World Applied Sciences Journal 10 (Special Issue of Computer & Electrical Engineering): 32-40, 2010
ISSN 1818-4952

Misuse Detection Based on Feature Selection by Fuzzy Association Rule Mining

[1]Mansour Sheikhan and [2]Maryam Sharifi Rad

[1]Department of Electrical Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran
[2]Department of Computer Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

Abstract: In this paper, a new framework is proposed for misuse detection in computer networks that shows the effectiveness of data mining techniques in this field. The proposed system is developed in two stages. At the first stage, a feature selection engine based on fuzzy association rule mining (FSE-FARM) is developed to reduce the dimension of input vectors to classifier. At the next stage, a fuzzy ARTMAP neural network is used to determine the attack category. By using FSE-FARM module, the dimension of input feature vector is reduced from 41 to 31. The performance of proposed system is investigated in terms of standard metrics such as classification rate of attacks, detection rate (DR), false alarm rate (FAR) and cost per example (CPE). Experimental results show that the proposed approach performs better in terms of DR, FAR and CPE as compared to some other machine learning methods and the reduced set of features decreases the computation time, as well.

Key words: Intrusion detection · Feature selection · Fuzzy association rule mining · Neural network

INTRODUCTION

With the growth of computer networks and emergence of e-commerce in recent years, computer security has become a priority. Vulnerabilities in common security components such as firewalls are inevitable. Intrusion detection systems (IDSs) are used as another wall to protect computer systems [1]. The main purpose of IDS is to find out intrusions among normal audit data and this can be considered as a classification problem [2]. Intrusion detection techniques can be categorized into misuse detection and anomaly detection. Misuse detection systems discover an intrusion by looking for an activity that corresponds to signatures of known attacks or vulnerable spots in the system. While anomaly detection systems attempt to detect intrusions by observing expected behavior of the systems or deviations from the established normal usage. Some IDSs combine qualities from two categories and are known as hybrid IDSs.

On the other hand, data mining approaches are relatively new techniques for intrusion detection. In the recent years data mining, which is known as knowledge discovery in databases, has established its position as a prominent and important research area. Mining association rules is one of the most important research problems in data mining [3]. Association rule algorithms discover relationships between attributes in the large datasets. In recent years, fuzzy association rules have been applied to intrusion detection systems, as well. For instance in [4], the authors used sets of fuzzy association rules that were mined from network audit data as models of "normal behavior" and they generated fuzzy association rules from new audit data to detect anomalous behavior and then computed the similarity with sets mined from "normal" data. In [5], the authors used a data mining algorithm to discover fuzzy rules from network traffic data. In [1], fuzzy association rules were exploited as descriptive models of different classes and then compatibility of any new sample with different class rule-sets was assessed by using matching measures. Then, the class corresponding to the best matched ruleset was reported as the label of sample.

Most of the existing IDSs use all of the features in network packet to evaluate and look for known intrusive patterns [6], while it is better to find a small subset for classification purposes. In this study, the feasibility of applying fuzzy association rules for feature selection in intrusion detection systems will be demonstrated. To do this, a feature selection engine based on fuzzy association rule mining (FSE-FARM) is developed and a fuzzy ARTMAP neural network is used for classification, as well.

Corresponding Author: Mansour Sheikhan, P.O. Box 11365/4435, Post-Graduate Center, South Tehran Branch, Islamic Azad University, Iran. Tel: +98 21 88605565, E-mail: msheikhn@azad.ac.ir

The rest of paper is organized as follows. In section 2, fuzzy association rules and induction algorithms are described in details. In section 3 and 4, the feature selection approach and fuzzy ARTMAP neural network are reviewed, respectively. In section 5, knowledge discovery and data mining group (KDD) dataset, as the training and test dataset in this work and its preprocessing procedure is presented. In section 6, the proposed framework for intrusion detection is introduced. In section 7, the experimental results, carried out on KDD dataset, are given to show the effectiveness of the proposed method. Finally, section 8 draws conclusions.

Theoretical Backgrounds

Association Rules Mining: Association rule mining is one of the important topics in data mining research. This approach determines interesting relationships between large set of data items. This technique was initially applied to the so-called market basket analysis, which aims at finding regularities in shopping behavior of customers of supermarkets [1]. Agrawal in [7] has proposed the Apriori algorithm to find quickly Boolean association rules. In contrast to Boolean association rules, which handle only simple item-based transactions, the next generation of association rules faced quantitative attributes which their values were elements of continuous domains such as real number domain R. But, the typical Apriori algorithm was not capable of dealing directly with such attributes. Therefore, in [8] an algorithm has been proposed to mine quantitative association rules. This algorithm starts by partitioning the attribute domains and then transforming the problem into a binary one. This method can solve problems introduced by quantitative attributes, but it causes the "sharp boundary" problem. The sharp boundary problem either ignores or over-emphasizes the elements near the boundary of intervals in the mining process. As a remedy to the sharp boundary problem, the fuzzy set concept, introduced by Zadeh [9], has been used more frequently in mining quantitative association rules. This approach is better than partitioning method, because fuzzy sets provide a smooth transition between members and non-members of a set and increase the flexibility of systems.

