SURVEY ON ESTABLISHING A HYBRID METHODOLOGY FOR IDS

Vanita Kewlani
Department of Computer Engineering
LJIET
Ahmedabad, India

B.Madhushree
Department of Computer Engineering, Assist. Professor
LJIET
Ahmedabad, India

ABSTRACT

In today's world, there has been a huge growth of computer networks which leads to a tremendous increment in the range of applications that depend on it. Hence, security in network has became very important. In addition, practically all computer methods in any organization endure from security vulnerabilities that are technically troublesome and economically expensive to be resolved by the manufacturers. Network intrusion Detection method is one of the principal accessories to observe and analyze the traffic to find out any viable assaults in the network. They're the security measurements of any network. NIDS plays an most important function in privateness protection.
This paper reviews few techniques used for IDS , thereby giving an idea of creating a hybrid approach based on genetic algorithm and neural network. The key thought is to take skills of classification knowledge of genetic algorithm and neural network for intrusion detection procedure. The proposed model has capacity to identify an attack, to distinguish one attack from the other i.e. Classifying attack, and the predominant, to observe new attacks with high detection rate and low false negative.

Keyword--Genetic Algorithm, Intrusion Detection System, Features Selection, Machine Learning, Neural Network.

I.INTRODUCTION

Today, the internet is utilized by majority of people for communication. As a consequence, they expect an ease of network or a protected channel for communication, specifically when the exchange includes sensitive or personal knowledge. Previously, a significant number of study reviews had been done within the subject of network safety to ensure the security and trustworthiness of the transmitted and stored Information. Intrusion detection system (IDS) is the most commonly used approach for assuring

Computer security. This system is a tool that the network administrators use for protecting their networks against the malicious activities.

OVERVIEW OF IDS

An intrusion detection system (IDS) is a software application or a hardware that continuously monitors network traffic and/or system activities for abnormal behaviour or policy violations and produces logs to an administration unit.[1]

The extensive use of the Internet connects a host/network to every other computer/network on this globe exposing it to every possible intrusion. An IDS is a security system that dynamically monitors and observes the target system (which can be file, folder, a host or a network) for any misuse and tries to handle the abnormal behaviour either by itself or by producing alarms to an administrative unit. The use of IDS becomes necessary because building a completely secure system is almost next to impossible. This is because the target system is usually invaded by two kinds of users: Legitimate users: Those users who are a part of the system but go beyond the scope of their confidence. Illegitimate users: Those users who are unknown to the system but try to breach the security of target system. Protecting a system against the outsiders may seem to be easy but then a large number of users also dwell within the boundaries of the target system. [1]

An IDS system generated logs which record the activities/events in a target system. A legitimate user of a privilege similar to root/system administrator can possibly work at lever lower than the level where audit trials run and therefore bypass the monitoring scheme. Hence, the security of a target system is more susceptible to an intrusion from a member of the system. Although there are counter measures for such issues also but this gives a general idea of a loophole in a target system even though we may protect it from illegitimate users.[1]

IDS can be classified into two groups based on the method of detection [4]

A. Misuse or signature based system

This is rule-based detection. In this approach, the system analyzes data from audit logs to create a rule or signature for

the attack. Therefore, it can detect only known attacks with fewer false alarms.

B. Anomaly based system

It relies on "learning" about past behavior of users. Analysis of audit logs every time determines what behavior is normal for users. Any deviations generate alerts.

II.TYPICAL ARCHITECTURE OF IDS

Here we will identify important building blocks of typical IDS. Although IDS when implemented have their own specialties in architectural style, but system here can be said as general reference for IDS.

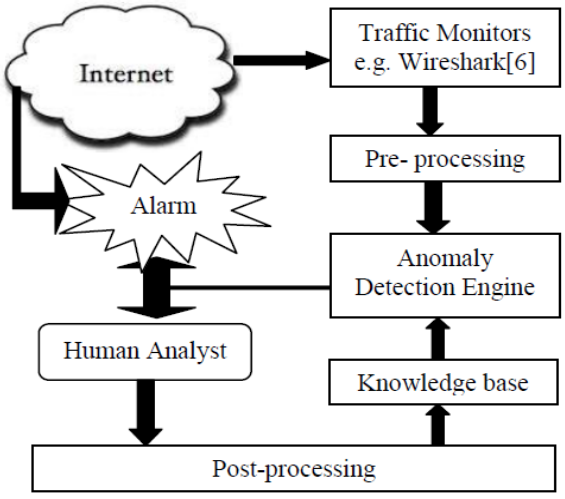

Figure 1:Typical Architecture of IDS

A. Network Traffic Monitors

It is very first component that captures data about Network activities. This can be considered as row data which need preprocessing to suit to specific system. This task can be accomplishes by tools like Wireshark.

