Journal of Engineering and Applied Sciences 15 (6): 1445-1450, 2020
ISSN: 1816-949X

A Genetic Fuzzy Model for Investigating Security and Trust in E-Commerce with Genetic Algorithm

[1]Neda Najaafi, [2]Mohammadsaeid Zahedi, [3]Fatemeh Mokhtari Esfidvjani and [4]Arash Hedayati
[1]Iran Industrial Design Company, Tehran, Iran
[2]Bioinformatics and Computational Biology Research Center,
Shiraz University of Medical Sciences, Shiraz, Iran
[3]Department of Computer Engineering, Islamic Azad University, Karaj, Iran
[4]Department of Technical and Engineering, University of College of Nabi Akram, Tabriz, Iran

Abstract: In e-commerce, the issue of security and trust has always been one of the most important principles. With the advancement of internet science, e-commerce has also become more advanced, always striving for greater security and trust. Of course, security and trust are not new issues in the area of e-commerce and a variety of techniques have been provided to improve the performance of this business. If an e-commerce business fails to meet the customer's expectations in terms of security and trust, it would be doomed to fail. Security in e-commerce is the protection of existing assets and information against attacks and unauthorized access to that e-commerce. In this study, it has been tried to prevent such attacks by preventive techniques and strategies and also by providing an artificial immune system based on agents and computational intelligence techniques including fuzzy control and Genetic algorithms. The rules generated by fuzzy logic are taught by Genetic algorithm and finally through these rules, the input patterns are categorized. In this study, the activities are classified into two categories of safe and unsafe and the incorrect boundaries between these activities are reduced by the proposed technique.

Key words: Security, trust, e-commerce, fuzzy-genetic, attacks

INTRODUCTION

Nowadays, due to the spread of internet science and information technology, we can do most of our day-to-day business through the internet such as daily internet purchases, internet payments, money transfers and other things, leading to emergence of a number of internet businesses called e-commerce. In fact, e-commerce means the exchange of information through the internet and computer networks. But any e-commerce must be secure and reliable enough to attract customers; in other words, it should be, so, secure and trustworthy that customers can log in and perform their purchase safely and without fear of being compromised. The principle of security and trust is the most important factor in the growth and development of e-commerce in which customers need to have complete confidence. As all transactions are done online in e-commerce, it can be an important place to steal information. Therefore, the intruders are constantly trying to penetrate the desired sites in a clever way and to do their damage through virus, robbery, fraud or manipulation. Activities carried out by a group to access secret resources are called intrusion (Heady *et al*., 1990). Therefore, to protect the system from intruders, a system must be developed to identify and prevent them from being accessed by the knowledge of the professionals who implement the attack program, so that, upon arrival of the intruders, they can be detected and alarm is sent to the database (Abadeh *et al*., 2011). But in traditional businesses, these were not the case because security was provided physically. So far, many preventative measures have been taken to improve the performance of this business and maintain security and trust, among which installing software, antivirus and encoders can be mentioned as the influencing ones. Nowadays, better optimization techniques have been developed that can provide more security through fuzzy Genetic algorithms, though these techniques require smarter infrastructure. Fuzzy Genetic algorithms have provided a powerful way to develop e-commerce security. In this research, it has been tried to discuss more about fuzzy Genetic techniques for optimizing security and trust in e-commerce.

The concept of security in e-commerce covers a wide range of areas. In examining the security of each system, according to the principles set out in standards, the system assets must first be identified and valued and then the risks to each asset need to be examined after which a strategy can be considered for each risk. In examining each of these and in order to provide appropriate solutions, we must first understand and carefully consider

Corresponding Author: Mohammadsaeid Zahedi, Bioinformatics and Computational Biology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

the dangers that threaten them (Abadeh *et al*., 2011). Trust has different definitions in various discourses but at the same time all these theories affirm the existence of trust as a value. Trust means a direct relationship between the guarantor and the reliant. Trust enables people to live together in an uncertain and risky environment and it also provides a means to reduce some of the confusion in this complex world. In fact, trust is one of the most influential factors that can drive the development and growth of e-commerce. Trust has become a cornerstone in communications and marketing theories. In fact, it is a valuable asset for any business and in addition it reduces risk, achieves satisfaction, creates commitment and provides long-term customer relationships (Debar *et al*., 1992).

e-commerce is the process of buying, selling or exchanging products, services and information through computer networks and the internet. e-commerce can be defined as any business which is performed online and through the internet. This technique has grown a lot in recent years, with e-commerce being any transaction that involves the purchase or sale of goods or services over the internet and leads to the import or export of goods or services. e-commerce usually has a wider application not only involving buying and selling over the internet but also other aspects of business such as purchasing, merchandising, product management, procurement and distribution and after-sales services. Of course, the broader concept of e-commerce is e-business (Lunt and Jagannathan, 1988).

