World Applied Sciences Journal 14 (5): 772-781, 2011
ISSN 1818-4952

Intrusion Detection Improvement Using GA-Optimized Fuzzy Grids-Based Rule Mining Feature Selector and Fuzzy ARTMAP Neural Network

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

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

Abstract: In this paper, a framework is used for intrusion detection that shows the effectiveness of data mining techniques in this field. The proposed system is developed in two main phases and also a supplementary optimizing stage. At the first phase, the most important features are selected using fuzzy association rules mining (FARM) algorithm to reduce the dimension of input features to the misuse detector. At the second phase, a fuzzy adaptive resonance theory-based neural network (ARTMAP) is used as a misuse detector. The accuracy of the proposed approach depends strongly on the precision of the FARM module parameters and also the training parameters of fuzzy ARTMAP neural classifier. So, the genetic algorithm (GA) is incorporated into the proposed method to optimize the parameters of mentioned modules in this study. Classification rate (CR) results show the importance role of GA in improving the performance of the proposed intrusion detection system (IDS). The performance of proposed system is investigated in terms of detection rate (DR), false alarm rate (FAR) and cost per example (CPE). Experimental results show that the proposed approach performs better in terms of FAR and CPE in comparison to other machine learning algorithms and its DR is 97.2%. Meanwhile, the reduced size of feature set decreases the computation load of the system.

Key words: Fuzzy grids · Rule mining · Intrusion detection · Feature selection · Neural network · Genetic algorithm

INTRODUCTION

With the increasing growth of computer networks in recent years, network security has become a priority in this field. Vulnerabilities in common security components such as firewalls are inevitable. Intrusion detection systems (IDSs) are used as an extra 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. 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.

Up to now, several researches and various methods of intrusion detection have been developed [2-7]. However, there is a growing interest in intrusion detection community toward the application of machine learning techniques in this field. Considering this trend and the extensive amount of data involved in intrusion detection problem, data mining approaches seem to be appropriate for this purpose [8, 9]. In the recent years, data mining that 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. Fuzzy association rules have been applied to intrusion detection systems, as well. For example in [1], fuzzy association rules have been exploited as descriptive models of different classes and then compatibility of any new sample with different class rulesets has been assessed by using matching measures. Then, the class corresponding to the best matched ruleset has been reported as the label

Corresponding Author: Mansour Sheikhan, Faculty of Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran. P.O. Box: 11365-4435. Tel: +98 21 33706159, Fax: +98 21 33717140.

of sample. In [8], the authors have used sets of fuzzy association rules that were mined from network audit data as models of "normal behavior" and they have generated fuzzy association rules from new audit data to detect anomalous behavior and then computed the similarity with sets mined from "normal" data. In [10], El-Semary et al. have used a data mining algorithm to discover fuzzy rules from network traffic data.

Most of the existing IDSs use all of the features in network packet to evaluate and look for known intrusive patterns, while it is better to find a small subset for classification purposes. Extra features increase the computational load and can impact the accuracy of the IDS. In this way, the feasibility of applying fuzzy association rules for feature selection in intrusion detection systems has been demonstrated by the authors in [11]. To do this, a feature selection engine based on fuzzy association rules mining has been developed and a fuzzy ARTMAP neural network has been used for classification. In the proposed method, size-adjustment (SA) parameter is an important factor, which is used to control the size of the feature subset. Also, minimum fuzzy support (Min FS) and minimum fuzzy confidence (Min FC) are important parameters in finding frequent itemset and generating fuzzy association ruleset, respectively. In this study, the genetic algorithm (GA) is incorporated into the mentioned method to determine the three mentioned thresholds. Also, the optimum values for the parameters of fuzzy ARTMAP are determined using GA.

The rest of paper is organized as follows. In section 2, the main concepts and preliminaries related to the methodology used in this work are described. In section 3, the framework for intrusion detection is introduced in details. In section 4, the results of the experiments carried out on KDD'99 dataset are presented and compared with some recent works in literature using the same dataset. Finally, section 5 draws conclusions.

