A Fuzzy Model for IT Security Investments

Guido Schryen

schryen@winfor.rwth-aachen.de

Abstract: This paper presents a fuzzy set based decision support model for taking uncertainty into account when making security investment decisions for distributed systems. The proposed model is complementary to probabilistic approaches and useful in situations where probabilistic information is either unavailable or not appropriate to reliably predict future conditions. We first present the specification of a formal security language that allows to specify under which conditions a distributed system is protected against security violations. We show that each term of the security language can be transformed into an equivalent propositional logic term. Then we use propositional logic terms to define a fuzzy set based decision model. This optimization model incorporates uncertainty with regard to the impact of investments on the achieved security levels of components of the distributed system. The model also accounts for budget and security constraints, in order to be applicable in practice.

1 Introduction

Emerging digital environments and infrastructures have rapidly generated new ways and services of communication, information sharing, and resource utilization for individuals, organizations, and societies in past years. For example, it has become common for individuals to use security services, such as *I2P Anonymous Network* and *TOR*. Organizations have started to explore the opportunities of web services, including storage services (e.g., *Amazon Simple Storage Service*) and computing services (e.g., Microsoft's Azure Services Platform and Google App Engine). While the aforementioned services are realized with cloud computing, services can also be requested from multiple administrative domains (*grid computing*). Even whole societies are involved in scenarios with shared information and transaction processing, as political elections with electronic voting systems show.

What all these services have in common is that some kind of distributed information processing and/or information sharing occurs, across private, organizational, or national boundaries. Often, consumers of these services have no control over their data, and they need to trust service providers not to violate their security policies. For example, scientific computation results can be modified or provided to third parties. In some cases, organizational, legal, and/or technical countermeasures have been taken in order to prevent or to mitigate the consequences of data abuse. For example, in Internet voting the separation of duties is quite common in order to realize the separation of voter's identity and his/her vote. In such cases, the abuse of data by a single party (insider abuse) and the compromise of systems by attackers (outsider abuse) do not disclose confidential information.

However, what happens when multiple parties maliciously cooperate and join their information, or when multiple system components are compromised jointly by attackers? This leads to scenarios where a voter's ID can be assigned to his/her vote, where the identity of a user is disclosed through the cooperation of parties of an anonymity mix net, etc. Consequently, when security investments in distributed systems are planned, the questions arise of (1) how important the security of particular system components is, and (2) how much should be invested in which component to increase the overall security of the distributed system. Thereby, we focus on the ex ante security assessment of distributed systems, and the support of security investment decision makers.

Beyond the challenge to address the aforementioned interdependencies between system components, decision makers also face budget constraints and various sources of uncertainty. Unfortunately, uncertainty is often not probabilistic so that the application of probabilistic approaches is of limited effectiveness. We thus draw on fuzzy set theory, which is a valuable uncertainty theory in the absence of probabilities and in the presence of subjective assessments.

The main purpose of this paper is to present a novel fuzzy set based decision support model for security investment decision makers. From the methodological perspective, we formally derive the decision model by proposing a formal security language and by applying propositional logic, decision theory, and fuzzy set theory. We further draw on computational complexity theory to analyze the complexity of the model.

The remainder of this paper is structured as follows: Section 2 presents related work. In Section 3, we describe our research framework. Section 4 proposes the formal security language, demonstrates its applicability, and shows how resilience terms of the security language can be mapped on propositional logic terms. Section 5 provides a brief introduction into uncertainty modeling and fuzzy set theory. In Section 6, the fuzzy decision support model is proposed and analyzed. Section 7 discusses implications and shows opportunities for further research.

2 Related work

The economics of information security investments has been analyzed in the literature at both the ex post level and the ex ante level (decision making). An example of the former perspective is the NIST *Performance Measurement Guide for Information Security* [NIS08], which focuses on ex post security measures. As the focus of this paper lies on decision making, we concentrate our overview on papers on the ex ante analysis of information security investments.

In their survey of economic approaches for security metrics, [BN08] analyze the literature through a methodological lens and identify two main areas of research, where one has its roots in investment and decision theory and is mainly pursued in the field of information technology-oriented business administration, and the other area of research has ancestors in micro-economics and deals with market concepts to gather security-relevant information. We adopt a different perspective and focus on theoretical approaches used to

address uncertainty in security investment decision making. Unsurprisingly, the literature is very much focused on probabilistic approaches and often adopts the risk-based perspective. [GL02] present an economic model that determines the optimal amount to invest to protect a given set of information and that uses probabilities that attacks are successful. [GJC09] propose metrics for measuring the price of uncertainty due to the departure from the payoff-optimal security outcomes under complete information, and assume that agents face randomly drawn probabilities of being subject to direct attacks. [GCC08] apply game theory to study how economic agents invest into security in different economic environments, and they assume that attacks arrive with a probability that remains constant over time. [CRY08] consider the decision-making problem of a firm when attack probabilities are externally given. In their approach to derive implications for security investment strategies based on attackers' decisions, [CN06] draw on the probability of (attackers') success given an amount of effort put into attacking a given target. [HHB06] propose an economic model that considers simultaneous attacks from multiple external agents with distinct characteristics, and derive optimal investments based on the principle of benefit maximization. In their model they draw on security breach probabilities.

