EFFICIENT INTRUSION ALERT REDUCTION MECHANISM USING FUZZY ARTMAP

Sudar Aishwarya [#1] Nagarajan Srinivasan [*2]

[#]M.Tech Advanced Computing, School of Computing, SASTRA University, Thanjavur, Tamil Nadu, India.

[*]Assistant Professor, Department of Computer Science, SASTRA University, Thanjavur, Tamil Nadu, India

[1] sudaraish@gmail.com

[2] nagarajan@cse.sastra.edu

***ABSTRACT* :-** The vast alert generation of IDS in the network is the major problem. It is the vital task to find solutions to reduce the alerts. Novel techniques namely Fuzzy Association rule and Fuzzy art map are proposed to identify attacks optimally and to reduce alerts. The execution time is reduced by placing the level of severity and importance. All alerts that are issued by IDSs are not on the same level of severity and importance. It would be great if the system can identify which alerts are highly important and which are not, so that the number of alerts that need to be dealt with can be reduced. The alert is reduced by finding out the attacks accurately using various methods. The Membership function is used to classify the attack as low, mid or high using continuous attribute. The rules are set for each attack using fuzzy association rule. The chi-square, confidence and support values are estimated for each rule and the minimum value will be set for all parameters .The Rules higher than the verge value are taken and the rules for each generation are updated. Then the rules are compared with test data set and calculated the match degree for each attack. The proposed fuzzy association rule is to obtain superlative features. The Fuzzy art map technique is used to classify the intrusion and normal data by calculating the match degree. Hence this technique aims to effectively reduce the alert rate when compared with existing approaches.

Keywords-Intrusion detection, Fuzzy Association rule, Fuzzy art map, Attack detection

I. INTRODUCTION

Computers are connected to furnish expedient services and perform tasks in an efficient manner. Today, safeguarding our data on the internet is the pivotal issue due to the activities of intruders and hackers. IDS plays a major role in addressing this problem. IDS is a device to find out the intruders. IDS is implemented to monitor both the network and the host level [1].

IDS produces more alerts for single attack instance. It is challenging for the human expert to handle the bountiful alert [2] .The research focus on how to reduce this bountiful alerts. This alert contains more false positive alerts which are not related to the attacks. The current alert processing techniques are statistical correlation, knowledge based correlation, and similarity correlations are used to reduce the alerts without missing any information [5].

The alert aggregation technique is used to reduce the alerts. The data stream approach [3] and generative modelling approach [4] based on probabilistic methods are used to reduce the alerts. Meta alerts contain information about the aggregated alerts.

This paper proposes Fuzzy association rule [11] and fuzzy ARTMAP [12] technique to reduce the number of alerts which is based on feature selection. The number of alerts is reduced by accurately find out the attacks. The fuzzy class-association rule mining generates rules for intrusion and normal data which successfully combines discrete and continuous values. Confidence, Support and chi-squared values are used to extract important normal class-association rules. The rules which are higher than the threshold values are taken. Match degree is calculated to classify the intrusion and normal data.

II. RELATED WORK

IDS are the system to find attack with high accuracy. The huge quantity of alerts created with IDS is the main problem. The current research focus on the interrelations of alerts. A Fredrik Valeur, et al proposed the general interrelation model which contains the set of units and core .It combines the complete units into a interrelation process. Alert fusion and alert verification is done by alert correlation approach. The proposed approach includes attack thread reconstruction [6]. Herv´e Debar and Andreas Wespi proposed the most comprehensive approach for alert correlation and aggregation. The alert interrelation unit contains standard data model intrusion detection and series of rules for processing the alerts. The alert reduction algorithm can find replicas and impact [7]. Alfonso Valdes and Keith Skinner proposed the probabilistic approach is used for alert

correlation .It furnishes the mathematical framework for correlating alerts. The single value is used to combine alert . The attributes in the dataset should match the uniformity to fuse alerts [8]. Fabien Autrel et Frederic Cuppens proposed the alert reduction by comparing with the corresponding operator .The IDMEF data format is used because it is supported by several IDS. The alerts can be represented as trees [9].

