IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

An Approach for Detecting Attacks Using Genetic Annealing in Intrusion Detection System

Laxmi Jonwar[1], Ashish Gupta[2]

NITM, Gwalior, M.P., India

ABSTRACT: Through proposed work, a new trend of machine learning based on evolution and genetics has been tried to be brought in an innovative way. IDS has been a part of our system development from a very long time, but, there has not been a very effective way to create a newand very efficient system. The first step towards it is to increase the capability of the system to detect the attacks and then, to encounter them. So, here a new approach has been proposed to detect the attacks with high precision and accuracy. The proposed gives a new term "genetic annealing" which is a hybrid of genetic algorithm and simulated annealing technique. The results show the effectiveness of the results and the proper working of the algorithm is explained in the paper.

KEYWORDS: IDS, Attack Detection, Genetic Algorithm, Simulated Annealing.

I. INTRODUCTION

In our day-to-day life, the use of computer networks is effectively increasing, so security also plays a vital role for a network. To overcome this problem, we develop protective software systems which can help us to identify the security goals [1]. The drawback was in terms of handling high dimensional data, false positives and false negatives [2]. The curse of dimensionality is that when we include something new in our result and we get performance degradation. That's why we use the number of features will be less than the original dataset so we can easily overcome with this problem and which also tends to reduce the dimensions of the dataset. One of the techniques we used is feature selection which uses limited features and leads to dimensionality reduction. For a strong security point of view, there are different soft computing techniques. IDS is a device that monitors suspicious programs, applications or activities of a system or software [3]. The most important component in this work was to set a benchmark of data set KDD Cup which is to be used to carry out the experiments. Therefore, the benchmark dataset KDD Cup data set is to be used as database. In this dataset we use 41 attributes which are used by each record to characterize network traffic behavior. [4].

Genetic Algorithm (GA) is an optimization technique from evolutionary computation.It uses the principle of survival of the fittest. It uses the mutation and crossover to provide the optimal number of features. This theory mainly defines that those individuals which are adaptable to changing environment thus fit in that scenario will move further and the rest of the weak population will be eliminated [5].

1.1 GA pseudo code

The initial population is chosen.

Initially many individual solutions are randomly generated to form an initial population. The population size depends on the nature of the problem, but typically contains several hundreds or thousands of possible solutions.

Fitness of every individual in the population is evaluated.Repeat Best adaptive individuals from the population are selected.

Breed new generation t with the help of crossover and mutation and give birth to newones Fitness evaluation of each offspring Worst ranked portion of population is replaced with offspringUntil <terminating condition>

In soft-computing technology[6-8], the neural network is usually applied in intrusion detection research. Hierarchical perception (MLP), hierarchical BP neural network model and self-organizing map (SOM) network have gained great experimental effectin the application of intrusion detection. Genetic fuzzy rule mining based intrusion detection mainly studies rule encoding, fitness function evaluation, etc. The basic framework of negative selection and genetic process remains unchanged.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

II. LITERATURE REVIEW

In this paper [9], an IDS using genetic algorithm based feature selection approach has been proposed..
In this paper [10], main focus was on a part of IDS which was anomaly detection. Theresearch focused on false positive rate (FPR).
In this paper [11], the authors worked towards finding the solution to intrusion detection by proposing a hybrid approach. The work was done on fuzzy and genetic algorithm. For feature selection to the processing and finding results, the fuzzy genetic algorithm was applied. Crossover and mutation rate were prefixed in this work.
In this work [12], the approach was hybrid but was focused on different databases. They applied their research to both discrete and continuous attributes.

III. PROPOSED WORK

The proposed work is based on genetic approach using simulated annealing as an important aspect affecting the intrusion detection system. There are few main phases in the process. These are:

i. The KDD Cup 99 dataset is used as the dataset. The data is loaded first.
ii. In the feature selection phase, out of 41 attributes, only the important attributes are selected using the soft set theory.
iii. The selected attribute based data is then saved in a file. The file is further processed using the genetic algorithm and simulated annealing concept to create file having only the data which need to be processed further toget the final result.
iv. The file is then processed in weka using the classification of the attributes.The algorithm used is tree based and is known as simple cart.
v. Finally, the results are obtained having the measurement parameters namedas: TP Rate, FP Rate, precision, recall, accuracy, etc.

The steps in the proposed algorithm are explained in detail:

1. Loading the dataset: the KDD Cup dataset is not loaded fully. Only the 10% of the dataset is used. For this, the dataset is divided using sql query and thenloaded in the MATLAB.
2. The next step is feature selection which aims at determining a minimal feature subset from the problem so as to find the results with high accuracy. The feature selection is a necessity as the data contains redundant and irrelevant data too. A feature is considered to be redundant if it is highly correlated with other features. This data need to be removed by various methods. The reduction in unnecessary data will lead to better, accurate and faster results.
3. The file is saved and soft set, GA and simulated annealing is applied further.

i. The soft set approach is applied on the data so as to reduce the redundant sets.
ii. The data is converted into binary form i.e. 0 and 1 to represent the data in form of chromosomes.

