INTERNATIONAL JOURNAL OF SCIENCE
ISSN: 2455-0108 | WWW.IJOSCIENCE.COM
VOLUME II ISSUE II APRIL 2016

Negative Selection Algorithm and CART Based Intrusion Detection System

Shikha Singh
Dept. of Computer Science and Engineering
NRI Institute of Information Science & Technology
Bhopal, INDIA
&
Sini Shibu
Dept. of Computer Science and Engineering
NRI Institute of Information Science & Technology
Bhopal, INDIA

Abstract— *In today's world of the raising technologies the security have become one of the major concern for the researchers of computers science field. So for providing the security to the system the concept of Intrusion detection mechanism is introduced by the researchers few years ago. Since the introduction of this approach till now several mechanisms for IDS have been exits which may enhance the performance of this concept. So this paper is objected to enhance the intrusion detection system's accuracy by introducing the concept decision tree. As this is a widely known algorithm so there is also exist some issues with the IDS approach like its low performance, high false alarm rate etc. Hence a requirement is there is to establish a new robust mechanism of decision tree to generate an efficient decision rules for providing security to the network or system.*

Keywords— *Classification, Intrusion Detection, Tree, CART, Negative Selection Algorithm.*

I. INTRODUCTION

The current scenario represents that the network is one of the vital component for the whole process of the communication. Each person may perform various activities on the internet by the use of many different types of applications. So the Internet have now provides the several types of benefits to the people but along with the several merits there are some sort of demerits also there with the internet concept. Some of the people can perform various new inventions with the help of internet and its services which are fruitful for the normal people where as some devil minded persons have used the services of internet for the hacking and damaging the other's resources like network, hardware, software etc which can be called as use of internet for committing crime [1].

So in this scenario there is huge requirement for a security mechanism in order to provide protection against all these types of damages. Hence one of the major aspects of security has now become a complicated issue that is needed in the incredible growth of the usage of computer networks. It have becomes very hard technically and also this concept is economically costly for the manufactures of the computers to secure them formal these types of external attacks.

This issue allows the chance in regards of offering the security to the host and network both by introducing a mechanism for security [2]. Hence the concept of Intrusion Detection System have been introduced which may be created as the technique, methods, tools and the resources in order to provide help in recognizing the attacks, assess and then report the unauthorized network activity to the system administers. The Intrusion Detection System is a type of program which received the information of what has happens or has to be happened while the execution process and this may attempt to obtain hints or indications such that the network or system is being misused by any unwanted user.

Hence this paper is purposed as the initiative for the intrusion detection approach, that has developed for the required persons or authorities or organizations in order to learn the security objectives of the intrusion detection system may perform, and how to choose and configure the intrusion detection systems for some particular system or the network environments, also how to manage the result of the intrusion detection systems, and finally how to merge the functions of intrusion detection along with the remaining departments of the organization's infrastructure [3].

Generally there are two major kinds of the intrusion detection systems that are used recently such as Network based which is a packet monitor system and another one is the Host based which is working for the instance at the system logs details for the mark of the malicious or any type of suspicious activity within the real time. Though, recent commercially provided intrusion detection systems are the signature based intrusion detection system. And the Signature based intrusion detection systems may performs the pattern matching approaches in order to match the pattern of attack that are regarding to the known attack patterns that are recorded within the database and which may generate the low false positives (FP), whereas it may needs the continuous updates of the rules or the signatures.

On the other hand, anomaly based IDS builds models of normal behavior and automatically detects anomalous behaviors. Anomaly detection techniques identify new types of intrusions as deviations from normal usage [4], but the drawback of these techniques is the rate of false positives (FP).

So the researchers may have suggested in this paper a decision tree based intrusion detection mechanism for resolving these issues. They have targeted on the feature selection approach and they have produced a neuro-tree in order to obtain a better accuracy of detection.

