



Research Center
Finance & Information Management



Project Group
Business & Information
Systems Engineering

Discussion Paper

IT-Enabled Risk/Return Management: Service-Oriented Infrastructures vs. Dedicated Systems

by

Hans Ulrich Buhl, Gilbert Fridgen



Europäische Union
„Investition in Ihre Zukunft“
Europäischer Fonds für
regionale Entwicklung

in: Proceedings of the Nineteenth European Conference on Information Systems
(ECIS 2011), Association for Information Systems (AIS), June 2011. Paper 99

WI-294

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: F.-v.-Schiller-Str. 2a, 95444 Bayreuth
Phone: +49 921 55-4710 (Fax: -844710)



IT-ENABLED RISK/RETURN MANAGEMENT: SERVICE-ORIENTED INFRASTRUCTURES VS. DEDICATED SYSTEMS

Buhl, Hans Ulrich, FIM Research Center, Department of Information Systems Engineering & Financial Management, University of Augsburg, Universitätsstr. 12, 86135 Augsburg, Germany, hans-ulrich.buhl@wiwi.uni-augsburg.de

Fridgen, Gilbert, FIM Research Center, Department of Information Systems Engineering & Financial Management, University of Augsburg, Universitätsstr. 12, 86135 Augsburg, Germany, gilbert.fridgen@wiwi.uni-augsburg.de

Abstract

Risk/return management has not only evolved as one of the key success factors for enterprises especially in the financial services industry, but is in the times of the economic crisis initiated by the financial markets crucial for the survival of a company. It demands powerful and at the same time flexible computational resources making it an almost ideal application for service-oriented computing concepts. An essential characteristic of service-oriented infrastructures is that computational resources can be accessed on demand and paid per use. Taking the estimation of covariances for a portfolio of risky investment objects as an example, we propose quantification for the economic value of fast risk/return management calculations. Our model then compares the cost structures of service-oriented infrastructures and dedicated systems in this domain. On this basis, we can determine under which circumstances the one or the other architectural strategy is superior.

Keywords: Risk/Return Management; Service-Oriented Infrastructures; Grid Computing.

1 Introduction

The ongoing financial and economic crisis (FEC) caused unprecedented damage for the economy worldwide. Its reasons and consequences illustrate the complex and dynamic environment in which companies act nowadays: Increasingly harmonized markets and the growing use of information technology (IT) enable companies to be part of globally connected value networks and therefore to realize quick return opportunities all over the world. The FEC definitely showed that these value networks are also risk (spreading) networks: Companies took too much risk on assets they believed to understand but whose risk contribution was in fact completely unknown. Similar to a domino effect, these risks spread rapidly all over the world, not only in the financial sector. Due to a lack of equity, banks did for example not renew credit contracts of companies in production or service sectors. Their respective bankruptcy brought suppliers and customers who relied on their business partners into liquidity issues, especially, when they had problems with their own credit renewals. Although the financial services industry pulled the trigger, the problems and solutions presented in this paper apply to all industries: Producing and service industries highly depend on the reliability of customers, suppliers, and many also of resource prices.

To enable decision support (particularly on capital market based hedging) the risk/return-position of the company has to be determined shortly before market actions, after market actions, and – due to dynamic markets – on a regular basis. Buhl et al. (2009) described, how to determine the optimal interval of these risk/return calculations. The computing capacity necessary can thereby be substantial, especially in globally acting enterprises. For a complete examination, the correlation between every possible pair of (relevant) assets has to be determined. This requires IT to be able to make extensive calculations in very short time, even when examining only parts of a company's assets. Innovative approaches of distributed computing, like grid computing, cloud computing, or service-oriented architectures, are principal topics of IT related science and business which are offering potentially suitable infrastructures.

We will speak of service-oriented infrastructures (SOI) as an instance of (the abstract principle of) service-oriented architectures where (mostly resource-intensive) distributed services are made available over a grid network. It is common practice in today's financial services industry to calculate the risk exposure within fixed time intervals (e.g. several days). SOI based on grid technologies can embrace resources of the whole enterprise and even of external resource providers. Therefore, they offer huge amounts of computing capacity that can be used to accelerate calculations dramatically. A Dedicated System (DS, e.g. a computing cluster) where a fixed set of resources is dedicated to the risk/return calculations could nevertheless be a sensible alternative to a SOI. Building upon the results of Buhl et al. (2009), this paper aims at answering the following research question: *Under which conditions is a SOI superior to a DS regarding calculations in risk/return management?* We formulate an analytical optimization model delivering a solution to this decision problem.

