RESEARCH ARTICLE OPEN ACCESS

Intrusion Detection System using Self Organizing Map: A Survey

Kruti Choksi*, Prof. Bhavin Shah**, Asst. Prof. Ompriya Kale***
*(Department of Computer Engineering, L.J .Institute of Engineering & Technology, Ahmedabad, India.)
** (M.C.A. Programme, L.J. Institute of Management Studies, Ahmedabad, India.)
*** (Department of Computer Engineering, L.J. Institute of Engineering & Technology, Ahmedabad, India.)

ABSTRACT
Due to usage of computer every field, Network Security is the major concerned in today's scenario. Every year the number of users and speed of network is increasing, along with it online fraud or security threats are also increasing. Every day a new attack is generated to harm the system or network. It is necessary to protect the system or networks from various threats by using Intrusion Detection System which can detect "known" as well as "unknown" attack and generate alerts if any unusual behavior in the traffic.

There are various approaches for IDS, but in this paper, survey is focused on IDS using Self Organizing Map. SOM is unsupervised, fast conversion and automatic clustering algorithm which is able to handle novelty detection. The main objective of the survey is to find and address the current challenges of SOM. Our survey shows that the existing IDS based on SOM have poor detection rate for U2R and R2L attacks. To improve it, proper normalization technique should be used. During the survey we also found that HSOM and GHSOM are advance model of SOM which have their own unique feature for better performance of IDS. GHSOM is efficient due to its low computation time. This survey is beneficial to design and develop efficient SOM based IDS having less computation time and better detection rate.

Keywords **-** Artificial Intelligence (AI), Growing Hierarchical Self Organizing Map (GHSOM), Hierarchical Self Organizing Map (HSOM), Intrusion Detection System (IDS), Network Security, Neural Networks (NN), Self Organizing Map(SOM).

I. INTRODUCTION

Today in the Internet era, Internet has become a routine in our life. Various personal and professional activities are carried out using Internet like online shopping, email, e-commerce, e-learning, e-governance and other. Important transactions and communication are done via Internet and are being attacked in order to know the secret information. Due to these reasons the network security in main concern today. Also according to the Symantec Internet Security Threat Report 2014 [20], 2013 was the Year of Mega Breaches. It has been surveyed by Symantec that there were major eight breaches in 2013, in which each individual breach exposed more than 10 million individual identity thefts, which proved dangerous for many organization and many government bodies, as sensitive data were stolen by the attackers.

An intrusion or attack can be defined as "any set of actions that attempt to compromise the security objectives" [1]. Anderson James P introduced first concept of Intrusion Detection System (IDS) in 1980. In 1984 Fred Cohen mentioned that the percentage of detecting an attack will increase as the traffic increases. Dorothy E. Denning introduced a model of IDS in 1986, which becomes the basic model of the current IDS models [3]. Recently various approaches are adopted to build IDS using different techniques mention in [2] like statistical models, Data Mining Base models, Signature analysis, Rule based systems, Genetic Algorithms, State transition based system, Expert based system and Petri nets. Now a days, the new approach for IDS are neural networks, which are able to detect anomaly base intrusions, while previous techniques where able to detect anomalies but with high false alarm rate. In neural networks one of the approaches is Self-Organizing Maps (SOM), which is proven technique for automated clustering, and visual organization and anomaly detection in IDS [1].

Section II and III provide the introduction about Intrusion Detection System (IDS) and Self Organizing Map (SOM) respectively. In Section IV various network attack are discussed. Section V covers the literature review and comparison of various approaches of IDS using SOM and its model. Section VI consists of comparison analysis and finally section VII provide the conclusion of the survey conducted.

II. INTRUSION DETECTION SYSTEM (IDS)

An intrusion detection system (IDS) is a software application or a hardware that continuously monitors network traffic and/or system activities for abnormal behavior or policy violations and produces logs to an administration unit [2]. The primary aim of IDS is to protect the availability, confidentiality and integrity of critical network information [14]. The working of the IDS is shown in the fig. 1.

