DESIGN OF EFFICIENT INTRUSION DETECTION SYSTEM WITH PARTICLE SWARM OPTIMIZATION

[1]Rajani S Kadam ,
[1]MEStudent ,
[1]Dept of Computer Engg,
[1]BSCOER,Pune,India

[2]Dr S N Gujar
[2]Professor
[2] Dept of Computer Engg
[2] BSCOER,Pune,India

Abstract: - In the computer age, the speedy growth of computer technology has lead the organizations to deal with huge amount of data every data. The organization with the aim of furnishing the customers, by offering online services has to use the network most of the time. Thus organizations are providing high efforts in securing customer personal data, sensitive data. Computer network security has become the major concern of computer environment because of the rapid growth in the field. Intrusion Detection System (IDS) is one of that tools that tries to protect the systems from an intruders. IDS are more prone to false alarm rate and false positive in high speed networks. Many algorithms are used in design of IDS. Each has their own advantages and disadvantages. Perfect IDS is thus a research topic. The project aims at optimizing the fast learning network for intrusion detection system using particle swarm optimization.

Index Terms:IDS,ANN,PSO,KDDcup99.

INTRODUCTION

We live in the 'computer age', were we deal with various and huge amount of data every day. In today's era of big data business is growing day to day at rapid speed. Each organization is trying to provide service to the customer at their foot step by offering online services. Thus, use of network has grown at rapid rate. The data usage over the network is also grown.

To provide online services, the organizations most of the time make use of customer personal data, confidential data. The growth in the field of computer technology have led to many possibilities, these include many deceiving actions, the systems can be remotely controlled and managed, the gateways can be opened to a fetch the information with online services. Thus due to this rapid growth, computer network security has gained the significant concern in computer environment. Network security at organizational has thus become a chief area of research.

Companies are always under constant pressure, they always need to keep their data safe and secured. The corporate employee or executive is always struggling with many tasks. He need to

- Update and grow as well as perform at the speed that satisfies the investors or shareholders.
- Keep making new products and pioneer himself in providing new services to meet customer demands.

The corporate has to

- make their employees happy, nourish them to become specialists so as to retain them.
- invest in the societies they activate in and also must be careful about their influence.
- provide the business without being affected by cyber attacks or other security issues.

On the other hand, the attackers, who need only little information to build the threat. Thus their growing rate is much faster than defenders. The companies have to struggle a lot with the security tasks. Continuous updating of security tools is required frequently. Security tools need to handle novel threats more effectively. Tools must be reliable and must provide accurate results. Security of data becomes the prime concern.

Goodarzi et al.[1] made the research on the problems that the organizations dealt in safeguarding their information, making it available and reliable. This has argued the motivation for building the systems which provide security from any external system, program, or person aiming at breaking the security line of the network. Many different tools and applications are developed to increase the security of the environments like computers, systems and networks. Intrusion Detection System (IDS) is one among those tools that tries to safeguard the machines from the invaders.
Intrusion Detection System (IDS) is one of that tools that tries to protect the systems from an intruder. IDS are more prone to false alarm rate and false positive. Many algorithms are used in design of IDS. Each has their advantage and disadvantage. Perfect IDS is a research topic. The work aims at optimizing the fast learning network for intrusion detection system using particle swarm optimization.

II. REVIEW ON THE RELATED WORK

Computer data security is the major area of research. Intrusion detection system , a prime tools with detects the intrusion on network is under continuous research. Many different IDS are developed using different algorithms for implementation.

In paper[6] titled "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection" authors Iftikhar Ahmad , Mohammad Basheri et al, compare the performance of 3 prominent algorithms namely Support vector machine, Random Forest and Extreme learning machine for modeling the intrusion detection system.The authors in [7] provide the survey on machine learning and deep learning algorithms for developing intrusion detection systems.MOHAMMED HASAN ALI et al,[10] developed intrusion detection system using fast learning network and optimization algorithm PSO.In [8], CHUANLONG YIN et al, modeled the intrusion detection system using deep learning algorithm named recurrent neural networks.

Machine learning and deep learning algorithms are most widely used for implementation of IDS because of their capacity to learn from actual examples. Each technique for IDS implementation has its own advantage and disadvantage.