Preliminaries

Foundations of Rule Mining: The objective of data mining is to obtain useful and non-explicit information from data stored in the large repositories. One important topic in data mining research is concerned with the discovery of interesting association rules. Association rules determine the interesting relationships between large set of data items. This technique has been initially applied to the so-called market basket analysis, which aims at finding regularities in shopping behavior of customers of supermarkets [1].

Given an itemset I and a transaction set T, where each transaction is a subset of I, an association rule is said to be an "implication" of the form "$A \Rightarrow C$" denoting the presence of itemsets A and C in some of the transactions, assuming that $A,C \subset I$, $A \cap C = \varnothing$; and $A,C \neq \varnothing$. The measures proposed in [12] for establishing an association rule's fitness are as follow: *support*(Supp($A \Rightarrow C$) which is the joint probability $p(A \cup C)$) and the *confidence*(Conf($A \Rightarrow C$) which is the conditional probability $p(C|A)$).

Apriori [12] is the best known basic 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 have faced quantitative attributes which their values were elements of continuous domains such as real number domain R. However, the typical Apriori algorithm was not capable of dealing directly with such attributes. Therefore, in [13] 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. In other words, it 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 [14], 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 the systems. In this study, the use of fuzzy association rules is considered as the key component in IDS structure because of the affinity with the human knowledge representation.

In other words, mining fuzzy association rules is the discovery of association rules, using fuzzy set concepts, such that the quantitative attributes can be handled. Let $I=\{i_1,\ldots,i_m\}$ be an itemset and T a fuzzy transaction set, in which each fuzzy transaction is a fuzzy subset of I. Given the transaction $t \in T$, we will use $t(i)$ to denote the membership degree of item I in the transaction t. Various proposals for fuzzy association rules can be found in the literature such as generalization of association rules when initial data are fuzzy [15-19]. An interesting in depth study into the extensions to quantitative attribute cases can be found in [19]. In this study, fuzzy grids based rules mining algorithm (FGBRMA) [19] is used to mine the fuzzy association rules. In this algorithm, each attribute is

viewed as a linguistic variable and the variables are divided into various linguistic terms. FGBRMA is an efficient algorithm since it scans database only once and applies the Boolean operations on tables to generate large fuzzy grids and fuzzy association rules.

Genetic Algorithm: GA is a method for solving optimization problems based on natural selection, the process that drives biological evolution. It is a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation and recombination. GA repeatedly modifies a population of individual solutions. At each step, GA selects individuals randomly from the current population to be parents and uses them to produce the children for the next generation. There are several methods for selecting parents such as stochastic uniform selection, remainder selection, uniform selection, roulette selection and tournament selection. In addition, to create the next generation from current population, GA uses: 1. crossover rules which combine two parents to form children for next generation, 2. mutation rules which apply random changes to the individual parents to form children, 3. selecting the best individual of fitness function as elite child. Over successive generations, the population evolves toward an optimal solution.

As it has been mentioned before, GA is incorporated into the proposed method to determine the optimum values of three thresholds in FARM-based feature selector (Min FS, Min FC and size-adjustment parameter) and also network training parameters of fuzzy ARTMAP (α, ρ and β).

Our target in using GA is automatically determining suitable threshold values. For this purpose, we use real-valued coding, where the chromosomes are represented as floating point numbers and their genes are the real parameters. In this study, we encode our problem to solve it by genetic algorithm. This encoding is get from [24]. First, we assign an index to each item. Then, we perform encoding of the $X \rightarrow Y$ rule according to Fig. 1, where j in this figure is an indicator that separates the antecedent from the consequent of the rule. That is, $X=\{A_1,\ldots,A_j\}$ and $Y=\{A_{j+1},\ldots A_k\}$; $0<j<k$. Therefore, a k-rule $X \rightarrow Y$ is represented by $k+1$ positive integers.