Giorgio Giacinto et al proposed a new strategy for alarm clustering .The online alarm clustering algorithm generates unified specification of attacks from multiple alarms. The alarms are examined and clustered when no more alarms can be clustered to the present group, the related Meta alarm is the output to the security administrator [10]. Manganaris S et al proposed the judgment making technique in huge data's because IDS produce large data's [1] . Axelsson S analyses the types of IDS and the methods used to find out the attacks [2]. Bishop C.M dealt with the calculation of pattern checking in large database using Markov method [3]. Henzinger M.R. et al proposed the randomized method to deal with large number of data and find solutions for max problem [4]. Safaa O provide the overview about the knowledge based, statistical and probabilistic method which are used to reduce the alerts [5]. Ashish Mangalampalli proposed the association rule method to deal with large data. The rules are created for all data. The alert is reduced by comparing these rules with new data [11]. Christina B proposed the ARTMAP method to lessen the alerts. It is based on attribute selection by match tracking[12].Anitha Devi .T et al proposed the GNP method to reduce the alerts which is based on fitness of the data's [13].

III. MATERIALS AND METHODS

III. A. EXPERIMENTAL SETUP:

DARPA dataset has been collected by the Lincoln Laboratory Massachusetts Institute of Technology. DARPA dataset is mainly used for the training as well as testing the intrusion detectors.

In fig1 Data extracted from DARPA dataset. Pre-processing the dataset and set the fuzzy membership values for each continuous attribute. Fuzzy class association rule is generated for each data's using attacks, service and fuzzy value. For each rule the chi-square, support and confidence are computed. The least value for confidence, chi–square and support is set by the user. Compare each rule with the least verge value.

The rules which are higher than the threshold values are taken. Update the fuzzy rules for each generation. Import the test dataset and set the fuzzy membership value for each continuous data.

Calculate the match degree by using the service and fuzzy membership value of the training data set. Classify the ordinary and intrusion data's from the dataset by using fuzzy ARTMAP [12] which compares the match degree of each data's with the vigilance parameter. The match degree greater than vigilance parameter is classified as normal data otherwise it is intrusion data.

