A Technical Review on Intrusion Detection System

Sejal K. Patel
Student at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Umang H. Mehta
Student at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Urmi M. Patel
Student at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Dhruv H. Bhagat
Student at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Pratik Nayak
Teaching Assistant at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Ankita D. Patel
Teaching Assistant at department of computer science and technology
Uka Tarsadia University, Bardoli, India

Abstract— Network security is a big issue at present for large organizations. There are many types of Intrusion Detection Systems like Host Based IDS, Network Based IDS, Anomaly Based IDS and Misuse Based IDS etc. There are various methods to implement IDS that are based on some artificial intelligence concepts which improve performance of IDS. Some of those techniques are Neural Network, Data Mining, Genetic algorithm, Fuzzy Logic etc. The paper reviews these techniques and their comparison in brief.

Keywords- Intrusion Detection Systems, Neural Network, Data Mining, Genetic algorithm, Fuzzy logic.

I. Introduction

Internet is a part of daily life nowadays. People use it for business, entertainment, education and so on. Security of using Internet is a big concern. There are many risks of network attacks when using the Internet. There are various systems designed to detect and block the Internet-based attacks. Specifically, IDSs helps the network to prevent from external attacks. The goal of IDSs is to prevent from attacks of computer systems on Internet. IDSs can be used to detect different types of malicious network communications and computer systems usage, whereas the firewall cannot perform this task.

An IDS is a system that monitors network or system activities for malicious activities or policy violations and produces reports to the management [1]. It is used to protect our network: from the outside environment, from the inside as well.

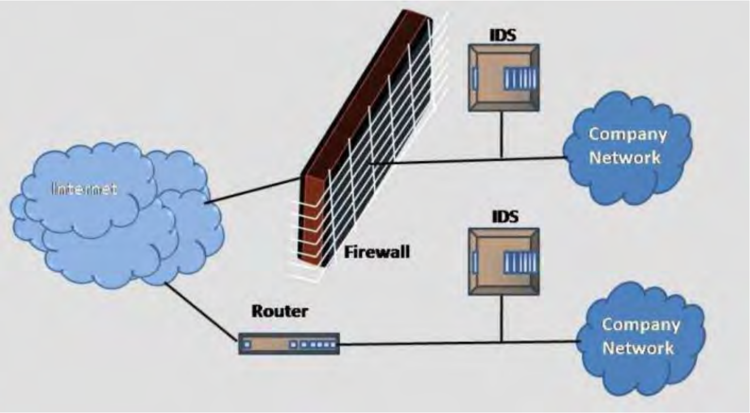

Figure 1 Typical location of an intrusion detection system [2]

Figure 1 describes how intrusion detection can be done. The information is retrieved from the internet, then it is checked by the firewall and finally it is protected by the IDS. After that it sends the information to the corresponding network.

There are various types (classifications) of IDSs described in [3], and which are as below:

Host Based Intrusion Detection System: It monitors the system and tracks changes done to important files and directories. Snap-shot of existing system files are taken and then they are matched with the previous snap-shot. If modification or deletion is found in critical system files, then the alert is sent to the administrator

Network Based Intrusion Detection System: It monitors and resolves traffic on its network segment to detect intrusion attempts. IDS may be made of few or many sensors. Each sensor monitors the traffic passing through its own segment. The sensors cannot monitor anything outside their own segment.

Misuse based Intrusion Detection System: It contains a database of know vulnerabilities. It monitors traffic and does pattern matching. It works similar to virus scanner, by searching for a known identity or signature for each specific intrusion event. It can be placed on a network to watch the network vulnerabilities or can be placed on a host.

Anomaly based Intrusion Detection System: It is also known as behavior-based system. It works on the fact that detection of intrusion can be done by observing variation from the expected behaviors of the monitored system. These “normal” behaviors can correspond to some observations done in past or to some preconceptions made by various techniques. Things that don’t match to “normal” can be said anomalous. The main process of such IDS is not for learning what is anomalous, instead it is for learning what is normal or expected. The main advantage of such system is that, it can detect unknown attacks, but, at the same time its disadvantage is that it leads to very high false alarm rate.