II. TREE BASED INTRUSION DETECTION

The decision tree is the classification approach within the Data mining. In this the classification approach is purposely understand in order to create a scheme from a reclassified data-set. Then the decision tree is got divided into the two classes of the normal and the attack. The purpose of the object is to distinguish the normal and the attack pattern based on the class of attack. And this process then got repeated for entire classes. Also the classifier has been prepared by the use of training data and the testing data both [5].

The technology of decision tree is the most common and the rapid type of classification approach. And it process of construction is the top-down approach it follow; divide-and-rule method is applied in this algorithm. Basically it is a type of greedy approach. Initiating from the root node, for every type of non-leaf node, initially it selects an attribute in order to test a sample set, then it may divide the training sample set among the various sub-sample sets based on the results of testing and every sub-sample set may constitutes in a new leaf node. Finally it may repeat the previous process of division, till they have obtained a particular condition at the end.

Within the complete procedure of building a decision tree, consists of choosing the attribute of testing and this may decide how to split the sample set which is crucial process. Various types of decision tree algorithm are there that uses the separate technologies. Generally due to the size of the sample set which is very large normally have the branches and the layers of the produced tree which are more in number. Additionally, the abnormality and the noise are present within the training-sample-set that may also create some abnormal type of branches therefore there is a requirement to prune the decision tree. The major benefit of the decision tree algorithm is there is no requirement for the users to learn so much background details within the learning process [6].

For the ID3 approach of decision tree, the concept which is applied to quantify the information is referred as entropy. And this entropy is applied in order to evaluate the volume of the uncertainty within the set of the data. If entire data within the set is associated with one class then the entropy is zero which means no uncertainty is present.

III. LITERATURE REVIEW

Here in this paper [7] suggested the combining approach of classification for the intrusion detection system. Within this paper a neural network model has been used and K- Nearest Neighbors approach has also been used along with the classification approach. These types of data mining approaches for classification of the Intrusion detection have been proved to be essential for the different types of knowledge gathering mechanism. This work also implemented within two stages along with the initial stage of neural network for the enhanced outcome and also to enhance the KNN classifiers and these both may have hold it for applying in the second phase in order to detect the class of the classification approach.

This paper [8] implemented the latest process of training that may be added very easily with the training data-set by not modifying the weights of available samples for training. And the performance of k-mean clustering approach is based on the value of the k, where for k is equal to 2, so the rate of detection is achieved the 96.6 percent very quickly with the low false-positive-rate. And this may build the k-mean clustering approach to be more feasible for the dynamic types of environments which are needed for the regularly updated training-data.

In this paper [9] presented the anomaly detection which is based on the behavior so the detectors have been trained within the network that is not feasible at the time it got installed within the network in which the behavior is slightly different. And those types of differences are very complicate to represent, since the behavior of user is based on the variables, like security rules for the appliances, available ports, network architectures or operating systems. However, the benefits of the anomaly detector are the detection of the zero-day attacks, confused variations and the insiders with their capability to decrease the range of search than with the emerging rules of the misuse detection that are the motivations for developing the system that are able to do self-learning within the network where the detector is installed.

Within this research [10] some of the new approaches have been investigated for the detection of intrusion and in order determine their performance based on benchmark KDD Cup 99 data-set. In this paper first the decision tree approach has been implemented for determining the intrusion detection. In this another approach called as support vector machines have also been examined which is then compared with the performance of the decision tree model. Since decision tree was implemented as the binary classifier, so here applied the five types of classifiers for each 5-class classification. Then the results of experiment represented that the decision tree have provides the better accuracy as compare to the SVM for the U2R,Probe and R2L classes while for the normal class both the approaches have provide similar accuracy also for the DOS the class decision tree have provides the bad accuracy as compare to the decision tree approach.

Within this paper [11], learning process has been supervised with the pre-processing phases for the intrusion-detection. In this database is produced by the use of stratified sampling approaches and then the classification approach is implemented over the samples. And the accuracy of the suggested approach is get analyzed by comparing them along with the available results of analysis to authenticate the validity and the accuracy of suggested approach.

