Traceable Quantitative Risk Assessment Applied to Investment Decision for Local Backups

Steffen Weiss[1,2], Martin Wahl[2], Michael Tielemann[1,2], Klaus Meyer-Wegener[2]
[1] DATEV eG, Paumgartnerstrasse 6-14, 90329
[2] University Erlangen-Nuremberg, Department for Computer Sciences, Institute of Computer Science 6 (Data management), Martensstrasse 3, 91058 Erlangen
E-mail: Steffen.Weiss@informatik.uni-erlangen.de

Abstract

Most organizations have critical data, i.e. customer data or large databases. If these data are lost, organization's existence is in danger. As a consequence, backups for the systems are produced. However, some data are stored only on the computers of employees. These data are usually less critical, but their value cannot be neglected. At DATEV eG in Nuremberg, the question has risen whether it is economically justified to install backup mechanisms for these data. To answer this question, we first present a model to perform an assessment. With the help of this model, we calculate the expected annual cost of repair employees' computers and data including all side effects. Finally the break-even point in cost-efficiency is computed for the example of DATEV eG.

1. Introduction

Data are one of the most important assets of an organization: no industry is able to continue its business if customer data, part lists, etc. are lost. Integrity of these data is very important. Thus, the data are stored on servers and are regularly backed up. For many reasons, some data are also stored on employees' desktop computers. For example, there is not enough space in the networked devices, or people do not want to store confidential data on the server. However, data on local devices also have some value to organizations. Thus, one must think about backup, too. Implementing the backup on the other hand is only useful if it is economically justified—meaning that the expected cost due to additional controls is smaller than the expected costs to reestablish the service as it was before the damage occurred.

DATEV eG is an association of tax-consultants, auditors, and lawyers with approx. 5,400 employees. Recently, the question of backing up employees' desktop computers has risen. An investigation was necessary to answer this question. It was required that the result is traceable—meaning that influences on results are explicitly named and also non-security experts can roughly understand whether the assessment is reliable or not. Moreover it was required that results are well understandable for management and bookkeeping. Thus quantitative assessment was needed.

After giving a short overview of related work (chapter 2), we will present the model we recently developed to solve this problem (chapters 3 and 4). Afterwards the given example of DATEV eG will be discussed (chapter 5). Results are summarized in chapter 6.

2. Related work

The IT Baseline Protection Manual (BPM) [1] contains standard security safeguards, implementation advice, and aids for numerous IT configurations which are typically found in IT systems today. Additionally, it contains threats and modules (typical areas which include IT assets). There is also a module for data security (3.4 Data backup policy) and a procedure to "assess" security and audit the result. However, the BPM is not directed to quantitative assessment of security but rather to technical modeling. In addition, the "auditor's judgment" finally determines whether controls are regarded adequate. Thus a traceable, quantitative assessment of security is not possible.

Similar problems occur with ISO/IEC 27002 (previously published as ISO/IEC 17799 [2]; also as BS 7799-1). The standard contains some hints for backup (10.5 BACK-UP) and there is also an aligning assessment standard—ISO/IEC 27001 [3] which has been developed from and is very similar to BS 7799-2.

Copyright by IEEE. DOI: 10.1109/ARES.2008.108
The original publication is available at http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4529441

This standard defines the fitness of the Information Security Management System (ISMS), that is, the security process of the organization is evaluated. However, assessment is performed indirectly and not quantitatively. More recent standards (e.g. BSI-Standard 100-1 [4]) have the same problem. None of these approaches allows a traceable and quantitative assessment of security.

[5] suggests security assessment by measuring security indicators. Examples mentioned are "mean time to recover," "elapsed time since last disaster recovery walk-through (days)," etc. NIST SP 800-55 [6] is a similar approach, but it does not say anything about data security. Although the indicators are quantitative, they do not help to quantify the (expected) loss. Thus, usefulness of these indicators is limited. Consequently, a traceable quantitative statement about an organization's security is not possible with these approaches.

An important concept for quantitative assessment in commercial information technology is ALE (annual loss expectance; see e.g. [7]). It gives a first, intuitive understanding of security evaluation. It is based on data that characterize the security risks of an organization. Yet there is a "data crisis," (compare e.g. [7], [8]) meaning that there are not enough or no adequate data for the input parameters. Consequently, data can only be obtained as a guess and results are not traceable.

Another important concept in (quantitatively) measuring security is (Information Security) Risk Assessment. Examples are NIST 800-30 [9] and Mehari [10]. Input values (threat-likelihood and threat-impact) as well as output (risk-severity) are assessed on scales with just a small number of units (typically three or five). Fixed calculation tables are used to calculate the results. Even quantitative risk assessment approaches (compare e.g. the IBM approach in [11]) align with this principle. [8] presents the idea of a process-based risk assessment, but details on how to proceed are missing. All together, input and output are not detailed enough for risk assessment. Moreover, the reasons why a specific scale value is used are very fuzzy. Thus, results are not traceable.