Our goal is to search the most interesting fuzzy association rules. Hence, the fitness function is very important for determining the interestingness of chromosome and it affects the convergence of the genetic algorithm. In this study, the fitness function is defined as follows [24]:

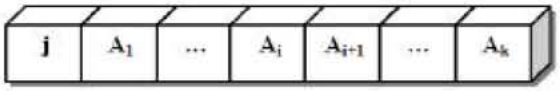

Fig. 1: Encoding of a k-rule

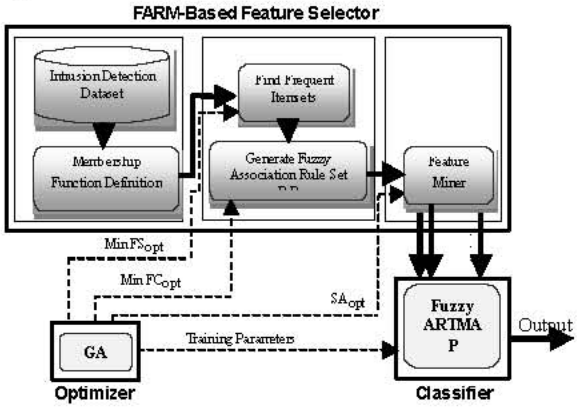

Fig. 2: Block diagram of proposed GA-optimized intrusion detection framework

$$fitness(c) = \frac{supp(A_1 \ldots A_k) - supp(A_1 \ldots A_j)\sup p(A_{j+1} \ldots A_k)}{supp(A_1 \ldots A_j)(1 - supp(A_{j+1} \ldots A_k))} \quad (1)$$

Where $c=(j,A_1,\ldots,A_j,A_{j+1},\ldots,A_k)$ is a given chromosome. The fitness of c is, in fact, the relative confidence of the corresponding association rule $\{A_1,\ldots,A_j\} \rightarrow \{A_{j+1},\ldots,A_k\}$.

IDS Model: In this study, the proposed model for intrusion detection system has been composed of three modules; FARM-based feature selector module, classification module and GA module for optimization of the parameters of two other modules. Figure 2 shows a schematic view of the proposed intrusion detection system. In the FARM-based feature selector module, the system uses a fuzzy data mining algorithm to generate the fuzzy association rules. The subset of the features discovered by the fuzzy data mining algorithm is used as the inputs of fuzzy ARTMAP-based neural classifier. In this study, the GA module is incorporated into the model to determine the optimum values of three important parameters of FARM-based feature selector (Min FS, Min FC and SA parameter) and training parameters (choice, vigilance and learning rate) of fuzzy ARTMAP classifier.

FARM-Based Feature Selector Module: In this study, knowledge discovery and data mining group (KDD) dataset [20] is used to train and test the intrusion detection system. The details of KDD'99 dataset is described in section 4.1. The FARM-based feature selector module comprises the following three stages:

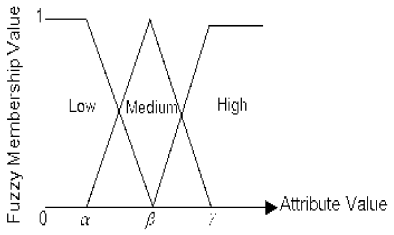

Fig. 3: Definition of the fuzzy membership function

Defining Fuzzy Membership Functions: In KDD'99 dataset, there are totally 41 features used to describe each session (see Appendix). To define the fuzzy membership functions, each feature value is transformed to three linguistic terms (Low, Medium and High). In other words, each feature is divided into three subfeatures with linguistic terms. A predefined membership function is assigned to each feature and the linguistic terms can be expressed by the membership function shown in Fig. 3. The parameters α, β and γ in a fuzzy membership function for feature F_i are considered as follow [21]:

β : Average value of feature F_i in the dataset;
γ : The largest value of feature F_i in the dataset;
$\alpha = 2\beta - \gamma$;

Search for Fuzzy Association Rules: As we have already mentioned, FGBRMA [19] is used in this study to mine fuzzy association rules. In this stage, the frequent itemsets are found by computing the fuzzy support counts of candidate itemsets. To check whether each candidate itemset is large or not, its fuzzy support 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 frequent itemset. After finding all of the frequent itemsets, fuzzy association rules are generated using the frequent itemsets. To check whether each rule r is effective or not, its fuzzy confidence 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 acceptable rule. As mentioned earlier, GA is used to determine the optimum values of Min FS and Min FC parameters in this work.

