Harsha Verma, 2021, 9:1
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

BFPSO Based Selected Features for Training of Intrusion Detection System

Harsha Verma, Asst prof. Mr. Atul Barve
Department of CSE
Oist Bhopal

Abstract- Digital world increases the dependency of users on communication network infrastructure. Security of network plays an important role for increasing the reliability and flexibility. Many of researchers are working on different network related issues and proposed various algorithms, protocols. This paper has resolved the network intrusion issue by developing a hybrid intrusion detection model of genetic algorithm and neural network. Butterfly Particle Swarm Optimization (BFPSO) was used in the model for finding the feature set that help in dimension reduction of input dataset. Cognitive, social parameter of butterfly increase the velocity updation function efficiency, so feature selection crossover operation in BFPSO get improved. Selected feature set was used for generating the training input matrix of Multiplayer neural network model. Optimized feature set of neural network model has increases the learning capacity of proposed intrusion detection model. Experimental work was done on MATLAB software and NSL-KDD dataset was used. Results shows that proposed model has increases the accuracy of detection.

Keywords: - Intrusion Detection System, Genetic Algorithm, Neural Network, Feature Selection, Network

I. INTRODUCTION

Providing web network protection to many web services on the internet, distinctive network foundations, communications arrange abundant means that has been taken similar to encryption, firewall, and virtual confidential network and so forth systematize Intrusion detection structure is a notable advance among those. Intrusion detection field increases up out of a large amount of current couple of years and built up a immense deal which utilizes the assembled information from a variety of sort of interruption assault, based on those distinguishing business and open source training items emerge to harden your network to improve safety of the miscellaneous correspondence, service providing networks.

As the amount of network customers and machine are growing step by step to give different sort of directions and smoothness for the efficiency of the world. Be that as it may, some unapproved customers or exercises from diverse sorts of attackers which may hidden attack or external attack keeping in mind the final aim to hurt the running structure, which are recognized as programmers. The elementary consideration procedure of such sort of programmer and gatecrashers is to slash down cumbersome networks and web directions.

Because of growth in eagerness of network safety of diverse forms of attack, frequent scientists has added their keenness for their field and broad collection of gatherings and in accumulation result has been formed by them, with a precise end objective to provide protected directions to the end customers. Along with diverse forms of attack interruptions is a sort of attack that build up a business scheme. Interruption detection structure is accessible for the insurance from disruption attack.

From the above exchange this work can lock the chief spot of the network Intrusion detection framework is to recognize all imaginable intermission which execute malicious movement, computer

International Journal of Science, Engineering and Technology

An Open Access Journal

assault, extend of infections, computer abuse, and so forth so a Internet intermission recognition structure investigations distinguishing data packages as well as monitor that movement over the Internet for such sort of spiteful act. So the smooth operation of universal network distinguishing server needs to resolve largely network which go about as network Intrusion detection framework that screen every one of the parcels developments and distinguish their behavior with the harmful exercises.

II. RELATED WORK

Yogitha et. al. [1] presented interruption discovery framework with Support Vector Machine (SVM). Confirmation is ended by organizing surveys on NSL-KDD Cup'99 information gathering which is reformer type of KDD Cup'99 data index. By using this NSLKDD Cup'99 information gathering they have reduced spacious time essential to shape SVM exemplary by attainment appropriate pre-training on information gathering. In this organization SVM made bunching of information. By compulsion suitable part gathering attack location rate is opened up and false positive rate (FPT) is pointed. In this planned work writer has used Gaussian Circular Basis.

A.R. Jakhale, et. al [2] In this exertion the writer depicts a anamoly discovery framework and its two phases chiefly are training and testing. The slipping window and gathering is familiar to tending the web network move by pulling out the recurring examples using computations. The estimation is as authentic and used as a division of regular monitoring. The standard multi-design communicable computation has elevated location rate. At long last, boost the recognition rate and compact the fake aware rate.

Research by **Jiefei, Lobo and Russo [3]** discovers the occasion of Multi-way steered assault where an attack is separated and sent over dissimilar courses to attempt to deception an IDS framework. This is unfair feasible due to multi way TCP (MPTCP) which enables communication to classes finished frequent ways between a foundation and objective.

Barolli et al [4] investigates the consumption of IDS using neural network for giving IDS arrangement in a Tor (The Onion Router) manage. Examinations did use a Tor server and client with back engendering NN to replicate exchanges over the Tor organizes while infectious for assessment. The structure planned is a ready ANN with data taken from Wireshark, at that position the server and client data are examined, and distinctions will identified an interruption or abuse. The conclusion from testing was productive in giving feasible accuracy when charged in the test situation.