The presentation of related work has shown that the number of approaches dealing with quantitative measurement is limited. In neither case of quantitative assessment, traceability is given. Thus, a new approach is needed, allowing quantitative assessment in a traceable manner.

3. Definition of terms

We have recently developed a UML class model for the collection of data needed in traceable quantitative assessment of security. Before presenting the model, six main terms regarding the model shall be defined:

1. Attacks: An attack is a deliberate act with the potential to harm a system or organization.

2. Incidents: An incident is a non-deliberate act with the potential to harm a system or organization.

3. Controls: According to [2], controls are means of managing risk—including policies, procedures, guidelines, practices or organizational structures—which can be of administrative, technical, management, or legal nature.

4. Damage: A damage is a reduction of the value for the owning organization. Damage can occur on physical assets (like computers or mobile phones), but also on non-physical ones like data (e.g. patents, or part lists). Moreover, also repair has to be taken in consideration. Thus, damages are measured as the cost to reestablish the service as it was before the damage occurred.

5. Scenarios: In our opinion a scenario is a successful incident or attack, meaning a damage can be assigned. An example for a scenario is "hard disk crashed."

Due to statistical understanding of security it is necessary to mention that terms are understood from an ex-ante point of view. So, e.g. attacks are potential deliberate acts to harm a system or organization.

4. Model for damages

4.1. Simplification of the model

The UML class model we developed covers more aspects than necessary for this particular problem. For example, modeling occurrence of incidents is not needed because its rate is known quite well in this case (compare section 5.5.). Moreover, we have one concrete example and thus we have not separated data which should be provided centrally (as it is the same for every organization) from data which must be provided individually (e.g. existence or non-existence of controls). Due to space limitation we omit these aspects and concentrate to the necessary parts, starting with modeling the *Damage* entity.

4.2. Overview of damages

Before modeling damage, it is important to further clarify the meaning of "damage."

Describing damages in great detail like “one file damaged” or “three files damaged” leads to a large amount of similar damages in the model. Thus building categories of damages and describing things in an abstract way is more appropriate—for example “some important data stored on a computer lost” or “all data stored on a computer lost“. Hence, damage is not an exact description of an adverse situation, but a rather abstract description of a “typical” situation.

Very abstract descriptions (e.g. “data lost”) do not help either, as costs of damages can only be guesses. Thus, we tried to reach some degree of detail. In the end, the degree of detail must be defined by modeling experts. They should take into account the following criteria:

- There can be **huge differences** in damage: A good example is whether important data are lost in a hard-disk crash or just not so important data.
- There can be **controls** influencing the damage: When talking about loss of data stored on a hard disk, existence of a backup determines whether there is large damage (e.g. all data have to be re-entered) or less damage (e.g. some hours of unavailability and a loss of the changes done after the backup).

Thus damage is rather abstract, while the level of detail is defined by modeling experts.

The model for damages will be described step-wise. First, the damage entity is described in detail (see section 4.3.).

4.3. The damage entity in detail

Most important entity to model is the *Damage* entity. It is a generalization of *ElementaryDamage* and *ComposedDamage*. This separation is necessary because they contain different sets of attributes. The *ComposedDamage*—being necessary to model coarse-grained damages—does not contain any attributes. In contrast, the *ElementaryDamage*—needed to model detailed damages—has an attribute *costTotal*. This attribute is an input parameter. The sum of all costs needed to recover from the damage (e.g. costs to restore) is assigned to this attribute. Note that assessment of *costTotal* is not part of the model any more, it should be provided e.g. by security experts based on existing data and statistics. The unit of *costTotal* is a currency (e.g. Euro (€)).For completion of the damage entity, two aspects are missing: First, the generalization is complete and disjoint. Second, the *Damage* entity has a virtual function *costTotal()*, being overloaded by *costTotal()* of *ElementaryDamage* and *ComposedDamage*.

4.4. Modeling influence of controls

In section 4.2. differences between damages have already been discussed: There are big differences between damages and controls influencing the damages. These differences already imply a first structure of modeling the damages. Regarding the logical process of an attack or incident causing damage, the first things to be modeled are the huge differences between damages. In a second step, damages are turned into more detailed damages, depending on the controls installed or—to be more precise—the states of these controls.

Thus, from the model point of view, there are two elements: First, a relationship between a scenario and a damage making it possible that there are many damages for one scenario. Second, a relationship allowing to turn a coarse-grained damage into detailed damages depending on state of controls. The first aspect will be discussed later on (see section 4.7.).