Extraction of Features: The aim of feature miner layer is to segregate the irrelevant and redundant features from original dataset. It finds the relationships among features in rule set R and then eliminates some unnecessary features. Suppose rule r form as $X \Rightarrow Y$; where X is the antecedent and Y is the consequence. In this rule, itemset Y depends on itemset X. Thus, all items in itemset Y can be eliminated because they are redundant.

begin
1) $S_f=\varnothing$; //The final feature subset
2) **if** (Rule Set R==$\varnothing$) **then break**
3) **else**
4) **for** each rule[r]$\in$R
5) **if** interesting(r) **then**
6) $S_f=S_f$+(extract_linguistic_variables_in_antecedent[r]);
7) Update R by deletion of all linguistic variables covered by the rule[r];
8) **end for;**
9) **end if;**
10) **Return** S_f;
end

Fig. 4: Feature mining algorithm

bool interesting(r)
1) **begin**
2) **if**(fuzzy confidence[r]$\geq$size-adjustment)
3) **return true:**
4) **else**
5) **return false;**
6) end

Fig. 5: Interesting(r) function

Figure 4 shows the details of feature mining algorithm. This algorithm employs *interesting(r)* Boolean function, for each rule r, to determine whether the rule r is interesting or not. If *interesting(r)* function (Fig. 5) returns "True", then the linguistic variables that only appear in the antecedent of rule r are extracted and all of them are added to the S_f. Then, a simple deletion procedure is performed that cancels all of the linguistic variables covered by the rule r from the ruleset R. At the final step, the feature set S_f will be the result of feature selection process. In *interesting(r)* function, the size-adjustment parameter controls the size of feature subset. Choosing smaller values for size-adjustment parameter leads to larger size feature subsets. In this study, GA is used to determine the optimum value of the size-adjustment parameter.

Classification Module: In this study, fuzzy ARTMAP neural network is used as the classification tool to measure the usefulness of FARM-based feature selector module. This network 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 [22]. It is composed of two fuzzy

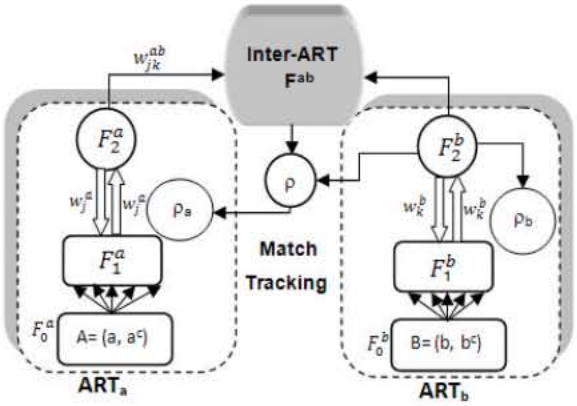

Fig. 6: Structure of the fuzzy ARTMAP

ART modules, ART_a and ART_b, interconnected by an inter-ART using an associative memory module as illustrated in Fig. 6. 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^a_2 layer is connected to the inter-ART module by the weights w^{ab}_{jk}.

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 $b = [b_1 \ldots b_{Mb}]$.

There are three fundamental parameters for the performance and learning of fuzzy ART network [23]: the choice parameter ($\alpha > 0$), the learning rate ($\beta \in [0,1]$) and the vigilance parameter ($\rho \in [0,1]$).

After the resonance is confirmed in each network, assume that 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:

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

Finally, 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 modules weight is given by:

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

In this study, the optimized values of three mentioned parameters are determined by GA to have the best correct identification rate.

