An Evolutionary Fuzzy System Architecture for Information Protection

A. ISAZADEH H. NOURI M. B. REZVAN
Department of Computer Science
Tabriz University
Tabriz, IRAN

Abstract: - This paper presents the architectural issues of a system being proposed for information security assurance check. The system is intended to extract certain parameters from network map, risk analysis, fault tree, user interface analysis, data sensitivity and gateways analysis. Using these parameters as input a fuzzy classification system would identify the normal and vulnerable points in a given information system. A genetic search technique would also be used to find the related and appropriate rules for a given set of inputs.

Key-Words: Information systems, information assurance, fuzzy classifiers, genetic algorithm

1 Introduction

The human society is undergoing a fundamental transformation, from an industrial society to the information society. In this information age, all human activities are changing and undergoing a vast transformation. Advances in information processing and communication are opening up exciting new possibilities. There is a shift from stand-alone systems to networked information and processes [2].

Growth of Internet or "network of networks", globalization of trade, and rise of information-based economy redefine the role of information [3, 4]. E-business and E-commerce are becoming more and more popular, day by day. Protecting the investments specially money and time are important parts of every business. Since information is the most important investment of digital firms, it is imperative that the information and information systems be kept secure and protected at all times. Even in the past, where no computers and no networks were in use, securing and protecting the information against various threats were some of the important tasks of every organization. Today, since the activities of organizations are more closely related to the information and since the accurate information is considered as a powerful tool to survive and compete in this era, protection of information is vital for organizations.

There are numerous threats with a wide range of harms against the integrity of information. These harms are not discrete; they must be quantified by some appropriate parameters. Fuzzy linguistic expressions are useful to represent these harms [5].

Using fuzzy rules gives the opportunity to make fuzzy classifiers and identify the miss-configured and vulnerable points. The rules are categorized with respect to the inputs, and each computation is performed based on an appropriate category of rules. We also use a genetic algorithm to search the rules space and find the related rules.

2 Why information is vulnerable?

When large amount of data are stored in electronic forms, they become real vulnerable to all different kinds of threats. These threats can stem from technical, organizational, and environmental factors compounded by poor management decisions.

Advances in telecommunications and software have increased the information threats. Distributed information systems, which are normally consists of some interconnected subsystems may be accessed by unauthorized users. With the growth of networking technologies, unauthorized access and abuse can take place at any access point of networks. The ever increasing complexity of hardware, software, and organizational arrangements create new opportunities for abuse. In wireless networks, using radio-based technology, radio frequencies can be easily scanned and changed. Because of the fact that Internet is originally designed for easy access to information, a new era of penetrations come along with Internet [3].

Finally, the fact that information is the most valuable commodity in today's *information society* makes it attractive for hackers and intruders to surf the internet, intercept communications, and find the information they want, with or without authorization. All these are why information is vulnerable. Figure 1 shows the common vulnerable points in the information systems.