B. Data Preprocessor

At this stage data captured from traffic monitors are converted to the format that can be handled by the anomaly detection engine. During preprocessing some other information to the data is inserted and some information is being reduced to make data appropriate to analyze.

C. Anomaly Detection Engine

This is the brain of any anomaly detection system. This engine employs different techniques for getting rid of normal traffic and if it finds any abnormality from normal flow, it generates alarm. It uses knowledge base to differentiate normal traffic from abnormal ones.

D. Alarm

It is responsible for registering and indicating /informing concern system for the abnormal behavior detected by the system.This acts as soon as anomaly detection engine gives indication.

E. Further Anaysis

This step normally executed by human analyst, who is responsible for interpreting alarm and the taking appropriate action. This kind of human intervention is needed when working with anomaly detection, and also it helps in updating knowledge base with appropriateness. Human intervention for updating knowledge base reduces false alarm rate.

F. Post Processing

This is for updating knowledge base from indicated events of abnormality. During post processing data abnormal data taken in consideration and then used to create rule base.

G. Knowledge base

This can be in the form of signature for misuse detection system and in the form of rule/profile for Anomaly detection system. This is being updated regularly/continuously after post-processing of network data.

H. System Configurations

This is the data about how each component of IDS works with each other and itself. This is used to set data that can increase performance or accuracy.

III. METHODOLOGY

Problems with Existing Systems.Most existing intrusion detection systems suffer from at least two of the following problems [2]:

First, the information used by the intrusion detection system is obtained from audit trails or from packets on a network.Data has to traverse a longer path from its origin to the IDS and in the process can potentially be destroyed or modified by an attacker. Furthermore, the intrusion detection system has to infer the behavior of the system from the data collected, which can result in misinterpretations or missed events.This is referred as the fidelity problem.

Second, the intrusion detection system continuously uses additional resources in the system it is monitoring even when there are no intrusions occurring, because the components of the intrusion detection system have to be running all the time. This is the resource usage problem.

Third, because the components of the intrusion detection system are implemented as separate programs, they are susceptible to tampering. An intruder can potentially disable or modify the programs running on a system, rendering the intrusion detection system useless or unreliable. This is the reliability problem.

GENETIC ALGORITHM

Genetic algorithm is is based on the natural principal of survival of fittest, which was first proposed by J.H. Holland in his Phd Dissertation thesis [10]. GA when applied to IDS starts with framing the problem in to chromosome data structure. Genes represents the Atomic unit in chromosome. Genes are evaluated for its fitness at each iteration and new population is created. Each iteration is composed of operation like mutation and crossover. In each iteration next generation/population of genes become more fitter then previous generation. This process of evaluation continue till the fittest generation of chromosome is achieved [11]. Following Figure 2 shows basic steps of GA.

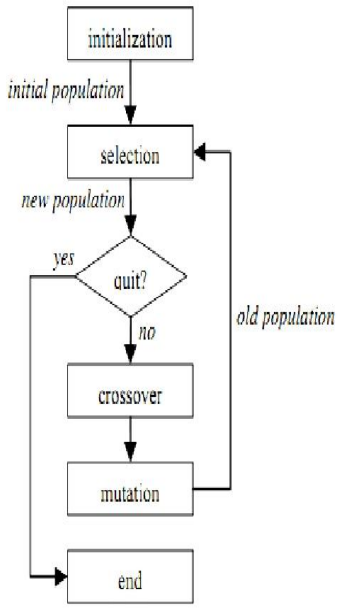

Figure 2: Basic Steps in GA

(1) Selection

This is the first step of each iteration. It selects individual chromosome and creates population.

(2) Evaluation

Each chromosome is evaluated against fitness function. Fitness function may be simple of it can be based on support-vector.