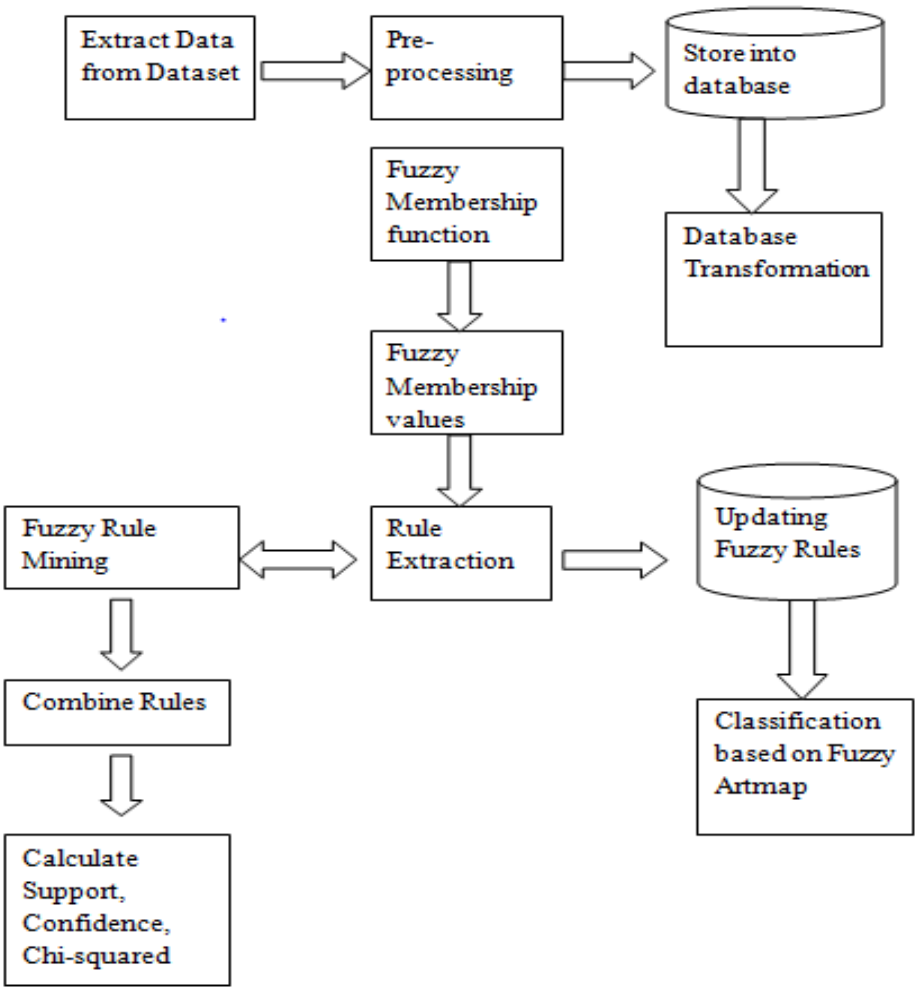


Fig.1: Flow diagram of proposed model

III. B.METHODS:

1. FUZZY MEMBERSHIP FUNCTION:

The Fuzzy membership function of a fuzzy set is an abstraction of the characteristic function in sets. It represents the level of truth. In fuzzy membership function each attribute value is transformed to three terms such as low, medium and high by comparing the maximum time value in the dataset. Alpha, beta and gamma parameters are calculated .Beta represents the mean value of attribute in the dataset. Gamma represents the largest value of attribute in the dataset.

2 . FUZZY ASSOCIATION RULE:

Association rules form an important class of patterns within data. The main goal is to set the rules to find out the intrusion data accurately. The endow co-occurrences are called associations. Each association rule has a support and confidence. It is used to identify the most important relationship. Support indicates how frequently the attributes appear in the database.

It is the percentage of transactions that demonstrate the rule. Confidence indicates how many times it satisfies the rules. Confidence is based on conditional probability. Chi–squared is calculated to measure the significance of association rules.

3 . UPDATING FUZZY RULES:

The least value for support, confidence and chi–square is set by the user. Compare the each item set value with the verge rate. The rules which are higher than the verge rate are taken as acceptable rule. The extracted rules are stored as an intrusion and a normal rule in a rule database. Hence the database is refreshing frequently and valuable rules are stored in the database.

4 . CLASSIFICATION:

Fuzzy ART MAP is used for classification .The match tracking technique is used to minimize the network error. Calculate the fuzzy membership values of the test data set. Compare the fuzzy

membership values and protocol of the test dataset with the existing intrusion and the normal rule pool dataset. The match degree is calculated for each test data set .Compare the match degree with the vigilance parameter. If match degree is greater than the vigilance parameter then classify it as intrusion data otherwise it is normal data.

III.C.ALGORITHM:

INPUT : DARPA intrusion dataset

OUTPUT: Accurately classify the normal and intrusion data

1) Calculate fuzzy membership value

Find out the max time

s=max time/2

For all continuous attributes a [ ]

If a [i] < s set as low

Else if a [i] > s AND a [i] < = s + s set as mid

Else a [i] >s set as high

2) Fuzzy association rules

For all attacks

Generate rules with the protocol and Fuzzy membership values

Generate rules for both intrusion and normal data's

3) Calculate support, confidence and chi-squared value for each rule.

For all rules

M->N if M presents in the transaction N will also be present

Support (M) =p.

Support (N) =q.

Support (M U N) = r.

The number of rows in the database is T.

Chi –squared $\chi 2$ calculated using p, q, r and T.

$\chi 2 = T (r– p*q) 2 / (p* q) (1 − p) (1 − q).$

4) Update the fuzzy rule

Set the minimum support as x, confidence as y and chi-squared as z .