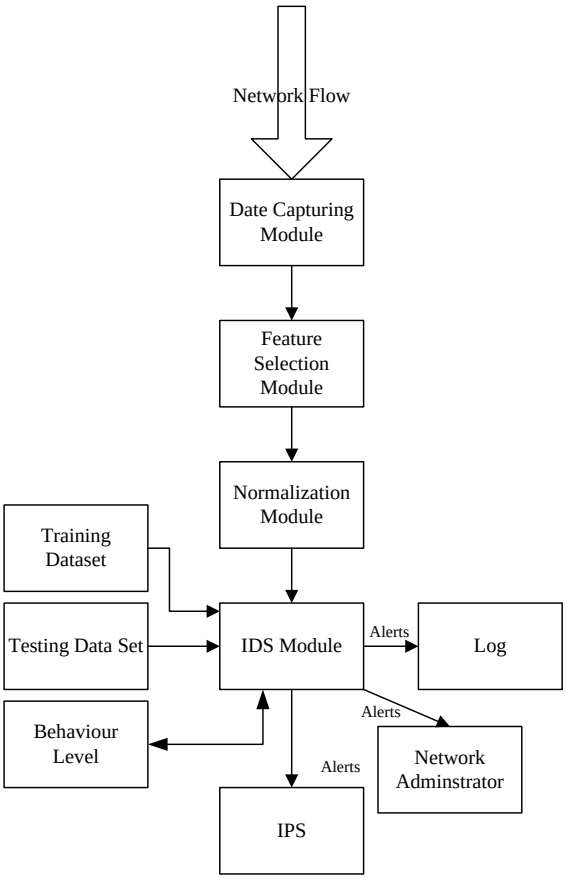

Figure: 1 General View of Intrusion Detection System

There are two main types of IDS (1) Host-based Intrusion Detection System (HIDS) which is designed to detect the attacks at host side. (2) Network-based Intrusion Detection System (NIDS) which is designed to detect attack of whole network. There are two basic techniques to detect an intrusion activity, namely anomaly detection and misuse detection. The anomaly detection technique detects the "known" attacks while misuse detection technique detects the "unknown" attacks.

III. TYPES OF NETWORK ATTACKS

As per the KDD cup dataset [21] there are four major categories of networking attacks are following.

1) Denial of Service (DoS):
It is an attack in which the attacker makes resources too busy in order to prevent legitimate user from using recourses.

2) Remote to Local attacks (R2L):
It is an attack in which attacker do not have the authority to access the system but illegally tries to gain the access.

3) User to Root Attacks (U2R):
It is an attack in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges.

4) Probing or Scanning:
It is an attack in which the hacker scans a machine or a networking order to gain information about the target machine.

IV. SELF ORGANIZING MAP (SOM)

The Self Organizing Map (SOM) is neural network model first described by the Finnish professor Teuvo Kohonen [15] and also referred as a Kohonen Map or Kohonen Neural Network or Kohonen Network. The aim of SOM is to build a topology which preserves the neighborhood relation of the point in the dataset [4]. SOM is used in various applications like image processing, voice recognition, speech recognition, spatial data mapping, data compression, pattern recognition, text mining and so on.

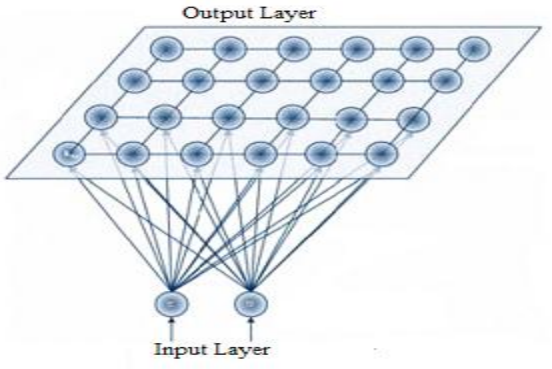

Figure: 2 Model of SOM [4]

Self Organizing Map Neural Network is a feedback network and is able to convert higher dimensional data to low dimensional data. SOM have pre-defined grid of neurons and it adopt competitive learning i.e. neuron with weight vector that is most similar to the input vector is adjusted towards the input vector. The neuron is said winning neuron or Best Matching Unit (BMU) who is found most similar to the input vector. The weights of neuron surrounding BMU are adjusted and the distance between them decreases with time.