Apriori Algorithm: As mentioned before, Agrawal proposed the Apriori algorithm that is known as a fundamental algorithm for mining frequent itemsets in a set of transactions. He defined the formal statement for association rules as follows: Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of items and D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. A set of items $X \subseteq I$ is called an *itemset*. It can be said that T contains an itemset X, if $X \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = \varnothing$. The rule $X \Rightarrow Y$ has *Support* S in the transaction set D, if $S\%$ of transactions in D contain $X \cup Y$ and it has *Confidence* C in the transaction set D, if $C\%$ of transactions in D that contain X also contain Y. Apriori algorithm works iteratively and has two steps: In the first step, it finds all the large itemsets and in the next step, it uses the large itemsets to generate effective association rules. A more detailed description of these two steps can be found in [7, 8, 10]. This algorithm should scan a database many times to find the large itemsets.

Fuzzy Association Rules Induction: Mining fuzzy association rules is the discovery of association rules, using fuzzy set concepts, such that databases with both categorical and quantitative attributes can be handled. In this study, fuzzy grids based rules mining algorithm (FGBRMA) is used to mine fuzzy association rules [11]. This algorithm consists of two phases: large fuzzy grids generation and fuzzy association rules generation. Let $I = \{i_1, i_2, \ldots, i_m\}$ be the itemset, where $i_j (1 \le j \le m)$ may be categorical or quantitative. At the first phase, this algorithm uses fuzzy partition method for each attribute so that both quantitative and categorical attributes are divided into K different linguistic values ($K=2, 3, 4, \ldots$). In fact, each attribute is viewed as a linguistic variable and the variables are divided into various linguistic values. Each linguistic value can be used to represent a candidate 1-dimensional fuzzy grid. Moreover, to generate a candidate high-dimensional fuzzy grid, we can apply two of large 1-dimensional fuzzy grids to construct one of candidate 2-dimensional fuzzy grids and so on. To check whether this fuzzy grid is large or not, its fuzzy support (FS) is computed. When its fuzzy support is larger than or equal to the pre-determined minimum fuzzy support (Min FS), it can be said that it is a large k-dimensional fuzzy grid. When all of the large fuzzy grids have been generated, the second phase will be started. In this phase, each R rule is generated by two large fuzzy grids. To check whether this rule is effective or not, its fuzzy confidence (FC) is computed. When its fuzzy confidence is larger than or equal to the pre-determined minimum fuzzy confidence (Min FC), the rule is considered as an effective rule. The FGBRMA is an efficient algorithm since it scans database only once and applies Boolean operations on tables to generate large fuzzy grids and fuzzy association rules. The brief description of this algorithm [11] is as follows (For more details see [12-14]).

Input: a. database; b. the pre-defined minimum fuzzy support; c. the pre-defined minimum fuzzy confidence.
Output: Large fuzzy grids generation (Phase I), Effective fuzzy association rules generation (Phase II).
Method:

Phase I. Large fuzzy grids generation	
1.	Perform the fuzzy partition
2.	Scan the database and construct the initial table FGTTFS
3.	Generate large 1-dim fuzzy grids
	3.1.Set K=1 and eliminate the rows of initial FGTTFS corresponds to candidate 1-dim fuzzy grids, which are not large.
	3.2. Reconstruct FGTTFS.
4.	Generate large K-dim fuzzy grids. Set K+1 to K. If there is only one (K-1)-dim fuzzy grid, then go to step 5.
	For any two unpaired rows, FGTTFS[u] and FGTTFS[v] (u ≠ v), correspond to large (K-1)-dim fuzzy grids do:
	4.1. From (FG[u] OR FG[v]) that corresponds to a candidate K-dim fuzzy grids c, if any two linguistic values are defined in the same linguistic variable, then discard c and skip steps 4.2 and 4.3 (that is, c is invalid.)
	4.2. If FG[u] and FG[v] do not share (K-2) linguistic terms, then discard c and skip steps 4.3 (that is, c is invalid.)
	4.3. Add (FG[u] OR FG[v]) to table FG, (TT[e_1] TT[e_2]…TT[e_k]) to TT and f_s to FS when f_s is larger than or equal to the min FS; otherwise, discard c.
	End.
5.	Check whether or not any large K-dim fuzzy grid is generated.
	If any large K-dim fuzzy grid is generated, then repeat by going to step 4, else continue to execute Phase II.
Phase II. Effective fuzzy association rules generation	
For two unpaired rows, FG[u] and FG[v] (u< v), correspond to large fuzzy grids L_u and L_v, do:	
1.	Generate the antecedent part of the rule
	1.1. Let temp be the number of nonzero elements in (FG[u] AND FG[v]).
	1.2. If the number of nonzero elements in FG[u] is equal to temp, then $L_v \subset L_u$ is hold and the antecedent part of one rule, R, is generated as L_u; otherwise skip steps 2 and 3.
2.	Generate the consequence of the rule
	Use (FG[u] XOR FG[v]) to obtain the consequence part of R.
3.	Check whether rule R can be generated [FC(R) = FS (L_v)/FS (L_u)] or not
	If FC(R) ≥ min FC, then R is effective.
End.	

