Information Security Risk Assessment, Aggregation, and Mitigation

Arjen Lenstra[1] and Tim Voss[2]

[1] Citigroup, Information Security Services
Technische Universiteit Eindhoven
1 North Gate Road, Mendham, NJ 07945-3104, USA
arjen.lenstra@citigroup.com
[2] Citigroup, Information Security Services
750 Washington Boulevard, 7th floor, Stamford, CT 06902, USA
tim.voss@citigroup.com

Abstract. As part of their compliance process with the Basel 2 operational risk management requirements, banks must define how they deal with information security risk management. In this paper we describe work in progress on a new quantitative model to assess and aggregate information security risks that is currently under development for deployment. We show how to find a risk mitigation strategy that is optimal with respect to the model used and the available budget.

Keywords: Risk management, risk assessment, risk aggregation, risk mitigation, Basel 2, multiple-choice knapsack problem

1 Introduction

Given the constantly changing vulnerabilities that may threaten the security of a company's data, how does the company decide where to spend its Information Security (IS) budget to limit as much as possible the damaging consequences of attacks? Traditionally, this decision-making process is mostly left to 'experienced' staff whose judgment, intuition, and taste is relied upon.

With the upcoming required compliance with the Basel 2 agreements, a stricter approach has to be adopted: banks must be able to quantify their operational risk, devise proper ways to contain it, and reserve an adequate budget to absorb potential damages. Since securing its data assets is part of a bank's operations, this applies to IS risk management as well.

In this paper we discuss the issues that need to be addressed to fulfil the IS risk management requirements and describe work in progress on the solution that is currently under development for deployment. It is generally recognized that subjectivity is inherent to any risk assessment methodology. Thus, there is no single 'a priori' correct way to approach this problem: the choice of a model and its various parameters can never be fully objective. An unrelated requirement is that the proposed solution must allow a user-friendly interface: our extremely varied world-wide user community of local security officers and

H. Wang et al. (Eds.): ACISP 2004, LNCS 3108, pp. 391–401, 2004.
© Springer-Verlag Berlin Heidelberg 2004

business managers must be able and willing to use it and have confidence in the outcomes.

The formalization provided by our prototype solution and the underlying model choices, combined with its easy-to-use automated reporting mechanism, takes away some of the subjectivity of the IS risk management process. The resulting quantitative approach is consistent on a company-wide basis, leads to results that are optimal with respect to the model's quantification parameters, and meets the needs and expectations of local and global security and business staff. The work described in the present paper represents work in progress and our prototype may not be ready to meet all requirements for compliance with the new regulations. For instance, a conscious decision was made not to include temporal dependencies in the prototype model, in order to make it easier to access and understand by uninitiated users. At a later stage, and depending on our findings with the present approach, refinements may be proposed and implemented (cf. Remark 4.4).

The IS risk management model proposed in this paper is very similar to the familiar *Annual Loss Expectancy* (ALE) approach to risk management [3]. What is novel in our paper is that we argue why an ALE-like approach is the only alternative for our type of application. Another innovative aspect of our paper is the way we derive the optimal risk mitigation strategy.

In Section 2 the IS risk management problem that is addressed in this paper is described in more detail. Various approaches to risk assessment and aggregation are reviewed in Section 3 along with their pros and cons with respect to our IS risk management application. This leads, in Section 4, to our quantitative IS risk management model. In Section 5 it is shown how the IS risk can be optimally contained in our model by solving a multiple-choice knapsack problem.

2 IS Risk Management

Our goal is to quantify the corporation's total Information Security risk and to find the most cost-effective way to contain the risk. We reiterate the comment made above that risk management is subjective. Different subjective choices made in the design or parameters of the model may lead to different risk containment strategies that may all be optimal with respect to their respective models. The best one can aim for is consistency within the model, overall soundness of the model, and an on average high level of user acceptance and appreciation of the results.