Within this paper [12], suggested the use of data-set called as Kyoto 2006+ which is prepared on the basis of real traffic data of over three years. In this suggested paper the approach of J48 decision tree have been used for the network intrusion detection purpose and also obtained the accuracy at nearly 97.23 percent. This approach is implemented by the use of WEKA 3.6.10 tool, here also prepared the decision-tree in order to find out the intrusion with n the given data-set that have received higher rate of true positive approx. 99 percent for the normal and the attack packets. By the analysis of the generated result, the tree may have correctly classified nearly 130931 instances among the 134665 instances that are 97.23percent approx. And the results of simulation have presented that this model is now capable to find out the unknown attacks also.

According to this paper [13] here provided an observation of the various effects of the imbalance over the improved tree which is dependent on the real-time intrusion-detection-system. So this suggested approach is a type of real-time-IDS that may be surely includes the imbalanced data, therefore this type of analysis have now becomes compulsory. In this the imbalance and their influences over the accuracy of classification also with its reliability have also been described. This paper presents an effective solution in order to manage the imbalance sampling in brief.

IV. NEGATIVE SELECTION ALGORITHM

The Negative Selection Algorithm belongs to the branch of AIS (Artificial Immune System). Negative Selection Algorithm is similar to another algorithm such as AIS(Artificial Immune Systems) such as the CSA(Clonal Selection Algorithm), and the INA(Immune Network Algorithm).

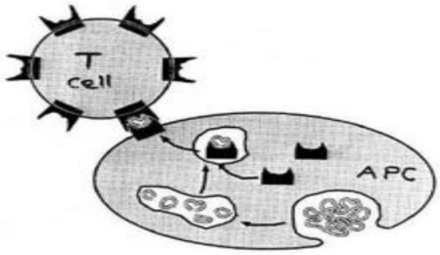


Figure 1: Negative Selection Process

Inspiration

The Negative Selection algorithm is stimulated by the self-nonself discrimination behavior seen in the mammalian acquired immune system. The CST (clonal selection theory) is responsible for the immune system of our body to be conducted in the adjusting manner and consider the selection and the conception of the body cell which are responsible for selecting the harmful (foreign cell) substance of our body. An interesting fact of this method is that the population of non selecting cell –for the tissue of the body is managed by this method. Specifically, it does not develop the immune cell which is reactive with self and is called as the auto-immunity. This identified problem is known as 'self-nonself discrimination' and it requires the readiness and the perpetuation of a immune cells in a way so that no one is auto-immune. This is accomplished by this method and is responsible for selecting and during cell creation and cell proliferation of those cells that are self reactive are removed. This process has been seen in the readiness of lymphocytes of category T, having naïve but due to the selection process of the negative and positive they become matured in thymus

V. PROPOSED WORK

This section deals with the proposed work. Main part of this work are based on

1. CART
2. Negative Selection Algorithm

Normalization: It is a method to convert numerical data of any range into range of 0-1.

$$New\ Value = \frac{Old\ Value - Minimum\ Value}{Maximum\ Value - Minimum\ Value}$$

Where
New Value = New Numerical Value
Old Value = Old Numerical Value of same attribute
Minimum Value = Minimum Value of same attribute
Maximum Value = Maximum Value of same attribute

The proposed work's algorithm is described here as follows:

```
Input: Dataset
Output: Detection of Intrusions

Procedure IDS_Work()
  A.  File read
      1. Input DS
      2. Scan DS and make DS
         n*m-1matrix
      3. Separate DS and C
      4. Apply Negative Selection
         Algorithm
  B.  Negative Selection Algorithm
         if DS_ij is C_ij(==1)  {DS is
            normal}
         else, DS_ij is C_ij(~=1) {DS is
            Attacks}
              if end
      Apply CART on DS_else
  C.  CART Classification
   1. Start at the root node.
   2. For each DS, find the set S that
      minimizes the sum of the node
      impurities in the two child
```