These chromosomes perform crossover at random locations among themselves. Few chromosomes turns out be the same after crossover is performed, for that, a process known as simulated annealing is performed. The chromosomes are changed totally and the new formed chromosomes are slowly taken to the point where they may vary to the other chromosomes but have at least 75% features same as others. This process improves the accuracy and diversification both.The coding can be represented as:

```
for i=1:row
for j=1:col
if(data< range)
chromosome = 0
end
end
else
chromosome = 1
if(chromosome(i) == chromosome(loc))matched = matched+1
%perform crossover
```

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

```
b1= [first(1, 1:cp), second(1,cp+1:end)];
b2= [second(1, 1:cp), first(1,cp+1:end)];
%perform SA
for(chromosome(i) <= chromosome(i+1)) first = chromosome(setall('1'));
s_chromosome(k,:)=first - '0';
end
```

1. The final results are calculated using WEKA. The algorithm is simple cart. The results generated are discussed further in detail. The results obtained are better than the previous results obtained from the base algorithm.

3.1 Measurement factors are:

The final results are shown in the form of confusion matrix. In IDS, the measurement parameters are represented using the confusion matrix having attributes as: TP (True Positive), FN (False Negative), FP (False Positive) and TN (True Negative). The matrix can be represented as:

Table 1: Confusion Matrix for TN, TP, FP and FN

Valid Record	Correctly Classified	Incorrectly Classified
	True Negative (TN)	False Positive (FP)
Attack Record	True Positive (TP)	False Negative (FN)

i. **Precision:** Precision refers to the closeness of two or more measurements to each other. Precision is defined as the number of true positives over the numberof true positives plus the number of false positives.

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (1)$$

ii. **Recall:** Recall is defined as the number of true positives over the number of true positives plus the number of false negatives.

$$\text{Recall} = \frac{t_p}{t_p + f_n} \quad (2)$$

iii. **Accuracy:** Accuracy refers to the closeness of a measured value to a standard or known value.

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (3)$$

iv. **Kappa: Cohen's kappa** is a measure of the agreement between two ratters who determine which category a finite number of subjects belong to whereby agreement due to chance is factored out. The two ratters either agree in their rating (i.e. the category that a subject is assigned to) or they disagree; there are no degrees of disagreement (i.e. no weightings).

$$\Pr(a) = \frac{t_p + t_n}{t_p + t_n + f_p} \quad (4)$$

$$\Pr(e) = \frac{(f_p + t_n) * (t_p + t_n)}{(f_p + f_n) * (t_p + f_n)} \quad (5)$$

$$k = \frac{(\Pr(a) - \Pr(e))}{1 - \Pr(e)} \quad (6)$$

v. **True Positive Rate (TPR):** The true positive rate shows the correctlyclassified attack types and thus the value should be higher.

$$\text{TPR} = \frac{T}{TP_4} \quad (7)$$

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

vi. False Positive Rate (FPR):

FPR = $\frac{F}{TN-}$ (8)

The tool used is MATLAB 2016 and weka. The feature selection, genetic algorithm and simulated annealing have been implemented on MATLAB while the final confusion matrix and the measurement parameters are calculated on weka. WEKA support various standard data mining job, more specifically, data pre-processing, clustering, classification, feature selection and visualization. We need to classify the KDD Cup dataset, hence, weka is used.

IV. RESULTS AND ANALYSIS

This part shows the experimental results calculated from Genetic annealing model of IDS along with its comparison with the fuzzy based IDS method. The two different algorithms when compared, shows that the proposed algorithm is better than the base algorithm in all terms. Both the algorithms consequently results in the data classification as attack type accordingly, Also, the accuracy obtained for Genetic annealing based IDS is raised appreciably than fuzzy based IDS

.

Table 2: Comparison of Base & Proposed Results

	BASE	**PROPOSED**
Correctly Classified Instances	99.887 %	99.927 %
Incorrectly Classified Instances	0.104 %	0.073 %
Kappa statistic	0.9968	0.9978
Mean absolute error	0.0006	0.0004
Relative absolute error	0.434%	0.3311 %
Total Number of Instances	32833	33162

Other factors are also better when compared with the base work. The results are compared further using the graphs that show the differences in the TPR rate along with precision and time factors, etc.

TP Rate: the increase in TP rate shows the increase in better results as this, results in finding better attributes in terms of performance and accuracy.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

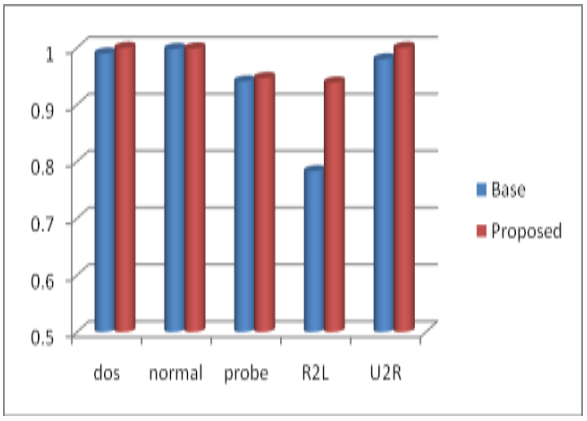

Fig.1 Comparison for TP rate of the base and proposed algorithm.