However, there are also dissenting voices, which doubt the appropriateness of using probabilistic approaches. For example, [WCR05] argue that risk-driven decision models are limited due to the difficulty of reliably estimating the potential losses from security breaches and the probability of these breaches. [HN10] find that risk assessment methods found in the literature tend to underestimate the risks associated with large-impact, hard-to-predict, and rare events.

We found two papers that suggest to apply fuzzy sets in the context of security investment decisions. [Lee03] presents a simple model that uses linguistic variables to represent criteria upon which investment decisions are made. [KSST09] suggest to use fuzzy sets in the context of evaluation processes. In their paper, fuzzy sets are used to express the extent with which security measures are implemented in an organization.

3 Research framework

Our approach (see Figure 1) assumes that the structure of a distributed system is known. We draw on this structure to derive a formal resilience term, which specifies which components and/or groups of components need to be secure with regard to a particular security requirement r (e.g. confidentiality, anonymity) so that the overall distributed system is secure with regard to r. The specification of r is important, because different security requirements can lead to different resilience terms. For example, in a system that implements a mixnet that routes messages sequentially through a set N of n anonymizing nodes, each node must be secure with regard to availability (n out of N), while only one node needs to be secure with regard to achieving anonymity (1 out of N). The formal security language that we propose in this paper draws on [HKS00], who use secret shares [BK05] and the concept that k out of n entities are required for revealing a secrecy. We adopt and adapt this concept, and we say: *"k out of N entities must be secure"*. In contrast to the aforementioned papers, which regard entities/components as homogeneous, we account

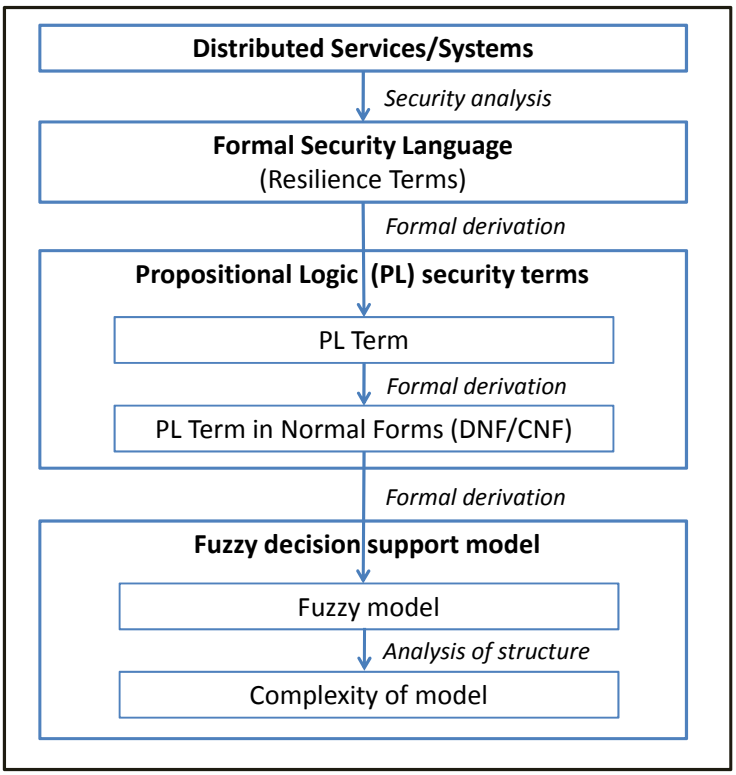

Figure 1: Research framework

for heterogeneity of entities by explicitly itemizing them in the set N.

While resilience terms are a useful representation of required security properties of a distributed system, they are less appropriate for analyzing systems with regard to weak points and strong points, and for making security investment decisions. We show that each resilience term can be mapped on a propositional logic term such that both terms are semantically equivalent, and we show that converting propositional logic terms into normal forms, such as the *conjunctive normal form* (CNF), is a useful way to identify such weak and strong points.

Accounting for the fact that security investment decision makers need to consider various sources of uncertainty and different types of constraints, we suggest a fuzzy decision support model. The goal function of this model is derived from the CNF representation of the particular resilience term, which links the introductory, theoretical parts of this paper with the proposed decision model. We finally analyze the structure of the decision model, and we discuss types of required data.

4 Formal security language and propositional logic terms

4.1 Formal security language

As our formal security language describes required security properties of distributed systems, we first define distributed systems: A distributed system is either an "atomic system" or is composed of other (sub)systems. We define a system as "atomic" if it contains only (atomic) components that are not being split any further. These components can be persons, computers, or even organizational units.

The definition of the security language (resilience terms) in terms of syntax and semantics follows the inductive definition of systems and is provided by definitions 4.1-4.4. In order to keep definitions short, we introduce the abbreviation "*wrts. r*" (with regard to security requirement r).

Let S be an atomic system with the set of atomic components $A = \{A_i\}_{i=1}^{n}$.

