AN INFORMATION SECURITY RISK ASSESSMENT MODEL BASED ON BAYESIAN NETWORK AND FUZZY INFERENCE SYSTEM[*τ]

BAYES AĞI VE BULANIK ÇIKARIM SİSTEMİ TABANLI BİR BİLGİ GÜVENLİĞİ RİSK DEĞERLENDİRME MODELİ

Sevilay BEKEN[i]
Mete EMİNAĞAOĞLU[ii]

Abstract

This study proposes a novel information security risk assessment approach based on Bayesian network and Fuzzy Inference System in order to evaluate and calculate both qualitative and / or quantitative risks. The proposed model is developed to analyse test processes for a software services company in order to evaluate the information security risks. Threats, vulnerabilities, risks, and their relations are constructed with a Bayesian network and marginal probabilities are calculated for each risk factor. Several fuzzy membership functions and fuzzy decision rules are designed and constructed for assets' values, risks' probabilities, and relative risk values. Finally, the impacts of risk values are calculated after the aggregation and defuzzification process. It is shown that this new model enables the business decision makers and managers to obtain more objective, reliable, and flexible information security risk assessment results.

Keywords: *Information Security Management, Risk Assessment, Fuzzy Inference System, Bayesian Networks*

Öz

Bu çalışmada, nitel ve / veya nicel riskleri değerlendirmek ve hesaplamak için Bayes ağı ve bulanık çıkarım sistemine dayanan yeni bir bilgi güvenliği risk değerlendirme yaklaşımı ortaya konmuştur. Önerilen model, bir yazılım şirketinin için test süreçlerini analiz etmek üzere geliştirilmiştir. Tehditler, güvenlik açıkları, riskler ve bunların bağlantılarının tanımlandığı bir Bayes ağı tasarlanmış ve her bir risk faktörü için bileşen olasılıkları hesaplanmıştır. Bilgi varlıklarının değerleri, riskleri, olasılıkları ve göreceli risk değerleri için bulanık üyelik fonksiyonları ve bulanık karar kuralları tasarlanmış ve oluşturulmuştur. Son aşamada da, risk değerlerinin etkileri, bulanık toparlama ve durulaştırma işlemleri ile hesaplanmış ve sıralanmıştır. Bu yeni model, kurumlardaki yöneticilerin daha objektif, güvenilir ve esnek bir şekilde bilgi güvenliği risk değerlendirme sonuçları elde etmelerine ve kullanmalarına olanak sağlamaktadır.

Anahtar Kelimeler: *Bilgi Güvenliği Yönetimi, Risk Değerlendirmesi, Bulanık Çıkarım Sistemi, Bayes Ağları*

[*] Makale Gönderim Tarihi:03.01.2019 Makale Kabul Tarihi: 18.01.2019

[τ] A simpler and different version of this study with a narrowed scope was presented at International Conference on Advanced Technologies, Computer Engineering and Science, Safranbolu, May 11-13, 2018.

[i] MSc Student, Dokuz Eylül University, Faculty of Science, Department of Computer Science, ORCID ID:0000-0002-8007-0302.

[ii] Assist. Prof. Dr., Dokuz Eylül University, Faculty of Science, Department of Computer Science, mete.eminagaoglu@deu.edu.tr, ORCID ID:0000-0003-2456-919X.

1. INTRODUCTION

Information and knowledge have become the most essential asset for all companies and because of this reason; information security has become an important concern in companies based on their size and complexity (Vercellis, 2009; TBD 4. Çalışma Grubu, 2006). Information security can be defined as the protection of data or information to prevent loss, unauthorized access, or misuse. Companies should ensure that systems and applications operate effectively while protecting their information assets with an acceptable level of risk that could be derived from any kind of theft or loss, misuse, unauthorized access, or modification (Pfleeger, 2007).

The fundamental objectives of information security are confidentiality, integrity, and availability (Pfleeger, 2007). Each objective focuses on different part of protection for information. National Information Assurance (IA) Glossary defines confidentiality as "the property that information is not disclosed to users, processes or devices unless they have been authorized to access the information" (Committee on National Security Systems, 2010). Confidentiality is "the property that data or information is not made available or disclosed to unauthorized people or processes that aims the protection of information against unauthorized access, uses, and disclosures". Integrity can be simply described as the property that information must not be altered or destroyed by unauthorized processes, people, or events. "Integrity factor indicates that information must protect against improper destruction or alteration of data and must provide appropriate backup in the event of a threat, hazard, or natural disaster" (Committee on National Security Systems, 2010). According to IA Glossary, availability can be defined as "the property that data or information is accessible and usable upon demand by an authorized entity". This factor indicates that authorized personal must be able to access to the information. In addition, disaster recovery and business continuity plans for business, governmental, educational, etc. operations should be identified and planned in order to keep the organization operational.

In order to sustain confidentiality, integrity, and availability objectives of information security, risk assessment methodologies should be applied for the organization's operations. The information security risk assessment can be defined as a process of determining the security risks, resolving security problems, and eliminating these risk factors to an acceptable level (Layton, 2007). Information assets, vulnerabilities, threats, and risk factors should be identified, analysed, and controlled within the scope of security risk assessment process according to ISO 27005 standard (ISO / IEC 27005, 2011). The assets are the main objects for organizations that need to be protected based on information security policies (Dhillon, 2007). Assets can be valuable information or resources such as computers, employees, internet connection and so on. The threat is defined as "the potential causes of accidents that may cause harm to systems or organizations", and vulnerability is described as "the weak link of an asset that may be exposed by the threat" (ISO / IEC 27001, 2013). Determining assets, threats, vulnerabilities, risk values, and likelihood is critical for information strategy (Layton, 2007). Information security risk assessment can provide the managers the strategic information and decisions they need to mitigate or control the information risks (Tipton and Krause, 2007).

This study focuses on constructing an accurate and effective information security risk assessment model. Hence, in this study, in order to evaluate and calculate both qualitative and quantitative risks, an information security risk assessment approach is proposed based on Bayesian network and fuzzy inference system. Information security risk factors can be

modelled in Bayesian network and risk probabilities can be calculated more accurately for business requirements with Bayesian network. Hence, Bayesian network model is selected for this study. Additionally, to obtain more reliable and less subjective approach to the risk assessment process and to combine quantitative and/or qualitative risk factors, fuzzy inference system is used.

The proposed model applied for a software services company according to their software test process. In our model, firstly the Bayesian Model was developed to analyse company's database security during the testing process. The assets, vulnerabilities, threats, risk factors and their relations were analysed with the experts and managers in the company. All risk and vulnerability factors were evaluated according to main three objectives of information security, which are confidentiality, integrity, and availability. After defining and analysing the assets, vulnerabilities, threats, risk factors and their relations in the system, Experts' opinions were collected in order to sustain needful data for assets, threats, vulnerabilities, and risk probability values in the Bayesian model. After establishing the Bayesian network model, fuzzy inference system was developed for risk management process. The risk factors were evaluated within the information security risk assessment scope based on the results for information risk factors.

It should be noted that a simpler and different version of this study with a narrowed scope was presented at an international conference (Beken and Eminağaoğlu, 2018). However, this previous model was a primitive architecture and it was used as a prototype. The scope of the risk assessment scope was smaller, there were less risks and fuzzy decision rules, all of the fuzzy operations and were done manually, MATLAB was not used, risk values versus different assets were not analysed, maximum value for each of the risks were not calculated, and the final risk values and rankings were entirely different.

The rest of this paper is organized as follows; information security risk assessment methods for are elaborated in the section. The proposed model based on Bayesian networks and fuzzy inference system explained and results are presented for selected company in the third section. The last section is the conclusion for evaluating the results according to final risk evaluation and prioritization.