Literature review: The methods of intrusion detection or attacks are divided into two main categories: anomaly detection and pattern abuse detection. Anomaly detection is based on user-safe behaviors and detects any deviations from these behaviors as intrusions or attacks. Statistical methods, expert systems (Feng, 2005) and neural networks (Tajbakhsh *et al*., 2009) are among the ways to do this. In this way, new or unsafe attacks are detectable but the percentage of error alarm is high. The pattern detection method can detect attacks that have previously been defined as pattern signatures for the system. This method is not capable of detecting new undefined attacks but the error alarm rate is low. Expert system methods (Al Naqshbandi and Samawi, 2012), Genetic algorithms (Shanmugavadivu and Nagarajan, 2011) and pattern matching methods have been used for this method.

One of the most common methods is to combine fuzzy logic with Genetic algorithm which results in fuzzy genetic systems. GFS is a fuzzy system that is taught by a learning process based on evolutionary computation such as Genetic algorithms and other evolutionary algorithms. To solve the intrusion detection problem, various studies have been performed to obtain the

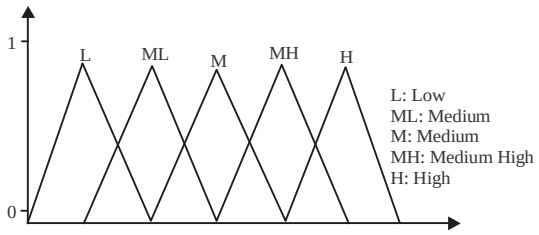

Fig. 1: Alphabetic properties are normalized by setting a random value of zero and one (Membership functions and fuzzy space for numerical features of the data set)

classification system. In designing and analyzing different types of fuzzy genetic systems for pattern recognition, the feasibility of fuzzy genetic systems in relation to intrusion detection is examined. In Olyaee *et al*. (2016) fuzzy rules are taught using particle swarm optimization algorithm. In a method using associative rules with fuzzy logic has been presented. Most of the things that happen to us in everyday life are ambiguous. The ambiguity may be associated with the shape, location, color, composition and content of the events and the meanings explain their identity. That is to say, using different meanings helps to describe and explain their existence and nature. The theory of fuzzy sets was first put forward by Professor Lotfizadeh in 1964. The foundation of fuzzy logic is based on fuzzy set theory, a generalization of classical set theory in mathematical science. In classical set theory, an element is either a member of the set or not. In fact, element membership follows a zero and one binary pattern. But fuzzy set theory extends this notion and proposes graded membership. In this way, an element can be to some degree not exactly a member of a set. The membership degree of an object to the fuzzy set is determined by the membership function. The range of this function is the range of the objects' values and their output in the interval [0-1]. The fuzzy space for the normalized numerical values of the dataset is shown in Fig. 1.

GENETIC FUZZY TECHNIQUE

The Genetic algorithm is inspired by genetic science and Darwin's theory of evolution and is based on the survival of the best by natural selection. Genetic algorithms is a special type of evolutionary algorithm that uses biological techniques such as inheritance and mutation. This algorithm was first introduced by John Holland. Later with Goldberg's efforts in 1989, this method found its place and today, due to its unique features is well positioned among other methods. In fact, Genetic algorithms use Darwin's natural selection principles to find the optimal formula to predict or adapt the pattern. A common use of Genetic algorithms is to use

it as an optimization function. Genetic algorithm is a useful tool in pattern recognition, attribute selection, image comprehension and machine learning. In Genetic algorithms, the genetic evolution of living things is simulated. Usually the application of this algorithm with fuzzy logic leads to algorithms that have an optimization approach. At each stage of the Genetic algorithm implementation, a bunch of search space points are randomly processed in a way that a sequence of characters is attributed to each point and genetic operators are applied to these sequences. The resulting sequences are then digged to obtain new points in the search space. Finally, the likelihood of their participation in the next step is determined by the value of the objective function at each point. Genetic algorithms can be considered as a random stochastic optimization technique that gradually moves to the optimum point. Regarding the features of the Genetic algorithm compared to other optimization methods, it can be said that it is an algorithm that can be applied to any problem without any knowledge of the problem, while no restrictions are applicable on the type of its variables and its effectiveness has been demonstrated in finding global optimum. The capability of this method is to solve complex optimization problems where classical methods are either inapplicable or unreliable in finding the global optimum. The main motivation for the Genetic algorithm can be stated regarding the fact that "gradual evolution" has been substantially complemented by the development of complex species and types through relatively simple mechanisms. A Genetic algorithm encodes the problem into a set of strings containing fine particles and then applies changes to the strings to stimulate the process of gradual evolution.

