# An ANN Based Technique to Improve the Detection Rate using Intrusion Detection System

Aakanksha Kori* and Harsh Mathur**

*Research Scholar, Department of Computer Science And Engineering,

IES College of Technology, Bhopal, (Madhya Pradesh), INDIA

**Associate Professor, Department of Computer Science And Engineering,

IES College of Technology, Bhopal, (Madhya Pradesh), INDIA

***Abstract: Nowadays, security of computer network has become major problem in most of everyone's lives. Intrusion Detection Systems monitor computer system to find out sign of security violations over network. When IDSs detects such sign triggers it has to report them to generate the alerts. These alerts tell the user about intruders. The alerts are presented to a human analyst then human analyst evaluates those alerts and initiates an adequate response. In Practice, IDS observe numbers of attacks per day. It also has to deal with different types of attacks. When IDS deals with network intrusions one of the important concerns is to generate true alarms, it means sometime it mistakenly generate an alarm for a legitimate user. Various soft computing techniques are used in Intrusion Detection System. In this paper, we propose a new approach such as Growing Self Organizing Map Algorithm for helping IDS to attain higher detection rate. The proposed approach is performed to detect intrusion has happened or not.***

***Keywords: Intrusion Detection, GSOM, Alert.***

## I. INTRODUCTION

In past years, the use of commercial intrusion detection system (IDS) technology has grown considerably, and IDSs are now standard equipment for large networks. Despite this enormous investment in IDS technology, no comprehensive and scientifically rigorous methodology is available today to test the effectiveness of these systems. There are different types of systems or programs are designed for the monitoring of different types of work of the computer systems. So, another system whose name is intrusion detection system is used in the field of computer networking for the sake of the monitoring of different components of the networks and checks the possibilities of infection of the system and maintenance of the policy of management also. Operations, which are primarily designed to protect the availability, confidentiality, and integrity of critical network information systems. These security management operations protect computer network against denial-of-service attacks, unauthorized disclosure of information, and the modification or destruction of data. Moreover, the automated detection and immediate reporting of these events are required in order to provide the basis for a timely response to attacks security management plays an important role in network management tasks. A secure network must provide the following:

***Data confidentiality*:** Data that are being transferred through the network should be accessible only to those that have been properly authorized.

***Data integrity*:** Data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.

***Data availability***: The network should be resilient to Denial of Service attacks.

## II. OPERATION CATEGORIZATION

Operations can be categorized in two groups: Static and Dynamic.

**Static mechanism**: Static mechanism is analogous to the fence around the premises of a building. In other words, Static mechanism operations are intended to provide barriers to attacks. Keeping operating systems and other software up-to-date and deploying firewalls at entry points are examples of static defense solutions. It is safe to assume that intruders are always one step ahead in finding security holes in current systems.This calls attention to the need for dynamic mechanism. [1]

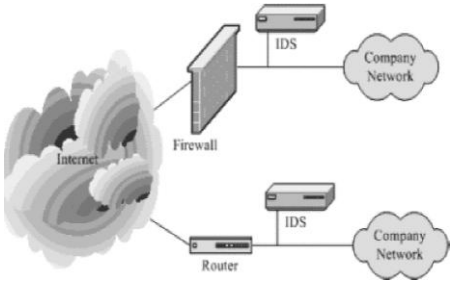


Fig 1: Static Mechanism

***Dynamic mechanism***: Dynamic mechanism is analogous to burglar alarms, which monitor the premises to find evidence of break-ins. IDS based on Dynamic Mechanism are best at detecting the following activities:

***1.Unauthorized outsider access:*** when an unauthorized user logs in successfully, or attempts to log in, they are best tracked with IDS based on Static mechanism.

However, detecting the unauthorized user before their log on attempt is best accomplished with network-based IDS [7].

***2.Bandwidth theft/denial of service:*** these attacks from outside the network single out network resources for abuse or overload. The packets that initiate/ carry these attacks can best be noticed with use of IDS based on Dynamic Mechanism.

## III. IDS FUNCTIONS

Functions of IDS are

- Monitoring users and system activity.
- Auditing system configuration for vulnerabilities and misconfigurations.
- Assessing the integrity of critical system and data files.
- Recognizing known attack patterns in system activity.
- Identifying abnormal activity through statistical analysis.
- Managing audit trails and highlighting user violation of policy or normal activity.
- Correcting system configuration errors
- Installing and operating traps to record information about intruders.