Experimental Results

Intrusion Dataset: As mentioned before, KDD dataset [20] is used to evaluate the proposed framework for intrusion detection. This dataset is a common benchmark for evaluation of intrusion detection techniques. KDD'99 consists of several components, that two of them are used in this work. This dataset contains a number of connection records where each connection is a sequence of packets containing values of 41 features. Also, attack types in this dataset fall into four main categories: denial of service (DoS), probe, user to root (U2R) and remote to local (R2L). In all experiments described below, '10% KDD' dataset is used for the purpose of training and 'Corrected' dataset is used as a test set. Several new and novel never-before-seen attacks have been used in 'Corrected KDD' in order to assess the generalization ability of IDSs. Statistical details of the two KDD components used here are summarized in Table 1.

Evaluation Criteria: 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) [25]. CPE is calculated using the following formula:

Table 1: Characteristics of KDD'99 components used for train and test

Dataset	Total attack patterns	Total normal patterns	Total patterns
10% KDD	396,743	97,278	494,021
Corrected	250,436	60,593	311,029

Table 2: Cost matrix values for KDD'99

	Predicted				
Actual	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 3: Specifications of GA parameters in simulations

Parameter	Value
Selection probability	0.90
Crossover probability	0.80
Mutation probability	0.01
Maximum size of population	100
Maximum iteration number	9000
Seed chromosome	(3,46,68,83)

Table 4: Optimum value of FARM algorithm parameters determined by GA

Parameter	Value
Min FS	0.4266
Min FC	0.7864
Size-adjustment	0.8311

$$CPE = \frac{1}{N_T}\sum_{i=1}^{m}\sum_{i=1}^{m} CM(i,j).C(i,j) \quad (4)$$

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 2 [20].

Experiments Setup and Results: We have used KDD CUP'99 dataset and conducted some experiments to assess the effectiveness of the proposed GA-optimized FARM-based feature selector. In our experiments, various values were selected and tested for crossover probability, mutation rate and population size as the important parameters of GA. Table 3 gives the parameters that result in the best achievements. So, we continue our simulations by using them.

After running the genetic algorithm, the optimum values of FARM algorithm parameters were obtained (Table 4). By applying feature mining algorithm, the linguistic variables that only appeared in the antecedent of each rule are extracted and all of them are added to the final feature subset.

When the proposed approach is implemented with Min FS=0.4266 and Min FC=0.7864, the total of 3295 rules were discovered. From these, there were 628 rules with two elements, 958 rules with three elements, 883 rules with four elements and 826 rules with five elements. In this way, some of the mined rules with different number of elements are listed in Table 5. As can be seen, the FS and FC values for these rules are greater than 0.4266 and 0.7864, respectively.

By using the proposed algorithm, the dimension of input feature space has been reduced and the most important features have been selected for classification. As it is mentioned in section 4.1, each network connection record in KDD'99 dataset consists of 41 features. This algorithm can result in considerable reduction in the size of feature set. In this way, if the dimensions of input feature space are reduced from 41 to 31, then the selected features are as shown in Table 6.

Optimized Neural Net Classifier: Before evaluating the system, we determine the optimum values of important parameters of fuzzy ARTMAP neural net by GA. The fitness function in fuzzy ARTMAP simulation is chosen as follows:

