A Fuzzy Approach to Risk Analysis in Information Systems

Eloy Vicente, Antonio Jiménez and Alfonso Mateos
Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo S/N, 28660, Boadilla del Monte, Madrid, Spain
e.vicentecestero@upm.es, {ajimenez, amateos}@fi.upm.es

Abstract: Assets are interrelated in risk analysis methodologies for information systems promoted by international standards. This means that an attack on one asset can be propagated through the network and threaten an organization's most valuable assets. It is necessary to valuate all assets, the direct and indirect asset dependencies, as well as the probability of threats and the resulting asset degradation. These methodologies do not, however, consider uncertain valuations and use precise values on different scales, usually percentages. Linguistic terms are used by the experts to represent assets values, dependencies and frequency and asset degradation associated with possible threats. Computations are based on the trapezoidal fuzzy numbers associated with these linguistic terms.

1 INTRODUCTION

Information Systems (IS) are composed of a set of data management elements designed to provide services and benefits in areas as far a part as public administration, industrial control, the banking or geographical and weather information.

Technological developments and the universal internet access has led to an increase in system vulnerabilities. Therefore, ISs have to be analysed with a view to risk minimization by means of well-planned actions to protect information, processes and services from possible threats. Threats range from act of terrorism, industrial espionage, etc., or even a simple unintentional human error by an operator.

Standards promoted by the *International Organization for Standardization* [ISO/IEC](2005, 2011) on IS security suggest three-stage risk analysis and management methodologies.

The *planning stage* establishes the necessary points for starting up the project, defines objectives, and identifies participants and competencies. The *analysis stage* identifies the IS assets, as well as their relations (dependencies), the threats to which they are exposed and their frequency and asset degradation levels. Finally, the *risk management stage* determines the safeguards and strategies that reduce impact and risk.

In this paper, we focus on the second stage, *analysis*. Assets are the IS or related resources, necessary for an organization's correct operation and for achieving the goals set by its manager. Assets can be data, applications, software, facilities, hardware, services...

The asset dependencies are usually represented in terms of percentages, signalling how likely the failure of an asset is to affect another.

Often only a few elements (*terminal assets*), usually data or services, account for the total value of an organization's assets. The value of these assets is transferred to other assets through the established dependency relations. Thus, non-terminal assets have no intrinsic values; they accumulate their value from terminal assets.

However, the methodologies based on international standards, such us (López Crespo, Amutio-Gómez, Candau and Mañas, 2006a, 2006b, 2006c), MEHARI [CSIF](2010), CRAMM [CCTA](2003), OCTAVE-S (Alberts and Dorofee, 2005) or NIST 800-30 (Stoneburner and Gougen, 2002), obviate the difficulty of correctly assigning asset dependencies, as well as terminal asset values or the impact on the entire system caused by the materialization of a threat to an asset. Moreover, these methodologies do not consider uncertainty concerning these assessments.

In this paper we propose a fuzzy risk analysis in IS as a solution to these deficiencies.

Section 2 reviews some operations on trapezoidal fuzzy numbers and introduces a fuzzy evaluation of asset dependencies. Section 3 provides a fuzzy five-component valuation of assets on the basis of five

components is provided. Threats and asset risk impact indicators are described in Section 4. In Section 5, we introduce the similarity function used to associate a linguistic term from a set with a trapezoidal fuzzy numbers. Finally, some conclusions and future research are discussed in Section 6.