The optimization procedure in the Genetic algorithm is based on a randomly guided process. This method is based on the theory of gradual evolution and Darwin's fundamental ideas. In this method, a set of target parameters is randomly generated for a constant number called the population. After running the program, a numerical simulator that represents the standard deviation or processing of that data set is assigned to that member of the population. This procedure is repeated for each of the created members, after which the next generation of Genetic algorithm is formed by calling Genetic algorithm operators including fertilization, mutation and selection, and this process continues until the convergence criterion is met. Genetic algorithms are usually implemented as a computer simulator in which the population, an abstract sample (chromosomes) of the solution candidates of an optimization problem leads to a better solution. In GA, a population of chromosomes is produced. Each chromosome represents a possible solution to the problem. These chromosomes evolve using genetic operators (selection, cutting, mutation). Figure 2 shows an overview of the proposed system. This system runs

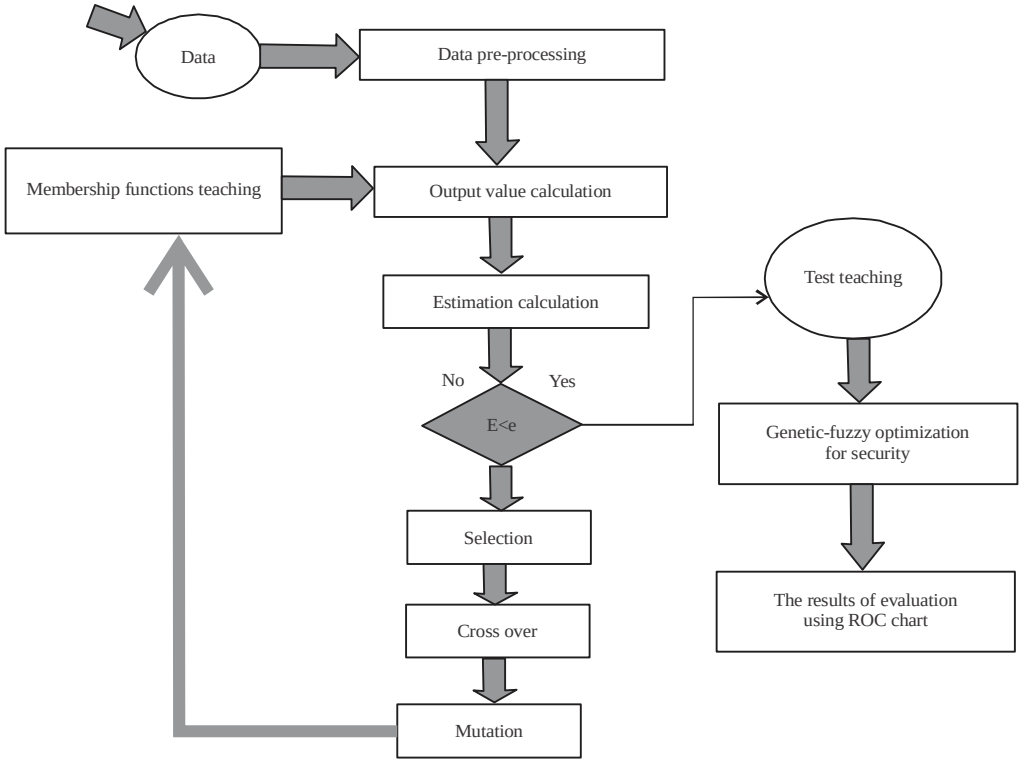

Fig. 2: Overview of the proposed system

(1)<condition>::=

<condition><operator><atomic_cond>

<atomic_cond><operator><atomic_cond>

(2)<atomic_cond>::==

<set><rel op><variable>

<operator>::=

<variable>::==

<rel op>::=

<set>::= L|ML|M|M|MH|H

Fig. 3: Genetic algorithm

Table 1: Criteria for classification of safe and unsafe activities

Output type	Predicted type	Actual type
True Positive (TP)	Safe	Safe
False Positive (FP)	Unsafe	Safe
True Negative (TN)	Unsafe	Unsafe
False Negative (FN)	Safe	Unsafe

separately for each pattern type. Each rule is represented as a chromosome. The following grammar is used to construct the rule section.