Definition 4.1 *A system S is (k out of N)-resilient, $k \in \{1, \ldots, |N|\}$, $N \subseteq A$, wrts. r*

$:\Leftrightarrow$ *At least k components out of N need to be secure wrts. r in order to make S meet r.*

In order to get more flexible representations of requirements on atomic systems, we define the following resilience terms:

Definition 4.2 *A system S is a) $((k_1 \oslash \ldots \oslash k_m)$ out of $(N_1, \ldots, N_m))$-resilient, b) $((k_1 \otimes \ldots \otimes k_m)$ out of $(N_1, \ldots, N_m))$-resilient, $k_i \in \{1, \ldots, |N_i|\}$, $N_i \subseteq A$ $\forall i$, wrts. r*

$:\Leftrightarrow$ *For a) each, b) any $i \in \{1, \ldots, m\}$, at least k_i components out of N_i need to be secure wrts. r so that S meets requirement r.*

With regard to non-atomic systems, we define resilience terms similarly: Let $\{S_i\}_{i=1}^{n}$ be (sub)systems of a system S, and let system S_i be l_i-resilient for all $i \in \{1, \ldots, n\}$.

Definition 4.3 *A system S is (k out of $\{l_{i_1}, \ldots, l_{i_m}\}$)-resilient, $k \in \{1, \ldots, m\}$, $\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$, wrts. r*

$:\Leftrightarrow$ *At least k systems out of $\{S_{i_1}, \ldots, S_{i_m}\}$ need to be secure wrts. r so that S meets requirement r.*

Definition 4.4 *A system S is a) $((k_1 \oslash \ldots \oslash k_m)$ out of $(N_1, \ldots, N_m))$-resilient, b) $((k_1 \otimes \ldots \otimes k_m)$ out of $(N_1, \ldots, N_m))$-resilient, $k_i \in \{1, \ldots, |N_i|\}$, $N_i \subseteq \{l_1, \ldots, l_n\}$ $\forall i$, wrts. r*

$:\Leftrightarrow$ *For a) each, b) any $i \in \{1, \ldots, m\}$, at least k_i systems out of the set of systems for which N_i contains resilience terms need to be secure wrts. r so that S meets requirement r.*

We now illustrate the security analysis and the determination of resilience terms with an example.

Example 4.1 *We use a web service scenario, in which a retailer uses three web services in order to identify customers' behavior. Service A offers data mining capabilities and stores sales data, including customer IDs. Service B is offered by a financial service provider, who provides credit ratings of customers. Service C provides storage capacities and stores master data on customers, including their customer IDs and identities. In this example, we consider secrecy with regard to information on which customer has bought what under which financial conditions. Secrecy is kept if one of the providers A and B is secure, or if one of B and C is secure. With regard to provider A, we assume that this provider accounts for secrecy by storing data on two components (A_3 and A_4) and implementing a secret share mechanism [BK05]. Components A_1 and A_2 are responsible for distributed computation in terms of data mining; both components get data from A_3 and A_4. With regard to financial service provider B, customer IDs generated by B (they differ from customer IDs stored at A) are stored on B_1 and B_2 together with financial data by implementing a secret share mechanism. Components B_3 and B_4 store names of customers and customer IDs (generated by B) redundantly. Analogous to A and B, storage provider C implements a secret share mechanism when storing customer data. Figure 2 shows the overall system S. Applying definitions 4.1, 4.2a, 4.2b, and 4.4b, we yield the following resilience terms:*

- A *is* $\underbrace{((2 \oslash 1) \textit{ out of } (\{A_1, A_2\}, \{A_3, A_4\}))}_{l_1}$*-resilient wrts. r. (def. 4.2a)*
- B *is* $\underbrace{((1 \otimes 2) \textit{ out of } (\{B_1, B_2\}, \{B_3, B_4\}))}_{l_2}$*-resilient wrts. r. (def. 4.2b)*
- C *is* $\underbrace{(1 \textit{ out of } \{C_1, C_2\})}_{l_3}$*-resilient wrts. r. (def. 4.1)*
- S *is* $((1 \otimes 1) \textit{ out of } (\{l_1, l_2\}, \{l_2, l_3\}))$*-resilient wrts. r. (def. 4.4b)*

4.2 Propositional logic terms

As example 4.1 shows, resilience terms can become complex, even for small systems. In order to yield representations that are comfortable to interpret for persons and appropriate for the computation of the uncertainty with which a system does not fulfill a specific requirement r, we transform resilience terms into propositional logic formulas. Particularly useful is the subsequent transformation of formulas into semantically equivalent formulas in normal form, such as the disjunctive normal form (DNF) or the conjunctive normal form (CNF). These normal forms show different strengths: while the CNF allows to determine "weak points", such as single points of failure, the DNF is useful for identifying "strong points", such as components or subsystems where security results in the security of the