The second aspect is modeled as a 3-way relationship called *Protection*, e.g. “important data lost“ can be transformed into “no data lost but restore needed” if the control “backup” is in the state “backup up-to-date”. The first endpoint of this relation is the *StatusOfControl* entity. The two other endpoints are damages, one taking part as coarse-grained damage and one taking part as detailed damage. The detailed damage can be either an *ElementaryDamage* or a *ComposedDamage*. The coarse-grained damage end of the relationship is always a *ComposedDamage*. Thus, the relationship must be drawn between *StatusOfControl*, *Damage*, and *ComposedDamage*.
The multiplicities of this relationship must be always 0..1 (for exact meaning of multiplicities in 3-way relationships see e.g. [13]).

4.5. Modeling protection against damages

One important aspect is not yet modeled: different states of controls occur with different frequency. For example, a backup is in status “backup up-to-date” in approx. 80% of the cases. Thus an additional attribute *percentageProtection* is introduced for entity *StatusOfControl*.

A question occurring is whether sum of all *percentage-Protection*s must add up to 1 for one coarse-grained damage. We think that this is usually the case, but there are cases, where the damage has such a small cost that it needs not to be modeled. Thus, we did not implement a restriction in the model.

We also investigated whether decomposition of controls is necessary. However, it has shown that different states already imply some decomposition and thus it does not make sense to model decomposition for *ControlDamage* entities.

Some details of the relationship between *ControlDamage* and *StatusOfControl* have to be discussed, too. A status of control belongs exclusively to one control and depends on it. Thus, this relationship is a composition with multiplicity 1 at the rhomb end. A *ControlDamage* has at least two states (working / non working), thus the multiplicity at the non-rhomb end is 2..*.

4.6. Restrictions on protection against damages

The general structure for the protection against damages is ready now. However, the model is not very restrictive: for example one can model a *StatusOfControl* “firewall not active” and a *StatusOfControl* “backup working” for the same *ComposedDamage*. It is intuitively evident that this does not make sense.

Thus further restriction is necessary: only states of control for one *ControlDamage* may be related to the same *ComposedDamage*. This can be achieved by introducing a relationship between *ComposedDamage* and *ControlDamage*.

Investigating multiplicities of this relationship, there is one or more *ComposedDamage* for a *ControlDamage* as it does not make sense to model the *ControlDamage* otherwise, and *ControlDamage*s can be reused for more *ComposedDamages*. Moreover, there is exactly one *ControlDamage* for a *ComposedDamage* as it does not make sense to model a *ComposedDamage* otherwise. Thus it is a 1..*/ 1 relationship.

One might expect that this is a hard restriction on the model because only one control can be modeled. However this is wrong: if there is more than one control, the second control can be modeled as control of the detailed damage and so on. Thus, it is not a restriction in the potential of the model but only a way to clarify the structure of the model.

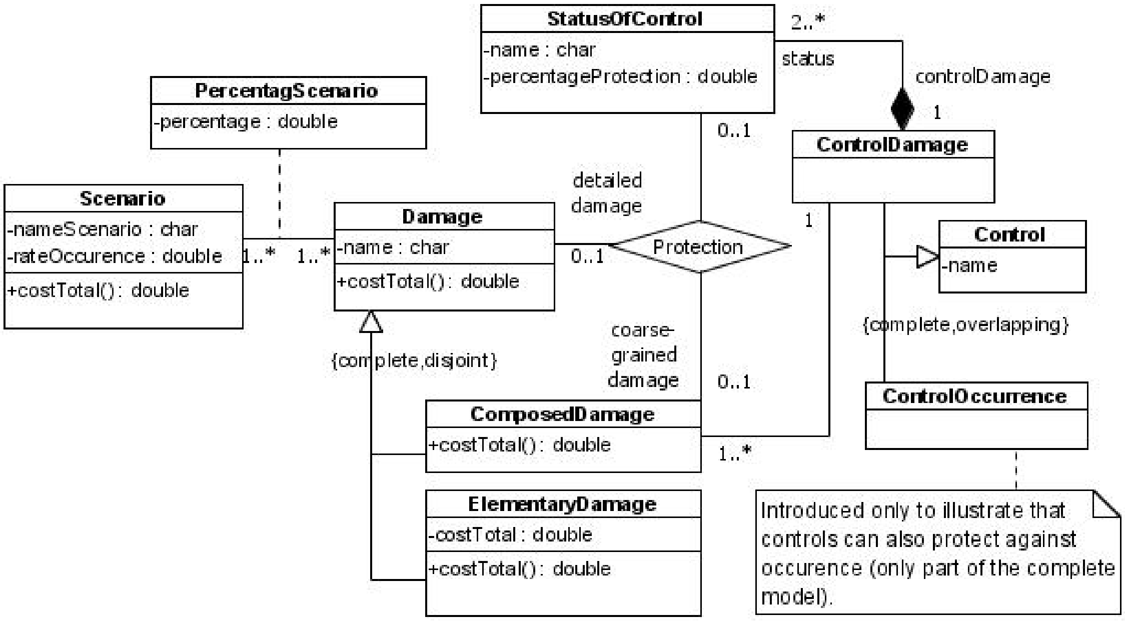

Fig. 1. UML class model as described during chapter 4

Probably in most cases, all states of control for one *Control* must take part in the *Protection* relationship. However, we are not sure whether this is always the case. Thus, this restriction was not implemented in the model.