## IV. LITERATURE SURVEY

A lot of research works have been carried out in the literature for intrusion detection sysem(IDS) and some of them have motivated us to take up this research. Brief reviews of some of those recent significant researches are presented below:

**Mansour M. Alsulaiman and et al. [01]** built an Intrusion Detection System using a well known unsupervised neural network, namely Kohonen maps. They proposed two enhancements that were able to solve one of the shortcomings of the available solutions, namely high value of false positive rate. The method called as Performance-Based Ranking Method was used. It works by deleting an input from the dataset and comparing the result before and after the deletion. They used the KDD data set.

**Stefan Axelsson and et al. [02]** counteract the two key deficiencies Low detection rates and a high rate of false alarms, they proposed an interactive detection system based on simple Bayesian statistics combined with a visualization component.

**Iftikhar Ahmad and et al.[03]** provided an approach to analyze denial of service attack by using a supervised neural network. The methodology used sampled data from Kddcup99 dataset, an intrusion database that is a standard for judgment of attack detection tools.

**Antonis Papadogiannakis and et al.[04]** presented selective packet discarding, a best effort approach that enables the Network Intrusion Detection System to anticipate overload conditions and minimize their impact on attack detection.

**Tie and Li [05]** used the BP network with GAs for enhance of Back Propagation algorithm, they used some types of attack with some features of KDD data. The detection rate of IDS for Satan, Guess-password, and Peral was 90.97, 85.60 and 90.79 consequently. The overall accuracy parameter of detection rate is 92.61 with false alarm rate of 7.35.

**Jimmy and Heidar [06]** used feed-forward Neural Networks with Back Propagation training algorithm, they used some feature from TCP Dump and the classification result is 25/25.

**Dima, Roman and Leon[07]** used Multilayer Perceptron algorithm and Radial Based Function (RBF) Neural Network for classification of five types of attacks or intruder, the accuracy rate of classifying attacks is 94.2 using RBF and 94.2 using MLP Neural Network, and the false alarm is 0.9%

**Iftikhar, Sami and Sajjad [08]** used Resilient Back propagation for detecting each type of attack along, the accurse detection rate was 96.93 used Back Propagation Neural Network with many types of learning Mukkamala, Andrew, and Ajith algorithm. The performance of the network is 95.0. The overall accuracy parameter of classification for RPBRO is 98.04 with false positive rate of 3.76% and false negative rate of 0.20.

**Andrew, and Ajith [09]** algorithm. The performance of the network is 95.0. The overall accuracy of classification for RPBRO is 98.04 with false positive rate of 3.76% and false negative rate of 0.20.

Jimmy and Heidar[10] used Neural Network for classification of the unknown attack and the result is 76% correct classification

**Vallipuram and Robert [11]** used back-propagation Neural Network, they used all

features of KDD data, the detection rate for experiment result for normal traffic was 98%, known attacks were 90%, and for unknown attacks were 70%.

**Tich Phu oc Tran[12]** have applied Machine Learning techniques to solve Intrusion Detection problems within network security. Due to complex and dynamic nature of computer networks and hacking techniques, identifying

malevolent activities remains a challenging task for security experts, that is, defense systems that were currently available suffer from low detection potential and high number of false alarms.

## V. IDS METHODOLOGIES

Different detection methodologies can be employed to search for the evidence of attacks. Two major categories exist as detection methodologies: Misuse and Anomaly detection.

### *5.1 Signature-based detection*

Signature-based detection is very effective at detecting known threats but largely ineffective at detecting previously unknown threats. Signature-based detection is the simplest detection method because it just compares the current unit of activity, such as a packet or a log entry, to a list of signatures using string comparison operations. Signature based detection technologies have little understanding of many network or application protocols and cannot track and understand the state of complex communications.[4] In this systems rely on the definitions of misuse patterns, which are the descriptions of attacks or unauthorized actions. A misuse pattern should summarize the distinctive features of an attack and is often called the signature of the attack. In the case of signature based IDS, when a signature appears on the resource monitored, the IDS records the relevant information about the incident in a log file. Signature-based systems are the most common examples of misuse detection systems.