Table 5: Fuzzy association rules with FS ≥ 0.4266 and FC ≥ 0.7864

Rule	FS value	FC value
$\{F_1=Low\}\Rightarrow\{F_8=High\}$	0.4961	0.8135
$\{F_3=Low\}\Rightarrow\{F_5=Medium\}$	0.4532	0.7911
$\{F_6=High\}\Rightarrow\{F_{13}=Medium\}$	0.5814	0.8367
$\{F_4=High\}\Rightarrow\{F_9=Low\}$	0.5836	0.8134
$\{F_{13}=Medium\}\Rightarrow\{F_{14}=Medium\}$	0.5902	0.9356
$\{F_{17}=High\}\Rightarrow\{F_{22}=High\}$	0.4435	0.7946
$\{F_{29}=Low\}\Rightarrow\{F_{33}=Low\}$	0.5317	0.9123
$\{F_3=Low \wedge F_4=High\}\Rightarrow\{F_{10}=High\}$	0.5804	0.9843
$\{F_3=Low \wedge F_{10}=High\}\Rightarrow\{F_{12}=Low\}$	0.4628	0.8732
$\{F_6=High \wedge F_{10}=High\}\Rightarrow\{F_{12}=Low\}$	0.5183	0.8437
$\{F_{10}=High \wedge F_{13}=Medium\}\Rightarrow\{F_{13}=Medium\}$	0.6133	0.9644
$\{F_{11}=Low \wedge F_{13}=Medium\}\Rightarrow\{F_{15}=Low\}$	0.4853	0.8647
$\{F_{14}=Medium \wedge F_{17}=High\}\Rightarrow\{F_{20}=Medium\}$	0.5713	0.8927
$\{F_6=High \wedge F_{10}=High \wedge F_{16}=Medium\}\Rightarrow\{F_{21}=Low\}$	0.4767	0.7953
$\{F_{11}=Low \wedge F_{13}=Medium \wedge F_{19}=Medium\}\Rightarrow\{F_{24}=Low\}$	0.5371	0.8763
$\{F_{19}=Medium \wedge F_{23}=High \wedge F_{24}=Low\}\Rightarrow\{F_{27}=High\}$	0.4727	0.7903
$\{F_{25}=High \wedge F_{26}=High \wedge F_{30}=Low\}\Rightarrow\{F_{32}=Medium\}$	0.5981	0.9677
$\{F_{32}=Medium \wedge F_{35}=High \wedge F_{36}=Low\}\Rightarrow\{F_{34}=High\}$	0.6002	0.9886
$\{F_{14}=Medium \wedge F_{16}=Medium \wedge F_{17}=High\}\Rightarrow\{F_{22}=High \wedge F_{23}=High\}$	0.5128	0.7913
$\{F_{31}=Low \wedge F_{35}=High \wedge F_{36}=Low\}\Rightarrow\{F_{33}=Low \wedge F_{34}=High\}$	0.4937	0.8624
$\{F_{35}=High \wedge F_{37}=High \wedge F_{38}=High\}\Rightarrow\{F_{40}=Medium \wedge F_{41}=Low\}$	0.4373	0.8031

Table 6: Selected features based on optimum values for parameters of FARM-based feature selector

Selected features	Size of feature subset
$F_1,F_3,F_4,F_5,F_6,F_{10},F_{11},F_{13},F_{14},F_{16},F_{17},F_{18},F_{19},F_{22},F_{23},F_{24},$	
$F_{25},F_{26},F_{27},F_{28},F_{29},F_{30},F_{31},F_{32},F_{35},F_{36},F_{37},F_{38},F_{39},F_{40},F_{41}$	31

Table 7: GA specifications for optimum parameter values of fuzzy ARTMAP

Population size	Selection function	Crossover function	Mutation function
50	Stochastic uniform	Scattered	Gaussian

Table 8: Specification of fuzzy ARTMAP, GA-optimized

Specification	Value
Learning rate (β)	0.9763
Vigilance parameter (ρ_a)	0.9856
Vigilance parameter (ρ_{ab})	0.4103
Choice parameter (α)	0.9934
Number of F_0 nodes	15
Number of F_1 nodes	400
Number of F_2 nodes	400

$$F = (pc)^2 \quad (5)$$

Where pc is the correct classification rate. The population size and selection/crossover/mutation functions that result the best fitness value are reported in Table 7. The specifications of fuzzy ARTMAP in our simulations are reported in Table 8, as well.

After determining the appropriate structure and parameter values for fuzzy ARTMAP (Table 8), we have evaluated the performance of proposed framework in terms of detection rate (DR), false alarm rate (FAR) and cost per example (CPE). The performance of proposed framework in term of CR of different attack patterns, DR, FAR and CPE is reported in Table 9. In this work,

Table 9: Performance of proposed GA-optimized IDS framework

	Classification rate								
Model	Normal	Probe	DoS	U2R	R2L	DR	FAR	CPE	Execution time (sec)
GA-optimized fuzzy ARTMAP (No feature selection)	99.71	79.13	98.35	18.93	58.41	94.86	0.35	0.1312	211.8417
GA-optimized FARM-based feature selector + GA-optimized Fuzzy ARTMAP	99.89	86.27	99.84	17.57	60.17	97.22	0.17	0.0924	196.7125