2 FUZZY VALUATION OF DEPENDENCIES

We use the following arithmetic for fuzzy numbers proposed in (Xu, Shang, Qian and Shul, 2010): If $\widetilde{A}_1 = (a_1, b_1, c_1, d_1)$ and $\widetilde{A}_2 = (a_2, b_2, c_2, d_2)$, then

$$\widetilde{A}_1 \oplus \widetilde{A}_2 = (a_1 + a_2 - a_1 a_2, b_1 + b_2 - b_1 b_2, c_1 + c_2 - c_1 c_2, d_1 + d_2 - d_1 d_2),$$
$$\widetilde{A}_1 \otimes \widetilde{A}_2 = (a_1 \cdot a_2, b_1 \cdot b_2, c_1 \cdot c_2, d_1 \cdot d_2).$$

As mentioned above, the assets in IS are connected by dependency relationships, and a failure of one asset may affect other assets. Asset A_j *depends* on the asset A_i (or A_i *influences* A_j), denoted by (A_i, A_j) (graphically $A_i \rightarrow A_j$), if a failure in asset A_i causes a failure in the asset A_j with any given probability. This probability is usually referred to as the *degree of dependency* of A_j with respect to A_i or the *influence* of A_i over A_j, which we denote it by $dd(A_i, A_j)$.

Proposed IS risk analysis methodologies (MAGERIT, MEHARI, OCTAVE...) assign just a percentage to indicate the degree of dependency between two assets, and sometimes even propose the use of a Boolean value indicating whether or not this dependency exists regardless of the degree of dependency. We propose the use of trapezoidal fuzzy numbers to represent these dependencies. Consequently, the experts can build a linguistic term set to intuitively define the dependency between two assets under uncertainty.

Our aim then is to compute the indirect asset dependencies since assets values are accumulated from terminal assets through these dependencies.

The degree of dependency of asset A_k with respect to A_i, $\widetilde{DD(A_i, A_k)}$, is computed as follows[1]. We denote by **P**=$\{P_1, ..., P_s\}$ the set of paths in the analysis of the influence of A_i over A_k. Then,

A) If all assets (excluding A_i and A_k) in the paths in **P** are influenced by only one asset, then

$$\widetilde{DD(A_i, A_k)} = \overset{s}{\underset{j=1}{\oplus}} \widetilde{DD(A_i, A_k|P_j)}, \qquad (1)$$

where $\widetilde{DD(A_i, A_k|P_j)} = \widetilde{dd(A_i, A_{j1})} \otimes \widetilde{dd(A_{j1}, A_{j2})} \otimes ... \otimes \widetilde{dd(A_{jn}, A_k)}$, and $P_j : (A_i \rightarrow A_{j1} \rightarrow A_{j2} \rightarrow ... \rightarrow A_{jn} \rightarrow A_k)$.

B) Otherwise, we assume that the first r paths in **P** are formed by assets (excluding A_i and A_k) influenced by only one asset, and the remaining $s-r$ paths include at least one asset influenced by two or more assets. Then, for the r first paths, we proceed as in *A)*, and we denote by **S** the set including the $s-r$ remaining paths. We proceed with **S** as follows:

(i) Compute the set of non-terminal assets in **S** influenced by two or more assets, denoted by I, and the subset of I including assets uninfluenced by any other asset in I, denoted by NI.

(ii) We consider an asset A_r in NI. Then, we simplify the paths in **S** that include asset A_r making $A_i \rightarrow A_r \rightarrow ... \rightarrow A_k$, with $\widetilde{dd(A_i, A_r)} = \widetilde{DD(A_i, A_r)}$ (computed as in *A)*).

(iii) Remove repeated paths from **S** and keep only one instance.

(iv) Build I and NI again from **S**.

(v) If NI is not empty, go to (ii). Otherwise, the algorithm finishes.

Let us denote the resulting set of paths by **S**= $\{P_1, ..., P_m\}$, with $m \leq s-r$. Then, the degree of dependency of A_k regarding A_i is

$$\widetilde{DD(A_i, A_k)} = \overset{r}{\underset{j=1}{\oplus}} \widetilde{DD(A_i, A_k|P_j)} \overset{m}{\underset{l=1}{\oplus}} \widetilde{DD(A_i, A_k|P_l)}. \qquad (2)$$

Note that transactions between trapezoidal fuzzy numbers representing linguistic terms from a set in $[0, 1]$ will remain in $[0, 1]$, and the results of these operations can be translated into one of the linguistic terms of the set by means of a *similarity function*. Furthermore, the operation $\oplus$ is consistent with the methodologies established for risk analysis and management in IS, allowing performances in probabilistic terms.