II. LITERATURE REVIEW

Traditional IDSs have many limitations like, time consuming statistical analysis, regular updating, non-adaptive, flexibility and accuracy. Various data obtaining methods like real traffic, sanitized traffic and simulated traffic can be used for obtaining data. Mainly IDS were tested on a standard dataset KddCup99. Training overheads (time consuming, regular update and unable to detect novel attack) and less optimization in performance (false positive, false negative, detection rate) are current issues in IDS using ANN [4].

Various methods can be used to develop IDS like, Data set generation, Train data, and Real-time prediction. In this way, a new system can be proposed in which data can be used as Real Time data set. The system can be used to implement the Real Time Host based attacks. This system is useful in online environment. Existing systems for Intrusion Detection are mostly for offline data and as a solution of it one system can be made, which might be able to work online [5].

From [6], it can be said that anomaly detection using neural network is also preferable. “Behavior” can be taken as parameters in anomaly intrusion detection using a back-propagation neural network. It can be observed if a neural network classifies normal traffic correctly, and detect known and unknown attacks without using a huge amount of training data. For the training and testing of the neural network, DARPA Intrusion Detection Evaluation data sets can be used. By doing experiments, it is possible to get classification rate of 88% for known and unknown attacks. Steps can be followed to build ANN based Anomaly Detection System which are like, Compose Training/testing dataset, Pre-process training/testing dataset, Determine the neural network structure, Train neural network, Test neural network. Neural networks can successfully be used as a method for training and learning an Intrusion Detection System. The main problem with today’s Intrusion Detection System is that they produce many false alarms, and this takes up much of a system administrator’s time and resources.

Moreover, when developing an IDS or ADS, it's not necessary to assign huge amount of training data to Neural Network to classify traffic correctly.

There are some limitations of Existing system like, it is based on off-line system which was less efficient and accurate and it's using static off-line database. An algorithm can be used to develop ANN is further divided into two parts: Training & Testing. If a system is developed based on proposed algorithm, it can have several advantages to IDS like, it can detect almost all types of attacks (intrusion), more secure, reliable, accurate and efficient than previous one, and execution time can be less [7].

Confidentiality, integrity and availability of the system can raise weakness of system to security threats, attacks and intrusion. There are different types of neural network used for intrusion detection namely, Self-Organizing feature map, Multilayered feed forward neural network, Elman back propagation, Cascaded forward back propagation. The dataset used for intrusion detection is the KDD CUP 99 data set. There are different types of attacks in KDD 99 dataset like, Denial of service attack, Prob, User to root attack (U2R), Remote to user attack (R2U). ELBP intrusion detection system shows very good classification rate, smaller false positive and false negative rates as compared to MLFF, CFBP, SOFM [8].

An ultimate goal of the IDS is to achieve accuracy. To do so, some machine learning techniques are needed to be created so that performance of the system can get improved. In [9], some commonly used machine learning techniques are discussed, which are Pattern Classification, Single Classifiers, Hybrid Classifiers and Ensemble Classifiers.

1. Pattern Classification:

It is an action to take raw data and activity on data category. The methods of supervised and unsupervised learning can be used to solve various pattern recognition problems. Supervised learning is based on using the training data to create a function, in which each of the training data contains a pair of the input vector and output (class label). The learning (training) refers to computation of distance between the input and output examples to create a classifier (model). When classifier is created, it can classify unknown examples into a learned class labels.

2. Single Classifiers:

Intrusion detection can be done using one single machine learning algorithm. Various machine learning techniques (for example, k-nearest neighbor, support vector machines, artificial neural network, decision trees, self-organizing maps, etc.) can be used to solve these problems.

3. Hybrid Classifiers:

The idea behind a hybrid classifier is to combine several machine learning techniques to improve performance of system. A hybrid approach consists of two functional components. The first one takes raw data as input and generates intermediate results. The second one takes the intermediate results as the input and produces the final results. The first level of hybrid classifiers can be based on supervised or unsupervised learning techniques. Hybrid classifiers can also be based on the integration of two different techniques in which the first one aims at optimizing the learning performance of the second model for prediction.