SOM is more suitable for real-time data classification due to high speed and fast conversion rate compared to other learning algorithms [3], hence suitable for IDS. As SOM found efficient, it is already used commercially in intrusion detection system [4].

V. LITERATURE REVIEW

During the literature survey, we come across various SOM based model. Model [4], [5] and [13] are traditional SOM while [3] and [12] are of Hierarchical Self Organizing Map (HSOM). And [6], [7], [8], [10], [11], [17], [18] and [19] are of Growing Hierarchical Self Organizing Map (GHSOM). To achieve the objective of the paper we have done the comparison of various models [3], [4], [5], [6], [7], [8], [10], [11], [12] and [13], which are more relevant to our scope of objective. Comparison between them is shown in table 2.

VI. COMPARISON ANALYSIS

The objective of the paper is to find the current challenges of IDS based on SOM. From table 2 we find two major challenges. (1) Poor detection rate of U2R and R2L attacks [6][11][12][13] and (2) Computational Time [3][4][12][13].

1) Poor detection rate of U2R and R2L attacks:
The reasons for poor detection rate of U2R and R2L attacks are following. First, adverse affect of normalization, as it harm the integrity of the data [5]. Second, there is great similarity between the normal flow and R2L attacks [6]. Lastly, KDD cup 1999 dataset have fewer records for U2R and R2L attacks compared to DoS and Probe attacks [21].

The data of normal and R2L attacks are according to some pre-defined standards. For this reason we cannot change data of normal and R2L attacks to make them different from each other. Similarly KDD cup data set is already developed by the UCI KDD Archive, so we cannot make changes in KDD cup 1999 to increase the number of record of U2R and R2L attacks. Here, we can generate own dataset for training and testing but creating new dataset is again a challenging task.

One the solution to improve the detection rate of U2R and R2L attacks is to handle normalization in proper manner. More care should be taken while normalization so that the integrity of the data is not lost and better performance can be achieved. One other hand we can avoid the normalization process to improve the detection rate of U2R and R2L attacks, but this can increase training time.

2) Computation Time:
The reason for more computational time in SOM and HSOM is that, both have more number of neurons, which increase the time of computation.

In SOM, pre-defined grid is used due to that there are few neurons in the grid which are not beneficial. Those are extra neuron or waste neuron, as they do not play vital role in detection but increases the computational time.

In HSOM, in order to have good performance various layers are used in combination. Due to more number of layers, there will be more number of neuron and more computational time. Another drawback of HSOM is that, great investigation is required to select the best layers of combination for best performance of IDS.

To overcome the problem of more computational time, a new approach of Growing Hierarchical Self Organizing Map (GHSOM) is introduced. GHSOM is adaptive nature, the topology grows with input. In GHSOM initial 2x2 grid is define, as the data is input into the grid, it expands horizontally and vertically according to thresholds. Due to this reason the neuron present in GHSOM grid as always useful for detection, not like in SOM predefined grid. GHSOM is better in computational time than SOM and HSOM, as GHSOM topology have neurons which are beneficial for the detection.

Further, discussing about dataset, KDD cup 1999 dataset is the benchmark for intrusion detection system based on SOM from observation. From [16], it can be said that dataset KDD cup 1999 dataset is better from NSL-KDD dataset for SOM. Also feature selection on KDD cup 1999 proved beneficial in order to achieve higher performance in some experiments.

Lastly, in order to build efficient IDS with performance GHSOM approach is proven efficient as it requires less time in comparison of traditional SOM and HSOM from the observation. The issues of poor detection rate in U2R and R2L attacks can be solve by proper normalization of data. From the observation it can be said that by using GHSOM approach and proper normalization technique we can achieve efficient performance with improved detection rate and less computational time.