Fig. 1: FGBRMA algorithm

In this algorithm a table structure, called FGTTFS, is implemented to generate large fuzzy grids. This table consists of the following substructures:

- Fuzzy grids substructure (FG): each row represents a fuzzy grid and each column represents a linguistic value.
- Transaction substructure (TT): each column represents a tuple t_p, while each element records the membership degree of t_p belongs to the corresponding fuzzy grid.
- Fuzzy support substructure (FS): stores the fuzzy support corresponding to the fuzzy grid.

The steps of the FGBRMA algorithm are listed in Fig. 1.

Fuzzy Clustering and Partition Generation: In this section, the fuzzy partition method for defining fuzzy membership functions is described. The membership generation techniques are usually obtained from expert knowledge or experimental approaches. Nevertheless, one way to determine membership functions of these linguistic values is based on Fuzzy C-Means (FCM) clustering algorithm. FCM method [15] is used in this paper. In FCM, every data point belongs to every cluster to a certain degree, μ, in the range [0, 1]. This algorithm tries to minimize the following objective function:

$$F_{obj}=\sum_{i=1}^{N}\sum_{j=1}^{N}\mu_{ij}^{m}\,\|x_i-c_j\|^2 \qquad (1)$$

Where μ_{ij} is the degree of membership of x_i in the cluster j, x_i is the d-dimensional measured data, c_j is the d-dimensional center of the cluster and ||*|| is the norm expressing the similarity between any measured data and the center. The fuzziness parameter, m, is an arbitrary real number (m>1). The brief overview of this approach is as follows: Let D be a set of transactions, where $t_1, t_2, \ldots, t_n$ are different crisp records and $I = \{i_1, i_2, \ldots, i_m\}$ be the set of attributes. Also, let $FP = \{FP_1, FP_2, \ldots, FP_m\}$, where $FP_m = \{f_1, f_2, \ldots, f_s\}$ be the set of fuzzy partitions of attribute i_m. Fig. 2 shows the algorithm for fuzzy partition generation using FCM.

Read fuzziness parameter m
For each $i_m \in I$
FP$_m$= apply_FCM(i_m)
For each partition f∈FP$_m$
Label 'linguistic value' appropriately

Function apply_FCM(i_m)
Read K (number of clusters)
Until max $\left\{\left\|u_{ij}^{(K+1)} - \mu_{ij}^{(K)}\right\|\right\} < \delta$
For each t_p∈D (p=1, 2, …, n)
For each cluster j (j=1, 2, …, K)
Calculate μ_{ij}

FP= set of fuzzy clusters (partitions) after completion of above iteration.
Return FP

Fig. 2: Fuzzy partition generation algorithm using FCM

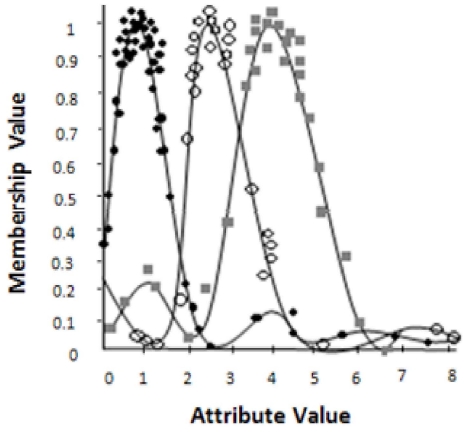

Fig. 3: Membership functions based on clustering

In this study, three linguistic values are defined for every attribute (i.e. the number of clusters is equal to 3) and triangular membership functions are used for each linguistic value. In Fig. 3, an example feature, i_m, is used to illustrate the function of FCM and creation of partitions. First, the values of $t_p[i_m]$ (for each p, such that $1 \le p \le n$) are partitioned into three clusters, using FCM. Then for each generated cluster, the membership degrees of samples are exploited to fit a triangle. The fitted triangle represents the membership function corresponds to that cluster. In Fig. 3 the membership degrees of the samples to three clusters, are plotted by "full dot", "circle" and "square" symbols, respectively.

