Intrusion Detection System by Using Data Mining Based On Class-Association-Rule Mining Using Genetic Network Programming

Mr. Shankar L. Tambe [1], Prof. Ms.Rasna Sharma [2]

[1]*Student, CSE, CIIT Indore, RGPV, Bhopal, India*
[2]*Assistant Professor, CSE, CIIT Indore, RGPV, Bhopal, India*

ABSTRACT

Consistently, there is the need to reestablish an installation of Intrusion Detection System (IDS) due to new attack methods or upgraded computing environments. Since many current IDSs are constructed by manual encoding of expert knowledge, changes to IDSs are expensive and slow. This paper describes a data mining framework for adaptively building Intrusion Detection (ID) models. Now security is considered as a major issue in networks, since the network has extended dramatically. Therefore, intrusion detection systems have attracted attention, as it has an ability to detect intrusion accesses effectively. These systems identify attacks and react by generating alerts or by blocking the unwanted data/traffic. The proposed system includes fuzzy logic with a data mining method which is a class-association rule mining method based on genetic algorithm. Due to the use of fuzzy logic, the proposed system can deal with mixed type of attributes and also avoid the sharp boundary problem. Genetic algorithm is used to extract many rules which are required for anomaly detection systems. An association-rule-mining method is used to extract a sufficient number of important rules for the user's purpose rather than to extract all the rules meeting the criteria which are useful for misuse detection. Experimental results with KDD99Cup database from MIT Lincoln Laboratory show that the proposed method provides competitively high detection rates compared with crisp data mining.

Keyword: - *Data Mining, Intrusion Detection System (IDS), Genetic Algorithm (GA), Network Security, Fuzzy Logic.*

1. INTRODUCTION

Security of network systems is becoming increasingly important as more and more sensitive information is being stored and manipulated online. Intrusion Detection Systems (IDSs) have thus become a critical technology to help protect these systems.

Most IDSs are based on hand-crafted signatures that are developed by manual encoding of expert knowledge. These systems match activity on the system being monitored to known signatures of attacks. The major problem with this approach is that these IDSs fail to generalize to detect new attacks or attacks without known signatures. Recently, there has been an increased interest in data mining based approaches to building detection models for IDSs. These models generalize from both known attacks and normal behavior in order to detect unknown attacks. They can also be generated in a quicker and more automated method than manually encoded models that require difficult analysis of audit data by domain experts. Several effective data mining techniques for detecting intrusions have been developed, many of which perform close to or better than systems engineered by domain experts.

An effective data mining-based IDS must address each of these three groups of issues. Although there are tradeoffs between these groups, each can generally be handled separately. IDS can also be divided into two groups depending on where they look for intrusive behavior: Network-based IDS (NIDS) and Host-based IDS. Network-based IDS refers to systems that identify intrusions by monitoring Traffic through network devices (e.g. Network Interface Card, NIC). Host- based IDS requires small programs to be installed on individual systems to be monitored.

IDS can also be divided into two groups depending upon behavior of IDS: Passive and Active type of IDS. Passive IDS simply detects and alerts the administrator whereas, active IDS will not only detect suspicious or malicious traffic can alert the administrator, but will take predefined proactive actions to respond to the threat.

1.1 Types of Networking Attacks

There are four major categories of networking attacks. Every attack on a network can be placed into one of these groupings [4].

- **Denial of Service (DoS):** A DoS attacks is a type of attack in which the hacker makes a memory resources too busy to serve legitimate networking requests and hence denying users access to a machine e.g. apache, smurf, Neptune, ping of death, back, mail bomb, UDP storm, etc.

- **Remote to User attacks (R2L):** A remote to user attack is an attack in which a user sends packets to a machine over the internet, and the user does not have access to in order to expose the machines vulnerabilities and exploit privileges which a local user would have on the computer, e.g. xlock, guest, xnsnoop, phf, sendmail dictionary etc.

- **User to Root Attacks (U2R):** These attacks are exploitations in which the hacker starts off on the system with a normal user account and attempts to abuse vulnerabilities in the system in order to gain super user privileges, 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 exploited so as to compromise the system. This technique is commonly used in data mining, e.g. Satan, saint, port-sweep, mscan, nmap etc.