A chromosome is a set of n genes. Each gene consists of an atomic condition and a fuzzy operator, the latter gene containing only the atomic condition. In the chromosome, the n value that represents the genes or indeed the number of traits, is randomly selected. The initial population is randomly generated, consisting of a set of chromosomes each of which has a different number of genes. The estimator function determines the degree of each chromosome in the current population; the following criteria are used in the activity classification problem to calculate the estimator function (Fig. 3 and Table 1).

FUZZY RELATIONS

Fuzzy relations are as follows:

$$TP = \sum_{i=1}^{p} Fuzzy\left(normal\_data_i\right)$$

$$TN = \sum_{i=1}^{q}\left[1 - Fuzzy\left(abnormal_{data_i}\right)\right]$$

$$FP = \sum_{i=1}^{q} Fuzzy\left(abnormal\_data_i\right)$$

$$FN = \sum_{i=1}^{p}\left[1 - Fuzzy\left(normal\_data_i\right)\right]$$

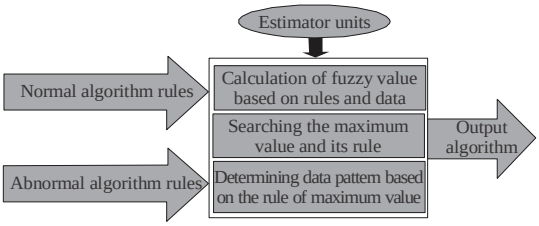

Fig. 4: Higher fuzzy value is known as a pattern category (A schematic view of classification)

Table 2: The number of files taken for the test and training phase

Training dataset	Values
Normal	000.25
DOS	000.25
Probe	4107
RLA	77
URA	42

$$Sensitivity = \frac{TP}{TP + FN}.\ Sensitivity = \frac{TN}{TN + FP}$$

$$Length = 1 - \frac{chrom - length}{100}$$

$$Fitness = w_1 * sensitivity + w_2 * specifity + w_3 * length$$

TP, TN, FP and FN indicate true positive, true negative, false positive and false negative, respectively.

Fuzzy genetic classification system: After teaching the system with secure and insecure training data, the set of rules for each of these types of patterns is obtained. Now in the testing phase, these rules are applied to the experimental dataset. Any rule that has a higher fuzzy value is known as a pattern category, as shown in Fig. 4.

This section describes the experimental results and performance evaluation of the proposed system. The proposed system has been implemented in MATLAB and the performance of the system has been evaluated using accuracy, calling and F measurement. For the empirical evaluation, 99 samples from the 37 KDD cup data were extracted, most commonly used to evaluate the performance of the intrusion detection system. To evaluate the performance, working with the proposed system in the 99 KDD cup dataset is very difficult because it is large scale, so, here, 10% of the 99 KDD Cup dataset has been used for training and testing. Table 2 show the number of files taken for the test and training phase.

The training dataset contains normal data as well as four types of attacks that are given to the proposed system to identify suitable attributes. The properties selected for the rule-making process are presented in Table 3. Then, using the fuzzy rules learning strategy, the system creates

Table 3: Selected attribute for rule generation

Selected attributes	Attribute feature
Duration	1
Src bytes	5
Dst bytes	6
Wrong fragment	8
Urgent	9
Hot	10
Num failed login	11
Num compromised	13
Nim root	16
Num file creation	17
Num shells	18
Num access files	19
Count	23
Sry count	24

Table 4: Experimental results

Type of alarm	Detection rate	Values
Normal pattern	01.95	04.0
Attack pattern	96.94	0367.0

a definite and infinite rule and eventually, fuzzy rules are generated from the specified rules. In the testing phase, the experimental dataset is given to the proposed system that classifies the input as usual or attack. Then the obtained result is used to calculate the overall accuracy of the proposed system. The overall accuracy of the proposed system is calculated on the basis of definitions, precision, call and F measurements which are commonly used to predict rare classes. The advantage is in elimination of the high recall above and also in precision. The F-criterion is the harmonic weight criterion that evaluates the trade-off between them.

Experiments on KDDCUP99 data: Experiments have been carried out on ten percent of the KDDCUP99 dataset on a laptop with a corei7 processor and GB4 memory. The results are presented in Table 4.