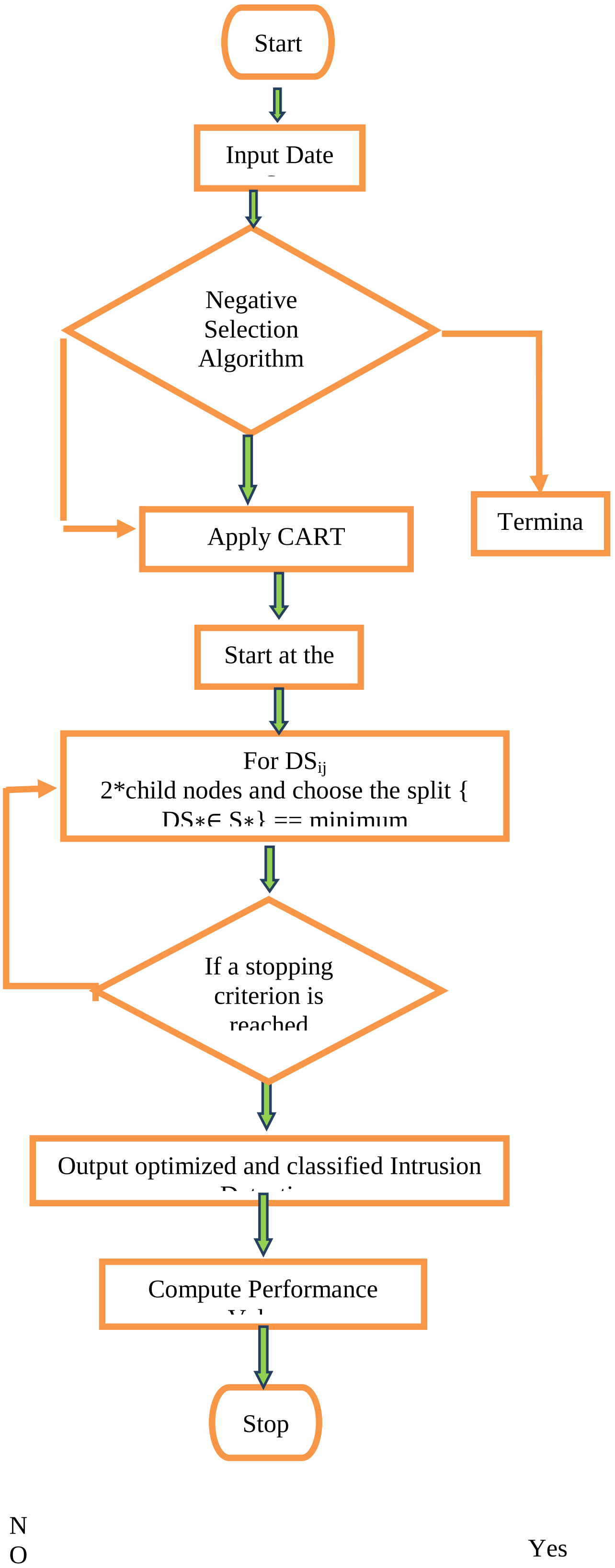

Start
Input Date
Negative Selection Algorithm
Termina
Apply CART
Start at the
For DS$_{ij}$ 2*child nodes and choose the split { DS*∈ S*} == minimum
If a stopping criterion is reached
Output optimized and classified Intrusion
Compute Performance
Stop
N O
Yes

Figure 2: Proposed Work Algorithm

Fig 3: Proposed Architecture

VI. RESULT ANALYSIS

This section deals with the results of the proposed work. These results are taken after execution of the implementation on a dataset. So before giving the details about results and it's analysis, this section discusses about the system on which these experiments are perform along with the used tool which is followed by the description of dataset used here.