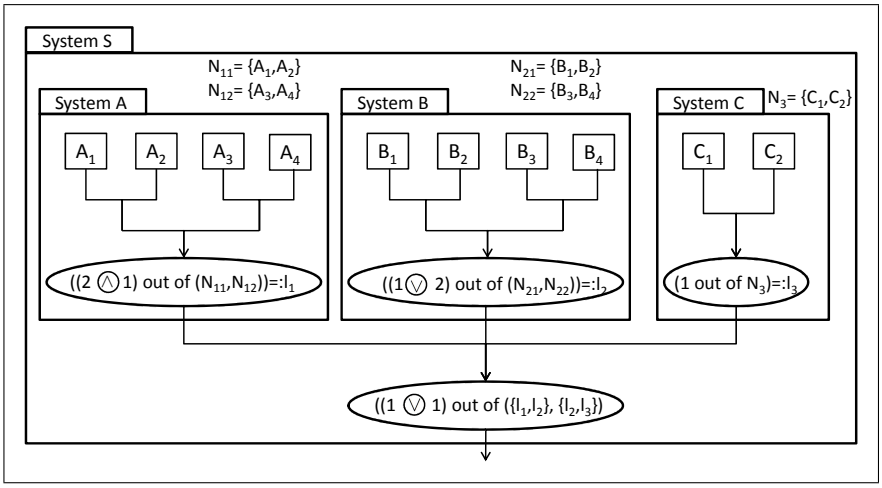

Figure 2: System structure and resilience values of Example 4.1

overall system, regardless of the security (levels) of other components and subsystems. Thus, both normal forms should be applied complementarily.

Theorem 4.1 *Let system S consist of basic components $A = \{A_1, \ldots, A_n\}$, and let $\{X_{A_1}, \ldots, X_{A_n}\}$ be literals with X_{A_i} =true $\forall i$, iff A_i is secure. Then, the resilience term l of S can be mapped on a propositional logic formula $f(l)$ such that S is secure iff $f(l)$ is true.*

Due to limitations of space, we provide a sketch of proof only: The principal idea of the proof is that we reformulate the expression "k out of a set L" by explicitly considering all combinations of elements of L, where L can be a set of basic components or of resilience terms of subsystems. The provision of such a mapping f (of resilience terms on propositional logic terms) proves the theorem.

We use the example shown in Figure 2 to illustrate how to determine the propositional logic formula of a particular resilience term.

Example 4.2

- *resilience term* $l_1 = ((2 \text{⩓} 1)$ *out of* $(\{A_1, A_2\}, \{A_3, A_4\}))$

$$\begin{aligned} \Rightarrow f(l_1) &= (f((2 \textit{ out of } \{A_1, A_2\}))) \wedge (f((1 \textit{ out of } \{A_3, A_4\}))) \\ &= ((A_1 \wedge A_2)) \wedge ((A_3) \vee (A_4)) = A_1 \wedge A_2 \wedge (A_3 \vee A_4) =: f_A \end{aligned}$$

- *resilience term* $l_2 = ((1 \text{⩔} 2)$ *out of* $(\{B_1, B_2\}, \{B_3, B_4\}))$

$$\begin{aligned} \Rightarrow f(l_2) &= (f((1 \textit{ out of } \{B_1, B_2\}))) \vee (f((2 \textit{ out of } \{B_3, B_4\}))) \\ &= ((B_1 \vee B_2)) \vee ((B_3) \wedge B_4)) = B_1 \vee B_2 \vee (B_3 \wedge B_4) =: f_B \end{aligned}$$

- *resilience term* $l_3 = (1 \textit{ out of } \{C_1, C_2\})$

$$\Rightarrow f(l_3) \quad = \quad (C_1) \vee (C_2) = C_1 \vee C_2 =: f_C$$

- *resilience term* $l = ((1 \circledcirc 1) \textit{ out of } (\{l_1, l_2\}, \{l_2, l_3\}))$

$$\begin{aligned} \Rightarrow f(l) &= (f((1 \textit{ out of } \{l_1, l_2\})) \vee (f((2 \textit{ out of } \{l_2, l_3\})) \\ &= (((f(l_1))) \vee ((f(l_2)))) \vee (((f(l_2))) \vee ((f(l_3)))) \\ &= (f(l_1)) \vee (f(l_2)) \vee (f(l_3)) = (f_A) \vee (f_B) \vee (f_C) \\ &= (A_1 \wedge A_2 \wedge (A_3 \vee A_4)) \vee (B_1 \vee B_2 \vee (B_3 \wedge B_4)) \vee (C_1 \vee C_2) \end{aligned} \tag{1}$$

We now convert the resulting propositional logic term into DNF and CNF. The DNF representation can be easily derived from (1) and is given by

$$(A_1 \wedge A_2 \wedge A_3) \vee (A_1 \wedge A_2 \wedge A_4) \vee B_1 \vee B_2 \vee (B_3 \wedge B_4) \vee C_1 \vee C_2 \tag{2}$$

Having available the DNF representation, we can easily derive the CNF representation, which is given by

$$\bigwedge_{\substack{X \in \{A_1, A_2, A_3\} \\ Y \in \{A_1, A_2, A_4\} \\ Z \in \{B_3, B_4\}}} (X \vee Y \vee B_1 \vee B_2 \vee Z \vee C_1 \vee C_2) \tag{3}$$

While the DNF representation in 2 shows that each of the components B_1, B_2, C_1, C_2 is a strong point, the CNF representation reveals that there is (fortunately) no single point of failure.