Two different attack types were included for this study: SYN Flood (Neptune) and Satan. These two attack types were selected from two different attack categories (denial of service and probing) to check for the ability of the intrusion detection system to identify attacks from different categories.

SYN Flood (Neptune) is a denial of service attack to which every TCP/IP implementation is vulnerable (to some degree). For distinguishing a Neptune attack, network traffic is monitored for a number of simultaneous SYN packets destined for a particular machine. Satan is a probing intrusion, which automatically scans a network of computers to gather information or find known vulnerabilities. The purpose of classifiers in IDSs is to identify attacks from all four groups as accurately as possible.

2. PROBLEM DEFINITION

Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPSes for other purposes, such as identifying problems with security policies, documenting existing threats, and deterring individuals from violating security policies. IDPSes have become a necessary addition to the security infrastructure of nearly every organization. Network Intrusion Detection System (NIDS) will be another wall for protection. Most of the existing commercial NIDS are signature-based but not adaptive. There are many have problems such as attack stealth ness: attackers try to hide their actions from either an individual in monitoring the system or a NIDS, novel intrusion: it is undetectable by signature-based NIDS; they can only be detected as anomalies by observing deviations from normal network behaviour. Whereas Anomaly detection approaches attempt to identify abnormal behaviour in patterns and can make use of supervised or unsupervised methods to detect the anomalies or attacks Does typically record information related to observed events, notify security administrators of important observed events, and produce reports. Some attacks that are aimed to be handled are:

Denial of Service (DoS) attack: A DoS attack is aimed at preventing authorized, legitimate users from accessing services on the network. The DoS attack is not aimed at gathering or collecting data.

Brute force attack: Brute force attacks simply attempt to decode a cipher by trying each possible key to find the correct one. This type of network attack systematically uses all possible alpha, numeric, and special character key combinations to find a password that is valid for a user account.
Project has done following attacks:-

2.1 Land attack:

Description: The Land attack is a denial of service attack that is effective against some older TCP/IP implementations. The only vulnerable platform used in the 1998 DARPA evaluation was SunOS4.1. The Land attack occurs when an attacker sends a spoofed SYN packet in which the source address is the same as the destination address [17]. An attacker sends forged stream of TCP SYN packets with the same source and destination IP address and TCP port numbers. The victim system will be confused and crashed or rebooted. Service providers can block LAND attacks that originate behind aggregation points by installing filters on the ingress ports of their edge routers to check the source IP addresses of all incoming packets. If the address is within the range of advertised prefixes, the packet is forwarded; otherwise it is dropped.

2.2 TCP SYN Flood Attack:

Description: TCP Reset is a denial of service attack that disrupts TCP connections made to the victim machine. That is, the attacker listens (on a local or wide-area network) for tcp connections to the victim, and sends a spoofed tcp RESET packet to the victim, thus causing the victim to inadvertently terminate the TCP connection. It takes advantage of a flaw in how most hosts implement the TCP three-way handshake. When Host B receives the SYN request from A, it must keep track of the partially opened connection in a "listen queue" for at least 75 seconds. Many implementations can only keep track of a very limited number of connections. A malicious host can exploit the small size of the listen queue by sending multiple SYN requests to ahost, but never replying to the SYN&ACK the other host sends back. By doing so, the other host's listen queue is quickly filled up, and it will stop accepting new connections, until a partially opened connection in the queue is completed or times out.

2.3 Attack Signature:

1) One way to detect the attack would be to look at the TCP session setup/takedown process, and note cases in which RESET packets appear to come from the machine that had initially attempted to begin the connection. (This might not be foolproof however, as there might be cases when this is a common/normal occurrence.)

2) Second way to detect the attack would be to look at the TCP session setup/takedown process, and note cases in which RESET packets appear to come from the machine that had initially attempted to begin the connection. If a particular Machine continuously tries to reset the field again and again, in our system we have kept the threshold limit of 20 SYN bit set.