2. LITERATURE REVIEW

M. C. Lee (2014) stated that "information security risk assessment process is the important prerequisite to achieve scientific and effective risk assessment". According to him, "information security risk assessment process includes following stages; preparation of risk assessment, asset identification, threat identification, vulnerability identification, and risk calculation" (Lee, 2014). Risk assessment process can be divided into six steps as follows (Fu and Xiao, 2012).

1) Determining assessment objects.
2) Deciding upon the appropriate assessment methodology and necessary tools.
3) Risk identification for the critical information assets, their vulnerabilities and related risk factors.
4) Risk analysis.
5) Risk assessment.
6) Risk control with different strategies and plans that are most suitable and feasible.

2.1. Information Security Risk Assessment Methods

Risk assessment and management are the most crucial parts of Information Security Management Systems. Therefore, "lots of risk analysis and evaluation research has been conducted and there are many publications in literature" (Takçı et al., 2010). Information security risk assessment methodologies can be categorized in three different groups as "qualitative", "quantitative" and combined methodologies (Tipton and Krause, 2007). Several quantitative and qualitative risk analysis and assessment methods may be included in the relevant standards such as ISO 27005 (ISO, 2011), ISO 27001 (ISO, 2013), and SP 800-30 rev. 1 (NIST, 2012).

Qualitative security risk assessment methods use qualitative estimates while quantitative assessment methods use numerical estimates. Risk is analysed with subjective evaluations/opinions and representative scalar values in qualitative risk analysis. Qualitative risk assessment methods assign levels for risk factors such as high, medium and low (Dhillon, 2007). These methods try to evaluate risk factors by personal and professional experience (Denys, 2006). There are different approaches and models for the implementation of qualitative methods such as, Dempster–Shafer theory as a general framework for reasoning with uncertainty (Sun et al., 2006), or "Modified Design Structure Matrix (MFDSM)" for the evaluation of information security risks (Lv et al., 2006).

On the other hand, quantitative risk analysis methods use mathematical and statistical tools to evaluate, measure, and calculate the risks. The value of each risk should be determined by quantifying risk factors with real and continuous variables such as monetary loss (Landoll, 2006; Tipton and Krause, 2007).

Some methods of quantitative security risk analysis can be described as "risk value, annual loss expectancy, safeguard value, and return of investment" (Fu and Xiao, 2012). One of the most well-known quantitative method in risk assessment is ALE (Annual Loss Expectancy) model. "ALE is based on the idea of expected loss, which is the product of probability of occurrence of events which have negative impact on IT and values of caused by them losses" (Fu and Xiao, 2012). Some common formulas to estimate the risk and other parameters in quantitative risk analysis can be shortly described as follows (Yuhan, et al., 2013):

- ✓ *Exposure Factor (EF)*: percentage of loss for an asset due to a specific risk.
- ✓ *Asset Value*: Monetary value of the information asset.
- ✓ *Single Loss Expectancy (SLE):* It is the monetary loss derived from the single occurrence of a risk.
- ✓ *Annualized Rate of Occurrence (ARO):* The estimated frequency for an event / risk that is expected to occur within one year.
- ✓ *Annualized Loss Expectancy (ALE):* The annual expected monetary loss due to the occurrence of a specific event.

The ALE approach is based on calculating the loss expectation and multiplying the possibility of loss for each attack over a period, which is given in the following equations.

ALE = (Probability of event) x (Value of Loss) (1)

$$\mathrm{ALE} = \sum_{i=1}^{n} I(O_i)F_i \quad (2)$$

where: $\{O_1, O_2, O_3 \ldots O_n\}$ is set of negative effects of event, $I(O_i)$ is value expressed loss resulting from event, F_i is frequency of event i.

Karabacak and Soğukpınar (2005) have proposed a new information security risk assessment method called "ISRAM", which is "based on a quantitative approach that uses survey results to analyse information security risks". "ISRAM" is performed by using opinions obtained by a survey. "ISRAM is a survey preparation and conduction process to assess the security risk in an organization" (Karabacak and Soğukpınar, 2005). On the other hand, "risk analysis methods based on qualitative measures, are more suitable for today's complex risk environment of information systems" (Karabacak and Soğukpınar, 2005).

Some of the qualitative methods are fuzzy comprehensive evaluation method, The Microsoft Corporate Security Group Risk Management Framework, and CRAMM. CRAMM (CCTA Risk Analysis and Management Method) is a qualitative risk analysis tool developed by UK government's Central Computer and Telecommunications Agency in 1985 (Enterprise, 2005). CRAMM methodology consists of three stages. These stages are defined as the identification and evaluation of assets, the identification of threats and vulnerabilities, and finally, the calculation of risk factors.

Moreover, there are some other methods in the literature that use combined qualitative and quantitative methodologies. For instance, Analytic Hierarchy Process (AHP) method is a popular approach for in security risk assessment that use qualitative methodology combined with quantitative approach. It is mentioned that "AHP can change from the qualitative index into quantitative index" (Award et al., 2011). It has been stated that there are usually four stages in AHP. The construction of the hierarchical structure, the construction of judgement matrix, calculation of the relative weight of the factors, and finally, the calculation of the entire weight of the factors (Omar and Herrera, 2002).

2.2. Bayesian Networks

Bayes' theorem is based on the formula that calculates the probabilities of hypotheses when prior evidence is given or known. It can be used to validate or combined with conditional probability formulas to solve a wide range of problems (Denys, 2006). The equation of Bayes' theorem is given in (3). In this equation, A and B denote the events; *P (A | B)* denotes the conditional probability that calculates the likelihood of A given that B is true. Similarly, *P (B | A)* represents the probability of B when A is known to have occurred.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \tag{3}$$

A Bayesian network can be defined as a graphical model that establishes the joint probability distribution for a set of random variables. It is a way of describing the relationships between causes and effects, and contains nodes and arcs. Bayesian networks are widely used in data mining, machine learning, and artificial intelligence for representing and reasoning about problems in which probability plays a role. A Bayesian network is a directed, acyclic graph whose nodes represent random variables and arcs represent direct dependencies (Chin et al., 2009). Each connected node's conditional probability and the joint probabilities are calculated within the network. For instance, if event A is directly connected to B, and event B is directly connected to C, then the probability of observing three events' occurrence can be calculated by the joint probability *P(A,B,C)* as follows:

$$P\ (A, B, C) = P(C|A, B)P(B|A)P(A) \tag{4}$$

It should be noted that *P(A)* is the probability of event A, *P(B|A) is the* conditional probability that calculates the likelihood of B given that A is observed, and *P(C|A,B)* is the conditional probability that calculates the likelihood of observing C given that both A and B is true or has occurred.

Bayesian Networks are also used for network security solutions. Frigault et al. (2008) have developed a model by using dynamic Bayesian network for measuring network security. They proposed a Dynamic Bayesian Network based model to incorporate relevant temporal factors, such as the availability of exploit codes or patches, into attack graph-based security metrics (Frigault et al., 2008). Additionally, Bayesian network model has also been used for "Earthquake Risk Management and Cyber Security Risk Assessment" (Bayraktarlı et al., 2005). They proposed that by using Bayesian Network models, the uncertainties associated with all elements of an earthquake from the source mechanism, site effects, structural response, damage assessments and consequence assessment (Bayraktarlı et al., 2005). Hence, Bayesian network models are useful for modelling today's complex and diverse information systems, some researchers proposed models that use Bayesian networks or dynamic Bayesian networks for information security risk assessment.

Wang et al. (2016) have used dynamic Bayesian network method for information security risk assessment. In their study, they created a "dynamic Bayesian network model based on the risk assessment process" (Wang et al., 2016). After this process, information system was analysed and the probability of the risk was calculated. Finally, they made an experiment "to analyse and compare the dynamic Bayesian network model with the static Bayesian network model" (Wang et al., 2016).