(3) Cross-over

During this operation best chromosomes are recombined to create new population.

(4) Mutation

This step varieties each chromosome by interchanging its properties in itself.

NEURAL NETWORK

Neural Networks (NNs) have attracted more attention compared to other techniques. That is mainly due to the strong discrimination and generalization abilities of Neural Networks that utilized for classification purposes [18]. Artificial Neural Network is a system simulation of the neurons in the human brain [19]. It is composed of a large number of highly interconnected processing elements (neurons) working with each other to solve specific problems. Each processing element is basically a summing element followed by an active function. The output of each neuron (after applying the weight parameter associated with the connection) is fed as the input to all of the neurons in the next layer. The learning process is essentially an optimization process in which the parameters of the best set of connection coefficients (weights) for solving a problem are found [20].

An increasing amount of research in the last few years has investigated the application of Neural Networks to intrusion detection. If properly designed and implemented, Neural Networks have the potential to address many of the problems encountered by rule-based approaches. Neural Networks were specifically proposed to learn the typical characteristics of system's users and identify statistically significant variations from their established behavior. Neural Network based models make the use of several computational interconnected elements (neurons); this parallel processing may imply time savings in malicious traffic analysis.

IV. COMPARATIVE ANALYSIS OF RELATED WORK

The survey has been done for the different types of methods adopted for Intrusion detection.The summary of this survey is as below:

Uzair Bashir, Manzoor Chachoo(2014)[1]:
This survey paper gives a description of some intrusion detection approaches based on two basic techniques.It shows that some approaches work better in one environment but then prove to be weak in other environments.A generic technique needs to be developed that can help us to secure our networks in any environment. This requires a detailed knowledge of already existing techniques and their loopholes so that researchers can propose ideas to overcome the weaknesses and develop a much stronger approach to deal with intrusions.

Nattawat Khamphakdee ,Saiyan Saiyod, Nunnapus Benjamas (2014)[2]:
This paper has improved Snort-IDS rules for the network probe attacks detection, and in addition also classifies the characteristics of network probe attacks. The results of the tested Snort-IDS rules confirm that the proposed Snort-IDS can correctly detect 100% of the network probe attacks based

on MIT-DAPRA 1999 data set. However, regarding to the comparative analysis with the notification Detection Scoring Truth, the detection number of the proposed Snort-IDS rules are more than the detection scoring truth. Because, some moments of the attacks had occurred in several times.However, the Snort-IDS rules can help the network administrators quickly analyze the patterns of attacks.In addition,they must also regularly update Snort-IDS rules.Because attackers try to find new methods with increasing the complexity of the network attacks to damage over time.

Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis(2014)[3]:

In this paper the author was interested in intrusion detection based on genetic algorithms to improve the search time in the audit data without losing the performance of the system. Satisfactory results were produced, in terms of very high detection rate (99%), reinforced by a low rate of false positives (3%). The results were obtained after several improvements of the approach used.

Satendra kumar ,Anamika Yadav(2014)[4]:

In this paper an artificial neural network based intrusion detection system is proposed which used Gradient descent with momentum backpropagation algorithm for learning. Although random patterns are selected for training but the proposed neural network is tested across complete "testing" data of KDD cup 99 dataset.The output is evaluated in terms of accuracy detection rate and false positive ratio and compared with latest reported works. The results shows the accuracy of the proposed NN based IDS for binary classification (Attack or normal) is high and detection rate for probe, R2L and U2R attacks are high as compared with other techniques.

Jabez J, Dr.B.Muthukumar(2015)[5]:

In this paper, they have presented the details of a new approach called Outlier Detection approach to detect
the intrusion in the computer network. Their training model consists of big datasets with distributed environment that improves the performance of Intrusion detection system. Their proposed approach is also been tested with the KDD datasets that are received from real world. The machine learning approaches detect the intrusion in the computer network with huge execution time and storage to predict the when compared to the proposed IDS system which takes less execution time and storage to test the dataset .Here in this study, the performance of proposed IDS is better than that of other existing machine learning approaches and can significantly detect almost all anomaly data in the computer network.