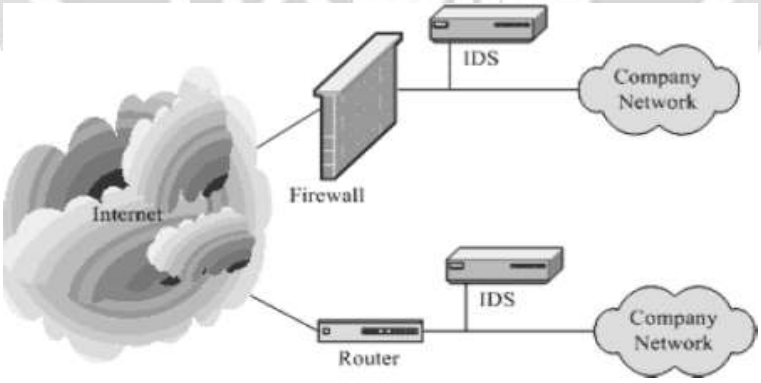

Fig -1: Intrusion Detection System

3. LITERATURE SURVEY

A brief description of related works is presented in this section.
Shingo Mau,et.al.,[1] this paper describes a novel fuzzy class association rule mining method based on GNP and its application to intrusion detection. By combining fuzzy set theory with GNP, the proposed method can deal with the mixed data base that contains both discrete and continuous attributes. Such mixed data base is normal in real-world

applications and GNP can extract rules that include both discrete and continuous attributes consistently. The initiative of combining as so citation rule mining with fuzzy set theory has been applied more frequently in recent years. The original idea comes from dealing with quantitative attributes in a database, where discretization of the quantitative attributes into intervals would lead to under or overestimate the values that are near the borders. This is called the sharp boundary problem. Fuzzy sets can help us to overcome this problem by allowing different degrees of memberships. Compared with traditional association rules with crisp sets, fuzzy rules provide good linguistic explanation.

Swati Dhopte et.al[2] proposed whole project in to another different way i.e describe the paper to creating the Genetic Algorithm(GA). Therefore here use the tree for creating the rule without the use of the graph. The drawback of using the tree in the creating the rule not generate with the help of the other rule i.e use of the gens. So it will be increase the complicity of this project.

Janatan Gomez et.al[3] In this paper, they show the applicability of genetic algorithms to evolve a simple set of fuzzy rules (fuzzy classifier) that can solve some well-studied intrusion detection problems. In this approach, genetic algorithms can find good and simple fuzzy rules to characterize intrusions (abnormal) and normal behavior of network systems. As the difference between the normal and the abnormal activities are not distinct, but rather fuzzy, fuzzy logic can reduce the false signal rate in determining intrusive activities.

Zohair Ihsan et.al[4] In this paper, the attributes can be qualitative or quantitative in nature. Attributes with large values significantly influence the performance of intrusion classifier making it bias towards them. Attribute normalization eliminates such dominance of the attributes by scaling the values of all the attributes within a specific range. The paper discusses various normalization techniques and their influence on intrusion classifiers such as Random Forest, Bayes Net, Naive Bayes, NB Tree and Decision Tree. Furthermore, the concept of hybrid normalization is applied by normalizing the qualitative and quantitative attributes differently. Experiments on KDDC up 99 suggest that the hybrid normalization can achieve better results as compared to conventional normalization.

Mohammad Sazzadul Hoque et.al[5] In this paper, they present an Intrusion Detection System(IDS),by applying genetic algorithm(GA) to efficiently detect various types of network intrusions. Parameters and evolution processes for GA are discussed in details and implemented. This approach uses evolution the oryto information evolution in order to filter the traffic data and thus reduce the complexity. To implement and measure the performance of our system we used the KDD99bench mark dataset and obtained reasonable detection rate.

Nivedita P. Chaudhari et.al[6] They describe Genetic Network Programming(GNP) is one of the fields from biological computation uses directed graph structure. To deal with both discrete and continuous attributes, fuzzy set theory of combination of triangular and triangular membership functions. Genetic operations such as crossover, mutation-1, and mutation-2 and fitness function are used to generate more number of strong hybrid rules. Rather than using conventional measurement methods for hybrid rules, new technique called X2 evaluation is used. Genetic operations changes fuzzy membership parameters changes.

Michale wiczs operator is used to overcome this problem in our methodology. At the end we have strong and robust hybrid rules without loss of information, which is used for classification purpose to match data using proposed classification algorithm.

4. LIMITATION OF EXISTING SYSTEM

An intrusion detection system is used to detect several types of malicious behaviours that can compromise the security and trust of a computer system. This includes network attacks against vulnerable services, data driven attacks on applications, host based attacks such as privilege escalation, unauthorized logins and access to sensitive files, and malware (viruses, Trojan horses and worms).