Feature Selection Approach: Feature selection is an important issue in intrusion detection. The aim of feature selection is to segregate the irrelevant and redundant attributes from a dataset, thus the dimension of dataset will be reduced. Extra features increase the computational load and can impact the accuracy of the IDS [16]. Feature selection in intrusion detection application has been investigated in several studies and various techniques including machine learning and statistical approaches have been used in the recent decade. In [2], taxonomy of feature selection algorithms in IDS was presented. In [4], genetic algorithm (GA) was used for feature selection. GA is a well-known feature selection method for dimensionality reduction and it is commonly used in the literacy. In [17], principal component analysis (PCA) method was used for feature selection, as well. Also in [6] and [16], the wrapper method was proposed for feature selection. Wrapper method exploits a machine learning algorithm to evaluate the feature set. In [6], integration of rough set and particle swarm optimization (PSO) was used to form a 2-tier structure of feature selection process. While in [16], floating search was applied to select the optimal or near optimal subset of features. In this paper, a novel method of feature selection, by applying fuzzy association rules, is proposed in intrusion detection application.

Fuzzy ARTMAP Neural Network: The fuzzy ARTMAP achieves a synthesis of fuzzy logic and adaptive resonance theory (ART) neural networks by exploiting a close formal similarity between the computations of fuzzy method and ART category choice, resonance and learning [18]. This network is composed of two fuzzy ART modules, ART$_a$ and ART$_b$, interconnected by an inter-ART using an associative memory module as illustrated in Figure 4. The inter-ART module has a self-regulator mechanism, match tracking, whose objective is to maximize the generalization and minimize the network error. The F_2^a layer is connected to the inter-ART module by the weights w_{jk}^{ab}. The steps of fuzzy ARTMAP algorithm are summarized below.

Input Data: The input pattern of ART$_a$ is represented by the vector $a = [a_1 \ldots a_{Ma}]$ and the input pattern of ART$_b$ is represented by the vector $b = [b_1 \ldots b_{Mb}]$.
Parameters: There are three fundamental parameters for the performance and learning of fuzzy ART network [19].

- The choice parameter, ($\alpha > 0$): acts on the category selection.
- Learning rate, ($\beta \in [0,1]$): controls the velocity of network adaptation.
- Vigilance parameter, ($\rho \in [0,1]$): controls the network resonance. The vigilance parameter is responsible for the number of formed categories.

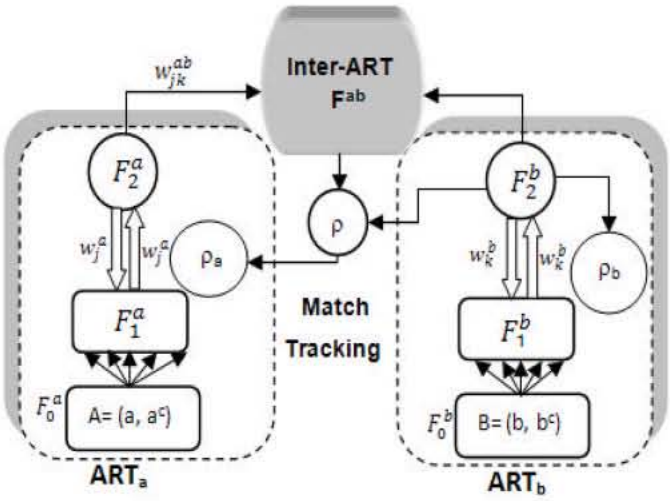

Fig. 4: Structure of the fuzzy ARTMAP

Algorithm Structure: After the resonance is confirmed in each network, J is the active category for the ART_a network and K is the active category for the ART_b network. The next step is match tracking to verify, if the active category on ART_a corresponds to the desired output vector presented to ART_b. The vigilance criterion is given by:

$$\rho_{ab} = \frac{\left| y^b \wedge w_{JK}^{ab} \right|}{\left| y_b \right|} \tag{2}$$

Learning: After the input has completed the resonance state by vigilance criterion, the weight adaptation is implemented. The adaptation of the ART_a and ART_b module weights is given by:

$$w_J^{new} = \beta\left(I \wedge w_J^{old}\right) + \left(1-\beta\right) w_J^{old} \tag{3}$$