Table I: System Configuration

Model:	Sony Vaio
Processor:	Intel® Core™ I5-2450M 2.5GHz
RAM:	4GB
System Type:	64 Bit Operating System
Windows Edition:	Windows 10 Home Basic
MATLAB	R2014a

Dataset:

Table II: Dataset Detail

Dataset Name	KYOTO 2006+	
Dataset weblink	http://www.takakura.com/Kyoto_data/	
Size	10.5mb	
Dataset FileName	20070117.txt	
S.N.	Datatype	Dataclass
1	Unknown Attacks	-2
2	Known Attacks	-1
3	Normal	1
Number of Attributes : 14 + 1 for class		
Number of Instances : 61745		
Number of Instances use : 38300 (62% of 20070117.txt)		

Precision:

Precision is a description of random errors, a measure of statistical variability.

$$\text{precision} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{false positives}}$$

True Positive Rate:

It measures the proportion of positives that are correctly identified

$$TPR = TP/P = TP/(TP + FN)$$

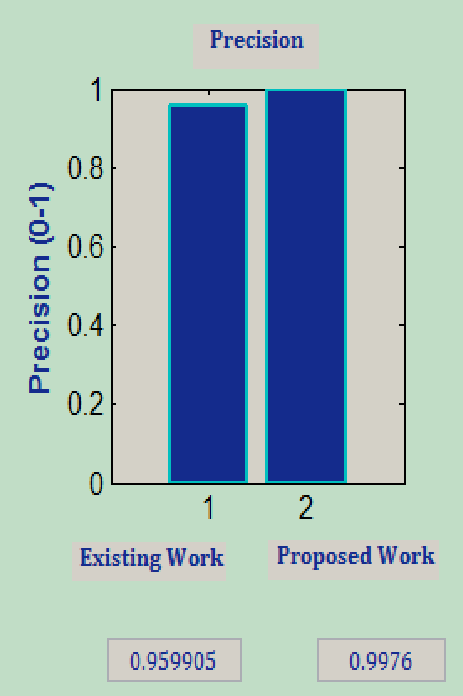
Where TP = True Positive
P = Positive
FN = False Negative

F measure:
It is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.

$$\text{F measure} \quad = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Table III: Values of Precision of proposed work with Existing Work

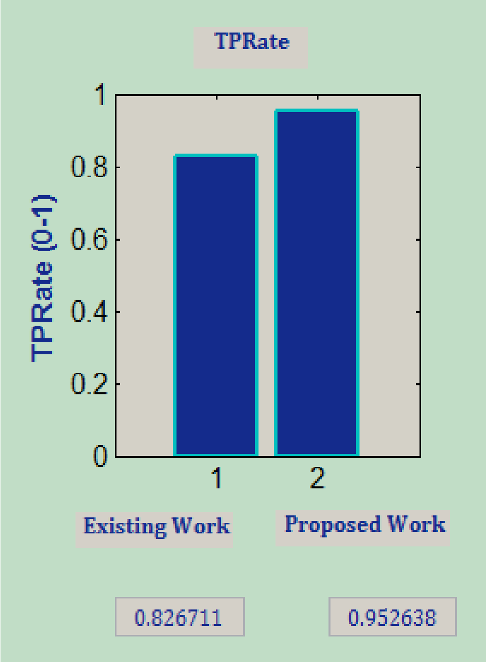
	Existing Work	Proposed Work
Precision	0.959905	0.9976



Graph 1: Precision of proposed work with Existing Work.

Table IV: Values of True Positive Rate of proposed work with Existing Work.

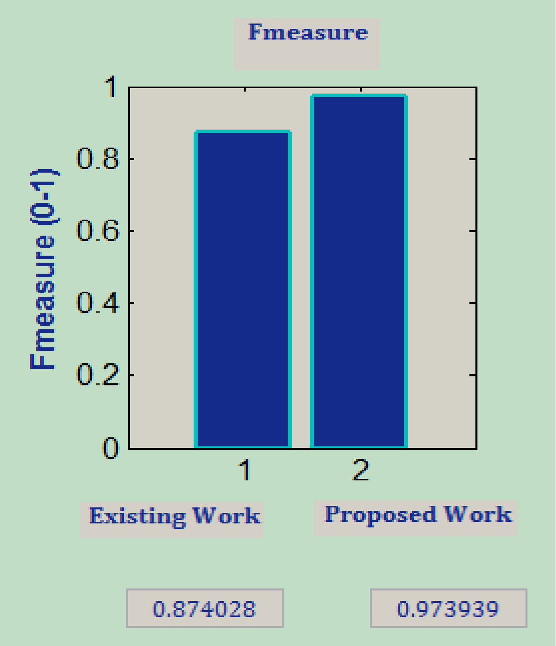
	Existing Work	Proposed Work
TPR	0.826711	0.952638



Graph 2: True Positive Rate of proposed work with Existing Work.