IDS can be composed of several components: Sensors which generate security events, a Console to monitor events and alerts and control the sensors, and a central Engine that record event logged by the sensors in a database and uses a system of rules to generate alerts from security events received. There are several ways to categorize IDS depending on the type and location of the sensors and the methodology used by the engine to generate alerts. In many simple IDS implementations all three components are combined in a single device or appliance. Intrusion

detection can allow for the prevention of certainty, attacks severity relative to different type of attack and vulnerability of components under attack the response may be kill the connection, install filtering rules, and disable user account.

Most of the NIDS products are signature-based but not adaptive. This paper focuses on adaptive learning and training the data. Current researches comprise of single engine detection systems, whereas paper focus on issues related to deploying a data mining-based IDS. It describes the approaches to address three types of issues: accuracy, efficiency and usability.

To improve accuracy, data mining programs are used to analyze audit data and extract features that can distinguish normal activities from intrusions; to produce more effective misuse and anomaly detection models. To improve efficiency, the computational costs of features are analyzed and high accuracy. To improve usability, adaptive learning algorithms are used to facilitate model construction and incremental updates; unsupervised anomaly detection algorithms are used to reduce the reliance on labeled data. It also presents an architecture consisting of sensors, detectors, a data warehouse, and model generation components. This architecture facilitates the sharing and storage of audit data and the distribution of new or updated models. This architecture also improves the efficiency and scalability of the IDS.

5. PROPOSED APPROACH

The proposed GA based intrusion detection using data mining approach contains two stages where each works in a different stage. In the training stage, using the GA and fuzzy-association rule mining algorithm, a set of classification rules are generated from KDD dataset. In the intrusion detection stage, the generated rules are used to classify in coming data from a test file. Once the rules are generated, the intrusion detection is simple and efficient. Following Figure shows the proposed system architecture.

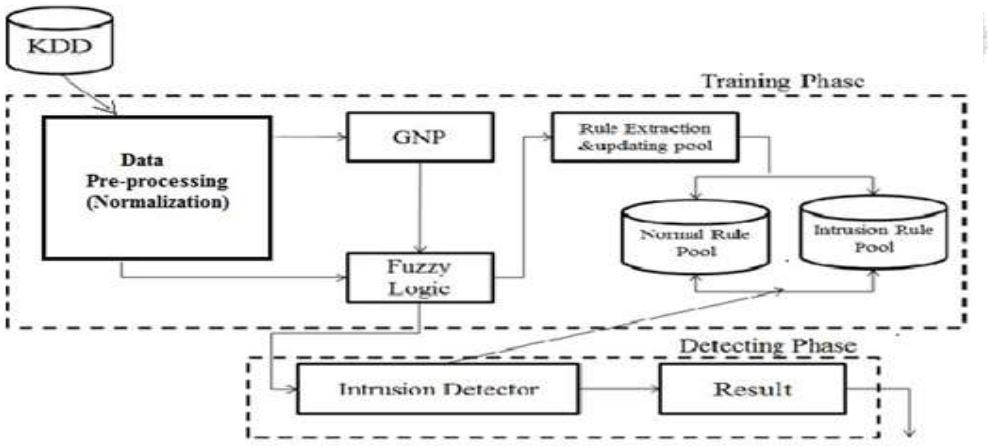

Fig -2: System Architecture

5.1 Data Pre-processing (Attribute Normalization):

Data pre-processing is the first step whilst analyzing the data. The data pre-processing is comprised of phases like dataset creation, data cleaning, integration, feature construction, normalization and feature selection. Reported that data pre-processing can take up to 50 of the overall process effort. This section provides a brief explanation of the normalization phase of data pre- processing. A dataset is collection of different attributes that describes various characteristics of the data. These attributes can be qualitative or quantitative in nature with different range of values. The nature and values of these attributes influences the data analysis process. For instance, Attributes with large values can dominate attributes with small value. The process of normalization can eliminate such dominance by scaling the mall within a specific range. Quantitative attributes can be directly normalized, whereas in case of

qualitative attributes, the nominal values first needs to be converted to numeric value before applying the normalization. The numeric values can be assigned based on certain criteria or simply replacing every nominal value with 1, 2, 3 n. Once the qualitative attributes have been converted to quantitative attributes, the normalization process can be applied to them.