ChuanLong [5] In this article, writer examine how to present an interruption recognition framework in light of thoughtful learning, and this exertion offer a thoughtful knowledge approach for interruption recognition using recurrent neural networks (RNN-IDS). In addition, this exertion inspect the execution of the model in balancing categorization and multiclass arrangement, and the amount of neurons and characteristic learning rate impacts on the implementation of the planned show. This effort compare it and those of J48, artificial neural network, arbitrary woodland, bolster vector machine, and further machine knowledge approach planned by history analysts on the standard information directory index.

III. TECHNIQUES OF SPAMMING

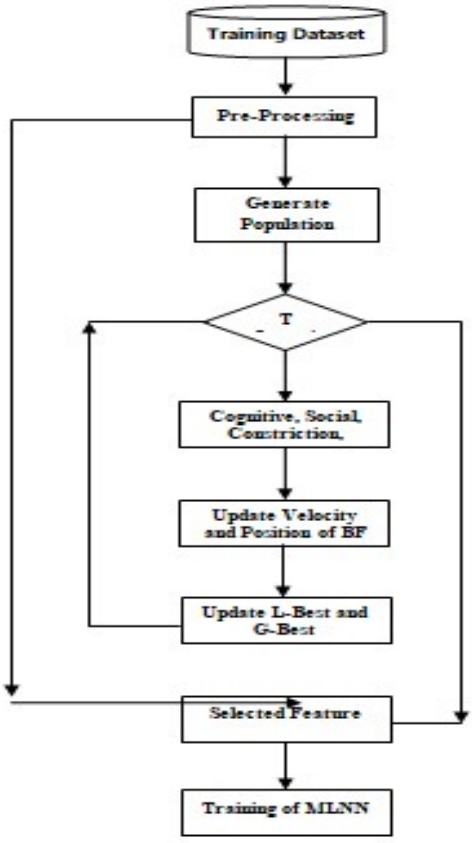

Fig 1. Proposed Work block diagram.

International Journal of Science, Engineering and Technology

An Open Access Journal

Whole work was divided two different modules base on the steps of training intrusion detection system. Fig. 4.1 shows whole proposed training model. First is selection of features for training of neural network by passing input dataset into Butterfly Particle Swarm Optimization (BFPSO). Then in second module learning of selected feature dataset was done by multi layer neural network for Intrusion detection system.

1. Pre-Processing:

As dataset consist of sessions which has normal behavior of network as well as abnormal behavior of network. Here pre-processing was done for the sessions where unnecessary information is remove and data is arrange for training. Instead of direct contribution of crude dataset to chose classifier, crude dataset is preprocessed in various approaches to beat diverse issues like preparing overhead, classifier perplexity, false cautions and recognition rate proportions.

1.1 Generate Population:

In this BFPSO (BFPSO Genetic Algorithm) set of probable solutions act as butterfly, so collection of butterfly is term as population. Hence butterfly having set of features in form of 1 or 0, where 1 means feature is available and 0 means feature not select in this probable solution. Hence butterfly generation function is shown in Eq. 2.

P←Generate Population (M, PD) -----Eq. 2

Where M is number of probable solution and PD is pre-processed dataset.

1.2 Fitness Function:

Estimation of fitness value of probable solution depends upon the input feature training matrix as per P. Hence this function performs training of MLNN from input feature matrix, than testing of trained neural network for same input feature matrix.

1.3 Sensitivity of Butterfly:

$$S = e^{-(M_r - C_r)/M_r}$$

Where S is sensitivity of rth iteration where Mr is maximum number of iterations takes place and Cr is current iteration of this BA-PSO algorithm.

1.4 Cognitive and Social parameters:

$$C_1 = y * \left(\frac{C_r}{M_r} + x\right) \qquad (4.7)$$

$$C_2 = x * \left(\frac{C_r}{M_r}\right) \qquad (4.8)$$

Where x, y are constant range between 0 to 1.