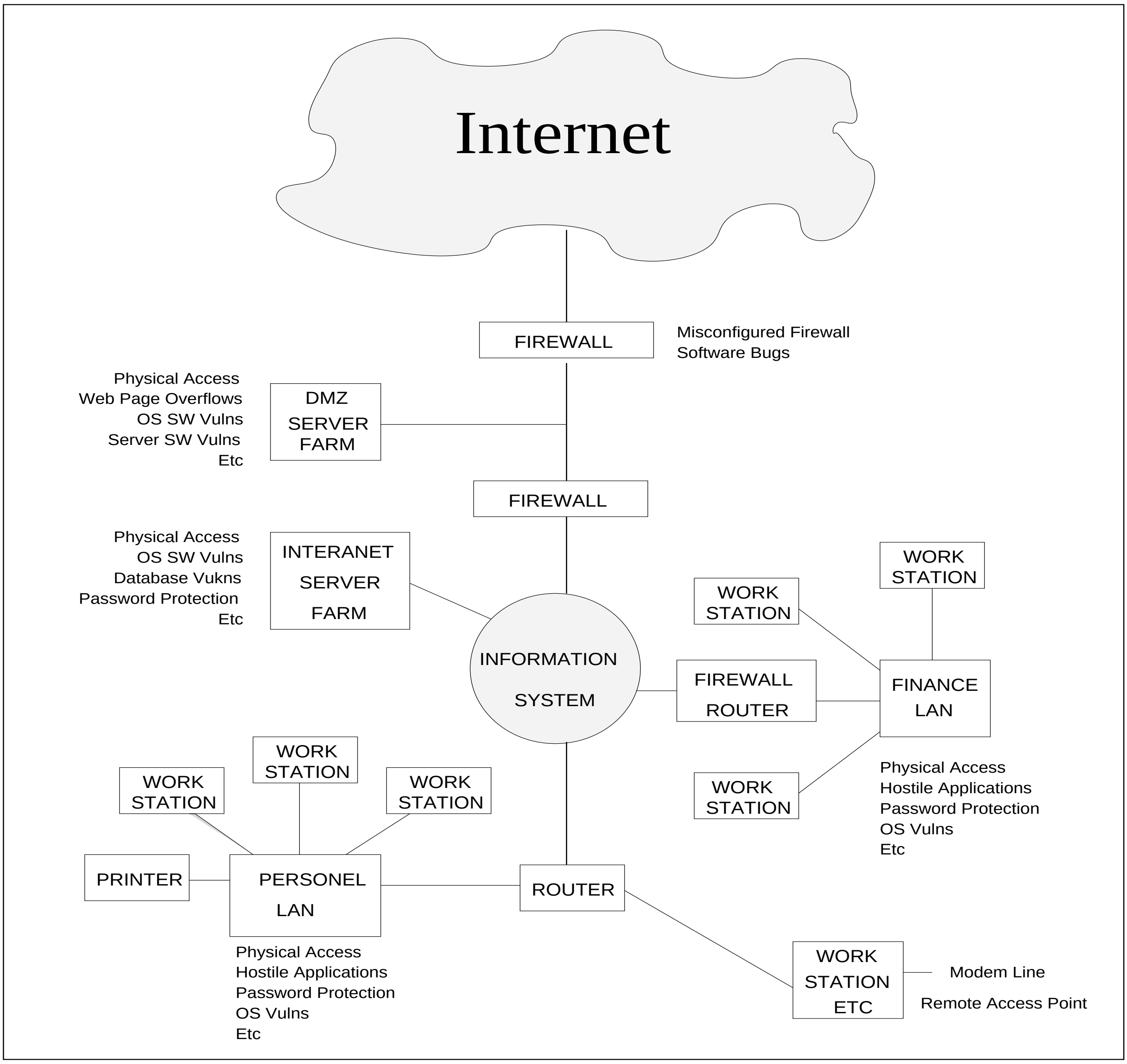

Figure 1: Information systems vulnerable points (From [1])

3 Fuzzy logic concepts

In fuzzy logic, *fuzzy sets* define the linguistic expressions and *membership functions* define the truth-value of such linguistic expressions [6].

Table 1 shows the difference between classic sets and fuzzy sets. The membership degree to a fuzzy set of an object defines a function where the universe (set of values that the object can take) is the domain, and the interval [0,1] is the range [7]. That function is called the membership function.

Figure 2 shows the most widely used membership function, the triangular membership function. In Figure 2, the object x has 0.5 degree of membership to the fuzzy set low.

x does not entirely belong to the set low, but x belongs to the fuzzy set and does not belong to the set at the same time. A collection of fuzzy sets, called fuzzy space, defines the fuzzy linguistic values or fuzzy classes that an object can belong to. A standard fuzzy space is shown in Figure 3.

Using fuzzy spaces, fuzzy logic allows an object to belong to different classes at the same time. This concept is helpful when the difference between classes is not well defined [8].

4 Genetic Algorithms

John Holland is often referred to as the “father of genetic algorithms.” He developed this brand of genetic programming during the 1960s and 1970s [9]. His student, David

Fuzzy sets	**Classic sets**
In fuzzy sets an object can partially be in a set.	In classic sets an object is entirely in a set or is not.
The membership degree takes values between 0 and 1	The membership degree takes only two values 0 or 1.
1 means entirely in the set, 0 means entirely not in the set, other values mean partially in the set.	1 means entirely in the set, 0 means entirely not in the set. Other values are not allowed.