Steps for pre-processing of attributes/features are shown in the following algorithm:

Algorithm: Classify KDD dataset, Feature extraction.

Input: KDD dataset Output: Dataset into two classes i.e. rule pool (Normal and attack)

1. Select KDD dataset
2. Transform attributes to numeric value
3. Find maximum value for each attribute/feature
4. Select important attribute/features
5. Store rules in rule pool

In algorithm, classification method data mining is used for classifying the whole dataset into two classes i.e. ”normal” & ”attack”. Feature selection is necessary as the use all available features are computationally in feasible.

5.2 Fuzzy Logic:

It utilizes advantage of fuzzy theory to have every continuous attribute value in [0, 1]. Fuzzy theory allows complex system to have linguistic description. In each continuous attribute in the data base is transformed into 5 linguistic terms (Vey low, Low, Middle, High, Very high). By using 5 linguistic terms for single continuous attribute, we will get more accurate membership value for corresponding continuous attribute.

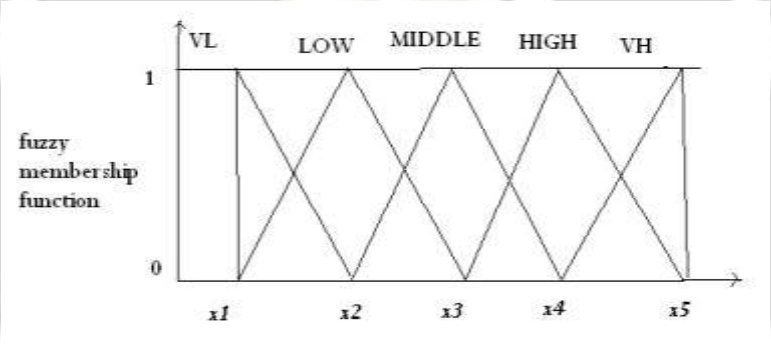

Fig -3: Fuzzy Membership functions

The linguistic terms are defined by the combination of trapezoidal and triangular membership functions. The parameters x1, x2, x3, x4 and x5 are also evolved along with the evolution of GNP. Each continuous attribute have its own membership value. The parameters for each continuous attribute are initialized by analyzing the distribution of data. During evolution process, the parameter value of fuzzy membership function should be adjusted generation by generation. Fuzzy membership values are used to determine the transition in GNP individuals while searching for association rules. Following table show small data base with two continuous attributes. Fig.3.3 shows the fuzzy membership function for attribute A1 and A2, respectively.

Table - 1: Small Database

TID	A1	A2
1	100	10000
2	200	8000
3	300	6000
4	400	4000
5	500	2000

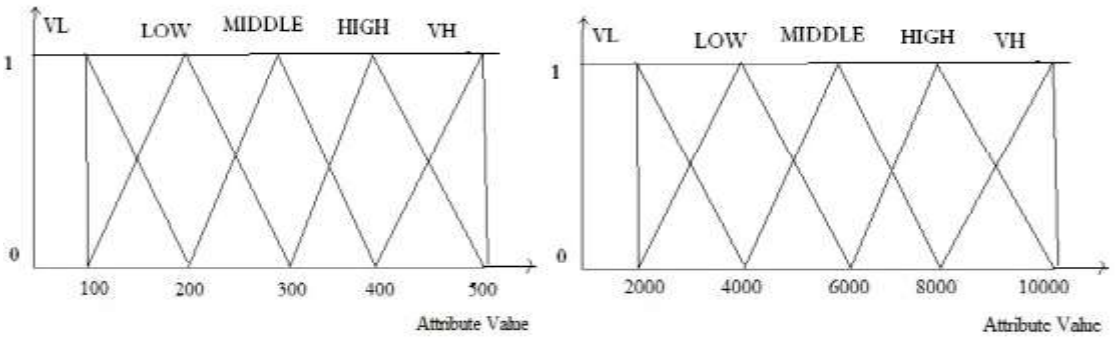

Fig -4: Membership functions for Attribute A1, A2

5.3 Genetic Network Programming:

Data pre-processing algorithm generates rules which are stored in the rule pool i.e. Normal rule pool contains normal records and attack rule pool contains records for intrusion. The following algorithm is common for both i.e. normal and attack rule pool and explains about the genetic algorithm and its operators.