Table V: Values of Fmeasure of Proposed work with Existing Work

	Existing Work	Proposed Work
F measure	0.874028	0.973939



Graph 3: Values of F measure of Proposed work with Existing Work

Note: Precision, F measure and TPR values are in between 0-1

VII. CONCLUSION

Some of the existing researches have also been discussed in this paper like some types of pre-processing approaches such as data mining, neural network models, artificial intelligence, have been examined for achieving the better rate detection within the intrusion-detection-system. In this paper for providing the security to the network from the intruders the latest mechanism of Negative Selection Algorithm with the CART approach for IDS has been introduced. In this paper also examine the huge volume of data from the network which is extremely difficult to observe and improve the performance of the IDS system by using the Negative Selection along with CART based approach for this determination. And the experimental result have presented that the intrusion detection algorithm dependent on the proposed NSA-CART is the most feasible and efficient approach which may has the higher rate of accuracy.

REFERENCES

[1] Huang Ming, Niu Wenying, Liang Xu".An improved decision tree classification algorithm based on ID3" and the application in score analysis.

[2] Deepthy K Denatious and Anita John. "Survey on data mining techniques to enhance intrusion detection". In Computer Communication and Informatics (ICCCI), 2012 International Conference on Digital Object Identifier, p. 1–5. IEEE, 2012.

[3] SURYA BHAGAVAN AMBATI, DEEPTI VIDYARTHI, ―A BRIEF STUDY AND COMPARISON OF, OPEN SOURCE INTRUSION DETECTION SYSTEM TOOLS‖ International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-1, Issue-10, Dec-2013

[4] Ammar Boulaiche, "A Quantitative Approach For Intrusions Detection And Prevention Based On Statistical N-Gram Models ", Procedia Computer Science 10 (2012) 450 – 457.

[5] Muamer N.Mohammed, "Intrusion Detection System Based On Svm For Wlan ",Procedia Technology 1 (2012) 313 – 317.

[6] V. Bapuji ,"Soft Computing and Artificial Intelligence Techniques for Intrusion Detection System", Department of Informatics, Kakatiya University, Warangal, India, Vol 2, No.4, 2012,pp 2- 30.

[7] Shalinee Chaurasia ,Prof. Anurag Jain, "Review: Ensemble Neural Network and KNN Classifiers for Intrusion Detection", International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013.

[8] Aakshi Choudhary, Sarbjit Kaur, "IDS Approach Using Data Mining Tool WEKA", Volume 4, Issue 6, June 2014 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering.

[9] Edward Guillen 1, Jeisson Sá nchez and Rafael Paez, "Inefficiency of IDS Static Anomaly Detectors in Real-World Networks", Future Internet 2015, 7, 94-109; doi:10.3390/fi7020094.

[10] Snehal Mulay, Urmila Kalshetti, G.V. Garje, Anshul Abhang, "Support Vector Machine and Decision Tree for Intrusion Detection", International Journal of Systems , Algorithms and applications, Volume 2, Issue 4, April 2012.

[11] Devendra kailashiya, Kanak Saxena, "Improve Intrusion Detection for Decision Tree with Stratified Sampling", International Journal of Electronics Communication and Computer Engineering, Volume 2, Issue 2. 2011.

[12] Shailendra Sahu, B M Mehtre, "Network Intrusion Detection System Using J48 Decision Tree" , 978-1-4799-8792-4/15/ 2015 IEEE.

[13] Dr.R.Balasubramanian1, S.J.Sathish Aaron Joseph, "Intrusion Detection on Highly Imbalanced Big Data using Tree Based Real Time Intrusion Detection System: Effects and Solutions", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 2, February 2016.

.