1.5 Constriction Factor C_{eq}

$$\alpha = C_1 + C_2$$

$$C_{eq} = 1 - \alpha - \sqrt{\alpha^2 - 4\alpha} \qquad (4.9)$$

1.6 Inertia Weight:

$$W_t = y + \frac{(M_r - C_r)}{M_r} \qquad (4.10)$$

1.7 Update velocity V and position X of each probable solution:

$$V_{i+1} = C_{eq} * (W_t * V_i + S \quad * (1 - P) * R * C_1 * (L_{best} - C_r) + P * R' * C_2 * (G_{best} - C_r) \qquad (4.11)$$

$$X = R*P*V_{i+1} \qquad (4.12)$$

In above equation V is velocity, X is position while R and R′ are random number whose values range between 0-1. P is probability of nectar for the butterfly selection. So as per X and V values crossover operation were performed.

1.7 Crossover:

In this work population P were modified as per x(t) values which range in 0 to maximum number of feature values n. Here crossover operation generates new combination. So selection of this common parent depends on fitness value. Here best fitness values butterfly act as common parent in all crossover operation. So other set of chromosome undergoes crossover by randomly replacing a feature presence or absence status as per common parent frog set.

1.8 Population Updation:

As crossover changes the chromosome of the population so retention of this butterfly depends on

fitness value. This can be understand if new butterfly have good fitness value as compared to parent frog fitness value than new butterfly was include in the population, otherwise parent butterfly will continue in population. Hence in all situation population size will never change from M number.

2. Multi layer Neural Network:
In this module cluster dataset of sessions are utilize to train the MLNN. Then trained dataset is utilize for the testing of unknown attack session.

2.1 Input Feature: Selected feature from the dataset as per BFPSO genetic algorithm was used to train the neural network Considering first input feature vector which consist of numeric values are arrange in the input matrix. While second desired output vector consist of the class of session which was obtained from the genetic algorithm.

2.2 Training of Multi layer Neural Network (MLNN): Here feature vector obtained are used as the input in the neural network while desired output make proper weight adjustment in the network. So with fix number of iteration or epochs work will get trained neural network

- A layer neural network is assume which have three layers.
- Input layer neuron were identified by I, while hidden layer neuron were identified by j. Output layer neuron is identified by k.
- Weights between neuron is represent by wij, where i and j are neuron layers.

Eq. 13 shows output of neuron as per weight and biasing value bj:

$$X_j = \sum x_i \, . \, w_{ij} \, b_j \text{ ------Eq. 13}$$

where, 1£ i £ n; n is the number of inputs to node j, and bj is the biasing for node j. Hence network will learn the weights between layers. This error need to be correct by adjusting the weight values of each layer. So estimation of error was done by eq. 8 [13].

$$e_k(n) = d_k(n) - y_k(n)$$

The MLNN weight updation was done by above matix of ∂W_i

$$w_{ij} = w_{ij} + \Delta w_{ij}$$

So end of above iteration steps over when error obtained from the output layer get nearer to zero or some constant such as 0.001.

3. Testing of MLNN:
In this step input session is preprocess as done in the training module, similarly selected feature vector is create for input in the neural network. Finally feature vector is input in the MLNN which give session intrusion class. Now analysis of that output is done that whether specified class is desired one or not.

IV. RESULTS

In this experiment comparison of proposed model was done with existing model of IDS proposed in [31]. Both algorithm work on detection of intrusion session from the input testing dataset where in [31] CNN (Cconvolutional neural network was used for the detection of intrusion in the input dataset. Convolutional steps act as pre-processing steps in the model for increasing the learning capacity of the intrusion. Proposed model use butterfly genetic algorithm for the pre-processing of the input dataset. Experiment was done on different size of dataset.

Table 1. Precision value based comparison of IDS.

Data-Set Size	Intrusion Detection System	
	Previous Work [31]	Proposed Model
5000	0.850509	0.97201
7000	0.850831	0.967523
8000	0.847056	0.964344

Above table 1 shows that proposed hybrid model BFPSO and MLNN has increased the precision value. This enhancement was achieved by use of BFPSO for classification of normal and attack sessions.

Above table 2 shows that proposed hybrid model BFPSO and MLNN has increased the recall value. This enhancement was achieved by use of MLNN for further identification of intrusion classes. Use of selected feature values for training of neural network has improved the parameter values.

Table 2. Recall value based comparison of IDS.

Data-Set Size	Intrusion Detection System	
	Previous Work [31]	Proposed Model
5000	0.929764	0.985171
7000	0.934855	0.983871
8000	0.932834	0.983258

Table 3. F-measure value based comparison of IDS.

Data-Set Size	Intrusion Detection System	
	Previous Work [31]	Proposed Model
5000	0.888372	0.978546
7000	0.890866	0.975628
8000	0.887878	0.973709

Above table 3 shows that proposed hybrid model BFPSO and MLNN has increased the precision value. This enhancement was achieved by use of BFPSO for classification of normal and attack sessions.