For all rules in the dataset

If (p > x AND q> y AND r > z)

Stored in the rule database

Update the database

5) Classification

Calculate match degree and set the vigilance parameter

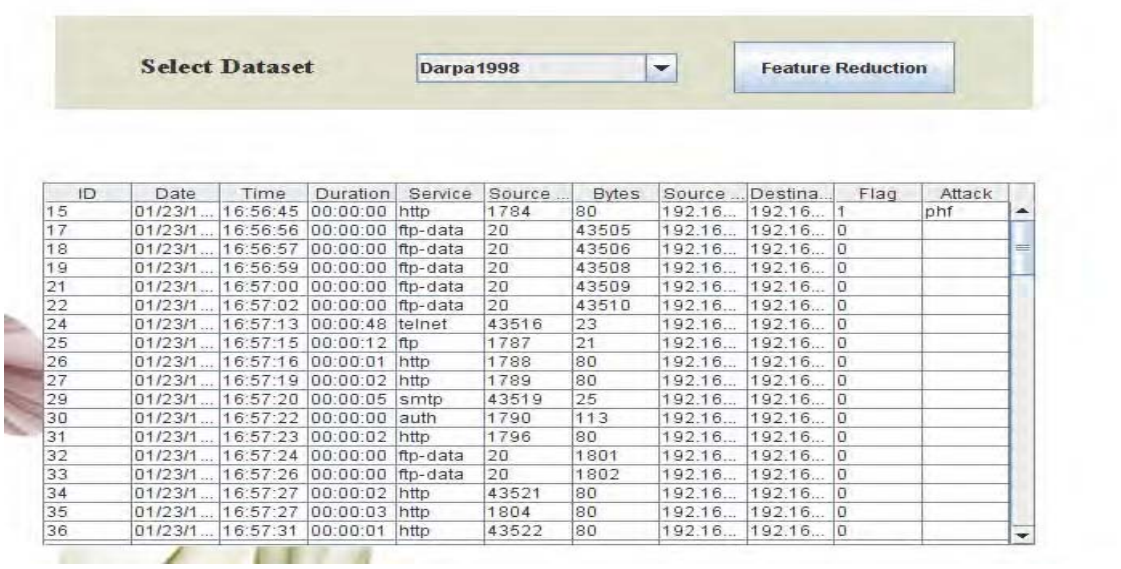
If match degree >vigilance

Set as Normal data

Else

Set as intrusion data

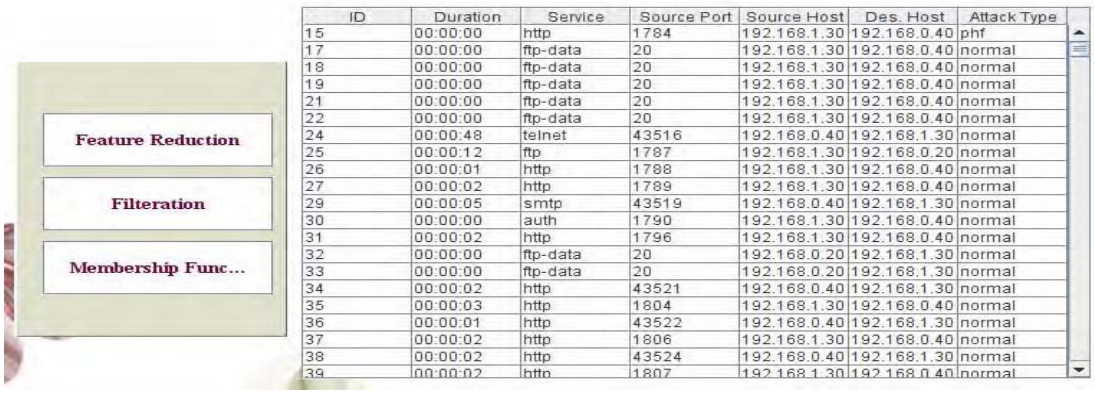
IV.RESULTS AND ANALYSIS



ID	Date	Time	Duration	Service	Source ...	Bytes	Source ...	Destina...	Flag	Attack
15	01/23/1...	16:56:45	00:00:00	http	1784	80	192.16...	192.16...	1	phf
17	01/23/1...	16:56:56	00:00:00	ftp-data	20	43505	192.16...	192.16...	0	
18	01/23/1...	16:56:57	00:00:00	ftp-data	20	43506	192.16...	192.16...	0	
19	01/23/1...	16:56:59	00:00:00	ftp-data	20	43508	192.16...	192.16...	0	
21	01/23/1...	16:57:00	00:00:00	ftp-data	20	43509	192.16...	192.16...	0	
22	01/23/1...	16:57:02	00:00:00	ftp-data	20	43510	192.16...	192.16...	0	
24	01/23/1...	16:57:13	00:00:48	telnet	43516	23	192.16...	192.16...	0	
25	01/23/1...	16:57:15	00:00:12	ftp	1787	21	192.16...	192.16...	0	
26	01/23/1...	16:57:16	00:00:01	http	1788	80	192.16...	192.16...	0	
27	01/23/1...	16:57:19	00:00:02	http	1789	80	192.16...	192.16...	0	
29	01/23/1...	16:57:20	00:00:05	smtp	43519	25	192.16...	192.16...	0	
30	01/23/1...	16:57:22	00:00:00	auth	1790	113	192.16...	192.16...	0	
31	01/23/1...	16:57:23	00:00:02	http	1796	80	192.16...	192.16...	0	
32	01/23/1...	16:57:24	00:00:00	ftp-data	20	1801	192.16...	192.16...	0	
33	01/23/1...	16:57:26	00:00:00	ftp-data	20	1802	192.16...	192.16...	0	
34	01/23/1...	16:57:27	00:00:02	http	43521	80	192.16...	192.16...	0	
35	01/23/1...	16:57:27	00:00:03	http	1804	80	192.16...	192.16...	0	
36	01/23/1...	16:57:31	00:00:01	http	43522	80	192.16...	192.16...	0	

Fig.2: Import Dataset