To the best of our knowledge, no similar model could be found in the literature that combines fuzzy inference systems with Bayesian network models about information security risk assessment methodologies, our model is different from other information security risk assessment models. In our study, GeNIe software is used for design and implementation of our Bayesian Network model. GeNIe is a tool that can be used to design and implement several Bayesian network models that calculates the conditional probabilities and marginal probabilities. The marginal probability can be simply described as the probability or likelihood of occurrence of any single event that is not conditioned on any other events.(Foroughi, 2008). They are defined as marginal since "they are calculated by summing values in a table along rows or columns and they are written the sum in the margins of the table" (Zhao et al., 2005). In other words, a marginal probability can be calculated by adding up the joint probabilities and this is named as "marginalization".

2.3. Fuzzy Inference Systems

Fuzzy inference systems can also be used for information security risk assessment. This approach is a qualitative method and it has several advantages (Mc Neill and Thro, 1994). One of its basic advantage is more variables can be assessed with less rules and decisions. In addition, linguistic / non-numerical values can be used. It establishes rigid / reliable relations between input and output data that enables people to design more accurate / meaningful / realistic systems (Mc Neill and Thro, 1994).

"Fuzzy Comprehensive Evaluation Method" is one of the common qualitative methodologies, which is based on the principle of fuzzy mathematics. Fuzzy comprehensive evaluation method

uses the "fuzzy statistical methods through considering a combination of relative factors for evaluation to determine the weight of various factors to make the evaluation of the pros and cons of the research objects" (Yuhan, et al., 2013). In addition, fuzzy methods and approaches can also be combined with AHP for different objectives in decision making and multi-criteria problems (Çiçekli and Karaçizmeli, 2013).

In order to evaluate information security risk assessment, Zhao et al. (2005) have proposed a model based on fuzzy logic and entropy theory. Long, Yong, and Qianmu (2008) used an approach that combination with AHP and fuzzy comprehensive method in their study. M. Lee (2014) reviewed "AHP model, neural networks, fuzzy logic, group decision making, software computing and hybrid models in information security risk problem". He stated that "the application areas include information security risk analysis, information security risk assessment, and information security management" (Lee, 2014). In his research, he also stated that "in order to improve performance of AHP method, fuzzy comprehensive evaluation method can be used for reduction the feature subset" (Lee, 2014). In addition, most of the researchers seem to prefer the AHP method (Altuzarra, Moreno-Jimnez, & Salvador, 2007); (Award et al, 2011).

3. PROPOSED MODEL

A new model is designed and developed for a software services company to assess the relevant information security risks during their critical business processes, which are based on the software testing procedures and systems. A Bayesian network and a fuzzy inference system are used together in the novel model that is proposed in this study. The company's clients are mostly working in the telecommunication sector. The company has seventy white collar employees and sixty-four of them are working in software testing projects. The testers are working in an open office which has two separate meeting rooms and a computer room with a server.

Firstly, the Bayesian model was designed based on the testing process in the selected company. Secondly, fuzzy membership functions were constructed for both assets and risk factors. After these steps, risk factors were evaluated within the information security risk assessment scope based on the results for information risk factors. Details about implemented approach were explained in the Design and Implementation section.

3.1. Materials

In order to collect expert opinions about probabilities in the Bayesian model, questions about threats, risks, and assets were asked to the experts in the company. Fifty-one experts who knows the system very well and are responsible for company's software test process contributed for analysing the system. The questions were arranged within four main categories, which are given below, and the answers / opinions were collected from the experts.

- Assets – what do we have?
- Vulnerabilities – is the asset at risk?
- Threats – what / who will attack / damage / destroy it?
- Risk factors – what are the main risks?

According to the scope of our risk assessment, several different assets, vulnerabilities, threats, and risks were identified and used as below.

1. Seven assets: “Data stored in Computers/Laptops”, “VPN Connection”, “Customer data stored in database”, “Test data”, “IT Users”, “Test Users”, and “Test Support Users”.
2. Twelve vulnerabilities: “Computers Vulnerable to Malfunctions”, “VPN Connection Vulnerable to Malfunctions”, “Vulnerable to Physical Problems/Damages”, “Vulnerable to Malicious Codes”, “Might Be Easily Lost/Stolen”, “Lack of Experience or Training”, “Lack of Awareness”, “Computers Vulnerable to Unauthorized Access”, “VPN Connection Vulnerable to Unauthorized Access”, “Disgruntled IT Users”, “Disgruntled Test Users”, and “Disgruntled Test Support Users”.
3. Eleven threats: “Technical Problems”, “Malicious Code”, “Blackout/Brownout”, “Fire”, “Earthquake”, “Thieves”, “Human Error/ Failure by IT Users”, “Human Error/Failure by Test Users”, “Human Error/ Failure by Test Support Users”, “Hackers/Cyber Criminals”, and “Resignation and Expel with Client Data”.
4. Eleven risks: “Unavailability of Computers”, “Network Connection Loss”, “Loss of Test Data”, “Database Crash”, “Disclosure of Confidential Information”, “Unauthorized Change or Damage of Data”, “Discontinuation of Testing Processes”, “Inaccurate Test Results”, “Prestige Loss”, “Labour Loss”, and “Penalty or Legal Issues”.

Twenty-three questions were prepared to define the possibilities of threats in the system. Three different levels (Low, Medium, and High) of scale were used in this phase. In order to analyse risk values, thirty-four questions were asked to the experts and the answers were evaluated by using a scale with five levels (Very Low, Low, Medium, High, and Very High). The final phase contains the questions for asset values in the system. Experts were asked to evaluate seven different information assets based on either their approximate monetary values or a scale of ten points (one, two … ten) where “one” represents the lowest value, “ten” represents the highest value. Collected data was analysed for calculating the probability of related threat, risk, and asset values. For each question and its corresponding criticality, the average scores were calculated based on the answers of experts. All of these average values were used in the Bayesian model to calculate the conditional probabilities of the related threat and risk.

3.2. Design and Implementation

After collection of data, threats and risk factors are evaluated one by one. Conditional and marginal probabilities are calculated in the GeNIe software (GeNIe Modeler, 2018) according to marginal probability distribution. The marginal probabilities for the threat “Human Error / Failure by Test Users” are denoted in Figure 1 and the marginal probabilities for the risk “Discontinuation of Testing Processes” are shown in Figure 2.