4. Ensemble Classifiers:

Such classifiers were proposed to improve the classification performance of a single classifier. "Ensemble" means combination of multiple weak learning algorithms or weak learners. The weak learners are trained on different training samples so that it can help to improve overall performance.

The genetic algorithms are a viable method for the detection of malicious intrusions. The comparison between various learning techniques will allow software professionals to find best machine learning technique to find clear, unambiguous knowledge about intrusion detection more effectively and efficiently.

IDSs can be implemented using one of the two techniques as described in [10], which are Anomaly Detection and Signature Detection. Methods like Statistical technique, Machine learning and knowledge based methods are used for the anomaly detection. Machine learning methods like Neural Network, Support Vector Machine, Genetic algorithm, Clustering, Fuzzy Logic can be used to implement IDS. Signature detection has high accuracy to detect known attack, but disadvantage is that they cannot detect any unknown attack. Signature detection uses pattern of unauthorized behavior to predict and detect similar subsequent attempt. These specific patterns are called as Signatures. Signature detection technique is efficient for the medium sized network which catches the standard threats. Anomaly detection technique is efficient for the large and diverse network, but it is more expensive technique to implement the IDS.

Genetic algorithm is programming technique that copy of biological evolution. It is based on Darwinian's principal of evolution. In [11], genetic algorithm is introduced for implementing IDS. KDD dataset was taken for performance measurement and detection rate was quite good. Various factors can be taken in account when genetic algorithm is to be used like, Fitness function, Representations of individuals, GA parameter. GA based algorithm classify all types of attack using training dataset with very low false positive rate.

Data mining can be helpful in implementation of IDS due to its abilities like: Manage firewall rules for anomaly detection, Analyze large volumes of network data, Same data mining tool can be applied to different data sources, Performs data summarization and visualization, Differentiates data that can be used for deviation analysis, Clusters the data into groups such that it possesses high intra-class similarity and low inter-class similarity [12].

As described in [13], Network Intrusion Detection Systems suffer from difficulties when its log files are high scale and dimensions so that new methods need to be developed for processing these huge data sources. Efficiency in terms of accuracy is one of the most critical measurements which are mostly defined by ratio of false positive and false negative alarms. Therefore, it is needed to design efficient algorithms whereas scan data once and extract hidden patterns inside it. Evolving data, visiting data once, accuracy in intrusion detections and space limitations are major issues in intrusion detection systems.

III. METHODS

Various processing mechanisms (methods) are there which can be used together with IDS. A brief of each is as below:

A. Neural Network:

Artificial neural networks (ANN) were developed as generalizations of mathematical models of biological nervous systems. After the introduction of simplified neurons by McCulloch and Pitts in 1943, people started taking interest in neural network. The main processing elements of neural networks are called artificial neurons (also called neurons or nodes). In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm [14]. Neural networks use learning algorithms to learn about connection between input and output vectors, and to generalize them for extracting new input/output relationships. The main objective of using neural network approach into intrusion detection is to learn the behavior of actors in the system [15].

B. Genetic Algorithm:

Genetic algorithm is an adaptive heuristic search method based on population genetics. Genetic algorithm was introduced by John Holland in the early 1970s [16]. The genetic algorithm is used to derive a set of classification rules from network audit data. The support-confidence framework is utilized as fitness function to determine each rule’s quality. The generated rules are used to detect or classify intrusions in real-time environments [17].

C. **Data Mining:**

Data mining is the research and analysis of large data sets, to discover meaningful pattern and rules. The main idea is to find effective way to combine the computer’s power to process the data with the human eye’s ability to detect patterns. The goal of data mining is to design and work efficiently with large data sets. Data mining is the component of wider process called knowledge discovery from database [18]. It is a set of techniques that use the process of extracting data which were unknown previously but useful from large stores of data. This method excels at processing large audit data or system-logs. They are not much useful for stream-analysis of network-traffic. One of the main data mining techniques used in intrusion detection is decision trees. Such models allow one to detect anomalies in large databases. Another technique is segmentation that allows pattern extraction for unknown attacks. A typical data mining technique is associated with finding association rules. It allows extracting previously unknown knowledge on new attacks or built on normal behavior patterns. Anomaly detection frequently generates false alarms. It is easy to correlate data related to alarms with mined audit data using data mining, so that it reduces the rate of false alarms [14].