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
Fuzzy class-association-rule mining based on genetic network programming [21]	NR[a]	NR[a]	NR[a]	NR[a]	NR[a]	98.7	0.53	NR[a]
PNrule [25]	99.5	73.2	96.9	6.6	10.7	91.1	0.4	0.2371
Winner of KDD in 2000 [26]	99.5	83.3	97.1	13.2	8.4	91.8	0.6	0.2331
Runner up of KDD in 2000 [27]	99.4	84.5	97.5	11.8	7.3	91.5	0.6	0.2356
ESC-IDS [28]	98.2	84.1	99.5	14.1	31.5	95.3	1.9	0.1579
Hybrid Elman/CPAR[b] [29]	97.4	91.5	97.0	33.3	31.8	92.6	2.6	NR[a]
FARM-based feature selector +Fuzzy ARTMAP [11]	99.8	84.9	99.7	17.5	59.3	96.8	0.18	0.0934
GA-optimized FARM-based feature selector + GA-optimized Fuzzy ARTMAP (Proposed)	99.9	86.3	99.8	17.6	60.2	97.2	0.17	0.0924

a: Not-Reported

b: Classification-based Predictive Association Rule

the simulations were running on a PC powered by a Pentium IV, 3.6 GHz of CPU and 2 GB of RAM. In this study, the feature selection process when using GA-optimizer module nearly takes 30 seconds on average. The duration of the training phase for fuzzy ARTMAP with 31 features was approximately 166.71 seconds. Using the same machine, the training took 211.84 seconds for 41 features. It can be seen that the reduced set of features decreases the computation time more than 7%.

Table 10 shows the performance of this method compared with some other machine learning methods. As shown in Table 10, the proposed system has higher classification rate for all of the classes, as compared to systems reported in [11, 21, 25-29]. This system performs better in term of DR, FAR and CPE as compared to the results reported in [25-28]. So, it can be inferred that the proposed approach increases the detection rate and decreases the false alarm rate and the cost per example, effectively. However, the DR of proposed model is slightly lower than the system reported in [21]. Also, it is noted that its performance in terms of FAR is better than the system reported in [21]. So, it is interesting to note that by using FARM algorithm and fuzzy ARTMAP ANN which have been equipped by GA-based optimizer, the proposed hybrid IDS performs well with a reduced size of feature set.

CONCLUSION

In this research, an intrusion detection framework based on fuzzy association rules and fuzzy ARTMAP neural network has been proposed. The proposed method objective is to find a set of fuzzy association rules, equipped by GA to automatically determine the optimum parameter values of proposed system. 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) has been used for finding fuzzy association rules to discover the most important features. In this way, the dimension of input feature space has been reduced from 41 to 31. Experimental results have shown that the proposed hybrid model performed better in terms of classification rate, FAR and CPE in comparison to some other machine learning methods.

APPENDIX

Name and type of 41 features 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