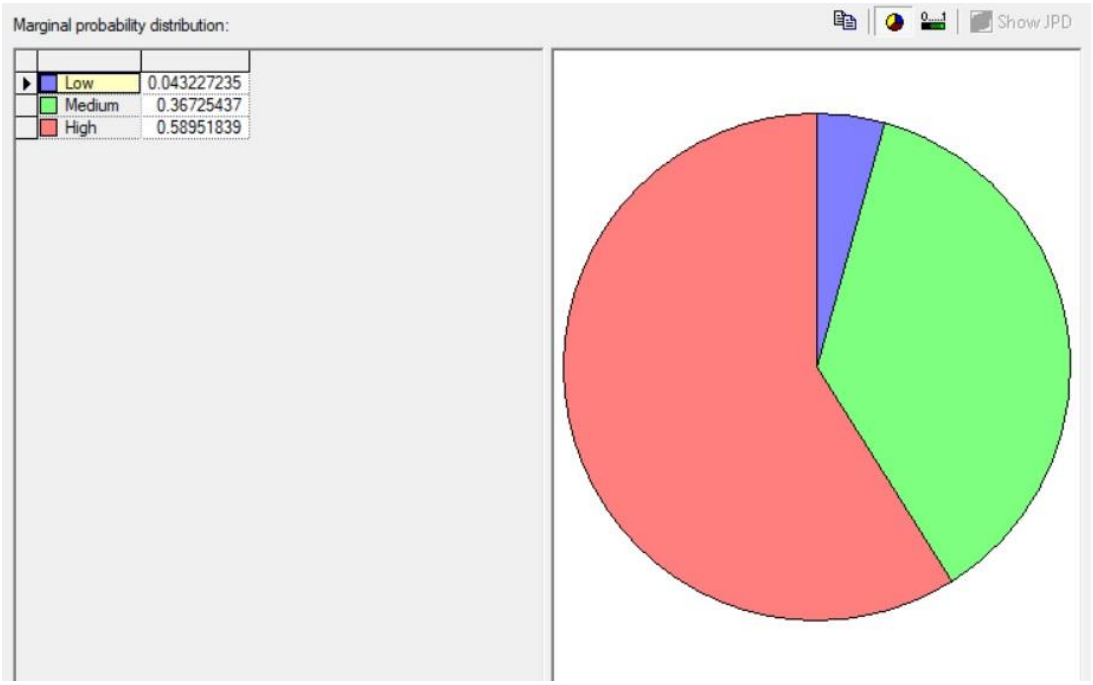

Figure 1: Marginal Probabilities for The Threat "Human Error / Failure by Test Users"

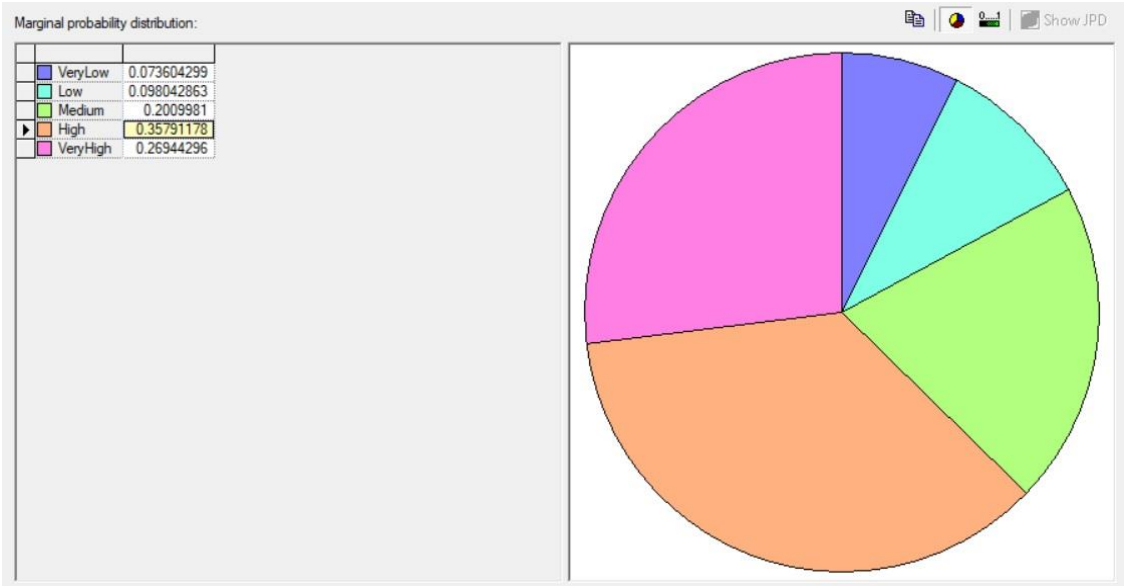

Figure 2: Marginal Probabilities for The Risk "Discontinuation of Testing Processes"

After defining all assets, vulnerabilities, threats, risk factors and their relations, Bayesian network model is constructed to calculate and retrieve the probabilities of each of information security risks in our model. A small excerpt from the Bayesian network used in this study is denoted in Figure 3. The entire Bayesian network is shown in Figure 4.

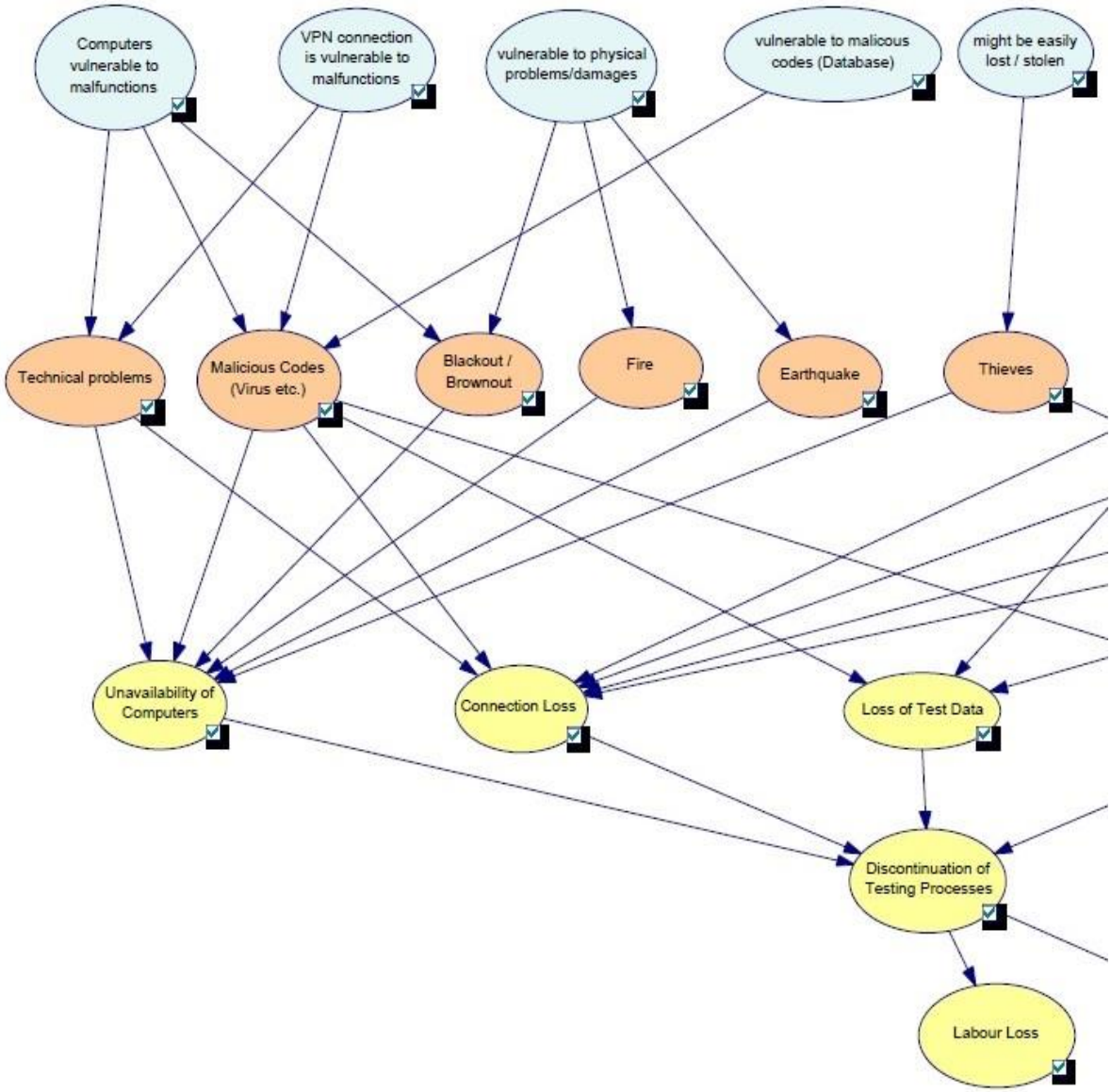

Figure 3: A Partial View of the Bayesian Network Model

Because of the uncertainty of the risk factors, the fuzzy logic method, and a fuzzy inference system is used in this study. First, membership functions are determined for both assets and risk values. Hence, it could be deduced that membership function is a curve showing a point mapping points of inputting data into membership values, whose interval is between zero and one (Ariyanti et al., 2010).