2 Grid Computing

Grid computing can be regarded as an infrastructure technology enabling the virtualization of physical resources. Various proponents have agreed with Foster and Kesselman (1998) that "A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities". Grid technologies are one possibility to realize a SOA consisting of so called "grid services". Grid services are based on specific web service standards, like the specifications "Open Grid Services Architecture" and "Web Services Resource Framework". They extend web services insofar as they imply the dynamic, yet for the user transparent, allocation of (physical) resources to services by a grid middleware and therefore are especially suited to fulfil resource intensive tasks. There is an extensive literature on service-oriented computing or grid computing in general (see e.g. Foster and Kesselman, 1998; Berman, 2005; Silva, 2006; Singh and

Huhns, 2004). Some publications even consider the application of these technologies for portfolio management, derivatives pricing or other areas of financial risk management (Brownlees et al., 2006; Crespo et al., 2006; Schumacher et al., 2006). These approaches describe how grid technologies can be applied, but do not quantify the resulting business value. In the context of grid computing also resource allocation mechanisms have been widely discussed—most often under the term “grid economics”. We refer to Buyya et al. (2000), Nabrzyski et al. (2003) and Wolski et al. (2004) which provide an overview of this area. They dwell either on the question which principles are appropriate to resource management or on technical issues connected with the development of resource management systems. Accordingly, in most of the existing approaches demand for computational resources is merely an external factor whereas in our approach it is subject to optimization.

In this context an important characteristic of SOI based on grid technologies is the on-demand access to distributed resources. When resources stem from an external provider this concept is often labelled utility computing, meaning that resources can be consumed and priced as easy as for instance electricity or water. Utility computing has been subject to research as well. Bhargava and Sundaresan (2004) analyze pay-as-you-go pricing scenarios where providers guarantee computing capacity, but users do not make a commitment towards actual use. Our paper takes the perspective of a service user and presents a rationale for decisions on computing capacity in the context of risk/return management. This may be accomplished by investing into an own internal SOI or by making use of utility computing offers by major infrastructure providers like HP, SUN or IBM (Bhargava and Sundaresan, 2004, p. 202) but also online service providers like Google, Amazon and Microsoft, whose services are today oftentimes marketed under the term “cloud computing”. The question underlying this paper is under which circumstances a SOI for risk quantification is favorable over (traditional) DS.

3 Risk/Return Management

The term “risk” is used heterogeneously in general speaking as well as in academic circles. While in the economic literature risk is often explained as the “possibility of missing a planned outcome” we will follow a more finance-related approach. From this point of view we define with Schröck (2001, p. 24) risk as “the deviation of a financial value from the expected value”. A positive deviation is often referred to as “chance” while a negative deviation is characterized as “danger”. Because of this two-sided perception of risk, variance or standard deviation of a risky value are suitable and well accepted measures of risk. We will use the standard deviation of historical portfolio returns as the risk measure later in this text. Synonymously we will speak of the volatility of a portfolio and define it as the “annualized standard deviation of percentage change in daily price” (Spremann, 2003, pp. 154).

Enterprises are investing capital into investment objects in order to generate cash inflows and subsequently to increase the return of the invested capital. Typically risk-averse management is making risky investments hoping to achieve an excess return over the risk-free rate. There is a general connection between risk and return of an investment object: higher return is systematically associated with higher risk. This connection is theoretically explained by economic models like the “Capital Asset Pricing Model” which was originally developed by Sharpe (1964), Lintner (1965) and Mossin (1966) and empirically evaluated later on (an overview of relevant empirical studies can be found e.g. in Copeland and Weston (1988, pp. 212). Following the argumentation of Wilson (1996, pp. 194) it is therefore crucial for the survival and success of an enterprise to be able to allocate the available capital to the right combination of investment objects. Investment objects in this context are not necessarily restricted to securities. Almost all business transactions are associated with uncertainty and thus contribute to an enterprises overall risk exposure.

One major goal of risk/return management in this context is the prevention of bankruptcy by restricting potential losses resulting from risky investment objects. The increasing importance of this goal is emphasized by the current economic crisis. A growing number of rules and regulations require enterprises to hold a part of their available capital to back their risky investments (Jackson et al., 1998, pp. 8). This share of the available capital then makes less or no contribution to the overall earnings. By

management decisions these restrictions are broken down along the organizational hierarchies into guidelines on business unit or departmental level. We are assuming in the following text that those guidelines are essentially representing limits for the maximum risk a department, business unit and consequently an enterprise is willing (or able) to take.

To evaluate whether an enterprise or department complies with a given risk limit, it is necessary to calculate the current risk exposure. We concentrate on one fundamental instrument in this paper: Covariances can be used for a comprehensive enterprise-wide risk/return management as described by Huther (2003, pp. 111) and Faisst and Buhl (2005, pp. 408) and as later picked up by Fill et al. (2007), Kundisch et al. (2007) and Gericke et al. (2009). Other publications are dealing with the empirical estimation and forecasting of covariances using historical data. For applicable techniques refer to Engle (1982), Kupiec (2007) and Hull and White (1998). An overview of the corresponding methods used in volatility and correlation forecasting can be found in Alexander (2008, pp. 89). Taylor (2005) presents recent approaches to volatility forecasting like conditional autoregressive Value-at-Risk (CAVaR) models. In our context, covariances are used for determining the overall risk position of an enterprise. Nevertheless, our methods are also applicable to other risk measures as long as they consider dependencies between single investment objects.