5 Fuzzy set theory

Fuzzy set theory goes back to Lotfi Zadeh [Zad65], who proposed fuzzy sets as means for dealing with non-probabilistic uncertainty. As it is far beyond the scope of this paper to provide an overview of this field, we briefly introduce the very basic ideas of fuzzy set theory [Zim96, BE02]. The key idea of fuzzy set theory is the extension of the (crisp) membership concept in traditional set theory by providing for a degree with which an element belongs to a set. The degree is specified by a membership function.

Definition 5.1 *Let Ω be some set. Then we define a fuzzy set A as follows:*

$$A := \{(x, A(x)) | x \in \Omega\}, \textit{ with } A(x) := \mu_A(x) : \Omega \rightarrow [0, 1] \textit{ being the membership function.} \tag{4}$$

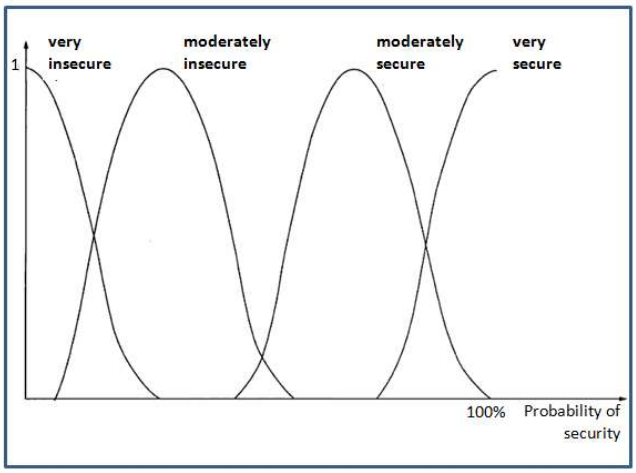

Figure 3: Linguistic variable *security* [Zim96, p. 132]

A particular type of fuzzy set is a *fuzzy number*:

Definition 5.2 *A fuzzy number is a fuzzy set A over $\Omega = \mathbb{R}$ such that $\mu_A(x)$ is piecewise continuous and it exists exactly one interval $[a, b]$ with $\mu_A(x) = 1\ \forall a \leq x \leq b$.*

For example, a fuzzy set can represent an *integer number close to 10*, where

$$A = \{(x, \mu_A(x)) | \mu_A(x) = (1 + (x - 10)^2)^{-1},\ x \in \mathbb{R}\}$$

Set-theoretic operations with fuzzy sets are pointwise defined over their membership functions [Zim96, BE02]. For example, the membership function $\mu_C(x)$ of the intersection $C = A \cap B$ can be defined by $\mu_C(x) = \min\{\mu_A(x), \mu_B(x)\},\ x \in \Omega$. Further set-theoretic operators, and relational operators and arithmetic operators for fuzzy numbers are presented in [Zim96, BE02].

Another powerful concept in the field of fuzzy set theory turned out to be linguistic variables [Zad73]. We present a formal definition provided by [Zim96, p. 131]:

Definition 5.3 *A linguistic variable is a quintuple $(x, T(x), \Omega, G, \overline{M})$, where (1) x is the name of the variable (e.g., security), (2) $T(x) = T$ denotes the terms of the variable (e.g., $\{$very insecure, moderately insecure, moderately secure, very secure$\}$), (3) Ω is some set, (4) G is a syntactic rule for generating terms, and (5) $\overline{M}(T)$ assigns a fuzzy set to term T.*

Example 5.1 *Figure 3 shows an example of the linguistic variable security.*

6 Fuzzy decision support model

6.1 The model

In order to keep the model simple, we do not consider more than one security requirement at the same time, thus receiving a model that contains only one goal function. If we need to address several security requirements contemporaneously, which differ with regard to their resilience terms, we get a multi-criteria decision model, which contains one goal function for each security requirement. We assume that the security description of a distributed system, which contains the set of components A, is given by the propositional logic formula (in CNF)

$$\begin{aligned} A &= (A_{11} \vee \ldots \vee A_{1n_1}) \wedge \ldots \wedge (A_{m1} \vee \ldots \vee A_{mn_m}) \\ &= \bigwedge_{i=1}^{m} \left(\bigvee_{j=1}^{n_i} A_{ij} \right), \; A_{ij} \in A \; \forall i,j, \; A_{ij} \text{ not necessarily different} \end{aligned} \tag{5}$$