Precision: it is an important factor to calculate the working efficiency of the algorithm. The precision of the algorithm should be high. A comparative graph based study is shown below to ensure the positive working of the proposed work:

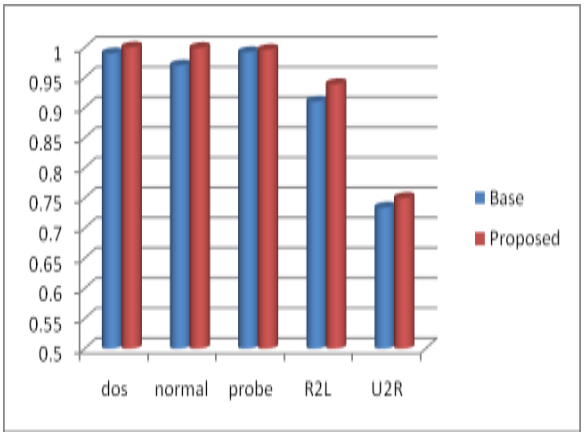

Fig.2 Comparison for precision of the base and proposed algorithm

Recall: recall of the proposed algorithm should be high and is also one of themeasures for the calculation of the working of the proposed algorithm.

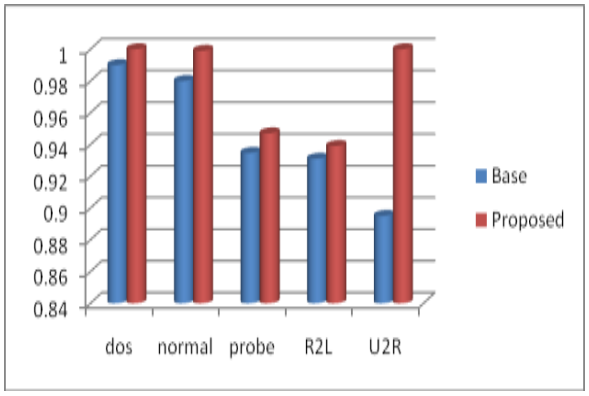

Fig.3 Comparison for recall of the base and proposed algorithm

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 7, July 2023 ||

| DOI:10.15680/IJIRSET.2023.1207189 |

V. CONCLUSION

The proposed research worked a lot over making better IDS. The requirement is increasing and so the research methodologies. The increasing need for a perfect system is leading to the research of the new techniques in this field. IDS is worth detecting system and thus, all the work for the betterment of the system is going on. The proposed research has shown better results on comparison. Simulated annealing has never been a part of the research in IDS. This is first time that IDS is created with the help of simulated annealing concept and has proved to be a better system.

REFERENCES

1. Anthony, Raj.A," A Study on Data Mining Based Intrusion Detection System", International Journal of Innovative Research in Advanced Engineering (IJIRAE)Volume 1 Issue 1 (March 2014), pp: 21-25.
2. Manikandan, R., Oviya, P., and Hemalatha, C., "A New Data Mining Based Network Intrusion Detection Model," Journal of Computer Application, Volume 5, Issue EICA2012-1, pp. 1-10 February 10, 2012.
3. Wenke, L., and Salvatore, J., "Data Mining Approaches for Intrusion Detection," Proceedings of the 7th USENIX Security Symposium San Antonio, Texas, January 26-29, 1998.
4. Mahbod, T., al., "A Detailed Analysis of the KDD CUP 99 Data Set", IEEE symposium on computational intelligence in security and defence applications, 2009.
5. Sugandha, R., and Rishi, S., "Privacy Preservation in Utility Mining Based on Genetic Algorithm: A New Approach", Proceedings of Fifth International Conference on Soft Computing for Problem Solving, Springer, pp 71-80.
6. Herawan, T., and Mustafa M., "A direct proof of every rough set is a soft set".Proceeding of International Conference AMS 2009, IEEE Press, 119-124.
7. Maji, P., Roy, A., and Biswas, R., "An application of soft sets in a decision making problem". Compututer and Mathematics with Application, 44, 2002.
8. Mukkamala, S., Janoski, G., & Sung, A. (2007). Intrusion detection using neural networks and support vector machines. In Neural Networks, 2007. IJCNN'02. Proceedings of the 2002 International Joint Conference on (Vol. 2, pp. 1702-1707). IEEE..
9. Senthilnayakiet, B., al., "Intrusion Detection Using Optimal Genetic Feature Selection and SVM based Classifier", 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN).
10. Dipika, N., and Ompriya, K., "Optimizing False Positive In Anomaly basedIntrusion Detectionusing Genetic Algorithm" , 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE).
11. Ganesh, R., and Sachi, N., "A Hybrid Approach for Network Intrusion Detection", 2015 Fifth International Conference on Communication Systems and Network Technologies.
12. Jabez., and Dr. Anadha, M., "A Study On Genetic-Fuzzy Based Automatic Intrusion Detection On Network Datasets", IEEE, 2012.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details