Mamdani Fuzzy Inference System (FIS) was used for fuzzification, defuzzification, and aggregation operations in our risk assessment model. For each fuzzy rule, the minimum or maximum of the fuzzy membership values in the antecedent part is taken according to ‘‘and’’ or ‘‘or’’ operators in the antecedent part of that fuzzy rule.

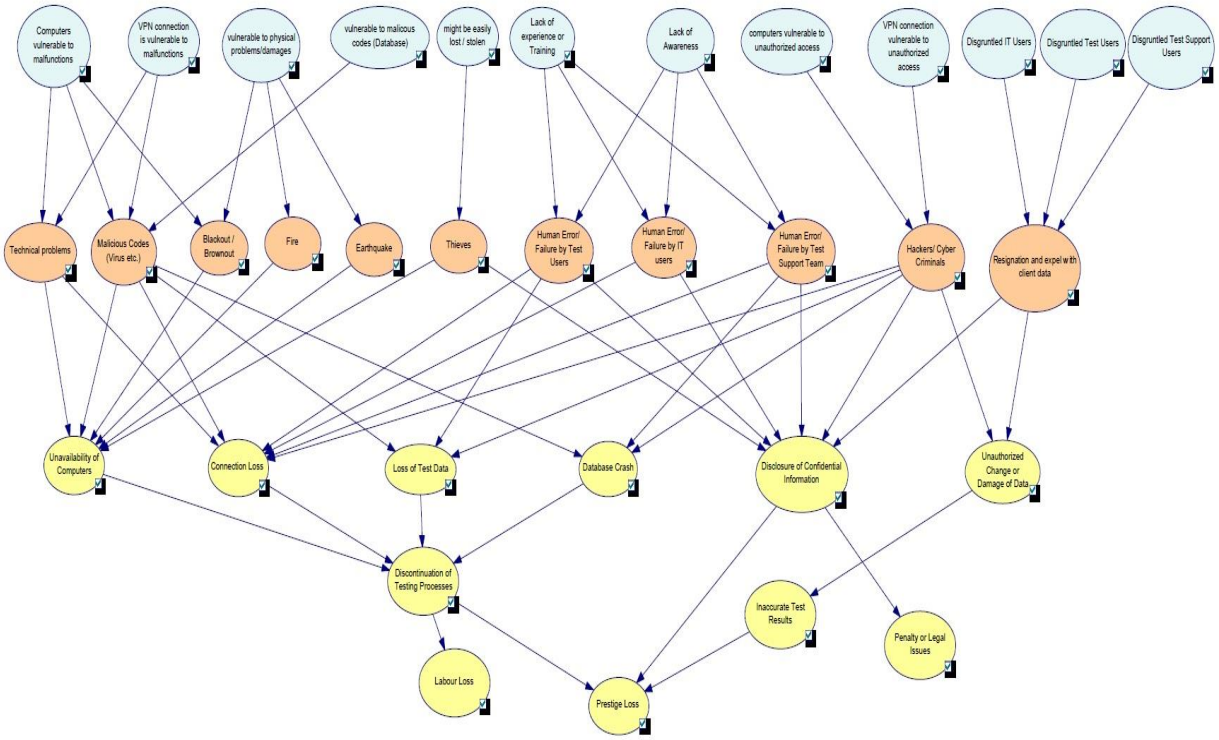

Figure 4: Bayesian Network Model Designed and Used in this Study

The membership value from operations on the predecessors / antecedents "truncates the membership function for the consequent or resultant part of that rule". This truncation or scaling is established for each rule, and then "the truncated membership functions from each rule are aggregated" (Ross, 2004). In order to create fuzzy membership function for assets and risk values, Trapezoidal membership function was selected and Gaussian membership function was used for risk probabilities. It should be noted that, "the final risk / impact value" in our model represents the overall risk or impact that is produced as the outcome of the relevant asset value and that risk's probability value. Hence, the outcome of a fuzzy rule in our model gives the fuzzy risk value using Mamdani FIS. Using these membership functions and adjusting the parameters, the fuzzy membership values for crisp asset values, risk values, and risk probabilities are calculated. All of the fuzzy rules, membership functions, fuzzy operators and fuzzy calculations were developed, implemented and calculated by using MATLAB software (MATLAB, The MathWorks, Inc., 2018).

The formula for Trapezoidal membership function that is used for calculating asset values and risk values is shown as in (5).

$$\mu_F(x,a,b,c,d) = \begin{cases} 0, & if\ x < a \\ \frac{x-a}{x-b}, & if\ a \le x \le b \\ 1, & if\ b < x < c \\ \frac{d-x}{d-c}, & if\ c \le x \le d \\ 0, & if\ d < x \end{cases} \quad (5)$$

We also used another fuzzy membership function, Gaussian function namely, to derive the fuzzy values for marginal risk probabilities that are continuous values between 0 and 1. These

probabilities were previously obtained from our Bayesian network. Gaussian membership function is shown in equation (6).

$$\mu(x, a, b) = e^{\frac{-(x-b)^2}{2a^2}} \quad (6)$$

The membership functions for asset values and risk probabilities are also denoted in Figure 5 and Figure 6, respectively.