Import the DARPA training dataset which contains information about time stamps, service, source port, Bytes , source IP address , Destination IP address , Flag and Attack name shows in Fig 2.



ID	Duration	Service	Source Port	Source Host	Des. Host	Attack Type
15	00:00:00	http	1784	192.168.1.30	192.168.0.40	phf
17	00:00:00	ftp-data	20	192.168.1.30	192.168.0.40	normal
18	00:00:00	ftp-data	20	192.168.1.30	192.168.0.40	normal
19	00:00:00	ftp-data	20	192.168.1.30	192.168.0.40	normal
21	00:00:00	ftp-data	20	192.168.1.30	192.168.0.40	normal
22	00:00:00	ftp-data	20	192.168.1.30	192.168.0.40	normal
24	00:00:48	telnet	43516	192.168.0.40	192.168.1.30	normal
25	00:00:12	ftp	1787	192.168.1.30	192.168.0.20	normal
26	00:00:01	http	1788	192.168.1.30	192.168.0.40	normal
27	00:00:02	http	1789	192.168.1.30	192.168.0.40	normal
29	00:00:05	smtp	43519	192.168.0.40	192.168.1.30	normal
30	00:00:00	auth	1790	192.168.1.30	192.168.0.40	normal
31	00:00:02	http	1796	192.168.1.30	192.168.0.40	normal
32	00:00:00	ftp-data	20	192.168.0.20	192.168.1.30	normal
33	00:00:00	ftp-data	20	192.168.0.20	192.168.1.30	normal
34	00:00:02	http	43521	192.168.0.40	192.168.1.30	normal
35	00:00:03	http	1804	192.168.1.30	192.168.0.40	normal
36	00:00:01	http	43522	192.168.0.40	192.168.1.30	normal
37	00:00:02	http	1806	192.168.1.30	192.168.0.40	normal
38	00:00:02	http	43524	192.168.0.40	192.168.1.30	normal
39	00:00:02	http	1807	192.168.1.30	192.168.0.40	normal

Fig.3: Preprocessing the dataset

Feature reduction and Filtration process takes place in pre-processing the dataset (Fig 3). The DARPA dataset contains more attributes in that some of the attributes are not essential .The important attributes are extracted in feature Reduction .Eliminating the missing values in the filtration process.