Table 1: Fuzzy sets and classic sets

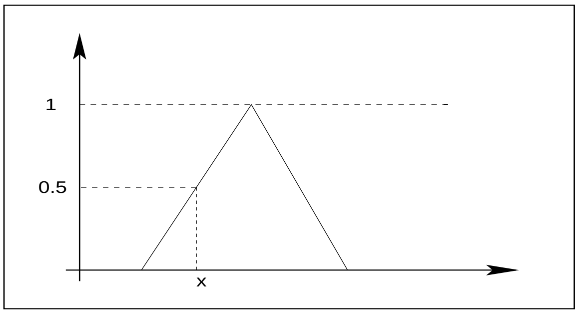

Figure 2: Triangular membership function

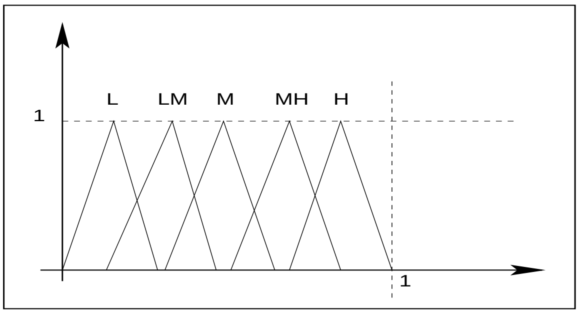

Figure 3: Fuzzy spaces of five fuzzy sets

Goldberg, popularized the method by solving a difficult problem involving the control of gas pipeline transmission in his dissertation [10]. Since that time, genetic algorithms have been applied to a wide variety of problems.

In using genetic algorithms (GA), the first step is defining an *objective function* with inputs and outputs. A binary GA encodes the value of each input parameter as a binary number. The parameter values are then placed side by side in an array known as a chromosome. A *population* is a matrix with each row representing a chromosome. The algorithm begins with a population, which is basically consist of some sequences of randomly generated ones and zeros. These random binary digits can be considered as representations of values of the input parameters. Next, the binary chromosomes are converted to continuous values which are evaluated by the objective function. *Mating* takes place between selected chromosomes. Mates are randomly selected with a probability of selection greater for those chromosomes yielding desirable output from the objective function (tournament or roulette wheel selection) [11]. Offspring (new chromosomes) produced from mating inherit binary codes from both parents. A simple crossover scheme randomly picks a crossover point in the chromosome. Two offsprings result by keeping the binary strings to the left of the crossover point for each parent and swapping the binary strings to the right of the crossover point. Crossover mimics the process of meiosis in biology. Mutations randomly convert some of the bits in the population from "1" to "0" or visa versa. The objective function outputs associated with the new population are calculated and the process repeated. The algorithm stops after finding an acceptable solution or after completing a preset number of iterations [12].

5 The Evolutionary Fuzzy Information Protection System

The main part of the proposed system is consist of a set of fuzzy rules. Fuzzy rules have the form $X \rightarrow Y$ [13]. Y is a fuzzy expression that represent the expected manner of the system. For each Y there is a set of prerequisites. Assuming that the set of prerequisites have the form $\{X_1, X_2, \ldots, X_n\}$, each X_i is a fuzzy expression. An atomic X_i has a *trust value* or a membership degree to the fuzzy sets in the fuzzy space. If X_i is a composed expression, we compute the trust value of the expression according to the table 2 [8].

logic operator	Fuzzy operator
$X_i \wedge X_j$	$min(X_i, X_j)$
$X_i \vee X_j$	$max(X_i, X_j)$
$\neg X_i$	$1 - X_i$

Table 2: Fuzzy operators

We classify the the problem domain in two-class classification. The first class is the normal points class and the second is the vulnerable points class of information system. In order to classify, we compute the trust value for system rules. Trust value of the rules are computed as follows: $R : X \rightarrow Y$
We assign for each rule a real number, denoting the *weight* which represents the confidence of the rule. As a fuzzy expression, X has a trust value. The trust value of R is computed as the product of the trust value of X by the weight of the rule; in other words,