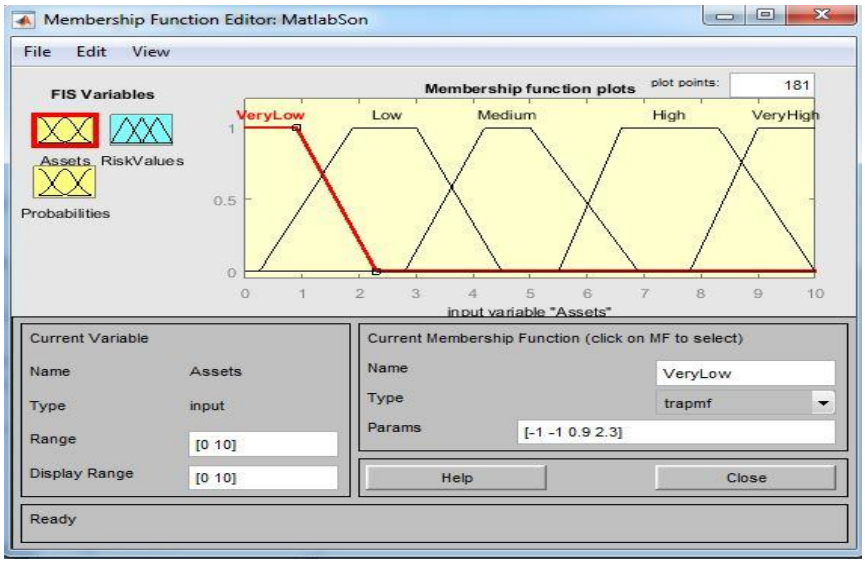

Figure 5: Trapezoidal Membership Function for Fuzzification of Asset Values

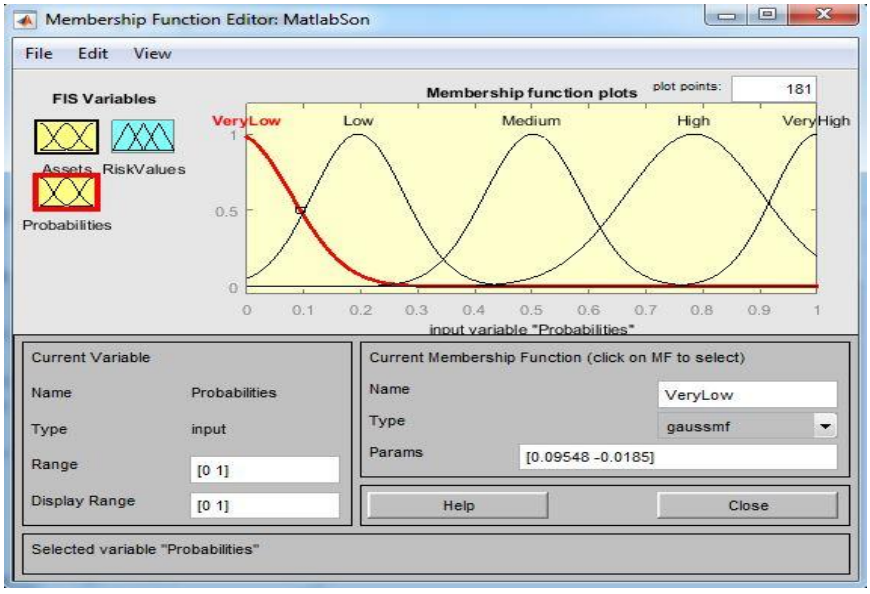

Figure 6: Gaussian Membership Function for Fuzzification of Risk Probabilities

After defining fuzzy membership functions, 50 fuzzy rules were constructed for the fuzzy inference process. For instance, suppose that a specific asset's crisp value is given as 4, and the risk probability is calculated as 0.5 for a specific risk related with that asset. If we use the fourth fuzzy rule in our model *("If Asset is Low and Probability is Low, then Risk is Low"*), then according to Mamdani FIS, the antecedent part of this rule will give a fuzzy value as 0.411. This is achieved by finding the minimum of the fuzzy membership values (0.411 and 0.862) of

asset and risk probability. All of the fuzzy decision rules were coded and executed in MATLAB and some of these rules are also denoted in Figure 7.

Aggregation was processed based on the fuzzy rules for each related risk. For calculating the crisp risk values during the defuzzification process, Center of Gravity method has been used. "If the output fuzzy set has at least two convex sub-regions, then the center of gravity (i.e., z^* is calculated using the centroid method) of the convex fuzzy sub-region with the largest area is used to obtain the defuzzified value z^* of the output" (Ross, 2004). This is given algebraically in equation (7).

$$z^* = \frac{\int \mu C_m(z) z dz}{\int \mu C_m(z) dz} \tag{7}$$

However, it should be noted that Eq. 5 is used when the membership values are continuous. If the values are discrete, the center of gravity method can be simply calculated as follows:

$$z^* = \frac{\sum_{i=1}^{n} x_i \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)} \tag{8}$$

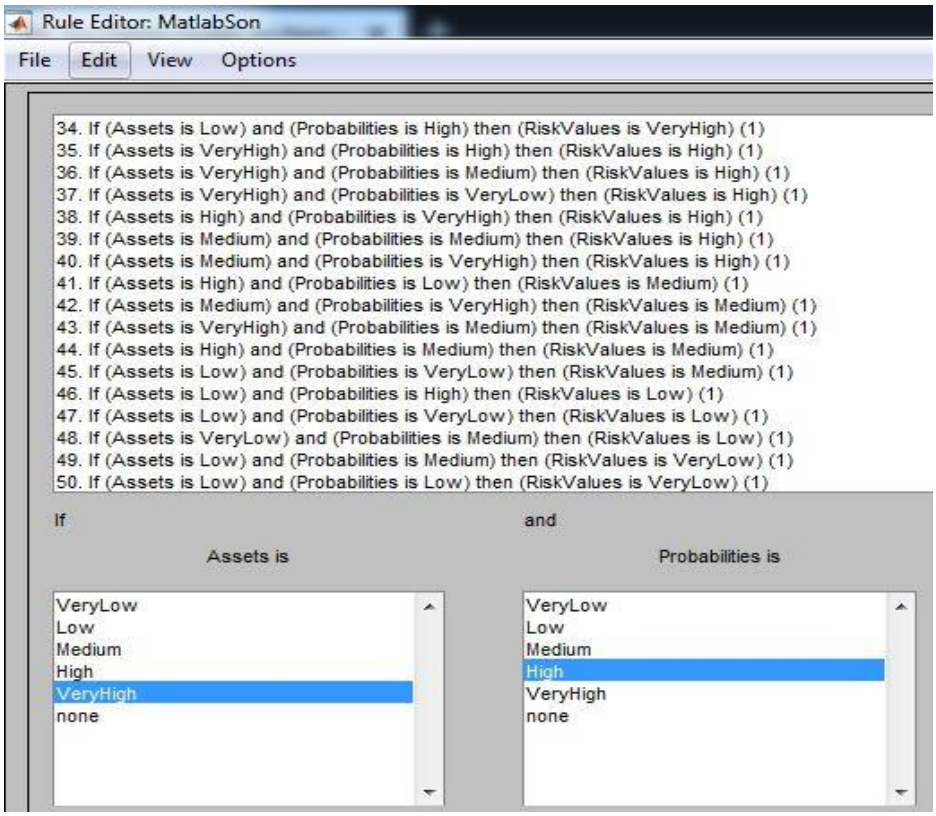

Figure 7: Some of the Fuzzy Decision Rules used in our Model

The defuzzification methodology can also be explained with a simple example. For instance, we can assume that a specific asset's crisp value is given as 5, and the risk probability is calculated as 0.73. Using the fuzzy membership tables in this study, and are aggregating all of the fuzzy rules mentioned in the previous page, then this risk's defuzzified value (by using the center of gravity method) can be calculated as follows:

$$R_d = \frac{(1x0.052) + (2x0.11) + (3x0.11) + (4x0.517) + \cdots + (8x0.617) + (9x0.167) + (10x0.167)}{(0.052 + 0.11 + 0.11 + 0.517 + 0.517 + 0.517 + 0.617 + 0.617 + 0.167 + 0.167)} = 6.064$$

The abstraction of the entire fuzzy inference system that was implemented in MATLAB is also shown in Figure 8.

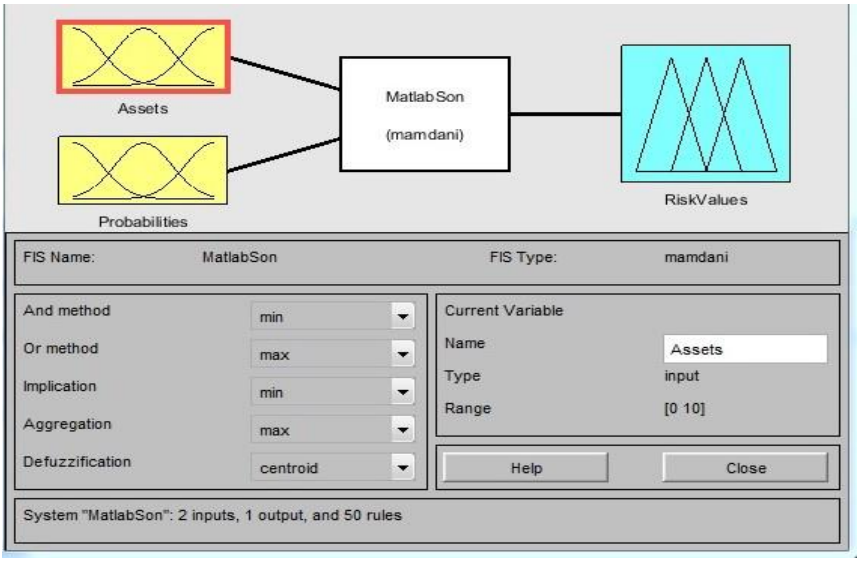

Figure 8: Some of the Fuzzy Decision Rules used in our Model

There were seven different assets within the scope of our information security risk assessment model and each of these were exposed to one or many risks. The relationship between assets and risks are denoted in Table 1 and if there exists a risk for a specific asset, then it is denoted with a checkmark sign.