Intrusion Database and Preprocessing of Features: In 1998, Defense Advanced Research Project Agency (DARPA) funded an "intrusion detection evaluation program (IDEP)" in Lincoln laboratory at the Massachusetts Institute of Technology which this dataset was prepared by Stolfo [20]. Since 1999, KDD'99 was built based on the data captured in DARPA'98. This dataset consists of three components and it contains a number of connection records where each connection is a sequence of packets containing values of 41 features and labeled as either normal or attack. Attack types in this dataset fall into four main categories: Denial of Service (DoS), Probe, User to Root (U2R) and Remote to Local (R2L) which is detailed in Table 1. In this research, we use KDD'99 [20] to evaluate our proposed framework. In this way, a subset of '10% KDD' dataset with the same distribution, is used for the purpose of training and a subset of 'Corrected KDD' dataset is used as the test set. Table 2 and Table 3 represent the distribution of normal and attack classes in the training and test dataset, respectively. The name and type of features in the KDD dataset are listed in Table 4, as well. Features in the KDD datasets have different forms: continuous, discrete and symbolic with significantly varying resolution and ranges. Most pattern classification methods are not able to process data in such a format. Hence, preprocessing is required.

Symbolic-valued features, such as protocol_type (3 different symbols), service (70 different symbols) and flag (11 different symbols) are mapped to integer values ranging from 0 to $S-1$, where S is the number of symbols. Continuous features having smaller integer value ranges, such as wrong_fragment [0,3], urgent [0,14], hot [0,101], num_failed_logins [0,5], num_compromised [0,9], num_root [0,7468], num_file_creations [0,100], num_shells [0,5], num_access files [0,9], count [0,511], srv_count [0,511], dst_host_count [0,255] and dst_host_srv_count [0,255], are also scaled linearly to the range [0,1].

Logarithmic scaling (base 10) is applied to three features spanned over a very large integer range (i.e. duration [0,58329], src_bytes [0,1.3billion] and dst_bytes [0,1.3billion]), to reduce the ranges to [0,4.77] and [0,9.11], respectively. Other features are either Boolean (e.g. logged_in), having binary values, or continuous in the range of [0,1] (e.g. diff_srv_rate) and

Table 1: Basic characteristics of KDD'99 components

Dataset	Normal	Probe	DoS	U2R	R2L
10% KDD	97277	4107	391458	52	1126
Corrected KDD	60593	4166	229853	70	16347
Whole KDD	972780	41102	3883370	52	1126

Table 2: Number of selected samples of '10% KDD' dataset for training

Class	Number of training samples	Distribution (%)
Normal	9727	19.69
Probe	411	0.83
DoS	39145	79.24
U2R	6	0.01
R2L	113	0.23
Total	49402	100.00

Table 3: Number of selected samples of 'Corrected KDD' dataset for test

Class	Number of test samples	Distribution (%)
Normal	6059	19.48
Probe	417	1.34
DoS	22985	73.90
U2R	7	0.02
R2L	1635	5.26
Total	31103	100.00

no scaling is needed for these features. So, each of the mapped features are linearly scaled to the range [0,1] [21].

Proposed Framework: The proposed framework for intrusion detection has composed of two blocks; FSE-FAR block and classification block. Fig. 5 shows a schematic view of the proposed intrusion detection system. As it is mentioned