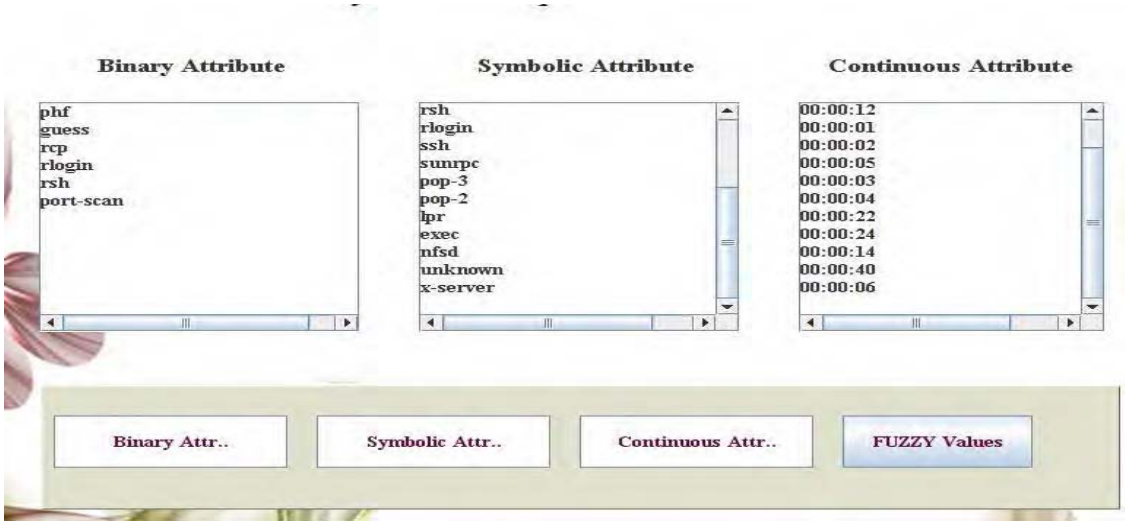


Fig.4: Splitting the attributes

In Fig 4 DARPA dataset attributes are split into three categories they are binary attributes, symbolic attributes and continuous attributes. Binary attribute denotes attack names. If attack presents it is 1 otherwise it is 0. Symbolic attribute denotes the service name. Continuous attributes denote the time stamp which indicates the duration of transmitted data

Fig.5: Fuzzy Membership Values

Continuous attributes are taken to calculate the fuzzy membership values . Alpha, Beta and Gamma parameters are calculated. Gamma represents the highest time value in the continuous attribute. Compare the gamma value with continuous attribute of each tuples and classify the each ID as low, mid or high in Fig 5.

Fig.6: Rule extraction

Fuzzy Association rules are used for rule extraction. Estimate the chi-square, confidence, and support values for all rules. The least support, confidence and chi-square values are set as the threshold values. Extract the rules which are higher than the threshold values. Updating the fuzzy rules for each generation is shown in Fig 6.