***Advantages of signature-based systems***

1. Very accurate at detecting known attacks, those are included in their signature database. Moreover, since signatures are associated with specific misuse behavior, it is easy to determine the attack type.
2. Typically signature-based approaches Result in fewer false alarms because they can be very specific about what it is they are looking for.
3. Because the IDS is looking for something Known, a lot of information regarding what the misuse is, the potential impact, And how to respond can be provided. This knowledge is extremely important in understanding what is occurring and effectively responding.
4. Efficiency is high and complexity is low.

***Disadvantages of signature-based Systems***

1. Their detection capabilities are limited to those within signature database.
2. As the new attacks are discovered, a signature database requires continuous updating to include the new attack signatures, resulting in potential scalability problems.
3. Many false positives: prone to generating alerts when there is no problem in fact.
4. Cannot detect unknown intrusions.

### *5.2 Anomaly based Detection*

Anomaly detection systems offer several benefits. First, they have the capability to detect insider attacks. For instance, if a user or someone using a stolen account starts performing actions that are outside the normal user-profile, an anomaly detection system generates an alarm. Second, because the system is based on customized profiles, it is very difficult for an attacker to know with certainty what activity it can carry out without setting off an alarm. Third,

an anomaly detection system has the ability to detect previously unknown attacks. An example of this would be if a user logs on and off a machine 20 times a day instead of the normal 1 or 2. Also, if a computer is used at 2:00 AM when normally no one outside of business hours should have access, this should raise some suspicions [3]. At another level, anomaly detection can investigate user patterns, such as profiling the programs executed daily. If a user in the graphics department suddenly starts accessing accounting programs or compiling code, the system can properly alert its administrators. The major benefit of anomaly-based detection methods is that they can be very effective at detecting previously unknown threats.

***Advantage of Anomaly-based system***

Because anomaly-based systems are capable of detecting misuse based on network and system behavior, the type of misuse does not need to be previously known. This allows for the detection of misuse a signature based system may not detect. Anomaly detection can detect novel attacks to increase the detection rate. Compared to supervised approaches, unsupervised approach breaks the dependency on attack-free training datasets. The performance of unsupervised anomaly detection approaches achieve higher detection rate over supervised approach. Also, unsupervised approach have high false positive rate over supervised approach. Using unsupervised anomaly detection techniques, however, the system can be trained with unlabeled data and is capable of detecting previously unseen attacks.

***Disadvantage of Anomaly-based system***

High false-alarm and limited by training data. Obviously, not all typical behaviors are attacks or intrusion attempts. This represents one drawback of intrusion detection methods based on clustering.

### *5.3 Hybrid method*

Through analyzing the advantages and disadvantages

between anomaly detection and misuse detection, a mixed intrusion detection system (IDS) model is designed. [10] First, data is examined by the misuse detection module, and then abnormal data detection is examined by anomaly detection module.

## VI. NEURAL NETWORK BASED METHOD FOR INTRUSION DETECTION SYSTEM

Artificial Neural Networks have been applied to many problems [3][11], and have demonstrated their superiority over classical methods when dealing with noisy or incomplete data. One such application is for data compression. Neural networks seem to be well suited to this particular function, as they have an ability to preprocess input patterns to produce simpler patterns with fewer components. This compressed information (stored in a hidden layer) preserves the full information obtained from the external environment. The compressed features may then exit the network into the external environment in their original uncompressed form. The main algorithms that shall be discussed in ensuing sections are the Back propagation algorithm and the Kohonen self-organizing maps.

### *6.1 Back propagation Neural Network*

The Back propagation (BP) algorithm [12] has been one of the most successful neural network algorithms applied to the problem of intrusion detection system[10]. The data compression problem in the case of the BP algorithm is posed as an encoder problem. The data or image to be compressed passes through the input layer of the network, and then subsequently through a very small number of hidden neurons. It is in the hidden layer that the compressed features of the image are stored, therefore the smaller the number of hidden neurons, the higher the compression ratio. The output layer subsequently outputs the decompressed image to the external environment. It is expected that the input and output data are the same or very close. If the image to be compressed is very large, this may sometimes cause difficulty in training, as the input to the network becomes very large. Therefore in the case of large images, they may be broken down into smaller, sub-images [9]. Each sub-image may then be used to train an individual ANN.

