© 2015 IJSRSET | Volume 1 | Issue 3 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Efficient calculation of fitness function by calculating reward Penalty for a GA-based Network Intrusion Detection System

Prof. Jahnavi. S. Vithalpura, Prof. H. M. Diwanji

Department of Computer-IT, L. D. College of engineering, Ahmadabad, Gujarat, India

ABSTRACT

Our network is facing a rapidly evolving threat landscape full of modern applications, exploits, malware and attack strategies that are capable of avoiding traditional methods of detection. Intrusion detection can perform the task of monitoring usability systems to detect any apparition of insecure states. To overcome above mentioned issues we have employed genetic algorithm to improve detection rate of intrusion detection system. To generate healthy rule pool we have focused in design of fitness function. We have proposed a new fitness function based on reward & penalty. This function make chromosome stronger by applying reward and remove weakness from it by deducting penalty. So such a healthy chromosomes generates a best fit population which is reducing false alarm rate and increasing a detection rate. In our work, we have classified a dataset as a normal record or attack record using seven network features and calculated detection rate and false alarm rate. Further we have classified DOS, Probe, and U2R and R2L type of attack from attack cluster. We measured improved efficiency of proposed system by observing improvement in detection rate and reduction in false alarm rate.

Keywords: Genetic Algorithm, Intrusion, Network Intrusion Detection System, Fitness Function, Reward Penalty.

I. INTRODUCTION

Intrusion detection system is a process of monitoring network activities [1], presented and available in computer network and investigate it's for detection of violating threats which could affect computer security strategies and security practices. Now-a-day our network system should play an important role in society to prevent computers from malicious threats. At a situation of transferring files and communication to be held on the network we should probably focusing network security. Intruders are more available in network to capture important files and materials and do some malicious activities in the computer system. Attacks are categorized into [2], four types. There are DOS, U2R, R2L and Probe. These attacks are classified by some optimization methodologies.

Normally, intrusions are causes damage to the computer system resources with the term of unauthorized activities (modifications) of important files and folders presented in the system. In [3], especially told information about intrusion detection and also told its two kinds of detection principles. There are anomaly detection and misuse detection. In [4] told methods of classification of intrusion detection system. Three kinds of methods to be introduced are audit source location, detection method and detection paradigm. In audit source location consider the factors of network packets, application log files, host log files and IDS sensor alerts. Behaviour and knowledge based policies are executed using detection methods. Finally, in detection paradigm to be considering factors of state based and transition based detections.

Usually, networking attacks are detecting and preventing by some classification methods. Normally used classification methods are svm, genetic and k-NN. Some of the network intrusion detection and prevention methods are discussed in [5]. In our proposed system we have design new the optimal fitness function solution here; fitness function can be calculated by reward penalty based model. Processes held in IDS are

i) Perform genetic algorithm to done the process of initial population, mutation and crossover.

ii) Fitness Calculation to be done with newly designed reward penalty based fitness function.

iii) Network details are collected by jpcap and wincap tools.

iv) Find attacks presented in our network and also obtain how percentages of attacks are presented.

The remaining section of this paper organized as follows. Section II completely describes our proposed system and Experimental Setup and Problems formulated in previous system can be explained in Section III. Section IV and Results are analyzed in Section V. Finally concludes this paper concept.

II. METHODS AND MATERIAL

Proposed System

A. Proposed Algorithm

PHASE A- Data Preprocessing

1. Capture KDD DATA SET.
2. Extract 7 network feature using SVM Classifier.

PHASE B- Training of data packets to define rule set.

Input: Audit data recorded from network, no. of generations size of population.

Output: classified rule set.

1. Encode network audit data as chromosome Using binary encoding
2. Initialize the population
3. N is the total number of records in the training set
4. For each chromosome (record) in the population
5. M = 0, MN = 0, X=Maximum value of MN, Y= Maximum value of M