III.THE PROPOSED WORK

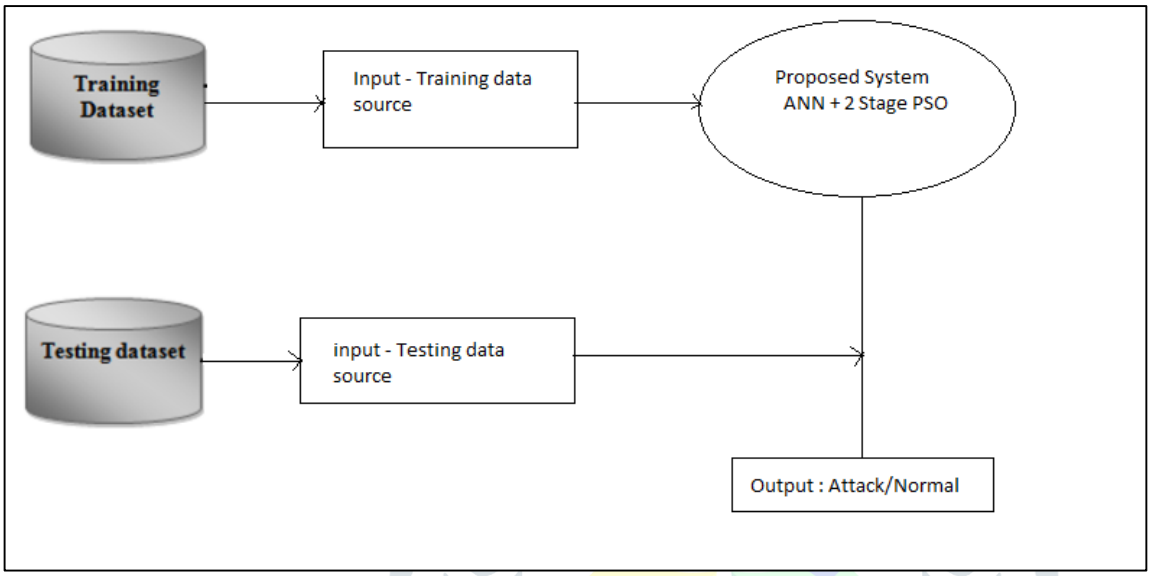

figure 1: proposed system architecture

In the work PSO optimization and ANN algorithm are used. The algorithms are implemented in java. The KDD99 cup data set is used to training and testing of algorithms. The work mainly focuses on optimizing the accuracy of IDS classification by using PSO optimization algorithm. To improve the accuracy, reduce the false alarm rates and false positive rates the optimization is applied. By this, the filtering of data is done so that the selected data are more likely to give the optimal result.

Dataset

The IDS system performance is tested using Dataset instead of live data from the network. KDDcup99 dataset has been used.

KDD Cup99 Dataset.

A dataset is the collection of information related to a particular subject. This consists of many elements that give the information about the particular subject. Dataset are represented in tabular form were each column provides the value of one particular attribute of multiple instance i.e., A particular variable in the dataset is represented by one column. Each column represents values of different samples for one. Thus numbers of column are depending upon number of variables in the dataset variable such as weight, height, color of an object or sample. Particular sample in the dataset is represented by one row. Thus number of rows is depending upon number of samples in the dataset. Each row represents values of different variable for one sample. Each value in the data set is called as Datum.

KDD Cup 99 is most widely used data set for design and evaluation of intrusion detection system. This dataset is built by DARPA 98. The data of DARPA '98 is used in Intrusion detection program. DARPA dataset has near about 400000 entries. There are 41 features and each entry is labeled as normal or an attack(specify attack type). These attacks are classified into 4 groups.

1) Denial of Service Attack (DoS):The DoS is most common form of attack. It is an attack in which the intended users make the machines so busy that the legal requests are denied from service. The attacker send the request to the machine such that either memory or processing system get stuck in serving these illegal requests ceaselessly.

2) User to Root Attack (U2R): This is a attack in which the intruder tries to exploit the resource. The intruder access the system as a normal user(they get access to the system stealing passwords, social engineering or by using some other attack technique) and do some mischievous acts to get the root access to the system.

3) Remote to Local Attack (R2L): the attacker sends the packets to machine remotely but does not have an account on that machine. The attacker does some vulnerable actions to gain access to the machine remotely.