Table 2: Comparison of SOM models for IDS

Model	**Objective**	**Dataset**	**No of Features**	**Detection Rate /False positive rate**	**Advantages**	**Disadvantage**
SOM[4]	To detect anomalies	DARPA	-	-	Topological Clustering	More no. of neurons in SOM which increase the computational time.
GSOM[5]	To use Grey correlation co-efficient in learning rules for co-relationship of neurons	DAPRA	-	DR: 97.794% (DoS)	Overall Average Detection rate rises by 4.064% compared to SOM	For four individual attacks detection rate is decline of 8.34% than traditional SOM.
HSOM[12]	To built best detector on based machine learning approach using unsupervised learning algorithm	KDD cup 99	41	DR: 90.40% FP: 1.38%	Achieved higher detection using HSOM.	Unable to detect u2r and r2l attack with more accuracy
HSOM[3]	Built for anomaly detection and also to reduce false positive rate	KDD	25	-	HSOM + PBRM increased the detection rate. HSOM showed that having more layers, increases the ability to handle the attacks more effectively.	Requires more investigation for choosing which combination of layer is effective.
GHSOM [11]	Built to improve detection rate of the attacks and train the system by probability labeling method	NSL-KDD	41	DR: 99.68% FP: 0.02%	Achieved the highest detection rate with lowest false positive, hence best overall performance	U2R attacks detection rate is worst amongst all other attacks
GHSOM [10]	GHSOM model with new metric in comporting both numerical and symbolic data is proposed for intrusion detection	KDD cup 99	41	1. DR: 99.99% FP: 5.41% 2. DR: 99.91% FP : 5.44%	Achieved almost high detection rate. Symbolic data is detail differently instead of converting into numerical form	False positive is high
A-GHSOM [6]	To increase the detection rate of unknown attack of ever-changing traffic using GHSOM with four enhancement: threshold based training, dynamic input normalization, feedback-based quantization error threshold adaptation and prediction confidence filtering and forwarding.	KDD cup 99	41	DR: 94.04% FP: 1.80%	Achieved high accuracy in unknown attack with high detection rate and low false positive rate. Topology growth is controlled.	Low detection rate of R2L attacks
GHSOM [7]	To reduce the false alarms	Own dataset	-	FP: 4.70%	Better than SOM in detecting alarm i.e. Reduced false positive from 15% to 4.7% and false negative from 16% to 4%.	Domain expert was required to analyze various scenarios used to identify the false alarms.
E-GHSOM [8]	Built in stable topology of GHSOM with meaningful initialization process, merging BMU to boost and stabilize the final topology and enhancing the training process by splitting criteria	NSL-KDD & Sec-Monet	-	High detection rate and low false positive rate	1. Stable topology 2. Minimum no. Of BMUs 3. High Performance Model	Various different data set and various different instance are take to prove the point but do not gives clear idea about the detection rate and false positive rate. No clear idea about performance.
SOM[13]	To increase the detection rate of u2r and r2l attack using vector hotspot and pruning vectors with SOM	KDD cup 99	41	DR: 13.8% (U2R)	Achieved increase in the detection rate of U2R attacks	Unable to detect R2L attacks

VII. CONCLUSION

This paper gives information regarding IDS using SOM and its models, and also comparison between them. The objective to find current challenge with IDS using SOM is fulfilled. The major challenges are poor detection rate of U2R and R2L attacks and more computational time. These two problems can be handled by proper normalization of dataset and GHSOM approach. This can improve the detection rate of U2R and R2L attack along with less computational time.

DECLARATION

The content of this paper is written by Author 1 (Kruti Choksi) while Author 2 (Prof. Bhavin Shah) had guided the work and Author 3 (Asst. Prof. Ompriya Kale) had reviewed this paper. Hence Author 1 is responsible for the content.