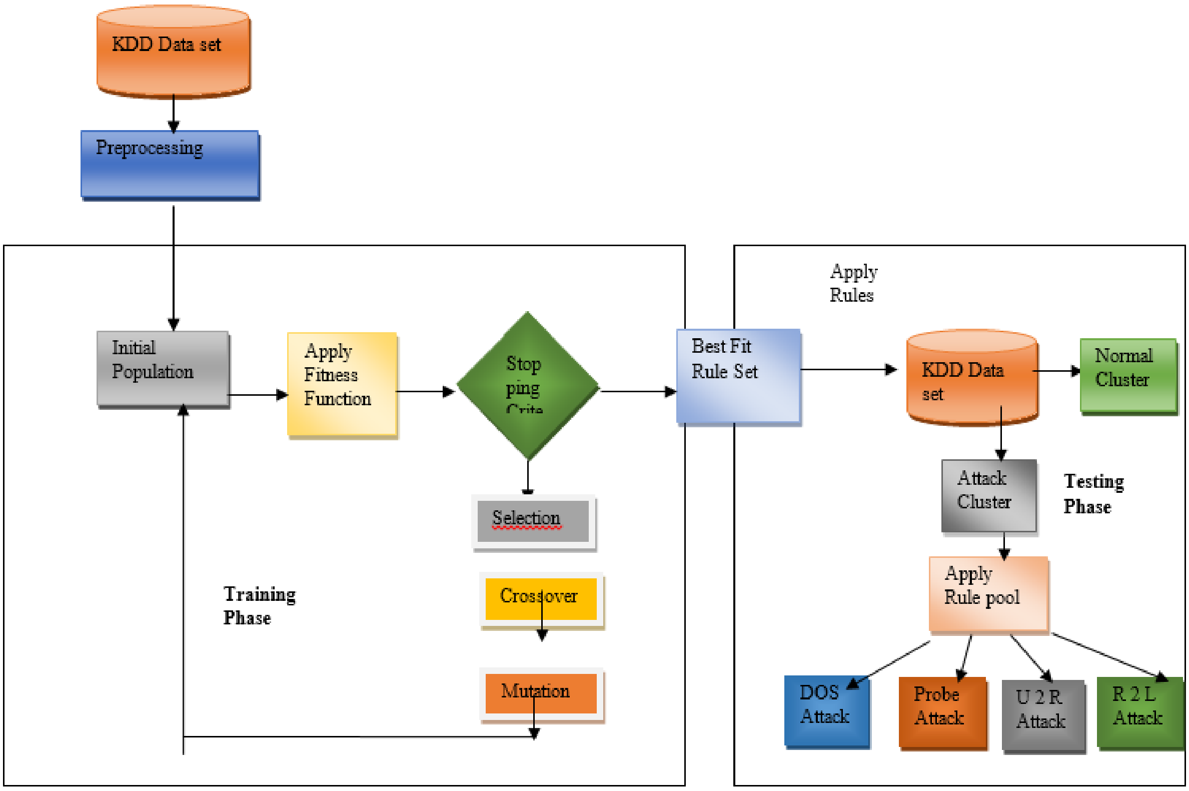

Figure 1: Proposed approach of GA based Intrusion detection system

6. Compare each record in the training set
7. If record matches with the chromosome
8. MN = MN + 1
9. End if
10. If the record matches only the “condition” part
11. M = M + 1
12. End if
13. End for Loop calculate X, Y.
14. Fitness F. =2+$\frac{MN-M}{MN+M}+\frac{MN}{X}-\frac{M}{Y}$
15. Apply roulette wheel selection method to select chromosome.
16. Select one chromosome into new population

```
17. End for Loop
18. For each of this chromosome in newly generated population
19. Implement single point crossover on the chromosome
20. Implement single point mutation on the chromosome
21. End for Loop
22. If no. of generations is left then, go to line 4
```

PHASE C TESTING OF SYSTEM

```
1: Capture KDD test data.
2 :Loop Forever {fetch packet from kdd test data.}
 3: for each rule in rule-base
 4: Match rule with network connection.
 5: if rules match the
6: Mark current connection as an intrusion(attack).
7: end if
 8: end for each
 9: end loop forever.
```

B. DISCUSSION OF NEWLY DESIGNED FITNESS FUNCTION

Formula of the newly designed function is as follows:

$$F=2+\frac{MN-M}{MN+M}+\frac{MN}{X}-\frac{M}{Y}$$

Where:

X is highest value of MN in the particular generation.
Y is the highest value of N in the particualr generation.

(MN/(MN+M)) = Resultant value of this formula will be increase strength of record.
(M/(MN+M)) = Resultant value of this formula will be decrease weakness of record.

To get good result about the strength of the record, one can deduct the weakness's value from the strength's value by performing calculation ((MN-M)/ (MN+M)), in the function designed.