In accordance with our assumption that security levels and security investment expenses can be appropriately represented by terms of linguistic variables and by fuzzy numbers, respectively, we suggest the following fuzzy decision model, which includes a fuzzy objective function, fuzzy variables, fuzzy parameters, and crisp constraints:

$$\max \; ((X_{11} \cup \ldots \cup X_{1n_1}) \cap \ldots \cap (X_{m1} \cup \ldots \cup X_{mn_m})) = \bigcap_{i=1}^{m} \left(\bigcup_{j=1}^{n_i} X_{ij} \right) \tag{6}$$

$$\text{s. t.} \quad X_{ij} \geq B_{ij}^0 \quad \forall i,j | A_{ij} \in A \tag{7}$$

$$X_{ij} \geq B_{ij}^* \quad \forall i,j | A_{ij} \in A \tag{8}$$

$$\sum_{i,j|A_{ij} \in A} c_{ij}(X_{ij}, B_{ij}^0) \leq b \tag{9}$$

$$\sum_{i,j|A_{ij} \in A^{(t)} \subset A} c_{ij}(X_{ij}, B_{ij}^0) \leq b_t \, \forall t \tag{10}$$

$$c_{ij}(X_{ij}, B_{ij}^0) \leq b_{ij} \quad \forall i,j | A_{ij} \in A \tag{11}$$

$$X_{ij}, B_{ij}^0, B_{ij}^* \in \overline{M}(T(\text{security})) \quad \forall i,j | A_{ij} \in A \tag{12}$$

$$b, b_{ij}, c_{ij}(X_{ij}, B_{ij}^0) \in \overline{\mathbb{R}} \text{ (fuzzy numbers)}, \quad \forall i,j | A_{ij} \in A \tag{13}$$

The fuzzy decision model aims at maximizing the overall security of the system described by A under security constraints (regarding single system components) (7-8) and budget constraints (9-11). The decision variables X_{ij} are fuzzy variables and can be assigned fuzzy sets of terms of the linguistic variable *security* (12) – $\overline{M}$ maps linguistic terms on fuzzy sets. For example, $T(\text{security})$ could consist of the terms *very insecure*, *moderately insecure*, *moderately secure*, *very secure*. The fuzzy goal function (6) combines all decision variables (security expressions) according to the logic-based security description A. In (6), the fuzzy set operators $\cup$ and $\cap$ are union and intersection operators, respectively,

and they correspond to the logical operators $\vee$ and $\wedge$. Here we can see how the binary differentiation between *secure* and *insecure* components is fuzzified. It should be noted that, in contrast to crisp decision models, it still needs to be specified how fuzzy sets are ordered in order to maximize the objective value.

The optimization underlies two types of security constraints: for each component A_{ij}, the security level X_{ij} after the investment $c_{ij}(X_{ij}, B^0_{ij})$ must be larger than or equal to its original level B^0_{ij} (7) and must also be larger than an exogenously given aspiration level B^*_{ij} (8), where B^0_{ij}, B^*_{ij} are fuzzy sets of linguistic terms (12). The optimization also underlies three types of budget constraints: the overall expenses to increase the security levels from B^0_{ij} to X_{ij} must be not larger than an overall fuzzy budget constraint b (9). Similarly, there may be a budget constraint for a single component (11) or for a group of components (10). It should be noted that the expenses for increasing the security level of a component depend on the current security level and the future security level (11). We model expenses as fuzzy numbers.

It should also be noted that while all constraints contain fuzzy sets, the decision of whether a constraint is met is sharp (in our model). This is due to the fact that (in our model) $\leq$ and $\geq$ define dichotomous relations between two fuzzy sets. Alternatively, we could define fuzzy relations. In this case, we would have to specify how constraints are combined. In this paper, we do not follow this path. We now present a simple example instance of our fuzzy decision model.

Example 6.1 *Let us assume that a distributed anonymizing system connects a source node S and a destination node D through two different paths (due to availability concerns), each of which contains two nodes (see Figure 4a). We use the linguistic variable security, which contains the terms T(security)=$\{$very insecure, moderately insecure, moderately secure, very secure$\}$. The security conditions of the nodes are shown in Figure 4a. Figure 4b shows graphically the membership functions of the fuzzy sets of the terms. Figure 4c provides the expenses for increasing the security level. For the purpose of simplicity, the fuzzy numbers given in the table refer to all components. We also require component A_{11} to become at least moderately insecure. The overall budget of all security investments is approximately 70,000 USD, and the security investment in component X_{12} should not exceed 25,000 USD.*

The resilience term of the system is $((1 \oslash 1)$ out of $(\{A_{11}, A_{12}\}, \{A_{21}, A_{22}\}))$, which results to the propositional logic formula (in CNF): $((A_{11} \vee A_{12}) \wedge (A_{21} \vee A_{22}))$. We get the following fuzzy decision model (the numbers in the model refer to the numbers of the

