estimate and analyze the effectiveness among the existing traditional methods and our proposed methods. Each and every attack related features are measured and the count of observed results of each attack is depicted as in figure 1.7.

The experimental results are based on the limitations like we have chosen 10% kdd cup labeled data for experimental study.

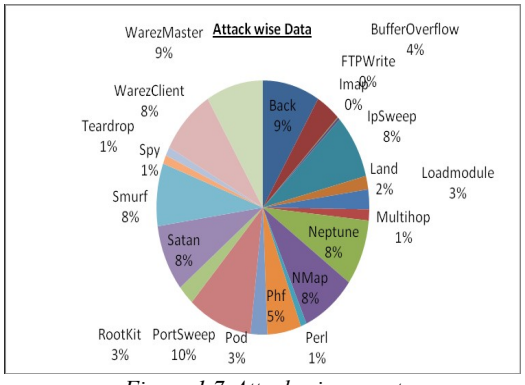

Figure 1.7-Attack wise count

7. CONCLUSION & FUTURE WORK

The EKBSVM technique has shown better results compared with existing methods. The technique has shown extensive efficiency in minimizing the false alarm rate and will reduce the administrator workload. This model has produced 13.24% higher sensitivity when compared with C4.5, 15.99% compared with SVM, 10.55% compared with C4.5+ACO, and 2.95% when compared with EDADT. The experimental result shows better accuracy compared with the existing system. The Future work is aimed to train the IDS to detect number of attacks, and the count can be increased from 23 to 40.

ACKNOWLEDGEMENT

I am thankful to those with whom I have had the delight to work amid this and other related undertakings. Each of the individuals from my Dissertation Committee has given me broad individual and expert direction and showed me an awesome arrangement about both logical research and life when all is said in done.

REFERENCES

[1]. Anderson, James P., "*Computer Security Threat Monitoring and Surveillance*," Washing, PA, James P. Anderson Co., 1980.

[2]. Bellovin, S.M. "Network firewalls", *IEEE Communications Magazine*, Vol. 32, pp. 50-57, 1994.

[3]. Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey. Intrusion detection using data mining techniques. In: *International conference on information retrieval and knowledge management*; 2010. p. 200–4.

[4]. Ching-Hao, Hahn-Ming L, Devi P, Tsuhan C, Si-Yu H. Semi supervised co-training and active learning based approach for multi-view intrusion detection. In: *ACM symposium on applied computing*, no. 9; 2009. p. 2042–7.

[5]. Denning, D.E. "An Intrusion-Detection Model", in *IEEE Transactions on Software Engineering*, Vol.13, No. 2, pp. 222-232, 1987.

[6]. Sethuramalingam S. Hybrid feature selection for network intrusion. *Int. J Computer Science Eng* 2011; 3(5):1773–9.

[7]. Mchugh, J. "Intrusion and Intrusion Detection", *International Journal of*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Information Security, Vol. 1, No. 1, pp. 14-35, 2001.

[8]. Prof. Ujwala Ravale, Prof. Nilesh Marathe, Prof. Puja Padiya, Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function, *International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015), Procedia Computer Science* 45 (2015) 428 – 435

[9]. Lee, W. and S. J. Stolfo, "Data mining approaches for intrusion detection", *In Proc. of the 7th USENIX Security Symp., San Antonio, TX.USENIX, 1998*

[10]. Gao Xiang, Wang Min. Applying semi-supervised cluster technique for anomaly detection. In: *IEEE international symposium on information processing*, no. 3; 2010. p. 43–5.

[11]. Mrutyunjaya Pandaa, Ajith Abrahamb, Manas Ranjan Patrac, a*,A Hybrid Intelligent Approach for Network Intrusion Detection, *International Conference on Communication Technology and System Design 2011, Procedia Engineering* 30 (2012) 1 – 9

[12]. Lane T. A decision-theoretic, semi-supervised model for intrusion detection. In: *International conference on machine learning and data mining for computer security*; 2006. p. 157–77.

[13]. Qiang Wang, Vasileios Megalooikonomou. A clustering technique for intrusion detection. In: *International conference on data mining, intrusion detection, information assurance, and data networks, security*, 5(12), 2005, p. 31–8.

[14]. Li Jimin, Zhang Wei, KunLun Li. A novel semi-supervised SVM based on tri-training for intrusion detection. *J Comput* 2010;5(4): 638–45.