REFERENCES

[1] Kumar, Gulshan, Krishan Kumar, and Monika Sachdeva. "The use of artificial intelligence based techniques for intrusion detection: a review." *Artificial Intelligence Review 34.4 (2010): 369-387.*

[2] Bashir, Uzair, and Manzoor Chachoo. "Intrusion detection and prevention system: Challenges & opportunities." *Computing for Sustainable Global Development (INDIACom), 2014 International Conference on. IEEE, 2014.*

[3] Alsulaiman, Mansour M., et al. "Intrusion Detection System using Self-Organizing Maps." *Network and System Security, 2009. NSS'09. Third International Conference on. IEEE, 2009.*

[4] Pachghare, V. K., Parag Kulkarni, and Deven M. Nikam. "Intrusion detection system using self organizing maps." Intelligent Agent & Multi-Agent Systems, 2009. IAMA 2009. International Conference on. IEEE, 2009.

[5] Wang, Chun-dong, He-feng Yu, and Huai-bin Wang. "Grey self-organizing map based intrusion detection." *Optoelectronics Letters 5 (2009): 64-68.*

[6] Ippoliti, Dennis, and Xiaobo Zhou. "An adaptive growing hierarchical self organizing map for network intrusion detection." *Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International Conference on. IEEE, 2010.*

[7] Mansour, Nashat, Maya I. Chehab, and Ahmad Faour. "*Filtering intrusion detection alarms." Cluster Computing 13.1 (2010): 19-29.*

[8] Salem, Maher, and Ulrich Buehler. "An Enhanced GHSOM for IDS." *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on. IEEE, 2013.*

[9] Ippoliti, Dennis, and Xiaobo Zhou. "A-GHSOM: An adaptive growing hierarchical self organizing map for network anomaly detection." *Journal of Parallel and Distributed Computing 72.12 (2012): 1576-1590.*

[10] Palomo, Esteban J., et al. "A new GHSOM Model applied to network security.*".Artificial Neural Networks-ICANN 2008. Springer Berlin Heidelberg, 2008. 680-689.*

[11] Ortiz, Andres, et al. "Improving Network Intrusion Detection with Growing Hierarchical Self-Organizing Maps." *University of De La Plata, Argentina (2011).*

[12] Gunes Kayacik, H., A. Nur Zincir-Heywood, and Malcolm I. Heywood. "A hierarchical SOM-based intrusion detection system." *Engineering Applications of Artificial Intelligence 20.4 (2007): 439-451.*

[13] Wilson, Ryan, and Charlie Obimbo. "Self-organizing feature maps for user-to-root and remote-to-local network intrusion detection on the KDD cup 1999 dataset." *Internet Security (WorldCIS), 2011 World Congress on. IEEE, 2011.*

[14] Bahrololum, M., E. Salahi, and M. Khaleghi. "Anomaly intrusion detection design using hybrid of unsupervised and supervised neural network."*International Journal of Computer Networks & Communications (IJCNC) 1.2 (2009): 26-33.*

[15] Kohonen, Teuvo. "The self-organizing map." Proceedings of the IEEE 78.9 (1990): 1464-1480.

[16] Ibrahim, Laheeb M., Dujan T. Basheer, and Mahmod S. Mahmod. "A Comparison Study For Intrusion Database (Kdd99, Nsl-Kdd) Based On Self Organization Map (SOM) Artificial Neural Network." *Journal of Engineering Science and Technology 8.1 (2013): 107-119.*

[17] Huang, Shin-Ying, and Yennun Huang. "Network forensic analysis using growing hierarchical SOM." *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, 2013.*

[18] Huang, Shin-Ying, and Yen-Nun Huang. "Network traffic anomaly detection based on growing hierarchical SOM." *Dependable Systems and Networks (DSN), 2013 43rd Annual IEEE/IFIP International Conference on. IEEE, 2013.*

[19] Zolotukhin, Mikhail, T. Hamalainen, and Antti Juvonen. "Online anomaly detection by using N-gram model and growing hierarchical self-organizing maps." *Wireless Communications and Mobile Computing Conference (IWCMC), 2012 8th International. IEEE, 2012*.
[20] Symantec Security Threat Report 2014. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf
[21] KDD Cup 1999 Data.http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html