More in detail, from an IS perspective the situation is as follows. The corporation relies on a number of business processes. Each business process is exposed to a certain *current IS risk*. As a consequence, the corporation is exposed to the combined current IS risks of its business processes: the *current aggregated IS risk*. Each business process uses a number of applications, where a single application may be used by more than one process. For each application any number of IS vulnerabilities may be identified, and for each such vulnerability any number of IS threats may exist that realize that vulnerability. The IS threats

are responsible for the corporation's aggregated IS risk. For each threat there may be any number of action plans of varying costs and degrees of effectiveness to counter the threat. Realization of an action plan against a threat mitigates the current IS risks of all processes using the application that was affected by that threat to their *residual IS risks*. The question is how, given a certain fixed budget, the action plans should be selected in such a way that the combination of all mitigated risks, the *residual aggregated IS risk*, is minimized.

Identification of vulnerabilities, threats, and action plans is done on a per application basis by the process owners, all relevant subject matter experts, and if possible representatives from the audit, risk, and review department. Because different processes may share applications, these application data may already have been provided by another business. It is the responsibility of the various businesses to coordinate and consolidate their views on their shared applications (cf. Remark 4.3).

A vulnerability would be, for instance, unencrypted customer data sent over a public network. One of the threats affecting that vulnerability would be an eavesdropper on the network, and each type of encryption of the data (symmetric using a system-wide shared key or using a customer-specific shared key, asymmetric, etc.) corresponds to an action plan countering the threat.

An important restriction on the solution to the above problem is that it has to work in a highly non-uniform environment. Central management is required (to be able to share vulnerability and threat data about shared applications), but data about business processes and applications must be entered by the businesses in many different countries on almost all continents. Inevitably the data will be colored by different regulatory and cultural influences. It is probably impossible to design a solution that is totally oblivious of all such effects and that works irrespective of the level of expertise or commitment of the staff that enters the data. But the robustness of the solution is strongly supported by making it easy to teach, understand, and use. We believe that the latter is a *conditio sine qua non* for an effective IS risk management solution for any even moderately large company. First, however, we have to discuss what we mean by risk.

3 Risk Assessment and Aggregation

The two most common approaches to risk assessment are qualitative risk analysis and quantitative risk analysis. The qualitative approach identifies events affecting a process (cf. *threats*) and a variety of corresponding controls that may mitigate the effects of the events (cf. *action plans*). Based on the perceived relational model between events and controls the risks are assessed and a strategy is decided upon, where it is often helpful to associate subjective qualitative rankings (such as High, Medium, Low) to the severity of the events or the effectiveness of the controls. As a result, the qualitative approach is mostly intuitive. It has the advantage that no probability data of past events are needed and that it leads to a reasonable decision model more or less built from scratch.

A qualitative approach may be applicable in simple situations where a vague indication suffices. For our IS risk management application it is not suitable. Its lack of precision makes consistency unachievable and, more importantly, the coarse-grained categorization makes aggregation virtually impossible and meaningless: what overall ranking should one assign the aggregate of four events with rankings High, Medium, Medium, and Low? And is it better or worse than all Medium? As a consequence, selecting an adequate set of action plans remains guesswork without any claims of mathematical rigor or soundness.

The quantitative approach employs as much as possible the distribution functions underlying the events and defines risk as a certain function of the distribution function. For instance, the distribution functions may be used to calculate the *Annual Loss Expectancy* for a single event, which can then be defined as the risk of the event. This type of risk can easily be aggregated over any number of events using standard statistical techniques. This is a consequence of the well known statistical fact that the expected value of a sum of distributions equals the sum of the expected values, irrespective of the type and potential dependencies or correlations of the distributions involved. This observation is important in circumstances where the precise distribution functions are not known but where expected losses can be estimated, since only the expected values are needed. Combined with projected costs and estimated effectiveness of controls, and defining risk as the expected loss, the best decision with respect to the residual aggregate expected losses, and as far as allowed by a given budget, can then be found using analytic methods.

In many industries risk analysis entails more than just minimization, with respect to a budget, of the expected losses. Although it is certainly relevant to know the expected losses, for capital management purposes it is also important to have accurate insight into the variability of the losses and in the *Value at Risk* (VaR), the probability that the losses exceed a given amount. But this more general quantitative approach (i.e., using more than just expected loss values) is not applicable in all situations. In the first place, it may be hard to collect so many data that the distribution functions can accurately be determined. This is in particular the case for so-called *heavy-tailed distributions* where high impact events occur with a very low probability; these typically occur in Information Security. It is illustrated by the observation that different companies often select different distribution functions for the same types of events [1]. Furthermore, collecting enough data to determine the distribution function underlying the behavior of a certain IS threat is most likely impossible given the fast and constantly changing IS environment. From this point of view IS risk management is quite different from more traditional insurance and stock portfolio risk management.