3 FUZZY VALUATION OF ASSETS

MAGERIT defines the *value of an asset* as the losses that would be sustained if the respective asset is no longer available. These can be losses of money, user confidence, the organizational prestige... Assets are usually evaluated taking into account the following five components (López Crespo et al., 2006a, 2006b, 2006c):

[1] To avoid ambiguity, we will write "DD" to refer to total dependency between two assets separated by other intermediate assets, and "dd" when they are directly connected.

- *Confidentiality.* How much damage would it cause if the asset is disclosed to someone it should not be? This is a typical data inspection.
- *Integrity.* How much damage would it cause if the asset is damaged or corrupt? This is a typical data inspection. Data can be manipulated, be wholly or partially false, or even missing.
- *Authenticity.* How much damage would it cause if we do not exactly know who has done what? This is a typical services (user authentication) and data (authenticity of the person accessing data to write or read) inspection.
- *Traceability.* How much damage would it cause if it is not known for whom the service is being provided?, i.e. who does what and when? How much damage would it cause if it is not known who accessed what data and what they did with them?
- *Availability.* How much damage would it cause if the asset is not available or cannot be used? This is a typical services inspection.

Only the terminal assets have an associated value for the above components. The other assets accumulate value from terminal assets on the basis of dependency relationships. We again use the set of linguistic terms that represent trapezoidal fuzzy numbers to represent uncertainty when valuating the terminal assets.

Let us denote assets by $\widetilde{v}_j = (\widetilde{v}_{j_{(1)}}, \widetilde{v}_{j_{(2)}}, \widetilde{v}_{j_{(3)}}, \widetilde{v}_{j_{(4)}}, \widetilde{v}_{j_{(5)}})$, where $\widetilde{v}_{j_{(i)}}$ is a linguistic term assigned by an expert for the ith value component in asset A_j. If we denote by TAS the terminal asset set, then the value of asset A_j with respect to terminal assets is:

$$\widetilde{v}_{j_{(i)}} = \bigoplus_{A_k in TAS} (\widetilde{DD(A_j, A_k)} \otimes \widetilde{v}_{k_{(i)}}).$$

4 THREATS

Next, we assess threats and estimate indicators of the impact on and risk to assets. A *threat* is an event that can trigger an incident in our organization, causing damage or intangible material loss to the assets, and an *attack* is any deliberate action aimed at violating the IS security mechanisms. MAGERIT suggests two threat assessment measures: *degradation*, the damage that the threat can cause to the asset, and *frequency*, how often the threat materializes.

We will again use fuzzy linguistic terms rather than percentages and probabilities to represent degradation and frequency. A threat is a vector $\overrightarrow{\widetilde{u}} = (\widetilde{D}, \widetilde{f})$ whose components are degradation and frequency. Note that the degradation has to be established for each the the five asset components described in the previous section,

$$\overrightarrow{\widetilde{D}} = (\widetilde{d}_1, \widetilde{d}_2, \widetilde{d}_3, \widetilde{d}_4, \widetilde{d}_5),$$

i.e., the threat causes a degradation $\widetilde{d}_i$ in the ith component of the asset.

Let us consider a threat on the asset A_j. When the threat is realized, each component is affected by the expression

$$\widetilde{I}_{j_{(i)}} = \widetilde{d}_i \otimes \widetilde{v}_{j_{(i)}}, \quad (3)$$

where $\widetilde{I}_{j_{(i)}}$ is the *impact* on the ith component of the attacked asset (A_j).