Table 1: Relationships between Assets and Risks

	Risks										
Assets	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**	**R7**	**R8**	**R9**	**R10**	**R11**
A1	✔					✔		✔	✔		✔
A2	✔		✔		✔	✔	✔		✔	✔	✔
A3				✔	✔				✔		✔
A4		✔	✔	✔		✔		✔	✔		
A5	✔	✔				✔			✔		
A6	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
A7	✔	✔	✔	✔	✔	✔	✔		✔	✔	✔

3.3. Results

We implemented our model by using the methodologies, and making the calculations that were described in the previous section. The asset values were obtained by averaging the expert opinions' evaluation scores, which are given in Table 2.

Table 2: Asset values evaluated by experts' evaluations

Asset name	Asset value
Data stored in computers/laptops	6.90
IT users (experience / knowledge)	7.74
Customer data stored in Database	7.55
Test data	7.21
VPN connection	8.40
Test users (experience / knowledge)	8.20
Test support users (experience / knowledge)	8.01

These asset values were given as crisp inputs to the fuzzy inference system. The marginal probabilities of risks were calculated from the Bayesian network and they are given in Table 3. These probabilities were also fed into the fuzzy inference system as the crisp input values. The fuzzy membership values of assets and risk probabilities were obtained by the membership functions. For each asset, relevant risk or risks were grouped, and then for each of these groups, all the fuzzy rules in the rule base were applied. Finally, aggregation and defuzzification operations were executed and risk / impact values were obtained. There were seven different assets that were exposed to one or many risks, which was mentioned in the previous section. Hence, there is a many-to-many relationship between the assets and risks which produces different defuzzified risk values. These different risk values corresponding to different assets are given in Table 4.

It is known that the final stage of an information security risk assessment should be the final evaluation of each risk with its unique value or score. Thus, we have to use a mechanism to find a single value for each of the eleven risks. For each of these eleven risks, we chose to use the highest risk value among its related assets' corresponding values. For instance, as it could be seen from Table 4, there exists four different defuzzified risk values for the R2 risk among the assets A4, A5, A6, and A7, namely. The highest R2 value belongs to asset A5, hence, the final R2 value is 6.59. The final risk values were calculated by using this maximum operator and the risks were ranked according to this mechanism, which is given in Table 5.

Table 3: Risks' Marginal Probability Values obtained from Bayesian Network

Risk name	Risk level	Marginal probability	Normalized marginal probability
Unavailability of computers	VL	0.138	0
	L	0.195	0.452
	M	0.238	0.786
	H	0.265	1
	VH	0.164	0.207

Table 3 (cont.d) : Risks' Marginal Probability Values obtained from Bayesian Network

Risk name	Risk level	Marginal probability	Normalized marginal probability
Network Connection Loss	VL	0.134	0.227
	L	0.228	0.701
	M	0.261	0.866
	H	0.288	1
	VH	0.088	0
Loss of Test Data	VL	0.162	0.141
	L	0.213	0.629
	M	0.226	0.763
	H	0.251	1
	VH	0.148	0
Database Crash	VL	0.144	0.351
	L	0.267	0.97
	M	0.272	1
	H	0.244	0.856
	VH	0.074	0
Disclosure of Confidential Information	VL	0.192	0.505
	L	0.235	0.859
	M	0.189	0.479
	H	0.252	1
	VH	0.132	0
Unauthorized Change or Damage in Data	VL	0.169	0.175
	L	0.180	0.224
	M	0.349	1
	H	0.171	0.182
	VH	0.131	0
Discontinuation of Testing Processes	VL	0.074	0
	L	0.098	0.086
	M	0.201	0.448
	H	0.358	1
	VH	0.269	0.689
Inaccurate Test Results	VL	0.053	0
	L	0.145	0.344
	M	0.226	0.644
	H	0.321	1
	VH	0.254	0.751
Prestige Loss	VL	0.026	0
	L	0.085	0.167
	M	0.137	0.315
	H	0.378	1
	VH	0.375	0.99
Penalty or Legal Issues	VL	0.041	0
	L	0.060	0.047
	M	0.114	0.180
	H	0.338	0.729
	VH	0.448	1
Labour Loss	VL	0	0
	L	0.020	0.047
	M	0.301	0.722
	H	0.417	1
	VH	0.262	0.627

Table 4: Defuzzified Risk Values for Different Assets

	Assets						
Risks	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**
R1	6.55	6.55			6.55	6.55	6.55
R2				6.38	6.59	6.51	6.42
R3		6.54		6.54		6.54	6.54
R4			6.42	6.42		6.55	6.46
R5		6.65	6.65			6.65	6.65
R6	6.56	6.56		6.56	6.56	6.56	6.56
R7		6.6				6.6	6.6
R8	6.67			6.67		6.67	
R9	8.07	8.07	8.07		8.07	8.07	8.07
R10		6.63				6.63	6.63
R11	8.09	8.09	8.09			8.09	8.09

It could be seen from Table 5 that the risk with highest score was “Penalty or Legal Issues” and the risk with the lowest value was “Loss of Test Data”. These results were also cross-checked and approved by the experts in the company as well as information security consultants. However, it was also observed that most of the risks’ values were slightly different or they were very close.

Table 5: The Final Risk Values

Risk number	Risk name	Final risk value
R11	Penalty or Legal Issues	8.09
R9	Prestige Loss	8.07
R8	Inaccurate Test Results	6.67
R5	Disclosure of Confidential Information	6.65
R10	Labour Loss	6.63
R7	Discontinuation of Testing Processes	6.60
R2	Network Connection Loss	6.59
R6	Unauthorized Change or Damage in Data	6.56
R4	Database Crash	6.55
R1	Unavailability of computers	6.55
R3	Loss of Test Data	6.54

This might have occurred due to the fact that in this study, asset values and risk probability values were not real continuous values but experts’ rankings between 1 and 10, which provided very similar resultant values. If real monetary values for the assets and real probability values could have been used, the risk values would not have been so close to each other. It should be noted that due to the conditional and marginal probabilities calculated by the Bayesian network might have also affected this closeness issue.

4. CONCLUSION

In this study, a new information security risk assessment model based on Bayesian network and Fuzzy inference system is proposed to evaluate and calculate both qualitative and / or quantitative risks in a more reliable, flexible, and objective manner. This information security risk assessment approach is different from other methodologies in the literature by combining the Bayesian network and fuzzy inference system.

The proposed model is developed to analyse test processes for a software company in order to evaluate the information security risks. First, the system's assets, threats, and vulnerabilities have been thoroughly analysed with the experts in the selected company. Then, information risk factors, threats, vulnerabilities, and their relations have been modelled in Bayesian network. Data is collected for our risk assessment model, with experts and managers based on the testing process in the company. Assets, vulnerabilities, threats, and related risk values are identified and analysed. Vulnerabilities, risks, and their relations are constructed with a Bayesian network and marginal probabilities for each risk are calculated. After Bayesian network is constructed, fuzzy membership functions are designed for assets' values, risks' probabilities, and risk values. In order to obtain more reliable and less subjective approach to the risk assessment process, fuzzy inference system has been used in this new model. Fifty fuzzy decision rules are constructed for some of the chosen risks by using the assets' values, relevant risk probabilities, and relative risk values. Finally, the risk impact values are calculated in the aggregation and defuzzification processes. Based on the final risk values, the risks are evaluated according to the relevant assets, maximum risk values were obtained, and they were finally ranked for information security risk assessment process.