An additional problem of general quantitative risk analysis is risk aggregation. Although loss variation and VaR can be determined per event based on its distribution function, aggregation of these and similar risk related quantities is much more difficult than aggregation of the expected losses. To mention some of the complications this type of risk aggregation runs into: it requires knowledge of the dependencies and correlations among the events, the problems are notori-

ously ill-conditioned (thereby requiring many more data points before anything can be said with any degree of reliability), and the results can be surprisingly counter-intuitive. To illustrate the latter point, examples have been published of identically distributed but independent events for which the aggregated VaR is larger than the sum of the individual VaRs [2]. Strange phenomena of this sort take place in particular when the distributions involved are heavy-tailed. As noted above, events to which this may apply are hard to recognize in practice and their precise distribution functions are, in practice, impossible to determine.

To summarize, the qualitative model is mostly intuitive, lacks any degree or claim of precision, but has proved to be a valuable tool. The finer points of risk analysis do not enter into the picture in the qualitative approach, with the pleasant side-effect that the intricacies and pitfalls of general aggregated risk analysis are also avoided. But it is not suitable for our application. The quantitative model, when supported by adequate amounts of data, lays greater claims to accurateness. Its mathematical underpinnings may inspire attempts to address more general risk questions. This carries the inherent danger that over-precise results are obtained and relied upon, while losing sight of the fact that the underlying distributions cannot be assessed with sufficient certainty. For general aggregated risk problems the analysis can easily be led astray by intuition. In particular the presence of hard-to-recognize and hard-to-pin-down heavy-tailed distributions makes any type of intuitive guesswork irresponsible. This implies that quantitative risk analysis in its full generality cannot be used for our IS risk management application. That leaves a single alternative for our application, namely the most simpleminded quantitative risk analysis where risk is defined as an expected loss value. As argued above, that approach does not require the actual event distributions or their interactions, reasonable estimates for the expected losses suffice, and aggregation is nothing but simple summation. This is further explored in the next section.

Obviously, our approach still requires risk quantification, which is admittedly a hard problem, but the degree to which it is needed is as small as can reasonably be expected for a quantitative approach.

4 IS Risk Management Model

The conclusion from Section 3 is that if we want to have a definition of IS risk that is workable in a rapidly changing environment and that allows meaningful aggregation, then IS risk must be defined as a simple expected value of some sort. This leads to the following slightly more formal approach to the setup described in Section 2.

Remark 4.1 The description below is not identical to the prototype that is actually implemented, but contains all relevant details. A desire for simplicity and backward compatibility with systems that are familiar to our user community has led to some choices that may be unexpected. They have no effect on the principle of the model. Once enough data are available for analysis, it will

be investigated to what extent the 'incongruent' aspects of the prototype design adversely affect the outcome. This may lead to small adaptations in later versions.

The IS risk faced by a business process are due to a breach of either confidentiality, integrity, or availability. For each of these categories the user enters an estimated loss amount, denoted for business process p by $L_c(p)$, $L_i(p)$, and $L_a(p)$, respectively. It is assumed that $\max(L_c(p), L_i(p), L_a(p)) > 0$.

The likelihood that these losses are actually incurred depends on the threats against the process (or rather: the threats realizing the vulnerabilities identified in the applications used in the process, cf. Section 2). To estimate this likelihood, the user characterizes a threat t by selecting three *type of threat* choices:

- **Source of threat**, with two possible choices indicating if the threat comes from a party *external* ($\text{Source}(t) = 1$) or *internal* ($\text{Source}(t) = 0.8$) to the company.
- **Access required for the threat**, with two possible choices indicating if *remote* access ($\text{Access}(t) = 1$) suffices to realize the threat or if *local* access ($\text{Access}(t) = 0.6$) is required.
- **Skill level required for the threat**, with four possible choices indicating the least level of skill required to realize the threat:
  - unstructured nontechnical ($\text{Skill}(t) = 1$);
  - unstructured technical ($\text{Skill}(t) = 0.9$);
  - structured nontechnical ($\text{Skill}(t) = 0.75$);
  - structured technical ($\text{Skill}(t) = 0.25$).