4) Probing Attack: is an action which tries to collect the information all about the computer network for breaking the security of the network to perform some vulnerable actions.

kddcup99 dataset. The data set consists of around 400000 tuples. The dataset contains some numeric values which can be used to identify and define the class to which the instance belong
Sample record of this dataset is shown below to understand how data will look like in dataset

0,udp,private,SF,105,105,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,255,1.00,0.00,0.00,0.00,0.00,0.00,0.00,0.00,snmpgetattack.

There are 41 attributes for each record in the dataset. 42th attribute gives the class type indicating where the record is a normal data or it is a type of attack. In the above record sets the last column indicates the class(eg. Snmpgetattack, normal). The 41 attributes are mentioned below

Attribute Information:

Duration	Su_attempted	Srv_serror_rate
Protocol	Num_root	Srv_rerror_rate
Service	Num_file_creations	Srv_diff_host_count
Src_byte	Num_shells	Dst_host_count
Dst_byte	Num_access_shells	Dst_host_src_count
Flag	Num_outbound_cmds	Dst_host_same_srv_count
Land	Is_hot_login	Dst_host_diff_srv_count
Wrong_fragment	Is_guest_login	Dst_host_same_srv_port_rate
Urgent	Count	Dst_host_diff_srv_port_rate
Hot	Serror_rate	Dst_host_serror_rate
Num_failed_logins	Rerror_rate	Dst_host_srv_serror_rate
Logged_in	Same_srv_rate	Dst_host_rerror_rate
Num_compromised	Diff_srv_rate	Dst_host_srv_rerror_rate
Root_shell	Srv_count	Class

Table 1. KDD99 Dataset Attributes

Flow of the proposed work

1. KDD99 dataset is used
2. Optimization algorithm PSO is applied
3. Filtered data is used for training the model using ANN
4. Test data is used to test the model performance.
5. Performance evaluation is done based on TP,TN,F,FN values

Evaluation Measures

Following 3 metrics will be used to evaluate the performance
P->positive tuples
N->negative tuples
TP->true positives—positive tuples that are correctly labeled
TN->ture negatives---negative tuples that are correctly labeled

FP->false positives—negative tuples that are labeled a positive
FN-> false negatives—positives tuples that are labeled as negative

1. *False alarm rate*: gives the number of outliers misclassified as normal data tuples

   False alarm rate=FP/ (FP+TN)

2. *Accuracy*: percentage of test set tuples that are correctly classified by the classifier

   Accuracy=(TP+TN)/(P+N)

3. *Error rate*: percentage misclassification rate

   Error rate= (FP+FN)/(P+N)

RESULT AND DISCUSSION.

1) One stage PSO with one hidden layer

1-PSO One Hidden Layer Result

TP	TN	FP	FN	Accuracy	FAR
3997	3998	4	5	99.88	0.0009995
3978	3978	24	24	99.4	0.005997001
3991	3992	10	11	99.73	0.002498751
3978	3978	24	24	99.4	0.005997001
3949	3950	52	53	98.68	0.012993503
				99.418	0.005697151

2) Two stage PSO with one hidden layer

2-PSO One Hidden Layer Result

TP	TN	FP	FN	Accuracy	FAR
1363	1364	0	0	100	0
1362	1362	1	2	99.98	0.000734
1357	1358	6	6	99.55	0.004399
1351	1352	12	12	99.11	0.008798
1351	1352	12	12	99.11	0.008798
				99.55	0.004546

3) One stage PSO with two hidden layer

1-PSO Two Hidden Layer Result

TP	TN	FP	FN	Accuracy	FAR
3951	3951	51	51	98.72	0.012743628
3934	3935	67	68	98.31	0.016741629
3912	3912	90	90	97.75	0.022488756
3927	3927	75	75	98.12	0.01874063
3918	3918	84	84	97.9	0.020989505
				98.16	0.01834083

4) Two stage PSO with Two hidden layer

2-PSO Two Hidden Layer Result

TP	TN	FP	FN	Accuracy	FAR
1311	1311	52	53	96.14	0.038151
1310	1310	50	54	96.18	0.036765
1311	1311	52	53	96.14	0.038151
1318	1308	51	51	96.26	0.037528
1311	1311	52	53	96.14	0.038151
				96.172	0.037749