It should be mentioned that the proposed novel model can be flexibly and easily adapted to different companies' and organizations' information security risk assessment processes within their diverse information security management scopes. Anyone can adapt and use this model for their organization by either quantitative values / scores, or qualitative measures, or both.

As a future study, in order to execute Bayesian network model and fuzzy inference system together and establish the integration between them, an automated tool with a user-friendly interface is planned to be developed with an appropriate programming language and a database management system. This tool is planned to be freely distributed to the decision makers and managers in companies and other organizations / institutions and the interface of our tool will be redesigned and improved according to the feedbacks from these users.

REFERENCES

Altuzarra, A., Moreno-Jimnez, J., and Salvador, M. (2007). "A Bayesian prioritization procedure for AHP-group decision making". *European Journal of Operation Research*, 18(1): 367-382.

Ariyanti, R., Kusumadewi, S., and Paputungan, I. (2010). "Beck Depression Inventory Test Assessment Using Fuzzy Inference System", Proccedings of IEEE Intelligent Systems. Modelling and Simulation 2010 International Conference, Liverpool, UK, pp. 6-9.

Award, G., Suitan, E., Ahmad, N., Ithnan, N., and Beg, A. (2011). "Multi-objective model to process security risk assessment based on AHP-PSO". *Modern Applied Science*, 5(3): 246-250.

Barber, D. (2011). *Bayesian Reasoning and Machine Learning.* Cambridge University Press, UK.

Bayraktarlı, Y., Ulfkjaer, J., Yazgan, U., and Faber, M. (2005). "On the Application of Bayesian Probabilistic Networks for Earthquake Risk Management", Proceedings of 9th International Conference on Structural Safety and Reliability (ICOSSAR 05), Rome, Italy.

Çiçekli, U. G. and Karaçizmeli, A. (2013). "Bulanık Analitik Hiyerarşi Süreci ile Başarılı Öğrenci Seçimi: Ege Üniversitesi İktisadi ve İdari Bilimler Fakültesi Örneği". *Ege Stratejik Araştırmalar Dergisi,* 4(1):71-94.

Beken S. and Eminağaoğlu M. (2018). "Information Security Risk Assessment using Bayesian Network and Fuzzy Inference System: A Case Study", ICATCES2018, Proceedings of International Conference on Advanced Technologies, Computer Engineering and Science, May 11-13, 2018, Safranbolu, Turkey.

Chin, K., Tang, D., Yang, J., Wong, S., and Wang, H. (2009). "Assessing New Product Development Project Risk By Bayesian Network With a Systematic Probability Generation Methodology". *Expert Systems with Applications*, 36(6): 9879-9890.

Committee on National Security Systems. (2010). *National Information Assurance (IA) Glossary.* Committee on National Security Systems.

Denys, P. (2006). "Efficiency of Risk Assessment Methods", Proceedings of IEEE Modern Problems of Radio Engineering, Telecommunications and Computer Science, Lviv, Ukraine.

Dhillon, G. (2007). *Principles of Information Systems Security*, John Wiley & Sons Inc., USA.

Foroughi, F. (2008). "Information Security Risk Assessment by Using Bayesian Learning Technique", Proceedings of the World Congress on Engineering, Volume 1, London, UK.

Frigault, M., Wang, L., Singhal, A., and Jajodia, S. (2008). "Measuring Network Security Using Dynamic Bayesian Network", Proceedings of the 4th ACM Workshop on Quality of Protection, Alexandria, USA.

Fu, S. and Xiao, Y. (2012). "Strengthening The Research for Information Security Risk Assessment", International Conference on Biological and Biomedical Science Advanced in Biomedical Engineering. 9: 386-392.

GeNIe Modeler, BayesFusion, LLC, https://www.bayesfusion.com/genie/, Erişim: 20.10.2018.

Insight Consulting, Siemens. (2005). "The Logic Behind CRAMM's Assessment of Measures of Risk and Determination of Appropriate Countermeasures". Technical report.

ISO/IEC 27001. (2013). *Information Security Management Systems.* Information Technology, Security Techniques, Geneve, Switzerland.

ISO/IEC 27005. (2011). *Information Security Risk Management.* Geneva, Switzerland.

Karabacak, B. and Soğukpınar, I. (2005). "ISRAM: Information Security Risk Analysis Method". *Computers & Security*, 24(2): 147-159.

Landoll, D. (2006). *The Security Risk Assessment Handbook.* Auerbach Publications.

Layton, T. P. (2007). *Information Security; Design, Implementation, Measurement and Compliance,* Auerbach Publications, USA.

Lee, M. (2014). "Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method". *International Journal of Computer Science & Information Technology*. 6(1): 29-45.

Lv, J. J., Qiu, W. H., Wang, Y. Z., and Zou, N. (2006). "A study on information security optimization based on MFDSM", Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, IEEE pub., Dalian, China.

Mc Neill, F. M. and Thro, E. (1994). *Fuzzy Logic: A Practical Approach.* Morgan Kaufmann Publications.

MATLAB, The MathWorks, Inc., https://www.mathworks.com/products/matlab.html, Erişim: 22.10.2018.

NIST National Institute of Standards and Technology. (2011). "Guide for Conducting Risk Assessments". Special Publication 800-30 rev.1, USA.

Omar, A. and Herrera, R. (2002). "Graphical Risk Analysis (GRA): A Methodology to Aid In Modeling Systems For Information Security Risk Analysis", pp.1-12.

Pfleeger, C. P. (2007). *Security in Computing, 4th edition*, Prentice Hall, USA.

Pitman J. (2006). *Combinatorial Stochastic Processes.* University of California, Berkeley, USA.

Ross, T. J. (2004). *Fuzzy Logic with Engineering Applications, 2nd edition*. John Wiley & Sons Ltd.

Sun, L., Srivastava, R. P., and Mock, T. J. (2006). "An information systems security risk assessment model under Dempster-Shafer Theory of belief functions", *Journal of Management Information Systems,* 22(4): 109-142.

Takçı, H., Akyüz, T., Uğur, A., Karabağ, R., Soğukpınar and Soğukpınar, İ. (2010). "Bilgi Güvenliği Yönetiminde Risk Değerlendirmesi İçin Bir Model". *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi,* 3(1): 47-52.

TBD 4. Çalışma Grubu (2006). "E-devlet Uygulamalarında Güvenlik ve Güvenilirlik Yaklaşımları". Türkiye Bilişim Derneği Kamu - Bilişim Platformu VIII, Sonuç Raporu.

Tipton, H. F., and Krause, M. (2007). *Information Security Management Handbook,* Auerbach Publications, USA.

Vercellis, C. (2009). *Business Intelligence; Data Mining and Optimization for Decision Making*. John Wiley & Sons, Ltd., UK.

Wang, J., Fan, K., Mo, W., and Xu, D. (2016). "A Method for Information Security Risk Assessment Based on the Dynamic Bayesian Network", IEEE International Conference on Networking and Network Applications, Hakodate, Japan.

Yong, Q., Long, X., and Qianmu, L. (2008). "Information Security Risk Assessment Method Based on CORAS Frame", IEEE International Conference on Computer Science and Software Engineering, Hubei, China, 3: 571-574.

Yuhan, H., Xiaoyan, C., Linqiao, D., Songsong, Z., Min, W., and Yanxiong, H. (2013). "The Reclamation Soil Suitability Study of the Highway Dumping Site Based on Fuzzy Comprehensive Evaluation Method". *Nature Environment and Pollution Technology*, 12(1): 51-56.

Zhao, D. M., Wang, J. H., Wu, J., and Ma, J. F. (2005). "Using fuzzy logic and entropy theory to risk assessment of the information security", Proceedings of IEEE International Conference on Machine Learning and Cybernetics, Guangzhou, China.