  A hacker, for instance, would be 'unstructured technical', but a script kiddie would be 'unstructured nontechnical'.

The *current likelihood indicator* $P(t)$ of threat t is defined as

$$P(t) = \text{Source}(t) * \text{Access}(t) * \text{Skill}(t).$$

These four numeric values remain hidden for the user. A qualitative ranking of $P(t)$, however, is presented to the user: High if $P(t) \geq 0.6$, Low if $P(t) < 0.2$, and Medium otherwise. This is done for compatibility and consistency with another business reporting tool (cf. Remark 4.1). The user gets the option to change the qualitative ranking; if done so the hidden likelihood indicator is changed: if the user specifies High and $P(t) < 0.6$, then replace $P(t)$ by 0.6; if the user specifies Medium and $P(t) \geq 0.6$, then replace $P(t)$ by $0.6 - \epsilon$ for some small $\epsilon > 0$; if the user specifies Medium and $P(t) < 0.2$, then replace $P(t)$ by 0.2; if the user specifies Low and $P(t) \geq 0.2$, then replace $P(t)$ by $0.2 - \epsilon$.

Remark 4.2 The various values and formulas used in the calculation of the likelihood indicators are not crucial to the model. They were chosen because of their ease of use and because the resulting qualitative rankings are consistent with the business tool referred to that the user community is already familiar with. They are by no means the unique values and formulas that achieve these

goals: additive versions can be made to work equally well and simple table look up would be just as effective. Our approach follows [5]. See also Remark 4.4.

To indicate what *type of loss* can be inflicted by a threat, the user enters three bits $T_c, T_i, T_a \in \{0, 1\}$, where $T_c = 1$ if and only if the threat may cause a breach in confidentiality (similar for T_i and T_a with respect to integrity and availability, respectively). Note that these bits depend just on the threat and not on the process they may affect (cf. Remark 4.1).

Remark 4.3 As indicated in Section 2, data about threats (as above) and action plans (as below) should be agreed upon by all businesses using that application. One business may originally have entered threat data and action plans for an application, but other businesses affected by the same threat may review the data provided and propose changes. It is the responsibility of all parties involved to come to an agreement on the proper values. A welcome side-result of this interaction is corporate-wide consistency of (and agreement on) the 'quantification' of the threats and action plans.

Given these values entered by the user, the *current IS risk indicator of process p with respect to threat t* is defined as

$$\mathcal{R}_{\text{cur}}(p, t) = \max(T_c L_c(p), T_i L_i(p), T_a L_a(p)) P(t).$$

Denoting by $\mathcal{S}(p)$ the set of applications used in process p and by $\mathcal{T}(A)$ the set of threats affecting application A, the *current IS risk indicator of process p* is defined as

$$\mathcal{R}_{\text{cur}}(p) = \sum_{A \in \mathcal{S}(p)} \sum_{t \in \mathcal{T}(A)} \mathcal{R}_{\text{cur}}(p, t).$$

If $\mathcal{P}$ is the set of all business processes, the corporation's overall (quantitative) *current aggregated IS risk indicator* is defined as

$$\mathcal{R}_{\text{cur}} = \sum_{p \in \mathcal{P}} \mathcal{R}_{\text{cur}}(p).$$

For an action plan α countering a threat t, denote by t_α the residual threat, i.e., what remains of t after action plan α has been carried out. For each action plan α countering a threat t the user characterizes the residual threat t_α by entering the three type of threat values $\text{Source}(t_\alpha)$, $\text{Access}(t_\alpha)$, and $\text{Skill}(t_\alpha)$, similar to $\text{Source}(t)$, $\text{Access}(t)$, and $\text{Skill}(t)$ above except that they now represent the values after action plan α has been carried out. This results in the *residual likelihood indicator*

$$P(t_\alpha) = \text{Source}(t_\alpha) * \text{Access}(t_\alpha) * \text{Skill}(t_\alpha).$$