D. Fuzzy Logic:

As described in [19], Fuzzy logic is an extension of Boolean logic that is often used for computer-based complex decision making. Fuzzy logic is used in intrusion detection since last two decades because it enables to deal with uncertainty and complexity which is derived from human reasoning. By the help of fuzzy variables or linguistic terms, intrusion detection features can be viewed easily and decision of normal and abnormal activity in the network are based on its fuzziness nature that can identify the degree of maliciousness of a node instead of yes or no conditions. IF-then-else based fuzzy rules are used to define all situations in the network for identifying the attacks or intrusions [20].

IV. METHOD COMPERISION

All the methods described for IDS have some pros and cons, which are as below:

Table 1: Method Comparison of IDS

Method	Pros	Cons	Reference
Neural Network	-Able to perform tasks that a linear program can't. -Even if an element of the neural network fails, it can continue without any problem due to their parallel nature. -Learns and does not need to be reprogrammed. -Can be implemented in any application without problems.	-Needs training to get operated. -It's architecture is different from the architecture of microprocessors therefore it needs to be emulated. -High processing time is required for large neural networks.	[21]
Genetic Algorithm	-Genetic algorithms are intrinsically parallel. -Parallelism allows genetic algorithm to implicitly evaluate many schemas at once. This make them well suited to solving problems where space of solution is huge. -Genetic algorithm based systems can be re-trained easily. It improves its possibility to add new rules and evolve intrusion detection system.	-Crossover rate is less. -Mutation rate is high. -The method of selection should be appropriate. -Writing of fitness function must be accurate.	[22], [23]
Data Mining	-Remove normal activity from alarm data to allow analysts to focus on real attacks. -Identify false alarm generators and "bad" sensor signatures. -Find anomalous activity that uncovers a real attack. -Identify long, ongoing patterns (different IP address, same activity).	-Because data mining in Intrusion Detection is a new concept, there can be obstacles in developing effective solution. -When there is huge amount of data, data mining can become quite computationally expensive.	[23], [24]
Fuzzy Logic	-Reasoning is approximate rather than precise. -Effective, especially against port scans and probes.	-High resource consumption Involved. -Reduced, relevant rule subset identification and dynamic rule updation at runtime is a difficult task.	[25]

V. CONCLUSION

In recent years, research on various techniques is being done to improve network security. Intrusion Detection System is an assured mechanism to secure the network but having limitations like, it is ineffective to update the audit data rapidly it involves human interference thus reduces the performances. There are so many methods (techniques) like, Neural Network, Genetic Algorithm, Data Mining and Fuzzy Logic. By studying these methods, it can be concluded that Neural Network is preferable approach over others. Neural networks have ability to derive meaning from complicated or unclean data that can be used to extract patterns and detect trends that are too complicated to be noticed by humans or other computer techniques, which other methods don't have. A trained neural network can work as an expert for analyzing data. Neural Network based IDS can learn to recognize users by what commands they use and how frequently and by using such an identification, intrusions can be detected in a network computer system. Neural Network Based IDSs are easy to train, also inexpensive to run because it operates off-line on daily logs. It formulates a promising and practical approach to intrusion detection in case where real-time detection is not required.

VI. REFERENCES

[1] Ms. Ruth D, Mrs. Lovelin Ponn Felciah M "A Survey on Intrusion Detection System with Data Mining Techniques" IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 3, May 2014.

[2] http://en.wikipedia.org/wiki/Intrusion_detection_system

[3] Devikrishna K S, Ramakrishna B B "An Artificial Neural Network based Intrusion Detection System and Classification of Attacks" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 3, Issue 4, Jul-Aug 2013, pp. 1959-1964.