The main disadvantage of Back propagation algorithm is

1. In this technique, the Detection rate is low
2. It take more time for detecting the intrusion
3. It is expensive technique

So, we apply the proposed approach such as GSOM Algorithm that will remove the above disadvantage and improve the compression ratio with quality and provide better result compared to traditional compression algorithm.

## VII. PROPOSED TECHNIQUES

### *7.1 Growing Self Organizing Map Algorithm*

A growing self-organizing map (GSOM) is a growing variant of the popular self-organizing map (SOM). The GSOM was developed to address the issue of identifying a suitable map size in the SOM. It starts with a minimal number of nodes (usually 4) and grows new nodes on the boundary based on a heuristic. By using the value called Spread Factor (SF), the data analyst has the ability to control the growth of the GSOM.

All the starting nodes of the GSOM are boundary nodes, i.e. each node has the freedom to grow in its own direction at the beginning. New Nodes are grown from the boundary nodes. Once a node is selected for growing all its free neighboring positions will be grown new nodes. In GSOM, input vectors are organized into categories depending on their similarity to each other. For data compression, the image or data is broken down into smaller vectors for use as input. For each input vector presented, the Euclidean distance to all the output nodes are computed. The weights of the node with the minimum distance, along with its neighboring nodes are adjusted. This ensures that the output of these nodes is slightly enhanced. This process is repeated until some criterion for termination is reached. After a sufficient number of input vectors have been presented, each output node becomes sensitive to a group of similar input vectors, and can therefore be used to represent characteristics of the input data. This means that for a very large number of input vectors passed into the network, (uncompressed image or data), the compressed form will be the data exiting from the output nodes of the network (considerably smaller number). This compressed data may then be further decompressed by another network. We take 50 neuron as a one input hidden layer and one output layer we take learning rate 0.5.the compression and decompression figure of GSOM Algorithm are following

### 7.2 Learning Algorithm of the GSOM:

The GSOM process is as follows:

**1. Initialization phase:**

1. Initialize the weight vectors of the starting nodes (usually four) with random numbers between 0 and 1.
2. Calculate the growth threshold ($GT$) for the given data set of dimension $D$ according to the spread factor ($SF$) using the formula $GT = -D \times \ln(SF)$

**2. Growing Phase:**

1. Present input to the network.
2. Determine the weight vector that is closest to the input vector mapped to the current feature map

(winner), using Euclidean distance. This step can be summarized as: find $q'$ such that $|v - w_{q'}| \leq |v - w_q| \, \forall q \in \mathbb{N}$ where $v$, $w$ are the input and weight vectors respectively, $q$ is the position vector for nodes and $\mathbb{N}$ is the set of natural numbers.

3. The weight vector adaptation is applied only to the neighborhood of the winner and the winner itself. The neighborhood is a set of neurons around the winner, but in the GSOM the starting neighborhood selected for weight adaptation is smaller compared to the SOM (localized weight adaptation). The amount of adaptation (learning rate) is also reduced exponentially over the iterations. Even within the neighborhood, weights that are closer to the winner are adapted more than those further away. The weight adaptation can be described by $w_j(k+1) = \begin{cases} w_j(k) & \text{if} j \notin \mathrm{N}_{k+1} \\ w_j(k) + LR(k) \times (x_k - w_j(k)) & \text{if} j \in \mathrm{N}_{k+1} \end{cases}$ where the Learning Rate $LR(k)$, $k \in \mathbb{N}$ is a sequence of positive parameters converging to zero as $k \to \infty$. $w_j(k)$, $w_j(k+1)$ are the weight vectors of the node $j$ before and after the adaptation and $\mathrm{N}_{k+1}$ is the neighborhood of the winning neuron at the $(k+1)$th iteration. The decreasing value of $LR(k)$ in the GSOM depends on the number of nodes existing in the map at time $k$.

4. Increase the error value of the winner (error value is the difference between the input vector and the weight vectors).

5. When $TE_i > GT$ (where $TE_i$ is the total error of node $i$ and $GT$ is the growth threshold). Grow nodes if i is a boundary node. Distribute weights to neighbors if $i$ is a non-boundary node.