Mrutyunjaya Panda, Ajith Abraham, Manas Ranjan Patra(2012)[6]:

In this paper,author proposed some novel hybrid intelligent decision technologies using data filtering
by adding supervised or un-supervised methods along with a classifier to make intelligent decisions in order to detect network intrusions. They used a variant of KDDCup 1999 dataset, NSL-KDD to build our proposed IDS. The performance comparison amongst different hybrid and combination of classifiers were
made in order to understand their effectiveness in terms of various performance measures. Finally, they
concluded that their proposed hybridization of END with nested dichotomies and random forest of 10 trees with out of bag error of 0.06 results almost 100% intrusion detection rate with 0% false alarm rate, which makes the approach as most efficient.But it resulted in higher cost which created an issue for the usage of this approach.

IV. PROPOSED WORK

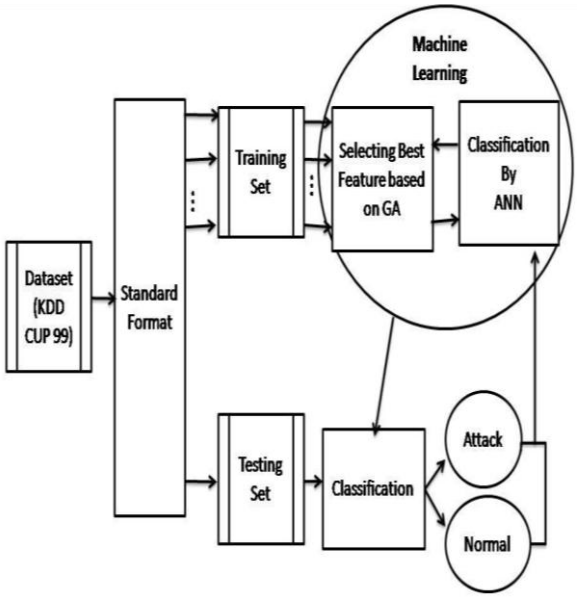

Fig.3
Proposed model

Figure 3 has illustrated the overall method and the main idea of the proposed work. First of all, this method will dive the dataset in a random pattern into two categories, the testing set and the training set. In the phase of training, the first task of the machine learning is to learn and select the most appropriate features and then in the phase of testing, the knowledge of the machine learning is tested by the machine learning and the features that had been selected in the phase of training are tested as well and then the data is classified into normal and attacks categories.

In the process of machine learning, the data is received by GA and then the features are made and selected for the classification of ANN. The classification of ANN is used for preventing the detection rate and the problem of over fitting from n tests, which their average is for receiving a value for fitness.

The genetic algorithm is a method that will search the global optimization and has the ability of simulating the evolution behavior and process in the nature. It means that every possible key will be trained in a vector type, which is known as chromosome. Each element of the vector represents a gene. The entire chromosome set will form a population and the projection of the population is based on the fitness function

[14]. A fitness value is used to measure the fitness of the chromosome. The initial populations of the genetic process are randomly developed. The genetic algorithm applies the operators for creating the next generation out of the generation that is currently being used: reproduction, crossover and mutation. The genetic algorithm omits the chromosomes with lower fitness. In addition, the chromosomes with a high fitness are prevented by the genetic algorithm [15]. All the process that has been mentioned above will be repeated, thus, the next generation will receive more chromosomes with high fitness. This process will continue until the detection of an individual good chromosome. A set of the primary individuals are turned into high quality individuals by genetic algorithm and every one of the achieved individuals operates as a problem solution. The individuals that have been mentioned above are referred to as chromosomes and some pre-determined genes are the elements that form those chromosomes [16]. The effectiveness of the algorithm is affected by the factors that will be mentioned in the following: the values of the parameters of the GA (the crossover rate, the mutation rate, the size of population, and the threshold of the fitness value), the fitness function selection, and the individuals' agency. These parameters have their own dependencies and their dependency results from the application.

ANN is known as one of the methods of machine learning and is based upon the statistical learning theory. ANN is considered as the best method and algorithm for the classification in the classification pattern. Warren and Walter (1943) considered the algorithms and the mathematics to create a model of computation for the neural networks for the algorithm of classification in recognizing the patterns. The basic idea of this paper for ANN is finding a hypothesis to ensure the detection rate in a high level for recognizing and separating the normal and the attack.

