A NOVEL APPROACH FOR INTRUSION DETECTION BY USING FUZZY GENETIC ALGORITHM

G.Sreekanth[1], B.S. Ayyappa Kumar[2]
1 M.Tech Scholar, Department of Computer science, Vignan’s Lara Institute of Technology and science, Vadlamudi, India
2 M.Tech, Assistant Professor, Department of Computer science, Vignan’s Lara Institute of Technology and science, Vadlamudi, India

ABSTRACT
Network security is of primary distressed now days for large organizations. The intrusion revelation systems (IDS) are becoming indispensable for suitable protection adjacent attacks that are constantly uncertain in magnitude and intricacy. With data probity, confidentiality and availability, they must be decisive, easy to dominate and with low allowance cost. Various modifications are being activated to IDS regularly to distinguish new attacks and handle them. This paper introduce a fuzzy eugenic algorithm (FGA) for intrusion detection. The FGA system is a napped classifier, whose knowledge vitiate is modeled as a fuzzy rule equivalent as “if-then” and corrected by a hereditary algorithm. The method is approved on the benchmark KDD’99 intrusion dataset and related with other extant techniques convenient in the literature. The results are promising and demonstrate the benefits of the determined accession.

Keywords: Intrusion Detection System Architecture, Detection types, Attacks, Protocols, KDD cup data set, ID3 algorithm, C4.5 algorithm, Decision trees, Classification.

I. INTRODUCTION
Intrusion detection system and prevention system are same. Both are used to detect the mischievous program which enters in our network or host. The only difference is the blockage system will give the response to malignant program by using firewall, anti spam and by blocking the malignant activity. We can perform the intrusion detection in network and host. There are two types of intrusion apprehension system. They are signature based and anomaly based detection methods. We can provide the intrusion prevention system with the proper software’s and hardware. Then only we can secure our system. Conjecturing modeling is used to predict the output based on historical data. Classification is used to envision the output by historical data. It has two processes. 1. We should physique the model and another one to see the culminate model. It is mainly used in customer segmentation, business modeling, credit risk and biomedical inquiry and drug responses modeling. 2. Intrusion Detection Systems Architecture an intrusion detection system is a software program which helps to analyze the malicious program which enter our system or in network. It helps to secure our system by counter to the malicious program. It is divided into two types. They are presenter based intrusion detection system and network based intrusion detection system. The active system will counter to the malicious program. But the passive system will encounter only whether any malicious packets entered the system or not.

Knowledge Discovery and Data Mining (KDD)

II. RELATED WORKS
The normal and abnormal behaviours [1] in networked computers are hard to predict, as the boundaries cannot be well defined. This diagnosis process usually beget fake alarms in many peculiarity based intrusion revelation systems. With the admittance of furry logic, the fake alarm rate in determining invasive behaviour can be diminished, where a set of fuzzy rules is used to illustrate the normal and unexpected behaviour [1] in a computer network. This system expected a technique to engender fuzzy rules that are able to ascertain malicious activities and some precise intrusions. This system presented an advance for the performance of accomplish fuzzy rules in classifying different types of infraction. In this system, I explained the assailing modes and point to the bounce of this assailing and its threats. From an attacker’s perspective, I analyse each of the attack’s modes, benefits and judicious conditions and think how to reform the attack by introducing the thought of fuzzy logic-based technique. Fuzzy set theory was received by Zadeh [10] in 1965 and it was categorically designed mathematically epitomize uncertainty and vagueness with formalized convictive tools for dealing with the imprecision intrinsic in many real world problems. Hassan [4], Baruah ([7], [8]), Neog and Sut [9] have forwarded a expanded definition of fuzzy set which approve us to define the correlate of a fuzzy set. Our proposed system concede with them as this new explanation satisfies all the estate respecting the
aggregate of a fuzzy set. Gong [2] presented an implementation of GA based entrance to Network Intrusion Detection using GA and exhibit software implementation. The approach imitated a set of classification rules and exploit a support-confidence framework to judge strength function. Xia, Hariri and Yousif [3] used GA to detect divergent network behaviours based on advice theory ([17], [18]). Some network features can be analyze with network attacks based on requited information between network lineaments and type of encroachment and then using these lineaments a linear structure rule and also a GA is imitative. The approach of using mutual advice and culminate linear rule intimate very effective because of the diminished complexity and higher revelation rate. The only problem is that it considered only the various features. Abdullah [6] showed a GA based performance opinion algorithm to network intrusion revelation. The approach uses information theory for squirt the traffic data. Lu and Traore [12] used historical network dataset using GP to achieve a set of classification [17]. They used support-confidence framework as the health activity and accurately classified several network infraction. But their use of genetic programming made the implementation agenda very challenging and also for training action more data and time is appropriate. Goyal and Kumar [13] portray a GA established algorithm to allocate all types of smurf attack accepting the training dataset with false conclusive rate is very low (at 0.2%) and detection rate is almost 100%. Li [14] described a method using GA to reveal anomalous network intrusion ([17], [18]). The advent includes both quantitative and categorical countenance of network data for acquire classification rules. However, the involvement of quantitative countenance can increase exposure rate but no experimental results are accessible.