Obviously, for an action plan to be any good, is should be the case that $P(t_\alpha) < P(t)$; it is assumed that this condition holds for all threats t and action plans α under consideration. As above, and using the same calculations, the qualitative ranking of $P(t_\alpha)$ is presented to the user, who has the option to change it, which may change the value $P(t_\alpha)$. If the resulting $P(t_\alpha)$ happens to be larger than

$P(t)$, which may happen if the user manually changed $P(t)$ or $P(t_\alpha)$ values, $P(t_\alpha)$ is set to $P(t)$; action plans for which this happens do not have to be further considered. The user also enters the *projected expense* $w(\alpha)$ of action plan α.

The type of loss bits are, in the present model, not affected by the action plans (cf. Remark 4.1). Therefore, the *residual IS risk indicator of process p with respect to threat t after action plan α is carried out* is defined as

$$\mathcal{R}_{\text{res}}(p, t_\alpha) = \max(T_c L_c(p), T_i L_i(p), T_a L_a(p)) P(t_\alpha).$$

We assume that either zero or at most a single action plan can be carried out per threat, that action plans cannot be carried out partially, and that different threats have different action plans. It is easily seen that this is not a restriction. In situations were it makes sense to consider a fractional combination of one or more action plans countering a single threat, one simply enters the relevant fractional combination of action plans with their partial or cumulative effects (and expenses) as an alternative action plan.

An *allowed set of action plans* is a set of action plans that contains at most one action plan per threat. Let $\mathcal{A}$ be an allowed set of action plans and let $w(\mathcal{A}) = \sum_{\alpha \in \mathcal{A}} w(\alpha)$ be the projected expense of $\mathcal{A}$. The *residual IS risk indicator of process p with respect to threat t after the action plans in $\mathcal{A}$ are carried out* is defined as

$$\mathcal{R}_{\text{res}}(p, t, \mathcal{A}) = \begin{cases} \mathcal{R}_{\text{cur}}(p, t) & \text{if } \mathcal{A} \text{ does not contain an action plan countering threat } t \\ \mathcal{R}_{\text{res}}(p, t_\alpha) & \text{if } \mathcal{A} \text{ contains action plan } \alpha \text{ countering threat } t \end{cases}$$

and the *residual IS risk indicator of process p under allowed action plan set $\mathcal{A}$* is defined as

$$\mathcal{R}_{\text{res}}(p, \mathcal{A}) = \sum_{A \in \mathcal{S}(p)} \sum_{t \in \mathcal{T}(A)} \mathcal{R}_{\text{res}}(p, t, \mathcal{A}).$$

Finally, the corporation's (quantitative) *residual aggregated IS risk indicator after allowed action plan set $\mathcal{A}$* is defined as

$$\mathcal{R}_{\text{res}}(\mathcal{A}) = \sum_{p \in \mathcal{P}} \mathcal{R}_{\text{res}}(p, \mathcal{A}).$$

Optimal risk mitigation consists of finding an allowed action plan set $\mathcal{A}$ that minimizes $\mathcal{R}_{\text{res}}(\mathcal{A})$. This is trivially solved by determining for each threat t the action plan α that minimizes $P(t_\alpha)$ (in case of conflict, select one), and by defining $\mathcal{A}$ as the set of those action plans (which will be allowed due to the construction). A more interesting problem is how to find an allowed action plan set $\mathcal{A}$ that minimizes $\mathcal{R}_{\text{res}}(\mathcal{A})$ under a budgetary constraint $w(\mathcal{A}) \leq \mathcal{W}$ on $\mathcal{A}$'s projected expense. That problem is addressed in the next section.

Remark 4.4 The current and residual aggregated IS risk indicators $\mathcal{R}_{\text{cur}}(p)$ and $\mathcal{R}_{\text{res}}(p, \mathcal{A})$ for a process p and allowed action plan set $\mathcal{A}$ must not and cannot be interpreted as the expected loss amount for p before and after $\mathcal{A}$. Any interpretation of that sort would at the very least require introduction of

a temporal dependency in the model. This may be done, if required, at a later stage. Similarly, a threat's likelihood indicator $P(t)$ should not immediately be interpreted as the probability that the threat is realized. It requires more threat related data and fine-tuning of the above parameter choices before the likelihood of a threat's occurrence can reliably be estimated based on the type of threat values. It may also be the case that for a reasonably accurate estimate more threat characteristics are required.