Algorithm: Rule pool generation using genetic network programming algorithm.

Input: Pre-processed dataset, number of generations (G), and population size.

Output: Large no. of rules in the Rule pool.

1. Initialize the population
2. N is population size, T (minimum fitness value) = 0
3. User input for number of generations (G)
4. Initialize individuals (I) =1
5. Initialize fitness counter (K) =1
6. Select two chromosomes (or rules) from population
7. Increment I by 2, K by 1
8. Apply cross over operator to the chromosome
9. Apply mutation operator to the chromosome
10. If rule is present in rule pool then go to step14 for next rule.
11. Else
    Calculate the number of connections Ntc correctly detected by ruler
    Calculate the number of connections in the training data Nt
    Calculate the number of normal connections Nni incorrectly detected by ruler
    Calculate the number of normal connections in the training data Nn
    Calculate Fitness value of new chromosome

$$\text{fitness}_r = \frac{Nt_c}{Nt} - \frac{Nn_i}{Nn}$$

12. If fitness is greater or equal to T then, Add newly generated chromosome to rule pool.
13. Else go to step 14 for next rule.
14. Repeat step 10 until K equals to 3.
15. Repeat step 5 until I equals to N/2.
16. Increment G by 1.
17. If number of generations is not reached, go to step 4.
18. Display number of rules generated for the input generations.
19. Go to next algorithm that is fuzzy rule extraction.

In the above algorithm, each rule is referred to as a chromosome or individual. In each generation, apply crossover and mutation to increase the number of rules. In a single point crossover, exchange of genes (attributes value)

between two individuals with respect to some point is carried out. Range of fitness value is [-1, 1], so threshold fitness is 0 in this approach. Once the individuals are selected for making a pair, avoid repeated selection find individuals to make other pairs. The above procedure is then repeated until no individuals for making pairs are remaining. At the end of this algorithm, a large number of rules will be available for further processing. For Anomaly detection, the quantity of rules matters more than quality.

5.4 Algorithm for Class-Association-Rule mining (CARM)

The following is a statement of association- rule mining. Let I= A1, A2, . . . , Al be a set of literals, called items or attributes. Let G be a set of tuples, where each tuple Tis a set of attributes such that T⊆I. Let TID bean ID number associated with each tuple. A tuple T contains X, a set of some attributes in I, if X⊆T.

An association rule is implication of the form X⇒Y, where X⊂I, Y⊂I, and XnY=φ. X is called antecedent and Y is called consequent of the rule. If the fraction of tuples containing X in G equals x, then we say that support(X) = x. The rule X⇒Y has a measure of its strength called confidence defined by support (XUY) / support(X).

Calculation of X2 value of rule X ⇒Y is shown as follows. Assume support(X) = x, support(Y)=y, support(X UY)=z, and the total number of tuples is N. We can calculate X2 as

$$\chi^2 = \frac{N(z-xy)^2}{xy(1-x)(1-y)}.$$

If X2 is higher than a cutoff f value, we should reject the assumption that X and Y are independent (3.84 at the 95 significance level or 6.64 at the 99 significance level).

Let Ai be an attribute in a data base with value1or 0, and k be class labels. Then, a class association rule can be represented by (Ap=1)∧ ⋯∧ (Aq=1)∧ (C=k)k∈ 0,1

As a special case of the association rule X⇒Y with fixed consequent.

B. class-association rules satisfying the following are defined as important rules:

$$\chi^2 > \chi^2_{\min}$$

$$\text{support} \geq \text{sup}_{\min}$$

$$\text{confidence} \geq \text{conf}_{\min}$$

Where X^2_{min}, Sup_{min} and $conf_{min}$ are the minimum X^2, minimum Support, and minimum confidence, respectively given in advance.