III. INTRUSION DETECTION OVERVIEW
The following category give a short audit of networking barrage, classifications and various factor of Intrusion Detection System.

A. NETWORKING ATTACKS :This section is an audit of the four major categories of networking attacks. Every attack on a network can comfortably be implanted into one of these groups [15] – Denial of Service (DoS): A DoS attack is a breed of attack in which the hacker makes a computing or memory backing too busy or too full to serve legitimate networking requests and hence denying users connection to a machine e.g. apache, smurf, neptune, ping of death, back, mail bomb, UDP storm etc. are all DoS attacks. Remote to User Attacks (R2L): A remote to user intervention is an attack in which a user sends packets to a machine over the internet, which he/she does not have access to in order to disclose the machines vulnerabilities and exploit privileges which a local user would have on the computer e.g. xlock, guest, xsnnoop, phf, sendmail dictionary etc.

User to Root Attacks (U2R): These attacks are exploitations in which the hacker starts off on the system with a regular user account and attempts to abuse vulnerabilities in the system in order to gain super user allowance e.g. perl, xterm. Probing: Probing is an attack in which the hacker scans a machine or a networking device in order to determine weaknesses or vulnerabilities that may later be abused so as to compromise the system. This technique is commonly used in data mining e.g. saint, portswipe, mscan, nmap etc.

B. CLASSIFICATION OF INTRUSION DETECTION: Intrusions Detection can be classified into two main categories. They are as follow: Host Planted Intrusion Detection: HIDSs evaluate information found on a single or multiple entertainer systems, including contents of operating systems, system and application files ([11], [16]). Network Based Intrusion Detection: NIDSs evaluate information apprehend from network communications, analysing the stream of packets which travel across the network ([11], [16]).

C. COMPONENTS OF INTRUSION DETECTION SYSTEM: An intrusion detection system regularly consists of three functional components [17]. The first component of an intrusion detection system, also known as the accident generator, is a data source. Data sources can be categorized into four categories namely Host-based monitors, Network-based monitors, Application-based monitors and Target-based monitors. The second elemental of an intrusion apprehension system is known as the analysis engine. This component takes clue from the data antecedent and examines the data for symptoms of attacks or other policy violations. The analysis engine can use one or both of the following hearing approaches: Misuse/Signature-Based Detection: This type of detection engine distinguish intrusions that follow III-knownpatterns of aggression (or signatures) that exploit known software vulnerabilities ([18], [19]). The main limitation of this advent is that it only looks for the known weaknesses and may not care about detecting unknown future intrusions [20]. Anomaly/Statistical Detection: An anomaly based apprehension engine will search for something rare or unusual [20]. They analyses system event streams, using

www.ijsret.org
statistical techniques to find decoration of activity that appear to be abnormal. The primary disadvantages of this system are that they are highly extravagant and they can recognize an intrusive behaviour as normal behaviour because of insufficient data the third component of an intrusion detection system is the reply manager. In basic terms, the response manager will only act when inaccuracies (possible intrusion attacks) are found on the system, by informing someone or something in the form of a feedback.

IV. GENETIC ALGORITHM

A. Introduction to Genetic Algorithm Genetic algorithms are a branch of evolutionary algorithms [8] used in exploration and optimization techniques. The three dominant behavior of a genetic algorithm i.e., selection, crasser and mutation correspond to the biological process: The survival of the competent (As shown in Figure 1). In a genetic algorithm, there is a population of strings (called chromosomes or the genotype of the genome), which cipher and indent solutions (called individuals, creatures, or phenotypes). [10] Traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution consistently starts from a population of randomly generated individuals and evolve over generations.

Fig 1: Structure of a Genetic Algorithm

In each generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current community (based on their fitness), & modified (recombined and possibly randomly mutated) to form a new population. The new community is then used in the next iteration of the algorithm. Commonly, the algorithms adjourn when either a maximum number of individuals are there in a generation, or a decent fitness level has been reached for the population. If the algorithm has terminated expected to a maximum number of individuals, a satisfactory solution may or may not have breathe reached.