6. Initialize the new node weight vectors to match the neighboring node weights.

7. Initialize the learning rate ($LR$) to its starting value.

8. Repeat steps 2 – 7 until all inputs have been presented and node growth is reduced to a minimum level.

**3. Smoothing phase.**

Reduce learning rate and fix a small starting neighborhood. Find winner and adapt the weights of the winner and neighbors in the same way as in growing phase.

## VIII. CONCLUSION

In this paper, we present an efficient technique for intrusion detection by making use of Neural Network based Growing Self Organizing Map Technique. The proposed method will be able to detect the attack on the basis of the behavior of the basic features of network. The proposed method used feature extraction and ranking based feature selection. Errors will be removed in proposed method by using neural network algorithm.

## REFERENCES

[1] Elike Hodo, Xavier Bellekens, "Threat Analysis of IoT networks Using Artificial Neural Network Intrusion Detection System" .

[2] Mansour M. Alsulaiman, "Statistical traffic modeling for network intrusion detection," in Proc. Model., Anal. Simulator. Computer. Telecommunication. Syst., 2000, pp. 466–473

[3] Stefan Axelsson, "Adaboost with Single and Compound weak classifier in Network Intrusion Detection", In Proceedings of International conference on Advanced computing, Networking & Security" , Vol. 1, pp 282-290, Dec-2

[4] Iftikhar Ahmad, "Rough set Based Personalized Recommendation in Mobile Commerce. 2009 International conference on Active Media Technology", Lecture Notes in Computer Science, 370-375, 2009

[5] Antonis Papadogiannakis, "Application of machine learning algorithms to KDD intrusion detection dataset with in misuse detection", context. In Proceedings of the international conference on machine learning: Models, technologies, and applications. pp. 209–215, 2003

[6] Tie and Li,"A Decision-theoretic generalization of online learning and an application to boosting," J. Computer. Syst. Sci., vol. 55, no. 1, pp. 119–139, Aug. 1997.

[7] Jimmy and Heidar, "Feature deduction and ensemble design of intrusion detection systems," Computer. Security., vol. 24, no. 4, pp. 295–307, Jun. 2005

[8] Dima, Roman and Leon, "Network-based anomaly intrusion detection system using SOMs," in Proc. IEEE 12th Signal Process. Communication. Appl. Conf., Apr. 2004, pp. 76–79.

[9] Iftikhar, Sami and Sajjad, "Results of the kdd99 classifier learning contest," SIGKDD Explor., vol. 1, no. 2, pp. 63–64, 2000.

[10] Andrew, and Ajith, "An intrusion detection method based on rough set and SVM algorithm," in Proc. Int. Conf. Communication., Circuits Syst., Jun. 2004, vol. 2, pp. 1127–1130

[11] Jimmy and Heidar, "Evolutionary neural networks for anomaly detection based on the behavior of a program," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 36, no. 3, pp. 559–570, Jun. 2006.

[12] Vallipuram and Robert, Network intrusion detection using an improved competitive learning neural network," in Proc. 2nd

[13] Annu. Conf. Communication. Network. Serv. Res., vol. 4, pp. 190–197., May 2004

[14] Tich Phu oc Tran , "Anomaly Detection based on Machine Learning: Dimensionality Reduction using PCA and Classification using SVM" , International Journal of Computer Applications (0975 – 8887) Volume 47– No.21, pp. 58-65, 201213. Gary Stein, "Using data mining to discover signatures in network-based intrusion detection," in Proc. Int. Conf. Mach. Learn. Cybern., vol. 1, pp. 13–17., 2002

[15] A.Y. Ng, M.I. Jordan and Y.Weiss, "On Spectral Clustering: Analysis and an Algorithm," IEEE Advances in Neural Information Processing Systems,vol. 14, pp. 849-856, 2002

[16] Leon, O. Nasraoui, J. Gomez, "Anomaly Detection Based on Unsupervised Niche Clustering with Application to Network Intrusion Detection," *IEEE Congres on Evolutionary Computation*, 2004.

[17] Zhang, Z. Xiong, X. Wang, " Disstributed Intrusin Detection Based on Clustering," IEEE Proc. of fourth Internation Conference on Machine Learning and Cybernetics, Guangzhou, Aug. 2005.