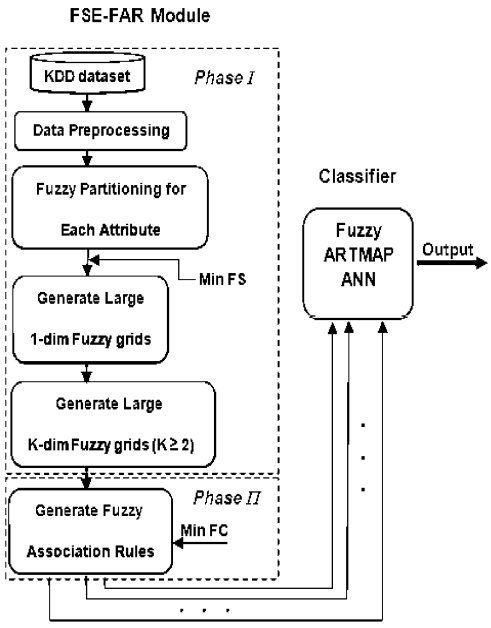

Fig. 5: Block diagram of proposed intrusion detection framework

Table 4: Name and type of 41 attributes in KDD dataset

Attribute number	Name	Type
1	duration	continuous
2	protocol_type	discrete
3	service	discrete
4	flag	discrete
5	src_bytes	continuous
6	dst_bytes	continuous
7	land	discrete
8	wrong_fragment	continuous
9	urgent	continuous
10	hot	continuous
11	num_faild_logins	continuous
12	logged_in	discrete
13	num_compromised	continuous
14	root_shell	continuous
15	su_attempted	continuous
16	num_root	continuous
17	num_file_creations	continuous
18	num_shells	continuous
19	num_access_files	continuous
20	num_outbound_cmds	continuous
21	is_host_login	discrete
22	is_guest_login	discrete
23	count	continuous
24	srv_count	continuous
25	serror_rate	continuous
26	srv_serror_rate	continuous
27	rerror_rate	continuous
28	srv_rerror_rate	continuous
29	same_srv_rate	continuous
30	diff_srv_rate	continuous
31	srv_diff_host_rate	continuous
32	dst_host_count	continuous
33	dst_host_srv_count	continuous
34	dst_host_same_srv_rate	continuous
35	dst_host_diff_srv_rate	continuous
36	dst_host_same_src_port_rate	continuous
37	dst_host_srv_diff_host_rate	continuous
38	dst_host_serror_rate	continuous
39	dst_host_srv_serror_rate	continuous
40	dst_host_rerror_rate	continuous
41	dst_host_srv_rerror_rate	continuous

in the previous section, each network connection record consists of 41 features; hence it is an important question that of all these features, which are the most useful and which are less significant. In this way, a feature selection engine based on fuzzy association rules is used. Association rules determine interesting relationships between large set of data items. Thus, they can be used to discover relationships between features in KDD dataset and then reduce the dimension of input vector [22, 23]. A fuzzy ARTMAP neural classifier is used in the classification block, too. In this study, we focus on two distinctive techniques to remove redundant inputs to fuzzy ARTMAP.

Table 5: Large itemsets for each class in IDS

Class	Large itemsets
Normal	1,4,5,6,10,11,13,16,17,18,19,22,23,25,26,27,28, 29,30,35,36,38,39,40,41
Probe	1,6,39
DoS	1,3,4,6,10,13,23,24,25,26,27,28,29,30,31,32,37,38,39,40,41
U2R	17,18,23,24,40
R2L	1,4,5,6,14,16,17,18,19,25,26,27,28,30,39,40,41

Technique 1: In this technique, the effective fuzzy association rules on KDD dateset are found. To achieve this purpose, the total input features (i.e. 41 features) and the total number of selected samples of '10% KDD' dataset (i.e. 49402 records) are applied to discover the relations between input features. In this method, the rules that have enough fuzzy support value and high fuzzy confidence value are found and then some inputs thanks to these rules can be removed. Suppose R rule form ($P \Rightarrow Q$); in this rule, Q itemset depend on P itemset. Thus, all items in Q itemset can be removed because they are redundant and not considered as the fuzzy ARTMAP inputs. In this study, after performing technique 1; no effective fuzzy association rule which has enough fuzzy support value and high fuzzy confidence value was found.

Technique 2: In this method, all of the input features are used and just large itemsets for each class are found. All of the instances in these large itemsets are the most important items for classification. If an item of a large itemset of any class is large in other classes, then that item must be used as the input of classifier. In this study, large itemsets for each class was found by using FGBRMA algorithm. The list of large itemsets for each class is reported in Table 5. According to Table 5, the dimension of input feature space is reduced from 41 to 31 by using the FSE-FAR module, so that the most important features for classification are selected as follow:

1-3-4-5-6-10-11-13-14-16-17-18-19-22-23-24-25-26-27-28-29-30-31-32-35-36-37-38-39-40-41

Experimental Results: In this work, the simulations were running on a PC powered by a Pentium IV, 3.6 GHz of CPU and 2 GB of RAM. Before discussing about the results of experiments, it seems necessary to mention the standard metrics that have been developed for evaluating IDS.

Detection rate (DR) and false alarm rate (FAR) are the two most common metrics. DR is computed as the ratio between the number of correctly detected attacks and the total number of attacks, while FAR is computed as the ratio between the number of normal connections that is incorrectly misclassified as attacks and the total number of normal connections. Another metric that is used here is the classification rate (CR). Classification rate for each class of data is computed as the ratio between the number of test instances correctly classified and the total number of test instances of this class. For the purpose of classifier algorithm evaluation, another comparative measure is defined which is cost per example (CPE) [24]. CPE is calculated using the following formula:

Table 6: Cost matrix values for KDD'99

Actual/Predicted	Normal	Probe	DoS	U2R	R2L
Normal	0	1	2	2	2
Probe	1	0	2	2	2
DoS	2	1	0	2	2
U2R	3	2	2	0	2
R2L	4	2	2	2	0

Table 7: Confusion matrix of fuzzy ARTMAP classifier

Actual/Predicted	Normal	Probe	DoS	U2R	R2L
Normal	6037	8	13	0	1
Probe	28	328	61	0	0
DoS	577	39	22366	3	0
U2R	2	1	2	2	0
R2L	690	4	3	0	938