4.7. Relationship to the scenario

The model for the *Damage*s is described in detail now, thus the relationship to the *Scenario* can be regarded. Before, the *Scenario* entity shall be described in short. As defined above (see chapter 3), a scenario is a successful incident or attack leading to some damage. Thus, it contains a statement about the damage, in this case a function *costTotal()*. Besides, it contains an attribute *rateOccurrence*, indicating how often the scenario occurs. (As already mentioned in section 4.1, this is a simplification of our model, but we do not need anything else to solve the particular problem of DATEV eG).

There is at least one *Damage* for a *Scenario* however there can be arbitrary many. There is at least one *Scenario* for a *Damage*, otherwise it does not make sense to model the damage. Moreover, there can be more than one *Scenario* for a *Damage*. An example for this is “loss of data” which can be caused by a virus as well as a defect of the hard disk besides others. As a result the relationship between *Scenario* and *Damage* is 1..* / 1..*. Some damages occur more often than others. To model this fact, the relationship *Scenario* / *Damage* must be attributed. The appropriate attribute is *Percentage* (in an attribute entity *PercentageScenario)*, indicating the percentage of scenario occurrence in which the according damage is likely to occur.

The complete UML model we developed can be seen in Fig. 1.

5. The particular example

In chapter 4 we described the model developed. In this chapter we want to explain how we proceeded to carry out the assessment. The results are subsumed in a UML object diagram (compare Fig. 2) which is an instance of the UML class diagram explained in chapter 4.

5.1. Backup methods and groups of persons

During the first discussions of this problem, it became obvious that there are different groups of persons. These different groups have different usage profiles and thus also cost and effect of data backup is different. We finally decided to include two groups in assessment: “important persons” (that is people of the board; 28 persons) and “internal employees” (approx. 5,400 persons with approx. 5,500 computers). The third theoretically possible group of “field service” will not be examined because details about this group were not available and effort of investigation should not be spent.

TABLE 1
ALTERNATIVES FOR ASSESSMENT

Alt.	Group	Description
(1)	Internal employees	No backup
(2)	Internal employees	Manual local backup to source drive; once a week
(3)	Internal employees	Manual local backup to other drive (external hard drive); once a week
(4)	Internal employees	Automatic local backup to second hard drive; once a day
(5)	Internal employees	Automatic network backup; once a day
(6)	Internal employees	Automatic copy to network drive (synchronisation); once a day
(7)	Internal employees	No additional backup: work done on network drive, automatic backup of network drive; once a day
(8)	Important persons	No backup
(9)	Important persons	Automatic local backup to second hard drive; once a day
(10)	Important persons	Automatic network backup; once a week
(11)	Important persons	Automatic copy to a network drive (synchronisation); once per week

The variety of backup mechanisms is very large. To limit the assessment effort, we restricted the number of alternatives to some useful alternatives for each group.

We finally identified a set of eleven alternatives—seven for internal employees and four for important persons (compare table 1). In each group, one alternative is “no backup.” This alternative is necessary as reference for installing controls.

5.2. Costs

Like every organization, DATEV wants to choose the economically best solution. This means that the sum of all (expected) costs should be minimized. All together, there are the following costs:

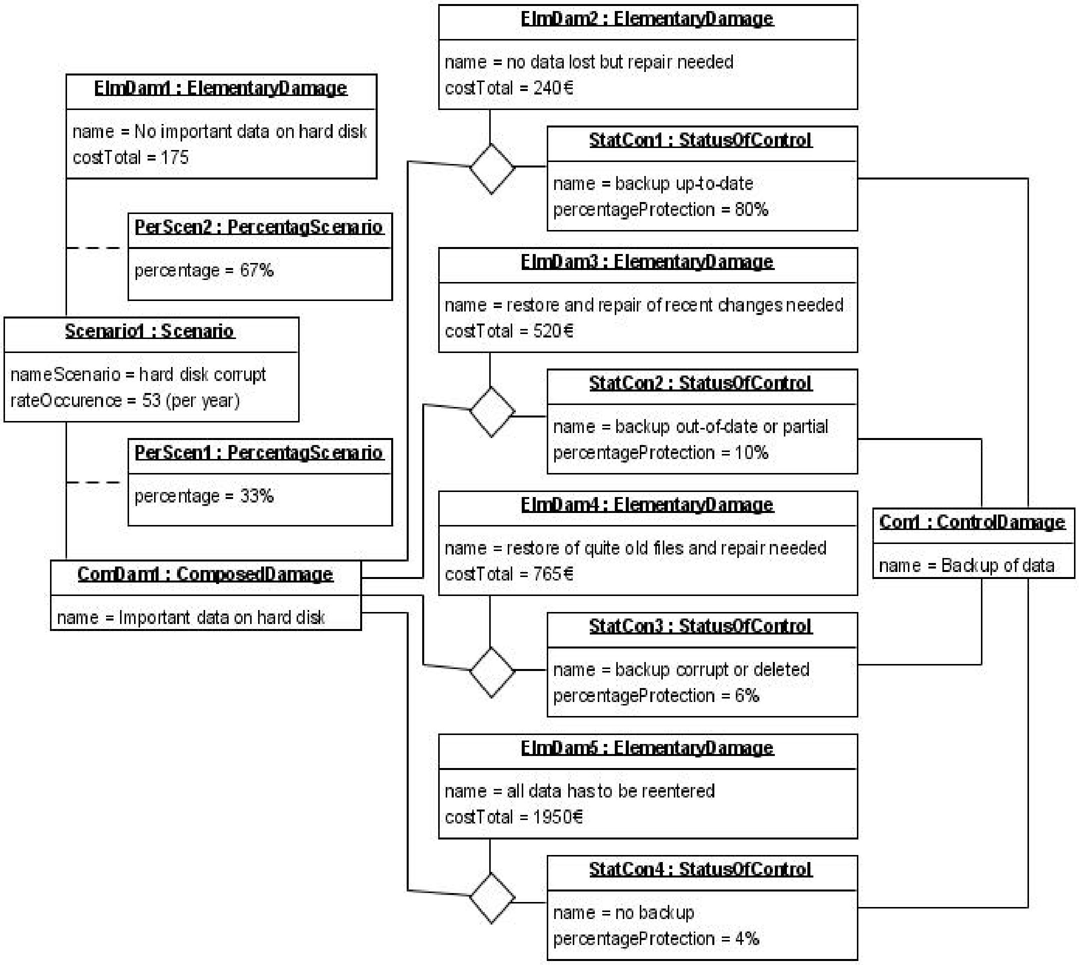
ElmDam2 : ElementaryDamage
name = no data lost but repair needed
costTotal = 240€
ElmDam1 : ElementaryDamage
name = No important data on hard disk
costTotal = 175
StatCon1 : StatusOfControl
name = backup up-to-date
percentageProtection = 80%
PerScen2 : PercentagScenario
percentage = 67%
ElmDam3 : ElementaryDamage
name = restore and repair of recent changes needed
costTotal = 520€
Scenario1 : Scenario
nameScenario = hard disk corrupt
rateOccurence = 53 (per year)
StatCon2 : StatusOfControl
name = backup out-of-date or partial
percentageProtection = 10%
PerScen1 : PercentagScenario
percentage = 33%
ElmDam4 : ElementaryDamage
name = restore of quite old files and repair needed
costTotal = 765€
Con1 : ControlDamage
name = Backup of data
ComDam1 : ComposedDamage
name = Important data on hard disk
StatCon3 : StatusOfControl
name = backup corrupt or deleted
percentageProtection = 6%
ElmDam5 : ElementaryDamage
name = all data has to be reentered
costTotal = 1950€
StatCon4 : StatusOfControl
name = no backup
percentageProtection = 4%

Fig. 2. UML object diagram for the example described in 5.4. to 5.6.

- one-time costs (implementation)
- annual costs of maintenance
- annual costs of damage

The costs of implementation for a specific alternative are easy to calculate. Choosing alternative (4) (Automatic local backup to second hard drive, once a day, and for internal employees) they consist of costs for

- testing the system to be installed (three working-days, 1,680€)
- training of hotline, and producing an information sheet (740€)
- reading of the information sheet and selection of data to be saved (duration: half an hour, cost: 35€ per computer, 5,500 computers)

All together, costs of 670,480€ have been identified in this case.

The annual costs of maintenance are easily collected, too. Sticking to alternative (4), this value adds up to 28,350€ and includes for example costs for software maintenance, hotline costs, energy costs for additional hard disks, and costs for replacement of hard disks.

Besides other values, table III contains the costs for implementation and the annual costs of maintenance for the eleven alternatives selected.

Thus, one-time costs and annual costs of maintenance are available; however annual costs of damage are missing. Calculation of these values is described in the next sections.

5.3. Scenarios

The first aspect in assessment of damage is definition of what to assess. After long discussions, we defined the following set of scenarios:

- (1) hard disk damaged
- (2) file with important data deleted or out of order
- (3) important file temporarily lost (but copy available)
- (4) mobile storage device lost
- (5) notebook lost
- (6) mobile backup device lost
- (7) natural incidents (fire, flood, etc.)

Besides others, we used the German Baseline Protection Manual [1] to define these scenarios. On basis of this extremely detailed document and according to our best knowledge, these scenarios cover all effects which might occur to the organization.