6. SYSTEM IMPLEMENTATION

The purpose of System Implementation can be summarized as follows: making the new system available to a prepared set of users (the deployment), and positioning on-going support and maintenance of the system within the Performing Organization (the transition). At a finer level of detail, deploying the system consists of executing all steps necessary to educate the Consumers on the use of the new system, placing the newly developed system into production, confirming that all data required at the start of operations is available and accurate, and validating that business functions that interact with the system are functioning properly. Transitioning the system support responsibilities involves changing from a system development to a system support and maintenance mode of operation, with ownership of the new system moving from the Project Team to the Performing Organization. A key difference between System Implementation and all other phases of the lifecycle is that all project activities up to this point have been performed in safe, protected, and secure environments, where project issues that arise have little or no impact on day-to-day business operations. Once the system goes live, however, this is no longer the case. Any miscues at this point will almost certainly translate into direct operational and/or financial impacts on the Performing Organization. It is through the careful planning, execution, and management of System Implementation activities that the Project Team can minimize the likelihood of these occurrences, and determine appropriate contingency plans in the event of a problem.

7. RESULT

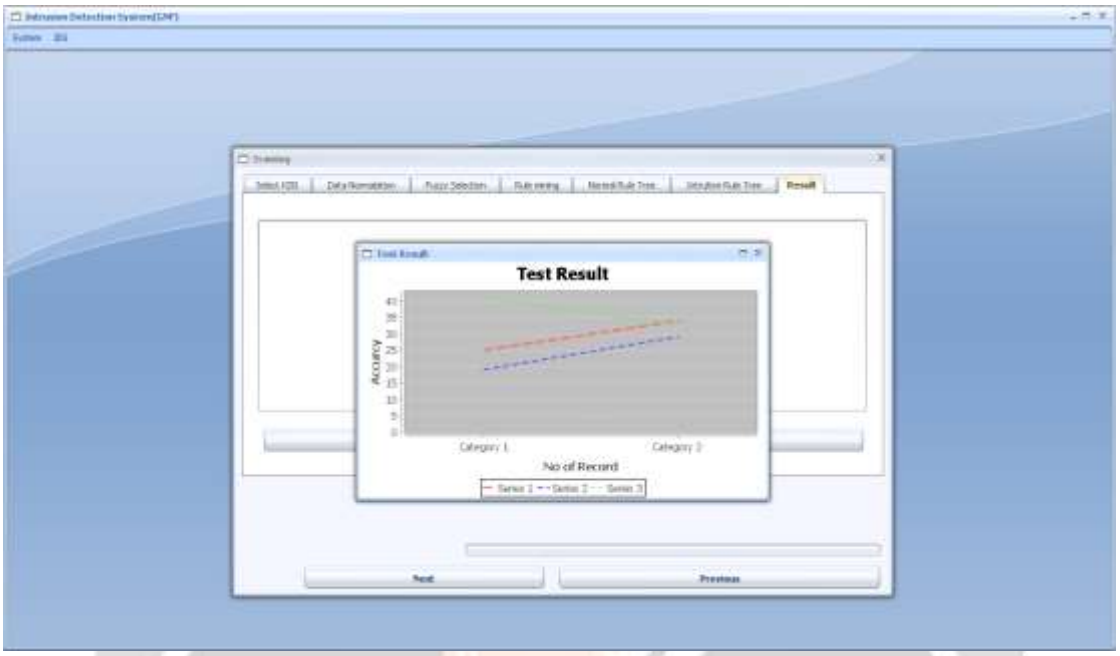

Fig- 5:Test Result

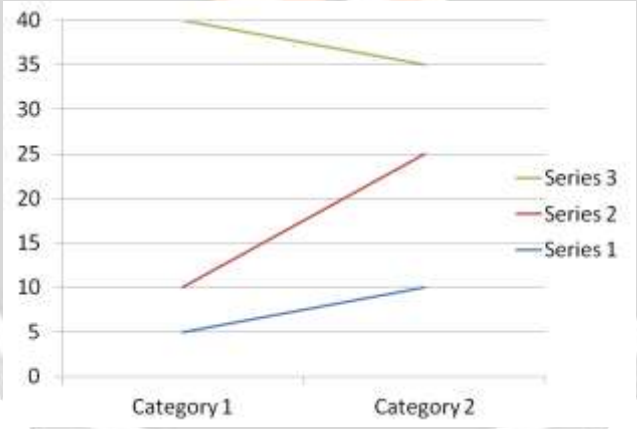

Chart- 1: Graph of Response Time

8. CONCLUSIONS

In this project report, a GNP-based fuzzy class-association-rule mining with sub attribute utilization and the classifiers based on the extracted rules have been proposed, which can consistently use and combine discrete and continuous attributes in a rule and efficiently extract many good rules for classification. As an application, intrusion-detection classifies for both misuse detection and anomaly detection have been developed and their effectiveness is confirmed using KDD99 Cup and DARPA98 data.