We use Eq. (4) below to compute the *risk* to the attacked asset

$$\widetilde{R}_{j_{(i)}} = \widetilde{I}_{j_{(i)}} \otimes \widetilde{f}. \quad (4)$$

After computing the impact caused by a materialized threat on an asset, we can compute the impact transmitted from the attacked asset to its dependent assets. If A_j is the asset on which the threat has materialized and the degree of dependency of A_j with respect to A_k is $\widetilde{DD(A_k, A_j)}$, then the attack on asset A_j has an impact on A_k of $\widetilde{I}_{k_{(i)}} = \widetilde{DD(A_k, A_j)} \otimes \widetilde{d}_i \otimes \widetilde{v}_{j_{(i)}}$. Thus, the risk to asset A_k is

$$\widetilde{R}_{k_{(i)}} = \widetilde{I}_{k_{(i)}} \otimes \widetilde{f} = \widetilde{DD(A_k, A_j)} \otimes \widetilde{d}_i \otimes \widetilde{v}_{j_{(i)}} \otimes \widetilde{f}. \quad (5)$$

5 SIMILARITY FUNCTION

A *similarity function* is required to associate the resulting trapezoidal fuzzy number with an element in the linguistic term set. This function can also be used at any step of the methodology to derive the linguistic terms associated with the respective trapezoidal fuzzy numbers output to represent dependencies, accumulated values...

Several authors have proposed different similarity functions, which are based on the centroid of a fuzzy number and the distance between the components of the fuzzy numbers, see (Lee, 1999; Chen and Chen 2001, 2007). Finally, a more recent similarity function was proposed in (Xu et al., 2010) and compared with the proposal reported in (Chen and Chen, 2007).

We use the function proposed in Vicente, Mateos and Jiménez (2012), which considers another parameter consisting of the ratio between the common area and the joint area under the membership functions of trapezoidal fuzzy numbers. Moreover, we use the distance l_∞ between centroids since the use of distances with non-rectangular spheres is inconsistent with the intuitive perception of similarity.

Given $\widetilde{A}$ and $\widetilde{B}$, the similarity function can be defined as

$$S(\widetilde{A},\widetilde{B}) = 1 - w_1\left(1 - \frac{\int_0^1 \mu_{\widetilde{A}\cap\widetilde{B}}(x)dx}{\int_0^1 \mu_{\widetilde{A}\cup\widetilde{B}}(x)dx}\right) - w_2\frac{\sum|a_i-b_i|}{4} - \\ - w_2\frac{\sum|a_i-b_i|}{4} - w_3 L_\infty[(X_{\widetilde{A}},Y_{\widetilde{A}}),(X_{\widetilde{B}},Y_{\widetilde{B}})],$$

where $w_1+w_2+w_3=1$, $(X_{\widetilde{A}},Y_{\widetilde{A}})$ and $(X_{\widetilde{B}},Y_{\widetilde{B}})$ are the centroids of $\widetilde{A}$ and $\widetilde{B}$, respectively, i.e.

$$X_{\widetilde{A}} = Y_{\widetilde{A}}(a_3+a_2)+(1-Y_{\widetilde{A}})(a_4+a_1) \text{ and}$$

$$Y_{\widetilde{A}} = \begin{cases} \frac{\left(\frac{a_3-a_2}{a_4-a_1}+2\right)}{6}, & \text{if } a_4-a_1 \neq 0 \\ \frac{1}{2}, & \text{if } a_4-a_1 = 0 \end{cases},$$

μ_χ is the membership function of χ,

$$\mu_{\widetilde{A}\cap\widetilde{B}}(x) = \min_{0\leq x\leq 1}\{\mu_{\widetilde{A}}(x),\mu_{\widetilde{B}}(x)\},$$
$$\mu_{\widetilde{A}\cup\widetilde{B}}(x) = \max_{0\leq x\leq 1}\{\mu_{\widetilde{A}}(x),\mu_{\widetilde{B}}(x)\},$$

and

$$L_\infty((x_1,y_1),(x_2,y_2)) = max\{|\,x_1-x_2\,|,|\,y_1-y_2\,|\}.$$

Note that w_1, w_2 and w_3 represent the relative importance of the three elements considered in the similarity function. Analysts will assign the values that best fits their own model.