REFERENCES

1. Tajbakhsh, A., M. Rahmati and A. Mirzaei, 2009. Intrusion Detection Using Fuzzy Association Rules. Applied Soft Computing, 9: 462-469.
2. Hoang, X.D., J. Hu and P. Bertok, 2009. A Program-Based Anomaly Intrusion Detection Scheme Using Multiple Detection Engines and Fuzzy Inference. J. Network and Computer Applications, 32: 1219-1228.
3. Saniee Abadeh, M., J. Habibi and C. Lucas, 2007. Intrusion Detection Using a Fuzzy Genetics-Based Learning Algorithm. J. Network and Computer Applications, 30: 414-428.
4. Özyer, T., R. Alhajj and K. Barker, 2007. Intrusion Detection by Integrating Boosting Genetic Fuzzy Classifier and Data Mining Criteria for Rule Pre-Screening. J. Network and Computer Applications, 30: 99-113.
5. Wang, G., J. Hao, J. Ma and L. Huang, 2010. A New Approach to Intrusion Detection Using Artificial Neural Networks and Fuzzy Clustering. Expert Systems with Applications, 37: 6225-6232.
6. Botha, M. and R.V. Solms, 2003. Utilizing Fuzzy Logic and Trend Analysis for Effective Intrusion Detection. Computers & Security, 22: 423-434.
7. Jiang, F., Y. Sui and C. Cao, 2010. An Information Entropy-Based Approach to Outlier Detection in Rough Sets. Expert Systems with Applications, 37: 6338-6344.
8. 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: 457-462.
9. Lee, W., S.J. Stolfo and K.W. Mok, 1999. A Data Mining Framework for Building Intrusion Detection Models. In the Proceeding of the IEEE Symposium on Security and Privacy, pp: 120-132.
10. El-Semary, A., J. Edmonds, J. Gonzalez-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.
11. Sheikhan, M. and M. Sharifi Rad, 2010. Misuse Detection Based on Feature Selection by Fuzzy Association Rule Mining. World Applied Sciences Journal, 10 (Special Issue of Computer & Electrical Engineering): 32-40.

12. Agrawal, R., T. Imieliński and A. Swami, 1993. Mining Association Rules between Set of Items in Large Databases. In the Proceedings of ACM SIGMOD Conference, pp: 207-216.
13. Srikant, R. and R. Agrawal, 1996. Mining Quantitative Association Rules in Large Relational Tables. In the Proceedings of ACM SIGMOD International Conference on Management of Data, pp: 1-12.
14. Zadeh, L., 1965. Fuzzy Sets. In Proceeding of Information and Control, 8: 338-353.
15. Berzal, F., I. Blanco, D. Sánchez and M.A. Vila Miranda, 2001. A New Framework to Assess Association Rules. In the Proceedings of the International Conference on Advances in Intelligent Data Analysis, Springer-Verlag, 2189: 95-104.
16. Delgado, M., N. Marín, D. Sánchez and M.A. Vila Miranda, 2003. Fuzzy Association Rules: General Model and Applications. IEEE Transactions on Fuzzy Systems, 11: 214-225.
17. Hong, T.P., C.S. Kuo and S.C. Chi, 1999. Mining Association Rules from Quantitative Data. Intelligent Data Analysis, 3: 363-376.
18. Kuok, C.M., A. Fu and M.H. Wong, 1998. Mining Fuzzy Association Rules in Databases. ACM SIGMOD Record, 27: 41-46.
19. 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.
20. 1999 KDD Cup Competition (Available on http:// kdd.ics.uci.edu/ databases/ kddcup99/ kddcup99.html).
21. Mabu, S., C. Chen, N. Lu, K. Shimada and K. Hirasawa, 2011. An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming. IEEE Transaction on Systems, Man and Cybernetics, Part C: Applications and Reviews, 41: 130-139.
22. Carpenter, G.A., S. Grossberg, N. Markuzon, J.H. Reynolds 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.
23. Carpenter, G.A., 2003. Default ARTMAP. In the Proceedings of the International Joint Conference on Neural Networks, 2: 1396-1401.
24. Yan, X., C. Zhang and S. Zhang, 2009. Genetic Algorithm-Based Strategy for Identifying Association Rules without Specifying Actual Minimum Support. Expert Systems with Applications, 36: 3066-3076.
25. 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, Technical Report TR 00-015, Report No. RC-21719.
26. Pfahringer, B., 2000. Winning the KDD99 Classification Cup: Bagged Boosting. ACM SIGKDD Explorations Newsletter, 1: 65-66.
27. Levin, I., 2000. KDD'99 Classifier Learning Contest: LLSoft's Results Overview. ACM SIGKDD Explorations Newsletter, 1: 67-75.
28. Nadjaran 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.
29. Sheikhan, M. and D. Gharavian, 2009. Combination of Elman Neural Network and Classification-Based Predictive Association Rules to Improve Computer Networks' Security. World Appl. Sci. J., 7 (Special Issue of Computer & IT): 80-86.