A. GENETIC ALGORITHM OVERVIEW:

A Genetic Algorithm (GA) is a programming technique that uses biological enlargement as a problem solving strategy [21]. It is based on Darwinian’s principle of evolution and survival of fittest to optimize a population of candidate solutions almost a predefined fitness [14]. The proposed GA based intrusion detection system contains two syllabus where each works in a different phase. In the training stage, a set of classification rules are generated from network analysis data using the GA in an offline environment. In the intrusion apprehension stage, the generated rules are used to classify incoming network connections in the real-time environment. Once the rules are developing, the intrusion detection system becomes simple, experienced and adequate one. GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination and alteration operators [14]. The process usually begins with randomly generated population of chromosomes, which perform all possible solution of a problem that are considered candidate solutions. From each chromosome different positions are encoded as bits, characters or numbers. These bearings could be referred to as genes. An evaluation function is used to calculate the correctness of each chromosome according to the desired solution; this function is known as “Fitness Function”. During the process of appraisal “Crossover” is used to simulate natural reproduction and “Mutation” is used to alteration of species [14]. For survival and combination the selection of chromosomes is partial fronting the fittest chromosomes. When I use GA for solving various problems three factors will have vital impact on the effectiveness of the algorithm and also of the utilization [2]. They are: i) the fitness function; ii) the representation of individuals; and iii) the GA parameters. The determination of these circumstances often depends on implementation of the system. In the following sections, I focus our discussions on acquire the set of rules using Genetic Algorithm.

B. FUZZY LOGIC:

It has been shown by Baruah [7] that a fuzzy number [a, b, c] is defined with associating to a membership function µ(x) lying between 0 and 1, a ≤ x ≤ c. Further, he has extended this definition in the following way. Let μ1(x) and μ2(x) be two functions, 0 ≤ μ2(x) ≤ μ1(x) ≤ 1. He has concluded μ1(x) the fuzzy membership function, and μ2(x) an associating function, such that (μ1(x) – μ2(x)) is the fuzzy membership value for any x. Finally he has characterized such a lurid number by \{x, μ1(x), μ2(x); x ∈ Ω\}. The complement of μx is always estimate from the ground level in Zadehian’s theory [10], whereas it actually counted from the level if it is not as zero that is the surface value is not always zero. If other than zero, the problem arises and then we have to count the membership value from the
surface for the complement of $\mu x$. Thus I could achieve the following statement – Complement of $\mu x = 1$ for the entire level Membership value for the complement of $\mu x = 1$. My system forwarded a definition of complement of an protracted furry set where the furry reference function is not always zero. The definition of complement of a fuzzy set proposed by Hassan [4], Baruah [7], Neog and Sut [9] could be seen a particular case of what I am giving. I shall use Baruah’s definition of the complement of a normal furry set in my article. In the two classes’ classification problem, there are two classes where every object should be classified. These classes are called positive (abnormal) and negative (normal). The data set used by the learning algorithms consists of a set of objects, each object with $n+1$ attributes. The first $n$ attributes define the object characteristics (monitored parameters) and the last attribute defines the class that the object correlates to the classification attribute. A fuzzy classifier for solving the two class classification problem is a set of two rules, one for the normal class and other for the abnormal class, where the condition part is characterized using only the monitored parameters and the conclusion part is an atomic interpretation for the classification attribute.

C. FLOWCHART: Figure 2 shows the operations of an accustomed genetic algorithm according to which GA is implemented in our system.

D. ALGORITHM OF THE PROPOSED SYSTEM:
Algorithm – Rule set generation using GA
Input – Network audit data, number of generations, and population size
Output – A set of classification rules
1. Initialize the population
2. Generate random population
3. $W_1=0.2$, $W_2=0.5$, $W_3=0.3$, $T=0.5$, $\text{chrom}\_\text{length}=9$
4. $N=\text{total number of populations to be generated}$
5. for each chromosome in the population
6. $TP=0$, $TN=0$, $FP=0$, $FN=0$
7. for each record in the training set
8. if the record matches the chromosome
9. increment membership value of $TP$
10. end if
11. if the records do not match the chromosome
12. increment membership value of $FP$
13. end if
14. end for
15. $\text{Fitness}=W_1*TP/(TP+FN)+W_2*FP/(FP+TN)+W_3*(1-\text{chrom}\_\text{length}/10)$
16. If $\text{Fitness}>T$ 17. if $N$
18. Break
19. else
20. select the chromosome into the new population
21. update the total number of population
22. $N=N-1$
23. end if
24. end if
25. end for
26. for each chromosome in the new population
27. apply crossover operator to the chromosome
28. apply mutation operator to the chromosome
29. end for
30. if the required number of generation is not reached, then go to step 5.

V. IMPLEMENTATION

Fig 2: Flow chart of Genetic Algorithm

Fig 3: The main screen for selecting the KDD dataset.
VI. CONCLUSION
In this paper, a method of applying ancestral algorithms with fuzzy logic is presented for network intrusion detection system to efficiently detect various types of network intrusions. To implement and measure the performance of the system I carried out a number of experiments using the accepted KDD Cup 99 benchmark dataset and obtained reasonable detection rate. To measure the fitness of a chromosome I used the fuzzy confusion matrix where the fuzzy membership value and fuzzy membership function for the complement of a fuzzy set are two different approaches because the surface value is not always counted from the ground level. The proposed detection system can upload and update new rules to the systems as the new encroachment become known. Therefore, it is cost effective and flexible. The method suffers from two aspects. Firstly, it generates false alarms which are very serious problem for IDS. Secondly, for high dimensional data, it is hard to generate rules that cover up all the attributes.

REFERENCES