However, we are not convinced that the disadvantage of the introduction of any extra complications (a steeper learning curve) would be outweighed by the potential advantages. At present the $P(t)$, $P(t_\alpha)$, $\mathcal{R}_{\text{cur}}(p,t)$, $\mathcal{R}_{\text{cur}}(p)$, $\mathcal{R}_{\text{res}}(p,t_\alpha)$, $\mathcal{R}_{\text{res}}(p,t,\mathcal{A})$, and $\mathcal{R}_{\text{res}}(p,\mathcal{A})$ values by themselves are simply not intended to be meaningful. What is relevant is the consistency that is achieved by this approach and the fact that the relative values are meaningful. That allows us to interpret terms such as $\mathcal{R}_{\text{cur}}(p,t)$ as expected values (of some value, up to an unknown and irrelevant constant scaling factor) and thereby to aggregate them into a quantitative IS risk indicator using simple summation, as in the definitions of $\mathcal{R}_{\text{cur}}(p)$, $\mathcal{R}_{\text{res}}(p,\mathcal{A})$, $\mathcal{R}_{\text{cur}}$, and $\mathcal{R}_{\text{res}}(\mathcal{A})$. It also allows us to find an optimal allowed set of action plans under a budgetary constraint, as described in the next section. Note that also the values $\mathcal{R}_{\text{cur}}$ and $\mathcal{R}_{\text{res}}(\mathcal{A})$ by themselves are hardly meaningful. What is meaningful is the quantity

$$\frac{100(\mathcal{R}_{\text{cur}} - \mathcal{R}_{\text{res}}(\mathcal{A}))}{\mathcal{R}_{\text{cur}}}$$

because it gives the percentage how much 'better' the situation is after carrying out the action plans in $\mathcal{A}$, with 0% indicating no improvement and 100% that there is no residual aggregated IS risk left (since $\mathcal{R}_{\text{res}}(\mathcal{A}) = 0$).

Remark 4.5 It may be tempting to include a weighting mechanism in the IS risks to account for 'relative importance' of the various business processes. However, this may be done only if the weights are not correlated to the loss indicator values, because a correlation would undermine the soundness of the aggregation method. If risk is no longer defined as the expected value of a linear function of a loss indicator (as would be the case if loss indicator correlated weights are included), risk aggregation can no longer be done by summation. Correct aggregation would require the distribution functions underlying the threats and their correlation behavior, leading to numerous complications and pitfalls (cf. Section 3 and [2]) and, if those can be solved and avoided, respectively, to considerably more involved definitions of $\mathcal{R}_{\text{cur}}(p)$, $\mathcal{R}_{\text{res}}(p,\mathcal{A})$, $\mathcal{R}_{\text{cur}}$, and $\mathcal{R}_{\text{res}}(\mathcal{A})$. Weights that reflect the relative importance of businesses may be used if they are independent of the amount of loss the businesses may incur due to IS failures. Obviously, this is only meaningful if the same set of weights is used in $\mathcal{R}_{\text{cur}}$ and $\mathcal{R}_{\text{res}}(\mathcal{A})$. Our current model does not use weights. Using weights would be one way to include a temporal dependency in the model.

5 Optimal Risk Mitigation and Multiple-Choice Knapsacks

With notation and definitions as in Section 4 the risk mitigation under budget constraint problem is as follows:

$$\text{minimize } \mathcal{R}_{\text{res}}(\mathcal{A})$$

$$\text{subject to the condition that } w(\mathcal{A}) \leq \mathcal{W}$$

$$\text{and that } \mathcal{A} \text{ is an allowed action plan set.}$$

This is a multiple-choice knapsack problem, as defined in [4]. For ease of reference we present the straightforward translation from the above formulation to the framework from [4].