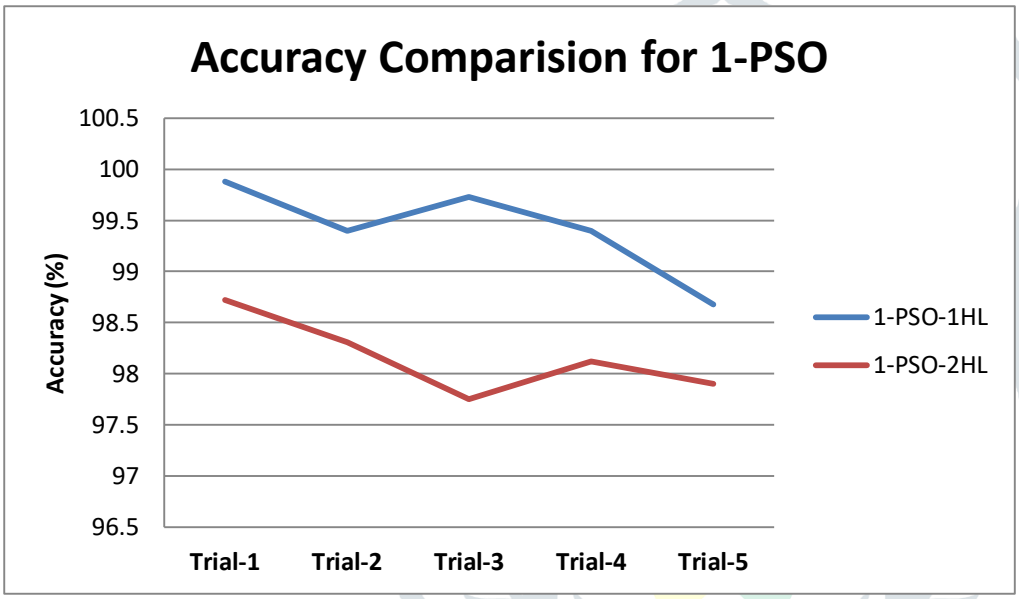

figure2: graph showing the accuracy of 1-stage pso with different hidden layers

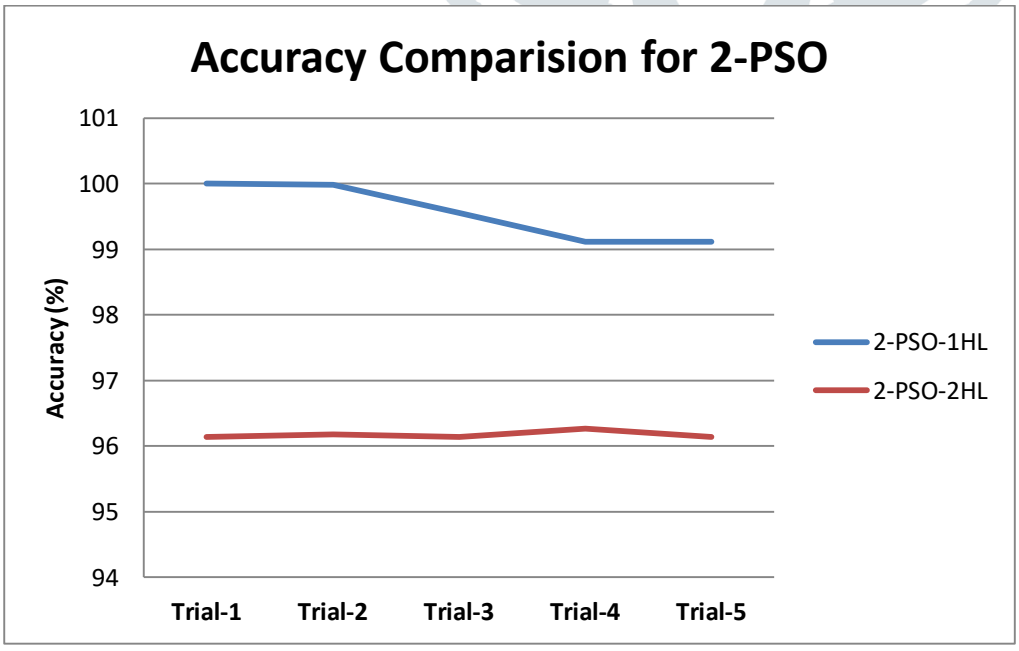

figure 3: graph showing the accuracy of 2-stage pso with different hidden layers

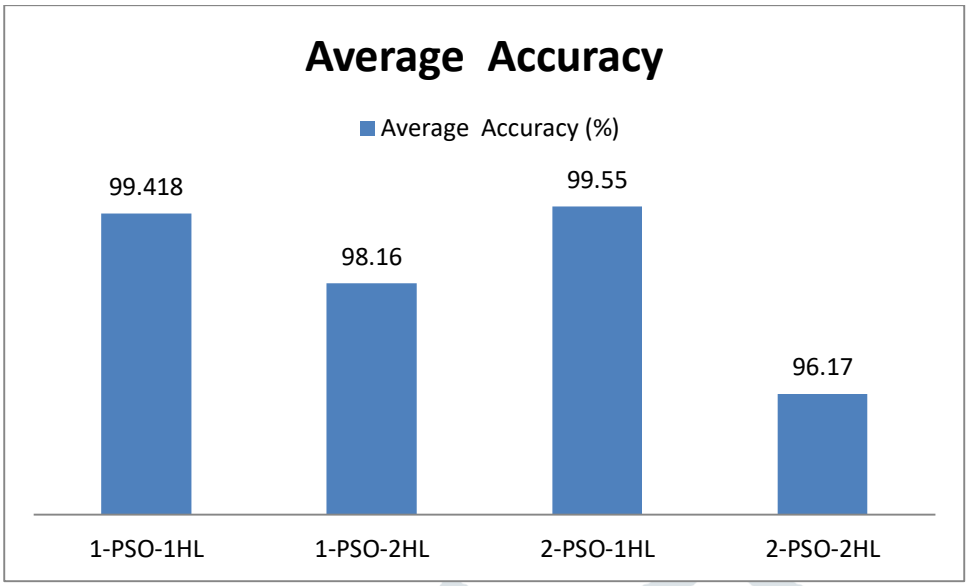

figure 4: graph showing average accuray of different stage pso with diffferent hidden layers

Result shows that PSO provides the solution towards optimal value. ANN with the PSO provides better result. As neurons increase the model gives the better result. With one hidden layer results are up to 99 %. The result gets affected with number of hidden layers. With increase in hidden layer consistent result is obtained.

CONCLUSION

The proposed system is tested with different test cases. ANN with the PSO provides better result. As neurons increase the model gives the better result. With one hidden layer results are up to 99 %. The result gets affected with number of hidden layer. With increase in hidden layer consistent result is obtained.

ACKNOWLEDGMENT

We would like to thank all the reviewers for their reviews, comments and suggestions

REFERENCES

[1] B. G. Goodarzi, H. Jazayeri, and S. Fateri, ‘‘Intrusion detection system in computer network using hybrid algorithms(SVM and ABC),’’ J. Adv. Comput. Res., vol. 5, no. 4, pp. 43–52, 2014.

[2] S. K. Gautam and H. Om, ‘‘Computational neural network regression model for host based intrusion detection system, ’’PerspectivesSci.,vol.8, pp. 93–95, Sep. 2016.