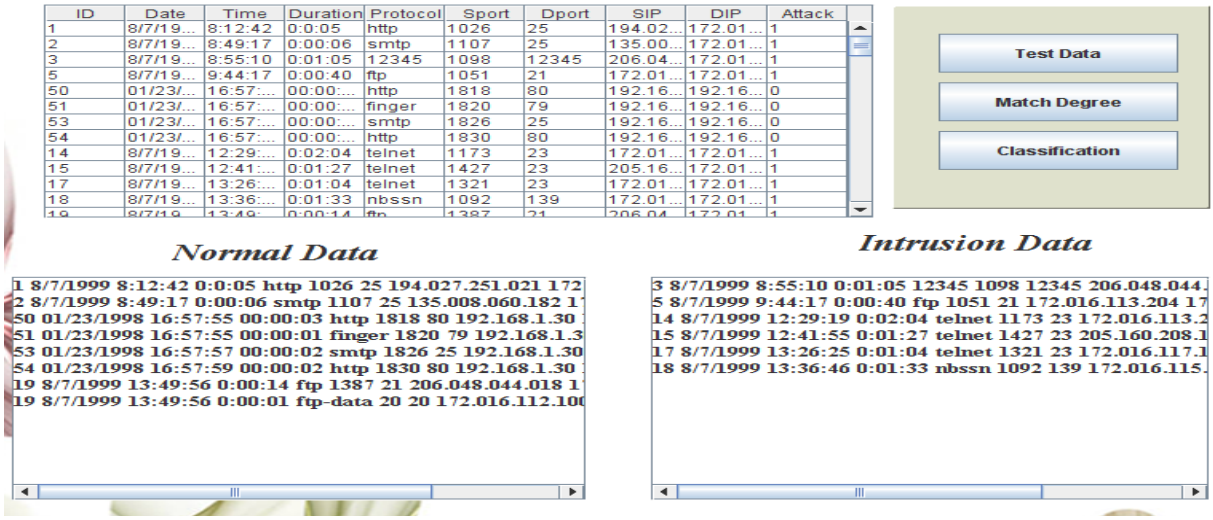


Fig.7: Classification

In Fig7 the test data set is imported. Compare the test data set time value with the training data set and classify as low, medium and high. Calculate the match degree by comparing with the rules. Classify the normal data and intrusion data according to match degree.

V.PERFORMANCE ANALYSIS

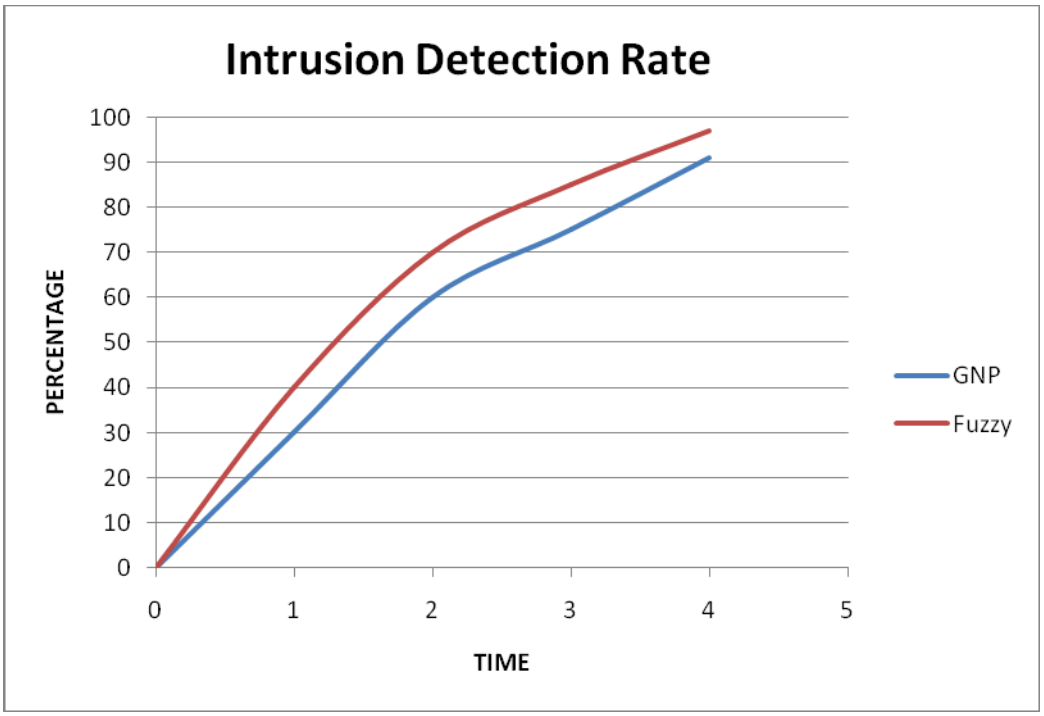


Fig.8: Intrusion Detection Rate

The proposed fuzzy technique intrusion detection rate is higher than the GNP technique [13] is shown in Fig 8.The fuzzy technique accurately finds out the attacks so the false positive alert rate is reduced. In Table 1 Normal1 and Normal 2 denotes the normal data's in the DARPA dataset and Normal1 and Intrusion 2 indicates the number of normal data represented as the intrusion data. The GNP technique achieves 94.2% detection rate.