Now, assume that there are two records with the following values of M and MN:

Table 1: Similar Fitness Values

Record	M	MN	Fitness = ((MN-M)/(MN+M))
Record1	0	2	1
Record2	0	6	1

But in a case study like above the resultant values is not perfect as it will treat both records as the similar strength record , but it is very known to us that second record is more stronger as compared to first record because of the value of MN. Here function should be supplied with other positive and negative values to apply reward and penalty to the records.

MN/X: It generates the value which affects the strengthens of the record depends on the strongest record in the population. So result will become 0 in the worst situation (when MN=0) and it becomes one in the best situation if the MN value of particular record is maximum MN value in the generation. So it will add to the equation to make more strong record.

M/Y: It generates the value which affects the weakness of the record depends on the most weakest record in the generation, the resultant value will become Zero in the best situation (when M =0) and will be equal to 1 in the worst situation if the M value of that record is the highest value in the generation, so the resultant value of M/Y should be deducted from the equation to give the penalty on that particular record.

Now, consider best case of the incoming network connection, so MN value of that record is the maximum MN value in the MN column, and M value is equal to Zero, this means that Fitness of this record becomes 2, in other hand, consider the record having worst situation, so M value of this record is the Maximum M value in the M column, and MN value is becoming Zero, this means now this fitness becomes -2, but the fitness value generated by this fitness function must make a non-negative cost to each candidate solution , so the constant value of 2 will be added to this function to generate fitness value become zero in the worst situation, and fitness value become four in the best case, in this way, it will become positive values of fitness function and in the range of 0 to 4 in any situation. So, here by designing new fitness function, we have discussed fitness calculation in both cases best and worst case. It is giving reward to make more healthy chromosomes and it is gives penalty to reduce weakness from that chromosome.

Table 2: Similar Analysis Of newly designed fitness function

Dos			Normal		
M	MN	Fitness	M	MN	Fitness
3	419	3.985	0	1	3.143
5	280	3.632	0	3	3.429
5687	18	0.049	50	7	1.246
23	5	1.365	44	4	0.858
R2L			Probe		
M	MN	Fitness	M	MN	Fitness
180930	11	0.016	130691	2	0.002
2114	714	2.493	242	12	1.107
0	4	3.006	0	856	4.000
0	16	3.022	0	6	3.007
			U2R		
			M	MN	Fitness
			134916	5	1.000
			1	3	3.267
			816	4	1.501
			11	2	1.348

III. RESULTS AND DISCUSSION

Genetic based intrusion detection system is implemented using below mention requirements

Software requirement:
Operating System: window 7 or higher
Programming Language used: JDK1.8 or higher versions
Database: SQL Server 2008
Datasets used: Kddcup 99 10% data set

Parameter defined for proposed System

Table 3: Parameter defined for the proposed system

Parameter name	Value
Dataset	KDD DATASET
Encoding	Binary
Initial population	20 to 40
Selection	Roulette wheel selection
Crossover	Single point crossover
Mutation	Single point mutation
No of generation	80 to 200

After thorough analysis of various intrusion detection system, and a detailed study of the genetic algorithm concepts. We analyzed that genetic algorithm will be best suited to design an efficient network intrusion detection system. In our Proposed work, we have designed a reward penalty based a new fitness function. This function calculates reward and penalty on the basis of strongest record and weakest record.

Standard Matrix used to evaluate proposed system
Detection Rate= True positive/(True Positive + False Negative)
FP Rate=False Positive/ (False Positive + True Negative)

We have done so many experiments to analyze performance of proposed GA base d intrusion detection system. Here we have listed three experiments results.

EXPERIMENT -1
We have done experiment two ways: 1) genetic algorithm using support confidence frame work in fitness function 2) genetic algorithm using proposed new fitness function. At first, we trained the genetic algorithm with the training dataset and test our algorithm with the testing dataset from KDD99 dataset. From our experiment, the detection rate of KDD99 dataset is 98.72% with low false negative rate.