[3] F. A. Anifowose and S. I. Eludiora, ‘‘Application of artificial intelligence in network intrusion detection, ’’ World Appl. Programm., vol. 2, no. 3, pp. 158–166, 2012.

[4] D. E. Denning, ‘‘An intrusion-detection model,’’ in Proc. IEEE Symp. Secur. Priv., vol. 2. Apr. 2012, pp. 118–131.

[5] Fang-Yie Leu , Kun-Lin Tsai , Yi-Ting Hsiao , Chao-Tung Yang , “An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques” IEEE Systems Journal,June2017

[6] Iftikhar Ahmad , Mohammad Basheri, Muhammad Javed Iqbal And Aneel Rahim,” Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection” IEEE access- special section on survivability strategies for emerging wireless networks ,May 2018,

[7] Yang Xin, Lingshuang Kong , Zhi Liu , (Member, Ieee), Yuling Chen,Yanmiao Li, Hongliang Zhu, Mingcheng Gao,Haixia Hou, And Chunhua Wang “Machine Learning And Deep Learning Methods For Cybersecurity”IEEE access, July 2018

[8] Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He," A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks" IEEE access Oct 2017

[9]Waleed Bul'ajoul,Anne James and Siraj Shaikh,"A New Architecture for Network Intrusion Detection and Prevention" IEEE access Dec 2018

[10] Mohammed Hasan Ali , Bahaa Abbas Dawood Al Mohammed ,Alyani Ismail, (Member, Ieee), And Mohamad Fadli Zolkipli," A New Intrusion Detection System Based On Fast Learning Network And Particle Swarm Optimization"IEEE march 2018