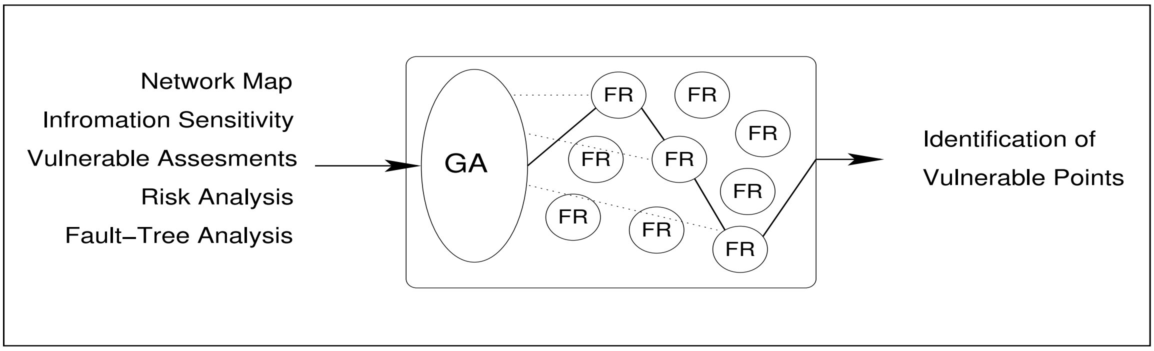

Figure 4: Information systems high risk nodes detector

$tv(R) = tv(X) * weight$

Now for each rule of the system we can compute the trust value. We use these values as the basis of classification. For each input we can make a fuzzy rule that represents the desired consequent. For example, assuming that X_i is an input and the desired result is Y, then the input rule can be stated as follows:

$R_i : X_i \rightarrow Y$

This rule has a pattern in system such as R_s that represents the minimum trust value for the result Y. if the trust value of the input rule is less than the trust value of the corresponding system rule, the input point is vulnerable, otherwise the point is normal.

$$class = \begin{cases} vulnerable, & if \ \ tv(R_i) < tv(R_s) \\ normal, & if \ \ tv(R_i) \geq tv(R_s) \end{cases}$$

We use a genetic algorithm to search the rules space to find the appropriate rules according to the inputs. Using genetic algorithm gives opportunity to test the various composition of rules to find the best compositions. The schema of proposed system is shown in Figure 4.

6 conclusion

In this paper we have discussed the possibility of using genetic algorithms and fuzzy expressions to identify the vulnerable points of information systems. We outlined the architecture of a system, by which information systems can be analyzed, the points of vulnerabilities can be detected, and possible information protection measures can be applied. We have shown that using fuzzy expressions is useful to represent the harms of numerous threats against information integrity. Using fuzzy classifiers reduces computations that need to detect the vulnerable points. A genetic algorithm finds the related rules and produces the combination of rules that give the best results.

References:

[1] K. Fox, Fuzzyfusion: Taking information assurance to the next level. Harris Corporation STAT Network Security Products, 2001.

[2] S. Stojakovic-Celustka, Building secure information systems, Master's thesis, Czech Technical University in Prague Faculty of Electrical Engineering Department of Computer Science & Engineering, 2000.

[3] K. C. Laudon and J. P. Laudon, *Management Information Systems*. Prentice Hall, 8 ed., 2002.

[4] J. A. O'Brien, *Introduction to Information Systems*. McGrawHill, 2001.

[5] J. E. A. Efraim Turban, Long Beach, *Decision Support Systems And Intelligent Systems*. Prentice Hall, 5 ed., 1999.

[6] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information Sciences*, Vol. 8, 1975, pp. 199–249.

[7] H. J. Zimmermann, *Fuzzy Set Theory and its Applications*. Kluwer Academic Publishers, 1996.

[8] J. Gomez and D. Dasgupta, Evolving fuzzy classifiers for intrusion detection, in *Proceedings of the 2002 IEEE Workshop on Information Assurance*, United States Military Academy, West Point, NY, 2001.

[9] J. H. Holland, Genetic algorithms, *Scientific American*, 1992.

[10] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[11] D. E. Goldberg and M. Rudnick, Genetic algorithms and the variance of fitness, *IlliGAL Report*, No. 91001, 1991.

[12] K. E. Kinnear, *Advances in Genetic Programming*. MIT Press, 1994.

[13] L. T. Koczy, Fuzzy if ... then rule models and their transformations into one another, *IEEE Transaction on Systems, Man and Cybernetics*, Vol. 26, No. 5, 1996.