The important function of the proposed method is to efficiently extract many rules that are statistically significant and they can be used for several purposes. For example, GNP can extract many rules of normal connections and known intrusion connections from the training database in this paper. When we use them for misuse detection, the matching of a new connection with the normal rules and the intrusion rules are calculated, respectively, and the connection is classified into the normal class or intrusion class. When we use the rules for anomaly detection, only

the rules of the normal connections are used to calculate the deviation of a new connection from the normal area. Therefore, many rules extracted by GNP cover the spaces of the classes widely. In another application of the data mining of GNP, even if the information of some attributes in some tuples is missing, GNP can extract rules by complementing such parts by other attributes.

9. FUTURE ENHANCEMENTS

In the future, we will focus on building distributions (probability density functions) of normal and intrusion accesses based on fuzzy GNP. By using the probability distributions, the data can be classified into normal class, known intrusion class and unknown intrusion class. In addition, the new data (testing data) can be labeled as normal or intrusion with a certain probability, e.g.,95 present reliability, by using the distributions.

10. REFERENCES

[1] Shingo Mabu, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hira- sawa," An Intrusion-Detection Model Based on Fuzzy Class-Association- Rule Mining Using Genetic Network Programming" IEEE Transactions On Systems, Man, And Cybernetics-Part C: Applications And Reviews, Vol.41,No. 1, January 2011

[2] Swati Dhopte,N. Z. Tarapore,"Design of Intrusion Detection System using Fuzzy Class-Association Rule Mining based on Genetic Algorithm" International Journal of Computer Applications (0975 -8887) Volume53- No.14, September2012.

[3] Jonatan Gomez and Dipankar Dasgupta," Evolving Fuzzy Classifiers for Intrusion Detection" Proceedings of the 2002 IEEE Workshop on Information Assurance United States Military Academy, West Point, NY June 2001.

[4] Zohair Ihsan, Mohd Yazid Idris and Abdul Hanan Abdullaha," Attribute Normalization Techniques and Performance of Intrusion Classifiers: A Comparative Analysis". Life SciJ 2013; 10(4): 2568-2576] (ISSN: 1097-8135).

[5] Mohammad Sazzadul Hoque, Md. Abdul Mukit and Md. Abu NaserBikas,"AN IMPLEMENTATION OF INTRUSION DETECTION SYS TEM USING GENETIC ALGORITHM" International Journal of Network Security Its Applications (IJNSA), Vol.4,No.2, March 2012.

[6] Nivedita P. Chaudhari and Dr.Leena Ragha," SMART NETWORK INTRUSION DETECTION SYSTEM USING HYBRID APPROACH "International Journal of Research in Advent Technology (IJRAT) Vol.1,No. 2, August 2013, ISSN: 23219637.

[7] J.G.-P.A. ElSemaray, J.Edmonds, and M.Papa, " Applying datamining of fuzzy association rules to network intrusion detection," presented at the IEEE Workshop Inf.,United States Military Academy, West Point, NY, 2006.

[8] A.S.S.Forrest, S.A.Hofmeyr, and T.A. Longstaff ,"A sense of self for unix processes," presented at the IEEE Symp. Secur. Privacy, Los Alamitos, CA, 1996.

[9] Kddcup 1999 data[Online]. Available: kdd.ics.uci.edu/ databases/kd- dcup99/ kddcup99.html.

[10] S.Manganaris, M. Christensen, D. Serkle, and K. Hermix,"A data mining analysis of rtidalarms,"presented at the 2nd Int. Workshop Recent Adv. Intrusion Detect., West Lafayette, IN,1999.

[11] D. E. Denning "An intrusion detection model," IEEE Trans. Softw. Eng.,vol. SE-13, no. 2,pp. 222-232, Feb. 1987.

[12] J. R. Koza,Genetic Programming, on the Programming of Computers by Means of Natural Selection. Cambridge, MA: MITPress,1992.

[13] J.R. Koza, Genetic Programming-II, Automatic Discovery of Reusable Programs Cambridge, MA: MITPress, 1994.