Each of these scenarios had to be assessed under each alternative. As this procedure is quite laborious, it is necessary to define information which is common to all alternatives and which has to be assessed for each scenario individually.

After short time, it became obvious that the "structure" of a scenario—that is e.g. which states of controls exist for a given damage—is the same for each alternative. In contrast, the (statistical) values used as input for the model vary (at least in most cases).

5.4. Building the structure of the scenarios

First of all, the structure of a scenario (common to all alternatives) had to be built. We will explain this for the example of scenario 1 (hard disk crashed).

If a hard disk crashed, there are principally two different situations: First, data are (temporarily) lost but no important data (in means of integrity) are on the disk. This situation is rather uncritical because damage is nearly exclusively limited to replacement of the hard disk and thus damage is small. Second, (temporarily) lost data are important data. This damage is an abstract description of a situation that has to be regarded in more detail. Depending on the controls installed, there are probably different costs to reestablish all services as it was before the damage occurred. Based on the different states of control "backup", we identified the following set of more detailed damages:

- status "backup up-to-date" implies a damage "no data lost but restore needed"
- status "backup out-of-date or partial" implies a damage "restore and repair of recent changes needed"
- "last backup corrupt or deleted" implies a damage "restore of quite old files and repair needed"
- "no backup" implies a damage "all data has to be reentered"

During our work we turned these damages into more detailed ones. Due to space limitation of this article and to preserve readability of the resulting UML object diagram, we limited the presentation to these damages.

5.5. Input values for calculation

Based on the structure of the scenario, (statistical) input values can be given. As already mentioned these data are (at least in most cases) only valid for one alternative. We again stick to alternative (4), scenario 1. Moreover we regard the particular situation of DATEV eG. Based on these conditions, we obtained the following values:

- Much data that are important for the organization are already stored on servers—e.g. project data. Based on the experience of the administrators of DATEV eG, only about one third (33%) of desktop computers contain important data (important with respect to integrity).
- Due to statistical experiences of the past with network backup, 80% of recoveries are assumed to be successful. This aligns with the data available in the literature (e.g. [15] refers to a source mentioning a 83% recovery rate).
- Also due to statistical experiences of the past with network backup, files are out-of-date or only partially available in 10% of the cases where data need to be restored.
- The remaining 10% have to be divided into "backup corrupt or deleted" and cases where no backups are available. Due to experiences, "backup corrupt or deleted" occurs slightly less frequent than "no backup available." Thus we assigned 4% to "backup corrupt or deleted" and 6% to "no backup."
- Each damage of a hard disk is reported to the central administration center. Thus, there is a very exact statistic of occurrence of hard-disks crashes. For the last year, we got a value of 53 (that is: 53/a). This value can be taken as basis for calculation.

5.6. Assessment of elementary damages

Besides the statistical values described in section 5.5., the detailed elementary damages have to be assessed. As an example, assessment of *ElementaryDamage* object "no data lost but restore

TABLE 2
ANNUAL COST OF DAMAGE FOR SCENARIOS, PROVIDED FOR SELECTED ALTERNATIVES

Scenario	Alternative (1)	Alternative (2)	Alternative (3)	Alternative (4)
(1)	20,394€	20,988€	15,293€	13,067€
(2)	78,400€	30,037€	29,596€	10,388€
(3)	568€	568€	568€	568€
(4)	2,994€	2,994€	2,251€	1,957€
(5)	10,127€	10,127€	9,652€	9,463€
(6)	340€	340€	340€	340€
(7)	372€	375€	354€	346€
(Total) annual cost of damage	113,200€	65,430€	58,050€	36,130€

needed" is explained.

First of all, the hard disk has to be replaced by a new one. Thus, there are hardware costs and costs to exchange the crashed hard disk by the new one. These costs were calculated as 170€. The specific situation described here implies that an up-to-date backup exists. Thus, data can be restored. However, some time is needed for this. Based on our knowledge, it takes approximately one hour until the system can be adequately used again. Thus, costs to restore integrity of data are 70€ (costs for one working hour). Thus, the total costs are 240€ and attribute *costTotal* of *ElementaryDamage* object "no data lost but restore needed" is assigned this value.

An analogous proceeding was also chosen for assessment of the other elementary damages.

5.7. Example of calculation

With the help of our model, we calculate the expectation of a damage. In some cases, a distribution can be more useful. For these cases, we refer to [12], telling something about useful distributions in similar situations. Going into details of calculation, the expected costs for a composed damage are calculated from the costs of elementary damages (attribute *costTotal* of *ElementaryDamage*), and the percentage of protection (*percentageProtection* of *StatusOfControl*).

Sticking to the example given above, the expected cost for the *ComposedDamage* "important data (temporarily) lost" is calculated from the following values:

- "no data lost but restore needed" occurs with a probability of 80% (see 5.5.) and causes a total damage of 240€ (see 5.6.)
- "restore and repair of recent changes needed" occurs with probability 10% and causes a damage of 520€.
- "last backup corrupt or deleted" occurs with a probability of 4% and causes a damage of 765€.
- "no backup" occurs with a probability of 6% and causes a damage of 1,950€.