Table 4: KDD DATASET RECORD Distribution

RECORD	Training Dataset	Testing Dataset
Normal	39387	39337
Intrusion	160147	160177
Total	199534	199514

Table 5: Experiment-1 using KDDcup 99 dataset

EXPERIMENT-1USING KDD CUP 99 DATASET				
Fitness function used in genetic algorithm	TESTING ATTCK	TESTING NORMAL	DR	FPR
Support confidence frame work	159650	39804	96.62%	1.23%
Proposed newly designed fitness function	160,117	39,337	98.72%	0.13%

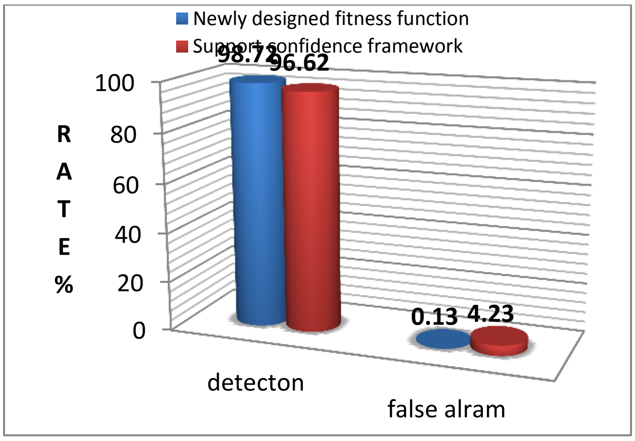

Figure 2 : Experiment-1 using KDD dataset

EXPERIMENT-2

Here in this experiment we have taken real dataset from network of L.D. College of Engineering we trained the genetic algorithm with the training dataset and test our algorithm with the testing dataset from Real network traffic. From our experiment, the detection rate of real data set is 98.72% with low false negative rate.

Table 6: Real DATASET RECORD Distribution

RECORD	Training Dataset	Testing Dataset
Normal	8000	16000
Intrusion	6300	10500
Total	14300	26500

Table 7: Experiment-2 using Real dataset

EXPERIMENT-2 USING REAL DATASET				
Fitness function used in genetic algorithm	TESTING ATTCK	TESTING NORMAL	DR	FPR
Support confidence frame work	11500	15000	95.52	2.03
Proposed newly designed fitness function	10500	16000	97.97	1.14

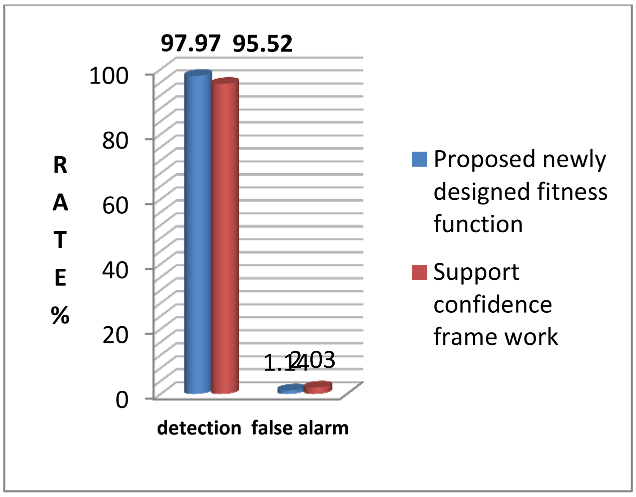

Figure 3 : Experiment-2 using Real dataset

EXPERIMENT-3

In this experiment, we used the training set of KDD99 dataset and "corrected" as testing dataset with our proposed genetic algorithm. In this case the training set contains 493342 records among which 97,250 are normal connection records, while the test set consists of 307700records among which 60,540 are normal connection records. Table 10 shows the distribution of each attack type in the training and the test set. Dataset with our proposed GA algorithm. The result shows in Table 11 that there are four types of attacks. , It means t representation of each attack type is quite different from the normal network activities. However, there are test case 4 (UDP flood) and test case 10 (Ipscan) having low detection rates which are 89.59% and 86.89%, respectively.

Table 8: Distribution of intrusion type in Datasets

DATASET USED	Normal	Probe	DOS	U2R	R2L	Total
Train Dataset ("kddcup.data_10_ percent")	97250	4075	391008	51	958	493342
Test Dataset ("corrected")	60540	3920	224853	228	14189	307700