[4] Iftikhar Ahmad,Azween B Abdullah, Abdullah S. Alghamdi "Artificial Neural Network Approaches to Intrusion Detection: A Review " Telecommunications And Iformatics book as ACM guide Included in ISI/SCI Web of Science and Web of Knowledge (2009).

[5] TejaswiniBadgujar, Prof. Priyanka More "A Review for an Intrusion Detection System Combined with Neural Network" International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 4, Issue 3, March 2014 (2014).

[6] Manoranjan Pradhan, Sateesh Kumar Pradhan, Sudhir Kumar Sahu "Anomaly Detection Using Artificial Neural Network" International Journal of Engineering Sciences & Emerging Technologies, Volume 2, Issue 1,April 2012 (2012).

[7] P. D. Somwanshi, S. M. Chaware "A Review on: Advanced Artificial Neural Networks (ANN) approach for IDS by layered method" P. D. Somwanshi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5129-5131 (2014).

[8] Afrah Nazir "A Comparative Study of Different Artificial Neural Networks Based Intrusion Detection Systems" International Journal of Scientific and Research Publications, Volume 3, Issue 7, July 2013 (2013).

[9] Manju Khari, Anjali Karar "Analysis on Intrusion Detection by Machine Learning Techniques: A Review" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

[10] Preeti Yadav, Divakar Singh "A Review on Network Intrusion Detection System" International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 9- Sep 2013.

[11] Mohammad Sazzadul Hoque, Md Abdul Mukit, Md Abu Naser Bikas "An Implementation Of Intrusion Detection System Using Genetic Algorithm" International Journal of Network Security & Its Applications (IJNSA), Vol4, No2, March 2012.

[12] Ms.Radhika S.Landge, Mr.Avinash P.Wadhe "Review of Various Intrusion Detection Techniques based on Data mining approach" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 3, Issue 3, May-Jun 2013.

[13] Madjid Khalilian, Norwati Mustapha, Md Nasir Sulaiman, Ali Mamat "Intrusion Detection System with Data Mining Approach: A Review" Global Journal of Computer Science & Technology, Volume 11 Issue 5 Version 1.0 April 2011.

[14] Ajith Abraham "129: Artificial Neural Networks" Handbook of Measuring System Design, edited by Peter H. Sydenham and Richard Thorn. 2005 John Wiley & Sons, Ltd. ISBN: 0-470-02143-8, Retrieved on 2nd January, 2015.

[15] http://www.windowsecurity.com/articles-tutorials/intrusion_detection/IDS-Part2-Classification-methods-techniques.html, Retrieved on 29th December, 2014.

[16] Manoj Kumar, Mohammad Husian, Naveen Upreti, Deepti Gupta "Genetic Algorithm: Review And Application" International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 451-454.

[17] Lamees Alhazzaa "Intrusion Detection Systems using GeneticAlgorithms" http://pdf.aminer.org/000/337/929/immunity_based_genetic_algorithm_for_classification_rule_discovery.pdf, Retrieved on 1st January, 2015.

[18] Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat, Amit Khaparde "A Review Paper on Various Data Mining Techniques" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014.

[19] Mostaque Md. Morshedur Hassan "Current Studies on Intrusion Detection System, Genetic Algorithm and Fuzzy Logic" International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.2, March 2013.

[20] A. Chaudhary, V. N. Tiwari, A. Kumar "Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Ad Hoc Networks" BIJIT - BVICAM's International Journal of Information Technology.

[21] http://www.slideshare.net/nilmani14/neural-network-3019822, Retrieved on 1st January, 2015.

[22] http://www.slideshare.net/karthiksankar/genetic-algorithms-3626322, Retrieved on 1st January, 2015.

[23] Vivek K. Kshirsagar, Sonali M. Tidke, Swati Vishnu "Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview" International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.

[24] http://www.sans.org/security-resources/idfaq/data_mining.php, Retrieved on 1st January, 2015.

[25] Jayveer Singh, Manisha J. Nene "A Survey on Machine Learning Techniques for Intrusion Detection Systems" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 11, November 2013.