Evaluation of separation method for two types of safe and unsafe activities: In diagnostic applications between multiple classes, there's a concern of imbalanced classes when statistical classifiers are applied. This concern is due to many early possibilities that exist between the classes, leading to poor classification performance. Accordingly, specific tests based on the integration matrix are used. One of these specific tests is the evaluation test of the receiver operating characteristic diagram which uses specific criteria to express the correct rate of detection. The AUC parameter which is the area under the ROC curve is used to measure the classification rate. This parameter provides a scalar number for comparing different classifiers. Since, AUC is part of a single square, its value is a number between 0 and 1. When classification is performed randomly, it has an AUC value of <0.5. The closer AUC is to 1, the better classification performance will be obtained. In this study, colitis is shown with code 1 and Crohn's disease is indicated with code zero. Then, the ROC curve is used to evaluate the

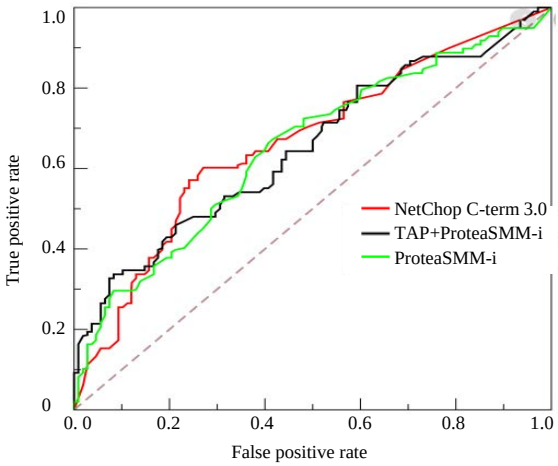

Fig. 5: Analysis of ROC diagram

combination classification theory with high-level consensus voting in the separation analysis of the two types of inflammatory bowel disease and the method is evaluated-regarding its precision and validity through ROC chart. According to Fig. 5, the diagram in MATLAB environment was executed using coding. Eventually the area under the AUC curve will show the accuracy of the model. The higher the area under the curve, the greater will be the accuracy of the model obtained. In this study, the target diagram for the safe and unsafe activities classification method was calculated with accuracy of 98%.

In Fig. 5, these parameters are used to evaluate the proposed algorithm. As can be seen, the value of AUC equals to 0.84. As such, this indicates that the fuzzy Genetic algorithm works better.

CONCLUSION

In this study, a fuzzy genetic classification system is developed that is capable of producing rules with variable length and number of properties. The proposed system can classify the type of input pattern as safe and unsafe. The detection rate for each pattern matching is 95% for the safe pattern and 94% for the unsafe pattern compared to other fuzzy Genetic algorithms, indicating good results.

REFERENCES

Abadeh, M.S., H. Mohamadi and J. Habibi, 2011. Design and analysis of genetic fuzzy systems for intrusion detection in computer networks. Expert Syst. Appl., 38: 7067-7075.

Al Naqshbandi, S.M. and V.W. Samawi, 2012. One-Rule Genetic-Fuzzy Classifier. Proceedings of the 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), May 25-27, 2012, IEEE, Zhangjiajie, China, pp: 204-208.

Debar, H., M. Becker and D. Siboni, 1992. A neural network component for an intrusion detection system. Proceeding of the Symposium Research Security and Privacy, May 4-6, 1992, Oakland, CA., pp: 240-250.

Feng, H.M., 2005. Particle swarm optimization learning fuzzy systems design. Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05; Vol. 1), July 4-7, 2005, IEEE, Sydney, NSW, Australia, pp: 363-366.

Heady, R., G. Luger, A. Maccabe and M. Servilla, 1990. The architecture of network level intrusion detection system. Technical Report, Department of Computer Science, University of New Mexico.

Lunt, T. and R. Jagannathan, 1988. A prototype real-time intrusion detection expert system. Proceeding of the IEEE Symposium on Security and Privacy, April 18-21, 1988, Oakland, CA., USA., pp: 59-66.

Olyaee, M.H., A. Yaghoubi and M. Yaghoobi, 2016. Predicting protein structural classes based on complex networks and recurrence analysis. J. Theor. Biol., 404: 375-382.

Shanmugavadivu, R. and N. Nagarajan, 2011. Network intrusion detection system using fuzzy logic. Indian J. Comput. Sci. Eng., 2: 101-111.

Tajbakhsh, A., M. Rahmati and A. Mirzaei, 2009. Intrusion detection using fuzzy association rules. Appl. Soft Comput., 9: 462-469.