Table 9: Detection rate for each type of attack

TYPE OF ATTACK	DETECTION RATE	FALSE POSITIVE RATE
DOS	89.1	3.23
PROBE	75.18	2.24
U2R	19.3	1.72
R2L	10.5	1.5

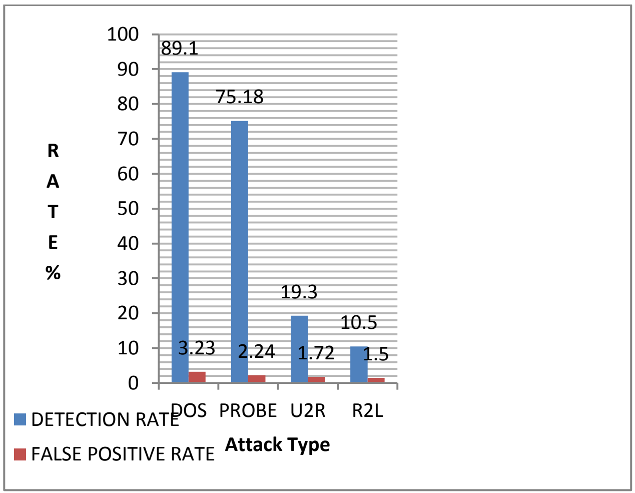

Figure 4 : Detection rate for each type of attack class

IV. CONCLUSION

In our Proposed research work, we have designed a reward penalty based a new fitness function. This function calculates reward and penalty on the basis of strongest record and weakest record in the population. It gives reward to make a chromosome stronger and gives penalty to reduce weakness of chromosomes .so such a healthy chromosomes generates a best fit population which is reducing false alarm rate and increasing a detection rate.

In our work, we have classified a dataset as a normal record or attack record using seven network features and calculated detection rate and false alarm rate. Further we have classified dos, probe, U2R, R2Ltype of attack from attack cluster. We have also calculated intrusion detection rate and false alarm rate for these attack type. We conclude that by implementing this newly designed fitness function we could make system to identify unknown attack and automatically it will learn attack and give global optimal solution.

In future we can combine other artificial intelligent technique with genetic algorithm to make more intelligent intrusion detection system.

V. REFERENCES

[1] R. Bace, and P. Mell, —Intrusion Detection Systemsǁ, National Institute of Standards and Technology NIST, 2001.USA

[2] M. Arvidson and M. Carlbark, "Intrusion Detection Systems: Technologies, Weaknesses, and Trends," 2003.

[3] S. Kumar, —Classification and Detection of Computer Intrusionsǁ, Ph.D. Thesis, Purdue University, 1995.

[4] S. Kumar, and E. H. Spafford, — A Software Architecture to support Misuse Intrusion Detectionǁ, Technical Report CSD-TR-5-009, 1995.

[5] Shelly Xiaonan Wu, Wolfgang Banzhaf, "The use of computational intelligence in intrusion detection system s: A review" Applied Soft Computing 10 (2010) 1–35 published by elsevier .com

[6] RC Chakraborty, Fundamentals of Genetic Algorithm: AI Course, June 2010, available at http://www.myreaders.info/09genetic_Algorithms.pdf

[7] S.Owais, V.Snasel, A.Abraham,"Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques" 7th Computer Information Systems and Industrial Management Applications IEEE, 2008

[8] Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi, "A Software Implementation of a Genetic

Algorithm Based Approach to Network Intrusion Detection", Proceedings of the Sixth International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed (SNPD/SAWN'05) 2005 IEEE

[9] Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis ,"Intrusion Detection System Using Genetic Algorithm/' Proceedings of the *Science and Information Conference 2014 IEEE*

[10] Zorana Bankovic, Dusˇan Stepanovic, Slobodan Bojanic,Octavio Nieto-Taladriz "Improving network security dsfusing genetic algorithm approach" computers and electrical engineering 33(2007)publish by elesevier pvt ltd.

[11] S.Owais, V.Snasel, A.Abraham,"Survey: Using Genetic Algorithm Approach in Intrusion Detection Systems Techniques" 7th Computer Information Systems and Industrial Management Applications IEEE, 2008

[12] V. Moraveji Hashmei, Z. Muda and W. Yassin, "Improving Intrusion Detection using Genetic Algorithm", International Technology journal 12(11) pp. 2167-2173, 2013

[13] Li, Wei, "Using Genetic Algorithm for Network Intrusion Detection", ,Proceedings of the United States Department of Energy Cyber Security Group,(2004)

[14] B. Uppalaiah, K. Anand, B. Narsimha, S. Swaraj, T. Bharat, "Genetic AlgorithmApproach to Intrusion Detection System", IJCST Vol. 3, Issue 1, Jan-March 2012

[15] Miss Priya U. Kadam, Mr. P. P. Jadhav,"An effective rule generation for Intrusion Detection System using Genetics Algorithm" ,International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 10, October 2013

[16] Bharat S. Dhak, Shrikant Lade, " An Evolutionary Approach to Intrusion Detection System using Genetic Algorithm" .ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 2, Issue 12, Dec. 2012