With the help of the *costTotal()* function of *ComposedDamage* "important data (temporarily) lost," the expected annual cost of damage of this object is calculated. We receive:

240€•0.8 + 520€•0.1 + 765€•0.04 + 1,950€•0.06 ~ 392€

A similar calculation (based on the *percentage* attribute of *PercentageScenario* and the result of *costTotal()* of *Damage*) can be performed for the scenario object. The resulting value is the expected damage of one incident. Multiplied with the expected number of incidents per year (that is attribute *rateOccurrence* of *Scenario*) the annual cost of damage for scenario 1 and alternative (4) is provided.

This value is one entry in table 2 (first row, rightmost column). The other six scenarios have to be assessed under the circumstances of alternative (4), too. Calculating is performed analogous to assessment of scenario 1, and results are written down in the according rows of the rightmost column of table 2.

Finally, the seven values in the rightmost column of table 2 are summed up, being the total annual cost of damages for alternative (4). For further illustration of this procedure, we included also other alternatives in table 2.

These sums are carried over to table 3, column "Annual cost of damage."

TABLE 3
EXAMPLES FOR RESULTS OF SCENARIO ASSESSMENT

Backup alternative	Cost of implementation	Annual cost of maintenance	Annual cost of damage	#Years to break-even
(1)	-	-	113,200€	-
(2)	195,480€	23,740€	65,430€	10.1
(3)	621,710€	182,970€	58,050€	Never
(4)	670,480€	119,990€	36,130€	15.8
(5)	195,480€	28,350€	36,130€	65.7
(6)	198,280€	74,000€	42,310€	Never
(7)	102,030€	74,000€	38,590€	169.3
(8)	-	-	4,850€	-
(9)	4,550€	590€	2,030€	4.0
(10)	490€	290€	2,530€	0.5
(11)	490€	290€	2,700€	0.5

5.8. Calculation of break-even

In section 5.2., costs of implementation as well as annual costs of maintenance were calculated. In section 5.7., costs of damages were calculated. Thus, all important costs are available now.

Only one aspect is still missing: calculation of the time passing until a backup method is cost-efficient—the so-called break-even point.

It was decided that the backup is not established at a single point of time on all computers. When old computers are replaced with new ones, the backup will be installed on the new machines. In DATEV eG this cycle for replacement is $a = 4$ years. We assumed that the replacement is linear over time, until all computers have been replaced. Therefore, the costs of implementation are linearly distributed over the replacement cycle.

We have developed two formulas: The first formula is used to calculate the break-even point if it is reached before the replacement is finished. The second formula is used if the break-even point is reached after completion of replacement. Therefore, only one of the two formulas can give a correct result lying in its domain.

The following formulas were developed:

$$t_1 = \frac{2e_i}{k - j_i} \quad (\mathrm{dom}(t_1) = [0;a])$$

$$t_2 = \frac{e_i}{k - j_i} + \frac{a}{2} \quad (\mathrm{dom}(t_2) =]a;\infty[\ )$$

while

a = time of replacement in years

e_i = one-time costs with selected backup alternative i

j_i = annual costs with selected backup alternative i (sum of annual cost of maintenance and annual cost of damage for alternative i)

k = annual costs without backup (alternative (1) for "internal employees" and alternative (8) for "important people")

Due to space limitation we refer to [14] for derivation of these formulas. The rightmost column of table 3 shows the results of the assessment for all alternatives.

6. Results

6.1. Results for DATEV eG

Regarding the resulting time of break-even in table 3, it is obvious that it exceeds 10 years for all alternatives of the group "internal employees." According to the experiences of responsible people of DATEV eG, a break-even time of approximately 5 years is adequate in this case. Everything above is very doubtful. Thus, it does not make sense to implement an alternative in this case. With a break-even-time of half a year alternatives (10) (automatic network backup) and (11) (automatic copy to a network device (synchronization)) provide good results for the group "important persons". Thus, one of these two alternatives should be implemented.

6.2. Benefit of the model

Besides calculation of the results for DATEV eG, applicability of the model was of interest because it was the first "in-the-field test" of our model. Most important, it was possible to model the real-world example we were faced with.

Moreover, the model helped much to structure our problem. If we were not able to use the model, it would have been unknown which input was needed and how to correlate it. The only solution would have been hiring a data security expert and asking him for a solution of the problem. Probably, he would have provided a solution; however reliability of the results would probably be dependent on the concrete person.

Even if a data security expert has more experiences with data backup, he would have problems with getting adequate values, too. All existing approaches (e.g. [7]) suffer from this problem. However, our model solves this problem. It allows calculation of the results, based on very detailed situations. Thus there is much traceability.