Let k be the number of threats (counted over all applications) and let N_i be the set of action plans for the ith threat, $i = 1, 2, \ldots, k$. Each action plan $\alpha \in N_i$ has an *IS risk reduction indicator* $p_{i\alpha}$ ('profit' in [4]) and a *projected expense* $w_{i\alpha} = w(\alpha)$ ('weight' in [4]): if t is the ith threat, then

$$p_{i\alpha} = \sum_{\substack{\text{processes } p \text{ for} \\ \text{which } t \text{ affects an} \\ \text{application used by } p}} (\mathcal{R}_{\text{cur}}(p,t) - \mathcal{R}_{\text{res}}(p,t_\alpha)).$$

For $0 < i \leq k$ we include a default 'free' action plan α in N_i with $w(\alpha) = 0$ and $P(t) = P(t_\alpha)$ (and thus $p_{i\alpha} = 0$) for the ith threat t, corresponding to not doing anything against t. The multiple-choice knapsack problem equivalent to our risk mitigation under budget constraint problem may then be formulated as:

$$\max \sum_{i=1}^{k} \sum_{\alpha \in N_i} p_{i\alpha} x_{i\alpha}$$

$$\text{subject to } \sum_{i=1}^{k} \sum_{\alpha \in N_i} w_{i\alpha} x_{i\alpha} \leq \mathcal{W},$$

$$\sum_{\alpha \in N_i} x_{i\alpha} = 1, \quad i = 1, 2, \ldots, k,$$

$$x_{i\alpha} \in \{0, 1\}, \quad i = 1, 2, \ldots, k, \quad \alpha \in N_i.$$

To see the equivalence of both formulations, note that it amounts to switching the order of the summations: in Section 4 we summed over all threats affecting a process p to define $\mathcal{R}_{\text{cur}}(p)$ and $\mathcal{R}_{\text{res}}(p, \mathcal{A})$, in the alternative formulation we sum over all processes affected by the ith threat to define $p_{i\alpha}$. Without loss of generality it may be assumed that all coefficients $p_{i\alpha}$, $w_{i\alpha}$, and $\mathcal{W}$ are non-negative integers (if necessary after appropriate scaling).

Although problems of this sort are known to be NP-hard, they can be solved quickly in pseudo-polynomial time. In [4] a particularly efficient method is presented that finds the optimal solution for any fixed budget constraint $\mathcal{W}$ using dynamic programming. We refer to [4] for a detailed description. For our purposes it is interesting to know that the LP-relaxation of the problem leads to an almost linear time optimal solution if small variations are allowed in the budget constraint $\mathcal{W}$. This follows by considering the sequence of weights encountered in the course of [4, Algorithm 1 *Greedy*], with respect to which the greedy solutions built so far are all optimal. Either way, finding the optimal spending strategy even for very large IS risk mitigation problems will be a matter of at most a few seconds.

6 Conclusion

We have presented an easy to use quantitative approach to IS risk management that allows a meaningful quantitative interpretation of the effect of risk mitigation and fast determination of the optimal risk mitigation strategy. The prototype model is sufficiently flexible that it allows fine-tuning and other more substantial refinements, if that is found to be desirable based on practical experience with the model. At this point in time the prototype's implementation is under development for imminent deployment on a world-wide scale.

Acknowledgments. We thank Carl Heybroeck, Satya Vithala, and Gary Word for their support and useful discussions.

References

1. Basel Committee's Risk Management Conference on Leading Edge Issues in Operational Risk Management, New York, May 29-30, 2003. Presentations available from `www.newyorkfed.org/pihome/news/speeches/2003/con052903.html`
2. P. Embrechts, A. McNeil, D. Straumann, *Correlation and dependence in risk management: properties and pitfalls*, August 1999; Chapter 7 in M.A.H. Dempster (ed.) *Risk Management, value at risk and beyond*, Cambridge University Press, January 2002
3. National institute of standards and technology, *Guideline for automatic data processing risk analysis*, FIPS PUB 65, August 1979
4. D. Pisinger, *A minimal algorithm for the multiple-choice knapsack problem*, Technical report 94/25, DIKU, University of Kopenhagen, Denmark; available from `www.diku.dk/~pisinger`
5. T.Voss, *A simple one-dimensional quantitative risk assessment model, v. 1.4*, internal Citigroup Information Security Office document, February 2002