V.CONCLUSION

In this paper a survey has been carried out to find out the best methods to undergo hybridization for intrusion detection sytem.It has been observed that GA is best to use for optimization problem.Also ANN usually results in more detection rates so the idea behind this paper to create a hybrid method combining the feature selection that has been based upon the GA with the ANN classification so as to improve the performance of the system by providing higher detection rates.

VI. REFERENCES

[1]Uzair Bashir, Manzoor Chachoo,"Intrusion Detection and Prevention System: Challenges & Opportunities",2014 IEEE,806-809

[2]Nattawat Khamphakdee,Saiyan Saiyod, Nunnapus Benjamas,"Improving Intrusion Detection System Based on Snort Rules for Network Probe Attack Detection",IEEE 2014, 69-74.

[3]Salah Eddine Benaicha, Lalia Saoudi, Salah Eddine Bouhouita Guermeche, Ouarda Lounis, "Intrusion Detection System Using Genetic Algorithm",Science and Information Conference 2014,564-568

[4]Satendra kumar ,Anamika Yadav, "Increasing Performance Of Intrusion Detection SystemUsing Neural Network", 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies(ICACCCT),546-550

[5]JabezJ,Dr.B.Muthukumar,"Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach", International Conference on Intelligent Computing, Communication & Convergence(ICCC-2015),338-346, ELSEVIER

[6] Mrutyunjaya Panda, Ajith Abraham, Manas Ranjan Patra," A Hybrid Intelligent Approach for Network Intrusion Detection", International Conference on Communication Technology and System Design (2011),1-9 , ELSEVIER

[7]J. Muna. M. and Mehrotra M., "Intrusion Detection System: A design perspective", *Proceeding of 2rd International Conference On Data Management, IMT Ghaziabad, India.,2009,265-372.*

[8] M. Panda, and M. Patra, "Building an efficient network intrusion detection model using Self Organizing Maps", *Proceeding of world academy of science, engineering and technology, 38, 2009,22-29.*

[9] M. Khattab Ali, W. Venus, and M. Suleiman Al Rababaa, "The Affect of Fuzzification on Neural Networks Intrusion Detection System", *IEEE computer society,2009, 1236-1241.*

[10]Man, Kim-Fung, Kit-Sang Tang, and Sam Kwong."Genetic algorithms: concepts and applications." IEEE Transactions on Industrial Electronics 43.5 (1996): 519-53.

[11]Khan,M.Sadiq Ali."Rule based Network Intrusion Detection using Genetic Algorithm" International Journal of Computer Applications 18.8 (2011)

[12]A. Chittur, "Model Generation for an Intrusion Detection System Using Genetic Algorithms", *Technical Report, Ossining, New York, 2001*.

[13] W. Li, "Using Genetic Algorithm for Network IntrusionDetection", *http://www.security.cse.msstate.edu, Department of Computer Science and Engineering, Mississippi State University, USA, 2004*.

[14]Y. Meng, "The practice on using machine learning for network anomaly intrusion detection," in *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on Machine Learning and Cybernetics*, 2011, pp. 576-581.

[15]H. Sarvari and M. M. Keikha, "Improving the accuracy of intrusion detection systems by using the combination of machine learning approaches," in *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of Soft Computing and Pattern Recognition*, 2010, pp. 334-337.

[16] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *Security and Privacy (SP), 2010 IEEE Symposium on Security and Privacy*, 2010, pp. 305-316.

[17] J. Gomez, D. Dasgupta, "Evolving Fuzzy Classifiers for Intrusion Detection", *Proceedings of the IEEE, 16(6), 2002, 1462-1475*

[18]D.Novikov, V. Roman Yampolskiy, and L. Reznik, "Artificial Intelligence Approaches For Intrusion Detection", *Proceeding of Systems, Applications and Technology Conference, IEEE Long Island, 2006, 1-8.*

[20] S. Lília de Sá, C. Adriana Ferrari dos Santos, S. Demisio da Silva, and A. Montes, "A Neural Network Application for Attack Detection in Computer Networks", *Proceeding of IEEE joint conference on Neural Networks, 2, 2004, 1569-1574.*

[21] P. Kukiełka and Z. Kotulski, "Analysis of Different Architectures of Neural Networks for Application in Intrusion Detection Systems", *Proceedings of the International Multi conference on Computer Science and Information Technology, IEEE, 2008, 807– 811.*