If we have to perform similar assessments in the future, it would be useful to reuse information we have already gathered during this assessment. The full version of our model contains the possibility to reuse "structural information" like scenarios and values for input (for details see [16]), and thus provides already a first step.

In addition we think that checklists would be useful. For example, they should contain keywords like "costs to replace the crashed disk" and "costs to restore and repair data". These checklists would help to not missing important aspects during assessment.

7. Conclusion

Motivation for the work we presented in this article was a problem of DATEV eG: it should be decided whether a backup for local hardware devices is economically efficient or not. Main requirement was that results are traceable and quantitative. Already at the beginning of the project, it was evident that the main problem is assessment of expected annual costs to repair employees' computers and data including all side effects. Approaches known in literature do not help to answer a question like this in a traceable and quantitative way; however a model we recently developed does. Thus, this model was used for the assessment of expected annual cost of damage. Due to space limitation, we limited presentation to those parts of the model which are relevant for the concrete problem.

Assessment for the concrete problem was carried out on basis of this model. We gave an example how the input values are assessed and how annual cost of damage is calculated. Besides, implementation and maintenance costs were calculated. Finally, the break-even point of different alternatives was calculated. For this, two different groups of employees of DATEV eG ("internal employees" and "important persons," that is people belonging to the board) were regarded. It became obvious that a backup of local hard drives is not economically efficient for the group "internal employees. However, there were two economically efficient alternatives for group "important persons."

Regarding evaluation of our model, we have seen that our model is applicable for real-world problems. The results are traceable and quantitative statements about costs due to damages are possible. Thus, a problem of all existing approaches (e.g. [7]) can be solved.

For reduction of assessment work in similar projects, reuse of information already gathered is necessary. We have already developed a model for this [16] but we think that this concept should be extended, e.g. by a checklist.

10. References

[1] Bundesamt für Sicherheit in der Informationstechnik, "IT Grundschutzhandbuch," Bonn, 2005.

[2] ISO/IEC, "Information technology – Security techniques – Code of practice for information security management (ISO/IEC 17799, final draft)," 2005.

[3] ISO/IEC, "Information security management system – Requirements (ISO/IEC 27001:2005)," 2005.

[4] Bundesamt für Sicherheit in der Informationstechnik (publisher), "BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS) Version 1.0," Bonn, 2005.

[5] A. Jaquith, Security Metrics Replacing Fear, Uncertainty, and Doubt, Addison-Wesley, 2007.

[6] M. Swanson, N. Bartol, J. Sabato, J. Hash, L. Graffo, "Security Metrics Guide for Information Technology Systems (NIST SP 800-55)," Washington, 2003.

[7] K. J. Soo Hoo, "How Much Is Enough? A Risk-Management Approach to Computer Security (Working Paper)," Consortium for Research on Information Security Policy (CRISP), Stanford University, June 2000.

[8] T. Nowey, H. Federrath, C. Klein, K. Ploessl, "Ansätze zur Evaluierung von Sicherheitsinvestitionen, " (Approaches to evaluation of investment in security.) In: H. Federrath (ed.), Sicherheit 2005, Beiträge der 2. Jahrestagung des GI-Fachbereichs Sicherheit, Lecture Notes in Informatics (P-62), Bonn: Köllen-Verlag, 2005, p. 15-26.

[9] G. Stoneburner, A. Goguen, and A. Feringa, "Risk Management Guide for Information Technology Systems (NIST SP 800-30)," Washington, 2002.

[10] Club de la securite des systemes d'information francais (Clusif) (publisher), "Mehari V3 Concepts and Mechanisms," 2004.

[11] K. N. Bhaskar, Computer security threats and countermeasures, NCC Blackwater, Manchester a.o., 1993.
S. Weiss, "Metric for IT security in an organization," Diploma Thesis, University of Erlangen-Nuremberg, Department of Computer Science, Computer Networks Group, 2005.

[12] S. Weiss, "Metric for IT security in an organization," Diploma Thesis, University of Erlangen-Nuremberg, Department of Computer Science, Computer Networks Group, 2005.

[13] M. Hitz, G. Kappel, E. Kapsammer, W. Retschitzegger, UML@work, Heidelberg: dpunkt-Verlag, 2005.

[14] M. Wahl, "Auswahl von Datensicherungskonzepten für Windows-Clients." (Selection of Data Security Concepts for Windows Clients.) Diploma Thesis, University of Erlangen-Nuremberg, Department of Computer Science, Data Management Group, 2007.

[15] D. M. Smith "The cost of lost data," Pepperdine University, http://gbr.pepperdine.edu/033/dataloss.html, 2007.

[16] S. Weiss, K. Meyer-Wegener, "Towards solving the data problem in measurement of organizations' security," University of Erlangen-Nuremberg, Department of Computer Science, Data Management Group, Nov. 2007, to be published at "Sicherheit 2008".