6 CONCLUSIONS

We have developed a fuzzy risk analysis model for information systems that conforms to international standards, particularly the MAGERIT methodology. The model is an improvement on this and other existing methodologies since it includes uncertainty about the assessments by means of linguistic terms, which have associated trapezoidal fuzzy numbers. The proposed methodology makes computations on the basis of trapezoidal fuzzy numbers to accumulate dependencies between assets and asset valuations and to determine impacts and risk from the threat degradation and frequency, respectively. Moreover, similarity functions can be used at any step in the methodology to derive a linguistic term for the trapezoidal fuzzy number output.

ACKNOWLEDGEMENTS

The paper was supported by Madrid Regional Government project S-2009/ESP-1685 and the Spanish Ministry of Science and Innovation project MTM2011-28983-C03-03.

REFERENCES

Alberts, C. and Dorofee, A. (2005). *OCTAVE-s Method Implementation Guide Version 2.0.* Pittsburgh: Canergie Mellon University.

Chen, S.-J. and Chen, S.-M. (2001). A New Method to Measure the Similarity between Fuzzy Numbers. *Proceedings of the 10th IEEE International Conference on Fuzzy Systems*, 208-214.

Chen, S.-J. and Chen, S.-M. (2007). Fuzzy Risk Analysis Based on the Ranking of Generalized Trapezoidal Fuzzy Numbers. *Applied Intelligence*, 26, 1-11.

CCTA *Risk Analysis and Management Method (CRAMM), Version 5.0.* London: Central Computing and Telecommunications Agency (CCTA), 2003.

ISO/IEC 17799:2005, *Information technology - Security techniques - Code of practice for information security management.* Geneva: International Organization for Standarization.

ISO/IEC 27005:2011, *Information technology - Security techniques - Information security risk management.* Geneva: International Organization for Standarization.

Lee, H.S. (1999). An Optimal Aggregation Method for Fuzzy Opinions of Group Decision. *Proceedings of the 1999 IEEE International Conference on Systems, Management and Cybernetics*, 314-319.

López Crespo, F., Amutio-Gómez, M.A., Candau, J. and Mañas, J.A. (2006a). *Methodology for Information Systems Risk. Analysis and Management (MAGERIT version 2). Book I-The Method.* Madrid: Ministerio de Administraciones Públicas.

López Crespo, F., Amutio-Gómez, M.A., Candau, J. and Mañas, J.A. (2006b). *Methodology for Information Systems Risk Analysis and Management (MAGERIT version 2). Book II-Catalogue of Elements.* Madrid: Ministerio de Administraciones Públicas.

López Crespo, F., Amutio-Gómez, M.A., Candau, J. and Mañas, J.A. (2006c). *Methodology for Information Systems Risk Analysis and Management (MAGERIT version 2). Book III-The Techniques.* Madrid: Ministerio de Administraciones Públicas.

Mehari 2010 - *Risk Analysis and Treatment Guide.* Paris: Club de la Sécurité de l'Information Francais (CSIF).

Stoneburner, G. and Gougen, A. (2002). *NIST 800-30 Risk Management. Guide for Information Technology Systems.* Gaithersburg: National Institute of Standard and Technology.

Vicente, E., Mateos, A. and Jiménez, A. (2012). A New Similarity Measure of Trapezoidal Fuzzy Numbers. *Expert Systems with Applications*, under review.

Xu, Z., Shang, S., Qian, W. and Shu, W. (2010). A Method for Fuzzy Risk Analysis based on the New Similarity of Trapezoidal Fuzzy Numbers. *Expert Systems with Applications*, 37, 1920-1927.