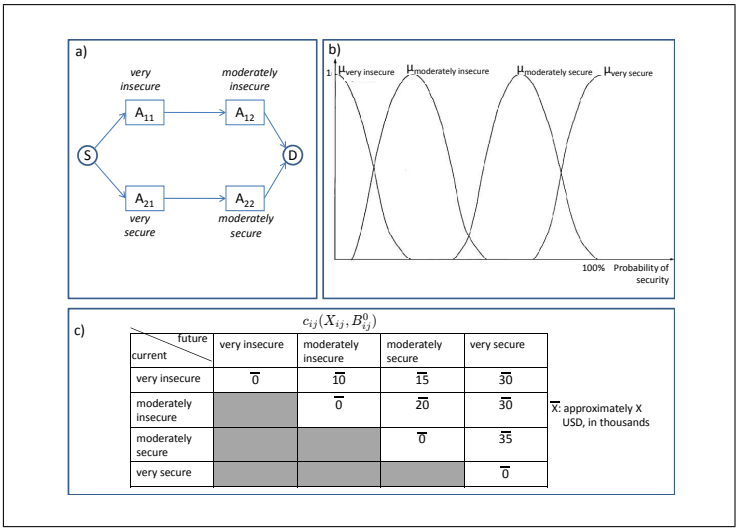

Figure 4: Data of example 6.1

generic fuzzy decision model presented above):

$$
\begin{aligned}
&\textit{max} && ((X_{11} \cup X_{12}) \cap (X_{21} \cup X_{22})) && (6)\\
&\textit{s. t.} && X_{11} \geq \overline{M}(\textit{very insecure}) && (7)\\
& && X_{12} \geq \overline{M}(\textit{moderately insecure}) && (7)\\
& && X_{21} \geq \overline{M}(\textit{very secure}) && (7)\\
& && X_{22} \geq \overline{M}(\textit{moderately secure}) && (7)\\
& && X_{11} \geq \overline{M}(\textit{moderately insecure}) && (8)\\
& && c_{11}(X_{11}, \overline{M}(\textit{very insecure})) + c_{12}(X_{12}, \overline{M}(\textit{moderately insecure}))+ && \\
& && c_{21}(X_{21}, \overline{M}(\textit{very secure})) + c_{22}(X_{22}, \overline{M}(\textit{moderately secure})) \leq \overline{70} && (9)\\
& && c_{12}(X_{12}, \overline{M}(\textit{moderately insecure})) \leq \overline{25} && (11)\\
& && X_{ij} \in \overline{M}(T(\textit{security})) \; \forall i, j | A_{ij} \in A && (12)
\end{aligned}
$$

6.2 Solving model instances

We now discuss which data and which specifications are necessary to solve an instance of the model.

1. The model requires knowledge of the security structure A. This structure can be derived by determining the resilience term and its representation as propositional logic formula in CNF.

2. In order to maximize the goal function, possible results need to be ordered, i.e. the decision maker needs to specify how fuzzy sets are ordered. There is not one single

solution of this problem; for example, we can draw on the relation of fuzzy sets as used in constraints (7) and (8). Alternatively, we can first defuzzify both fuzzy sets and then compare the resulting crisp values.

3. As the model uses a preference relation of two fuzzy sets ((7)-(11)), the decision maker needs to specify a concrete preference relations s/he applies (see Subsection 5).

4. The decision maker needs to know budget constraints. While the overall budget is often known (at least approximately), approximate budgets for components and groups of components are not always given. In case budget constraints of the latter type are not desirable, the respective constraints can be removed from the model.

5. The linguistic variable *security* needs to be defined, including the definition of the terms T(security) (e.g., *very insecure*) and their membership functions. Figure 4b provides such a linguistic variable.

6. For each component, the expenses to increase the level of security need to be specified in terms of a fuzzy number. For example, the decision maker may find that it cost about 50,000 $ to make a *very insecure* component *moderately secure*. Figure 4c shows an example of expenses for increasing the security of a component.

Each instance is a discrete optimization problem, where

- the number of decision variables n_V is the number of components ($n_V \leq \sum_{i=1}^{m} n_i$)[1],
- the solution space contains $n_S = |T(security)|^{n_V}$ elements, i.e. the size of the solution space increases polynomially in the number of linguistic terms, but it increases exponentially in the number of components,
- the number of constraints n_C is $n_V(7) + n_V(8) + 1(9) + n_V(11) = 3 \cdot n_V + 1$. However, we can easily reduce the number of constraints to $2 \cdot n_V + 1$, when we merge constraints (7) and (8) and substitute these with $X_{ij} \geq \max\{B^0_{ij}, B^*_{ij}\}$.

As the size of the solution space increases exponentially in the number of components –four security levels and ten components lead to a solution space that consists of $n_S = 4^{10} \approx 10^6$ elements –, solving a problem instance through enumeration becomes computationally infeasible. Even worse, the decision model turns out to be NP-hard so that the application of heuristic procedures becomes necessary for large instances. Again, due to space limitation we cannot provide details of the formal proof of NP-hardness in this paper. The guiding idea of the proof is the demonstration how the satisfiability problem 1-3-SAT, which is NP-complete [KL99, p. 59], is reducible to the fuzzy decision problem in polynomial time.

[1] A_{ij} do not need to be pairwise different.

7 Discussion

We now discuss some limitations and drawbacks of our approach and show how they can be addressed in further research.

While a key advantage of using a fuzzy decision support model lies in the dispensability of historic, probabilistic data, which are often unavailable, the solution of an instance of our (generic) fuzzy model requires to specify membership functions of fuzzy sets, terms of linguistic variables, and fuzzy operators in such a way that the model mirrors the attitudes and assumptions of the decision maker with regard to security investments. More precisely, what needs to be specified is the order of fuzzy sets with regard to the objective function and with regard to the constraints, the terms and membership functions of the linguistic variable *security*, upper budget bounds and lower security bounds, the current security level of components, budget constraints, and the function c, which maps a pair of (current) security level and (future) security level on an amount of investment. Empirical work would need to identify these attitudes and assumptions of decision makers. In addition, one could also draw on security standards, such as the Common Criteria [ISO09], to specify under which conditions a component is how secure and to determine terms of the linguistic variable *security*.