Table 8: Confusion matrix of FSE-FAR + Fuzzy ARTMAP classifier

Actual/Predicted	Normal	Probe	DoS	U2R	R2L
Normal	6048	8	2	0	1
Probe	28	354	33	0	2
DoS	14	10	22920	0	41
U2R	2	1	2	1	1
R2L	664	0	2	0	969

$$CPE = \frac{1}{N_\tau}\sum_{i=1}^{M}\sum_{j=1}^{M} CM(i,j)C(i,j) \quad (4)$$

where CM and C are confusion matrix and cost matrix, respectively. N_T represents the total number of test instances and M is the number of classes in classification. CM is a square matrix in which each column corresponds to the predicted class, while rows correspond to the actual classes. An entry at row i and column j, $CM(i,j)$, represents the number of misclassified instances that originally belong to class i, although incorrectly identified as a number of class j. The entries of the primary diagonal, $CM(i,i)$, stand for the number of properly detected instances. Cost matrix is similarly defined, as well and entry $C(i,j)$, represents the cost penalty for misclassifying an instance belonging to class i into class j. Cost matrix values employed for the KDD'99 classifier learning contest are shown in Table 6 [20]. The confusion matrix when using fuzzy ARTMAP classifier without FSE-FARM module is reported in Table 7. The confusion matrix for the hybrid structure of FSE-FARM+Fuzzy ARTMAP is

Table 9: Performance of proposed IDS framework

	Classification rate								
Model	Normal	Probe	DoS	U2R	R2L	DR	FAR	CPE	Training time (sec)
Fuzzy ARTMAP	99.65	78.83	97.31	18.90	57.37	94.37	0.36	0.1341	224.0230
FSE-FAR+Fuzzy ARTMAP	99.82	84.93	99.72	17.52	59.28	96.81	0.18	0.0934	178.8333

Table 10: Performance of proposed IDS framework as compared to other machine learning models

	Classification rate							
Model	Normal	Probe	DoS	U2R	R2L	DR	FAR	CPE
Winner of KDD in 2000 [25]	99.5	83.3	97.1	13.2	8.4	91.8	0.6	0.2331
PNrule [24]	99.5	73.2	96.9	6.6	10.7	91.1	0.4	0.2371
Runner up of KDD in 2000 [26]	99.4	84.5	97.5	11.8	7.3	91.5	0.6	0.2356
ESC-IDS [27]	98.2	84.1	99.5	14.1	31.5	95.3	1.9	0.1579
Hybrid Elman/CPAR[b] [28]	97.4	91.5	97.0	33.3	31.8	92.6	2.6	NR[a]
FSE-FAR+Fuzzy ARTMAP	99.8	84.9	99.7	17.5	59.3	96.8	0.18	0.0934

[a] Not Reported

[b] Classification-based Predictive Association Rule

shown in Table 8, as well. The performance of proposed framework in term of CR, DR, FAR and CPE has been shown in Table 9 and also the performance of this method has been compared with some other machine learning methods in Table 10. The duration of the training process for fuzzy ARTMAP with 31 features was approximately 178.83 seconds. Using the same machine, the training took 224.02 seconds for 41 features. It can be seen that the reduced set of features decreases the computation time more than 20 percent.

As shown in Table 10, the proposed system has higher classification rate for all the classes, as compared to systems reported in [24-27]. This system performs better in term of DR, FAR and CPE, as well. So, it can be inferred that the proposed approach improves the detection rate and decreases the false alarm rate and the cost per example, effectively.

CONCLUSION

In this research, an intrusion detection framework based on fuzzy association rules and fuzzy ARTMAP neural network was proposed. Fuzzy association rules mining is able to sufficiently handle large amounts of data and it can discover important relationships between large set of data items. In the proposed model, fuzzy grids based rules mining algorithm (FGBRMA) was used to reduce the dimension of input feature vector to classifier. In this way, the dimension of input feature space was reduced from 41 to 31. Experimental results showed that the proposed hybrid model performed better in terms of classification rate, DR, FAR and CPE as compared to some other machine learning methods.

REFERENCES

1. Tajbakhsh, A., M. Rahmati and A. Mirzaei, 2009. Intrusion Detection Using Fuzzy Association Rules. Applied Soft Computing, 9: 462-469.
2. Chen, Y., Y. Li, X.Q. Cheng and L. Guo, 2006. Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System. Springer-Verlag Berlin Heidelberg LNCS, 4318: 153-167.
3. Jain, V., L. Benyoucef and S.G. Deshmukh, 2008. A New Approach for Evaluating Agility in Supply Chains Using Fuzzy Association Rules Mining. Engineering Applications of Artificial Intelligence, 21: 367-385.
4. Florez, G., S.M. Bridges and R.B. Vaughn, 2002. An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection. In the Proceedings of the North American Fuzzy Information Processing Society Conference, pp: 27-29.
5. Semary, A.E., J. Edmonds, J.G. Pino and M. Papa, 2006. Applying Data Mining of Fuzzy Association Rules to Network Intrusion Detection. In the Proceedings of the IEEE Workshop on Information Assurance, pp: 100-107.
6. Zainal, A., M.A. Maarof and S.M. Shamsuddin, 2007. Feature Selection Using Rough-DPSO in Anomaly Intrusion Detection. Springer-Verlag Berlin Heidelberg LNCS, 4705 (Part I): 512-524.