[15]. G.V. Nadiammai, M. Hemalatha. The effective approach toward Intrusion Detection System using data mining techniques In: *Egyptian Informatics Journal* (2014) 15, 37–50, ISSN: 1110-8665.

[16]. Ghosh, A. and Schwartzbard, A. "A Study in using Neural Networks for Anomaly and Misuse detection", in *Proceedings of the Eighth USENIX Security Symposium,* Vol. 8, pp. 443-482, 1999.

[17]. Zhang Fu, Marina Papatriantafilou, Philippas Tsigas. Off-thewall: lightweight distributed filtering to mitigate distributed denial of service attacks. In: *IEEE international symposium on reliable distributed systems*, no. 31; 2012. p. 207–12.

[18]. SivathaSindhu, S.S., Geetha, S. and Kannan, A. " Decision Tree based Light Weight Intrusion Detection using a Wrapper Approach", in *Journal of Expert Systems with Applications*, Vol. 39, pp. 129-141, 2012.

[19]. Zhang Fu. Marina Papatriantafilou, Philippas Tsigas. CluB: a cluster based framework for mitigating distributed denial of service attacks. In: *ACM symposium on applied computing*, no. 26; 2011. p. 520–27.

[20]. Heady, R., Luger, G., Maccabe, A. and Servilla. M. "The Architecture of a Network Level Intrusion Detection System", *Technical report, Computer Science Department, University of New Mexico*, 1990.

[21]. Hesham Altwaijry , Saeed Algarny, Bayesian-based intrusion detection system, Journal of King Saud University – *Computer and Information Sciences*, (2012) 24, 1–6

[22]. Jian Pei, Shambhu J. Upadhyaya, Faisal Farooq, Venugopal Govindaraju. Data Mining for Intrusion Detection – Techniques, Applications, and Systems. *Data Mining Techniques for Intrusion Detection and Computer Security*

[23]. Zhang Fu. Marina Papatriantafilou, Philippas Tsigas, Wei Wei. Mitigating denial of capability attacks using sink tree based quota allocation. In: *ACM symposium on applied computing, no*. 25; 2010. p. 713–18.

[24]. Li Hanguang, Ni Yu, Intrusion Detection Technology Research Based on Apriori Technique, 2012 *International Conference on Applied Physics and Industrial Engineering, Physics Procedia* 24 (2012) 1615 – 1620

[25]. Zhang Fu. Marina Papatriantafilou, Philippas Tsigas. CluB: a cluster based framework for mitigating distributed denial of service attacks. In: *ACM symposium on applied computing,* no. 26; 2011. p. 520–27.

[26]. Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang. Semi supervised learning for false alarm reduction. In: *Industrial conference on data mining,* no. 10; 2010. p. 595–605.

[27]. Neminath Hubballi, Vinoth Suryanarayanan. False alarm minimization techniques in signature-based intrusion detection systems: *A survey, Computer Communications* 49 (2014) 1–17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[28]. PremaRajeswari, L., and Kannan, A. "An Intrusion Detection System based on Multiple Level Hybrid Classifier using Enhanced C4.5", *IEEE International Conference on Signal Processing, Communications and Networking*, pp. 75-79, 2008.

[29]. Vincenzo Gulisano, Zhang Fu, Mar Callau-Zori, Ricardo Jim Enez-Peris, Marina Papatriantafilou, Marta Patino-Martınez. STONE: a stream-based DDoS defense framework. In: *Technical report no. 2012-07, ISSN 1652-926X, Chalmers University of Technology*; 2012.

[30]. Zhang Fu, Marina Papatrianta Filou, Philippas Tsigas. Mitigating distributed denial of service attacks in multiparty applications in the presence of clock drifts. *IEEE Trans Depend Secure Computing* 2012;9(3):401–13.

[31]. Li Jimin, Zhang Wei, KunLun Li. A novel semi-supervised SVM based on tri-training for intrusion detection. *J Comput* 2010;5(4): 638–45.

[32]. Monowar H. Bhuyan, Bhattacharyya DK, Kalita JK. An effective unsupervised network anomaly detection method. In: *International conference on advances in computing, communications and informatics*, no. 1; 2012. p. 533–9.

[33]. Catania Carlos A, Garino Carlos. Automatic network intrusion detection: current techniques and open issues. *Elsevier Comput Electr Eng* 2012; 38(5):1062–72.

[34]. KDD Cup99 intrusion Detection Dataset. Available from: *<http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>*.