Table 4. Accuracy value based comparison of IDS.

Data-Set Size	Intrusion Detection System	
	Previous Work [31]	Proposed Model
5000	0.888037	0.977674
7000	0.889622	0.974405
8000	0.886014	0.972253

Above table 4 shows that proposed hybrid model BFPSO and MLNN has increased the accuracy value. This enhancement was achieved by use of MLNN for further identification of intrusion classes. Use of selected feature values for training of neural network has improved the parameter values.

Table 5. Execution time (Seconds) value based comparison of IDS.

Data-Set Size	Intrusion Detection System	
	Previous Work [31]	Proposed Model
5000	50.6476	46.7822
7000	66.2998	64.2385
8000	101.477	100.208

Above table 5 shows that proposed hybrid model BFPSO and MLNN has reduced the testing time value. This enhancement was achieved by use of MLNN for further identification of intrusion classes. Use of selected feature values for training of neural network has reduced the parameter values.

V. CONCLUSIONS

Detection of intrusion in a network is an important issue as number of researchers have proposed different model for its detection. As variety of users have different requirement, so proper identification of safe network is required.

Proposed IDS utilize BFPSO Genetic Algorithm and Multi layer neural network. Genetic algorithm was used for detecting the good feature set from the training dataset. These selected features train the neural network. So this combination of genetic and Neural network increase the detection accuracy of intrusion with less number of training features. Hence raining of MLNN by this reduced feature set has increase the detection of intrusion in network session.

This paper has shows that proposed hybrid model BFPSO and MLNN reduced the testing time value. Experiment was done on real dataset NSL-KDD while comparison was done by existing methods. Results shows that proposed BFPSO&NN model has increase the precision by 12.24%, while accuracy was enhance by 8.91%. In future researcher can reduce training feature vector by other genetic algorithm.

REFERENCES

[1] Yogita B. Bhavasar, Kalyani C. Waghmare "Intrusion Detection System Using Data Mining Technique: Support Vector Machine" 2013 International Journal of Emerging Technology and Advance Engineering volume 3, Issue 3, March 2013.

[2] A.R. Jakhale, G.A. Patil, "Anomaly Detection System by Mining Frequent Pattern using Data Mining Algorithm from Network Flow", International Journal of Engineering Research and Technology, Vol. 3, No.1, January 2014, ISSN. 2278-0181.

[3] Aljurayban, N.S.; Emam, A. (21-23 March 2015). Framework for Cloud Intrusion Detection System Service. Web Applications and Networking (WSWAN), 2015 2nd World Symposium on, p1-5

International Journal of Science, Engineering and Technology

An Open Access Journal

[4] Barolli, Leonard; Elmazi, Donald; Ishitaki, Oda, Tetsuya; Taro; Yi Liu, Uchida, Kazunori. (24-27 March 2015). Application of Neural Networks for Intrusion Detection in Tor Networks. Advanced Information Networking and Applications Workshops (WAINA), 2015 IEEE 29th International Conference on, p67-72.

[5] Koushal Kumar, Jaspreet Singh Batth "Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 150 – No.12, September 2016

[6] R. Karthik, Dr.S.Veni, Dr.B.L.Shivakumar "Improved Extreme Learning Machine (IELM) Classifier For Intrusion Detection System" International Journal of Engineering Trends and Technology (IJETT) – Volume-41 Number-2 - November 2016

[7] Premansu sekhara rath, 2manisha mohanty, 3silva acharya, 4monica aich "optimization of ids algorithms using data mining technique" International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-4, Issue-3, Mar.-2016

[8] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey "Intrusion Detection Using Data Mining Techniques", 978-1-4244-5651-2/10/$26.00 ©2010 IEEE

[9] YU-XIN MENG," The Practice on Using Machine Learning For Network Anomaly Intrusion Detection" Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, 978-1-4577-0308-9/11/$26.00 ©2011 IEEE

[10] Liu Hui, CAO Yonghui "Research Intrusion Detection Techniques from the Perspective of Machine Learning" 2010 Second International Conference on MultiMedia and Information Technology 978-0-7695-4008-5/10 $26.00 © 2010 IEEE

[11] Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He. "A Deep Learning Approach For Intrusion Detection Using Recurrent Neural Networks" current version November 7, 2017. Digital Object Identifier 10.1109/ACCESS.2017.2762418