7. Agrawal, R. and R. Sricant, 1994. Fast Algorithms for Mining Association Rules. In the Proceedings of the 20th International Conference on Very Large Databases, pp: 487-499.
8. Sricant, R. and R. Agrawal, 1996. Mining Quantitative Association Rules in Large Relational Tables. In the Proceedings of the ACM SIGMOD International Conference on Management of Data, pp: 1-12.
9. Zadeh, L.A., 1965. Fuzzy Sets. Proceeding Information Control, 8: 338-353.
10. Agrawal, R., T. Imielinski and A. Swami, 1993. Mining Association Rules Between Sets of Items in Large Databases. In the Proceedings of the ACM SIGMOD International Conference on Management of Data, pp: 207-216.
11. Hu, Y.C., R.S. Chen and G.H. Tzeng, 2003. Discovering Fuzzy Association Rules Using Fuzzy Partition Methods. Knowledge-Based Systems, 16: 137-147.
12. Hu, Y.C., 2006. Determining Membership Functions and Minimum Fuzzy Support in Finding Fuzzy Association Rules for Classification Problems. Knowledge-Based Systems, 19: 57-66.
13. Hu, Y.C., R.S. Chen and G.H. Tzeng, 2002. Mining Fuzzy Association Rules for Classification Problems. Computers and Industrial Engineering, 43: 735-750.
14. Hu, Y.C., R.S. Chen and G.H. Tzeng, 2003. Finding Fuzzy Classification Rules Using Data Mining Techniques. Pattern Recognition Letters, 24: 509-519.
15. Bezdek, J.C., R. Ehrlich and W. Full, 1984. FCM: The Fuzzy C-Means Clustering Algorithm. Computers and Geosciences, 10: 191-203.
16. Vilakazi, C.B. and T. Marwala, 2006. Application of Feature Selection and Fuzzy ARTMAP to Intrusion Detection. In the Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp: 4880-4885.
17. Gao, H.H., H.H. Yang and X.Y. Wang, 2005. Principal Component Neural Networks Based Intrusion Feature Extraction and Detection Using SVM. Springer-Verlag Berlin Heidelberg LNCS, 3611: 21- 27.
18. Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynold and D.B. Rosen, 1992. Fuzzy ARTMAP: A Neural Network for Incremental Supervised Learning of Analog Multidimensional Maps, IEEE Transactions on Neural Network, 3: 689-713.
19. Carpenter, G.A., 2003. Default ARTMAP. In the Proceedings of International Joint Conference on Neural Networks, pp: 1396-1401.
20. Stolfo, S.J., 1999. KDD-99 dataset, (Available on http://www.kdd.ics.uci.edu/databases/kddcup99/kddcup99.html).
21. Sheikhan, M. and A. Sha'bani, 2009. Fast Neural Intrusion Detection System Based on Hidden Weight Optimization Algorithm and Feature Selection. World Applied Sciences Journal, 7 (Special Issue of Computer and IT): 45-53.
22. Karabatak, M. and M.C. Ince, 2009. An Expert System for Detection of Breast Cancer Based on Association Rules and Neural Network. Expert Systems with Applications, 36: 3465-3469.
23. Karabatak, M. and M.C. Ince, 2009. A New Feature Selection Method Based on Association Rules for Diagnosis of Erythemato-Squamous Diseases. Expert Systems with Applications, 36: 12500-12505.
24. Agrawal, R. and M.V. Joshi, 2000. PNrule: A New Framework for Learning Classifier Models in Data Mining (A Case-Study in Network Intrusion Detection). IBM Research Division, Report No.RC-21719.
25. Pfahringer, B., 2000. Winning the KDD 99 Classification Cup: Bagged Boosting. Journal of SIGKDD Explorations, 1: 65-66.
26. Levin, I., 2000. KDD Classifier Learning Contest: LLSoft's Results Overview. J. SIGKDD Explorations, 1: 67-75.
27. Nadjarian Toosi, A. and M. Kahani, 2007. A Novel Soft Computing Model Using Adaptive Neuro-Fuzzy Inference System for Intrusion Detection. In the Proceedings of the IEEE International Conference on Networking, Sensing and Control, pp: 834-839.
28. Sheikhan, M. and D. Gharavian, 2009. Combination of Elman Neural Network and Classification-Based Predictive Association Rules to Improve Computer Networks' Security. World Applied Sciences J., 7 (Special Issue of Computer and IT): 80-86.