A further assumption of our decision model is that the structure of the distributed system is known. This is not always the case; for example, in several anonymizing networks the participating components are determined during process execution.

Due to the NP-hardness of the decision model, solving large instances optimally becomes computationally infeasible. Consequently, further research needs to develop heuristic algorithms.

If the decision maker needs or wants to distinguish between different security requirements, s/he would have to use one (fuzzy) goal function per requirements. The resulting model is a fuzzy multi-criteria problem, which can be solved with methods proposed in the literature (e.g., [Zim96, p. 303ff]).

Despite the aforementioned challenges with regard to the application of the fuzzy decision support model, we argue that a fuzzy set based perspective on security investment situations is a valuable means for practitioners, who need to deal with uncertainty in the absence of (reliable) probabilities.

References

[BE02] James J. Buckley and Esfandiar Eslami. *An Introduction to Fuzzy Logic and Fuzzy Sets*. Advances in Soft Computing. Physica-Verlag, 2002.

[BK05] Robert Blakley and Gregory Kabatiansky. *Encyclopedia of Cryptography and Security*, chapter Secret Sharing Schemes, pages 544–545. Springer, 2005.

[BN08] Rainer Böhme and Thomas Nowey. Economic Security Metrics. In Irene Eusgeld, Felix C. Freiling, and Ralf Reussner, editors, *Dependability Metrics*, volume 4909 of

Lecture Notes in Computer Science, pages 176–187, 2008.

[CN06] Marco Cremonini and Dmitri Nizovtsev. Understanding and Influencing Attackers' Decisions: Implications for Security Investment Strategies. In *Workshop on the Economics of Information Security*, 2006.

[CRY08] H. Cavusoglu, S. Raghunathan, and W. Yue. Decision-theoretic and game-theoretic approaches to IT security investment. *Journal of Management Information Systems*, 25(2):281–304, 2008.

[GCC08] Jens Grossklags, Nicolas Christin, and John Chuang. Security investment (failures) in five economic environments: A comparison of homogeneous and heterogeneous user agent. In *Workshop on the Economics of Information Security*, 2008.

[GJC09] Jens Grossklags, Benjamin Johnson, and Nicolas Christin. The Price of Uncertainty in Security Games. In *Workshop on the Economics of Information Security*, June 2009.

[GL02] Lawrence A. Gordon and Martin P. Loeb. The Economics of Information Security Investment. *ACM Transactions on Information and System Security*, 5(4):438–457, 2002.

[HHB06] C. Derrick Huang, Qing Hu, and Ravi S. Behara. Economics of Information Security Investment in the Case of Simultaneous Attacks. In *Workshop on the Economics of Information Security*, 2006.

[HKS00] T. Hofmeister, M. Krause, and H.U. Simon. Optimal k out of n secret sharing schemes in visual cryptography. *Theoretical Computer Science*, 240:471–485, 2000.

[HN10] Kjell Jorgen Hole and Lars-Helge Netland. Toward Risk Assessment of Large-Impact and Rare Events. *IEEE Security and Privacy*, forthcoming, 2010.

[ISO09] ISO/IEC. Common Criteria for Information Technology Security Evaluation, July 2009.

[KL99] Hans Kleine Büning and Theodor Lettmann. *Propositional Logic: Deduction and Algorithms*. Cambridge University Press, 1999.

[KSST09] Philipp Klempt, Hannes Schmidpeter, Sebastian Sowa, and Lampros Tsinas. Business Oriented Information Security Management A Layered Approach. In *Proceedings of the OTM Confederated International Conferences CoopIS, DOA, ODBASE, GADA, and IS*, volume 4804/2009 of *Lecture Notes in Computer Science*, pages 1835–1852, 2009.

[Lee03] Vincent C.S. Lee. A Fuzzy Multi-criteria Decision Model for Information System Security Investment. In *Proceedings of the Intelligent Data Engineering and Automated Learning*, volume 2690/2003 of *Lecture Notes in Computer Science*, pages 436–441, 2003.

[NIS08] NIST. Performance Measurement Guide for Information Security, May 2008. NIST Special Publication 800-55 Revision 1.

[WCR05] J. Wang, A. Chaudhury, and H.R. Rao. An extreme value approach to information technology security investment. In *Proceedings of the International Conference on Information Systems*, Las Vegas, NV, 2005.

[Zad65] L. Zadeh. Fuzzy sets. *Information Control*, (8):338–353, 1965.

[Zad73] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Memorandum ERLM 411, Berkeley, Calififfornia, 1973.

[Zim96] H.-J. Zimmermann. *Fuzzy set theory - and its applications*. Kluwer, Boston, Dordrecht, London, 3rd edition, 1996.