Table 1. Intrusion Detection Table

	Normal2	Intrusion2	Total
Normal1	83	1	84
Intrusion1	3	32	35
Total	86	33	119

The detection rate (DR) and False Positive Rate (FPR) of proposed algorithm is calculated as

DR=83+32/119=96.64%

FPR=1/84=10.9%

The detection rate of the proposed fuzzy algorithm is 96.64% which is higher than the existing algorithm and 10.9% which is lower than the existing GNP approach.

VI.CONCLUSION:

The main aim of this paper is to classify the intrusion and normal data effectively to minimize the alerts. DARPA 1998 training and testing intrusion datasets are taken to classify the attacks. The process of an alert reduction by using fuzzy association rule and fuzzy ARTMAP in this paper reveals a comparatively efficient system than the other prevailing practices. The attack detection rate is comparatively higher than the existing approach. The false positive alerts are reduced by accurately detecting the attacks.

REFERENCES

[1] S. Axelsson, "Intrusion Detection Systems: A Survey and Taxonomy," Technical Report Dept. of Computer Eng., Chalmers Univ. of Technology, 99-15, 2000.
[2] Manganaris S, Christensen M , Zerkle , Hermiz K." A data mining analysis of RTID Alarms ," Computer Networks: The International Journal of Computer and Telecommunications Networking Volume 34 (4) pp.571-577,2000.
[3] C.M. Bishop, "Pattern Recognition and Machine Learning". Springer, 2006.
[4] M.R. Henzinger, P. Raghavan, and S. Rajagopalan, Computing on Data Streams. Am. Math. Soc., 1999.

[5] Safaa O. Al-mamory, Hong Li zhang: School of Computer Science, Harbin Institute of technology," A Survey on IDS Alerts Processing Techniques,"6th WSEAS International Conference on Information Security and Privacy, Tenerife, Spain, December 14-16,2007.
[6] F. Valeur, G. Vigna, C. Kru gel, and R.A. Kemmerer, "A Comprehensive Approach to Intrusion Detection Alert Correlation," IEEE Trans. Dependable and Secure Computing, vol.1, no. 3,pp. 146-169,July-Sept. 2004.
[7] H. Debar and A. Wespi , "Aggregation and Correlation of Intrusion-Detection Alerts," Recent Advances in Intrusion Detection, eds., pp. 85-103, Springer, 2001.
[8] A. Valdes and K. Skinner, "Probabilistic Alert Correlation," Recent Advances in Intrusion Detection, W. Lee, L. Me, and A. Wespi, eds.pp. 54-68, Springer, 2001.
[9] F. Autrel and F. Cuppens, "Using an Intrusion Detection Alert Similarity Operator to Aggregate and Fuse Alerts," Proc. Fourth ConfSecurity and Network Architectures, pp. 312-322, 2005.
[10] G. Giacinto, R. Perdisci, and F. Roli, "Alarm Clustering for Intrusion Detection Systems in Computer Networks," Machine Learning and Data Mining in Pattern Recognition, P. Perner and A. Imiya, eds. pp. 184-193, Springer, 2005.
[11] Ashish Mangalampalli , Vikram Pudi ," Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets" *IEEE International Conference on Fuzzy Systems* (*FUZZ-IEEE*) Jeju Island, Korea ,2009.
[12] Christina B. Vilakazi and Tshilidzi Marwala ," Application of Feature Selection and Fuzzy ARTMAP to Intrusion Detection" Computing & Processing ;Robotics & Control Systems ,Publication Year: 2006 .
[13] T.AnithaDevi, K.Ruba sounder " An Efficient Model for Network Intrusion Detection System based on an Evolutionary Computational Intelligence Approach" International Conference on Recent Trends in Computational Methods, Communication and Controls,2012.