Following the covariance approach, we can represent risky investment objects by random variables. Considering the portfolio risk and return merely on an aggregated level is not satisfactory because all information about the risk attributable to a single investment object is lost. It is crucial to separate the portfolio and decompose it into data per investment object. Only then the enterprise can perform economically rational investment decisions on different aggregation levels. With σ_i^2 and Cov_{ij} denoting variance and covariance of investment objects respectively we can determine the overall risk of a portfolio σ_p^2 , consisting of $n \gg 0$ investment objects (numbered from 1 to n), as $\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n Cov_{ij}$. The so defined matrix is called *covariance matrix*. An important characteristic of covariances is that $Cov_{ij} = Cov_{ji}$. This makes the matrix symmetric and thus not all of its values have to be calculated. The total number of covariance calculations necessary is given by $n(n + 1)/2$.

4 A Valuation Model for fast Risk Quantification

We consider an enterprise which has access to (and is possibly engaged in) a set of risky investment objects as well as to a risk-free investment alternative. It is frequently (re)calculating its overall risk position by estimating the covariance matrix of its portfolio. Our main hypothesis for the valuation of benefits is: the faster risk/return management calculations can be executed the higher will be the return of the enterprise because given risk limits can better be exploited.

Since the enterprise is acting in an uncertain and dynamic environment its risk position is changing willingly (by investment decisions) or unwillingly (by “movement” of the markets). Because the estimation of risk cannot be accomplished in real-time the covariances at hand are always outdated. We are in the following recurring to the fact that the enterprise is adjusting its risk position to a value somewhere below a certain threshold thus constituting a “safety margin”. In the regulatory context this is often called “haircut”, like described by the Basel Committee on Banking Supervision (2004). The financial and economic crisis made clear, that – depending on the assets invested in – a high safety margin is necessary to account for changes of parameters, especially in liquidity. It is doing so by using the capital allocation between the risky investment objects and the risk-free alternative for balancing their overall risk position. Our model is therefore *not* addressing the evaluation of the efficient set of investment objects or portfolio optimization (both require covariances), but the aggregation and management of the risk position of an enterprise. Whenever covariances are available the safety margin can be adjusted immediately in a way that the resulting (and over time changing) overall risk position of the enterprise does not exceed the given risk limit at any time. Hence, the faster covariances are available, the smaller the safety margin can be. We will use this effect to quantify the benefits of fast estimations depending on the time needed for the completion of one covariance matrix.

The time interval we consider consists of equidistant periods. $t = 0$ denotes the beginning of the first period. σ_t e.g. indicates the value of a model parameter at the end of period t . (Dis)investment decisions take effect at the end of each period. If not described differently, all variables are $\in \mathbb{R}$. The enterprise is equipped with a total capital of $K > 0$ which is always completely allocated to the risky portfolio and/or the risk-free alternative. We denote the risky portion of K with $x \geq 0$ and furthermore use x for risk adjustment. T indicates the length of the calculation time frame, $T \in \mathbb{N} \setminus \{0\}$ with \mathbb{N} as the set of natural numbers. At the end of each time frame we choose x in a way that the risk limit is “probably” not exceeded during the next time frame. We will formulate more precisely what is meant by “probably” later on. Future returns of the portfolio are modelled as independent random variables. Their probability distribution for each period can be characterized by mean and standard deviation. This implies that the investment objects can be marked to market, i.e. there is a price attached to them. We additionally need a set of assumptions for the deductions following thereafter.

Assumption 1 *The enterprise is generally risk-averse and striving for efficient combinations of investment objects. A sufficient number of investment objects (adjacent to the safety margin) are perfectly divisible, liquid, and traded on a no-frictions market.*

Assuming risk aversion is probably well justified by the experiences of the FEC. As liquidity was an issue there, we only assume investment objects to be liquid in the area adjacent to our safety margin.

Assumption 2 *The risky part of the enterprise’s capital yields the expected return μ , the risk-free investment pays the time-invariant risk-free interest rate i , which is equal to the borrowing rate. We always have $\mu > i > 0$.*

With the so defined parameters and x as decision variable we can determine the overall expected return $\mu^U(x)$ and risk $\sigma_t^U(x)$ of the enterprise according to common rules of statistics as

$$\mu^U(x) = x\mu + (1 - x)i = i + x(\mu - i) \text{ and } \sigma_t^U(x) = x\sigma_t \quad (1)$$

The overall risk of the enterprise is expressed by the portfolio risk and thus changes over time driven by the varying σ_t . Note that due to our focus on changing risk we do not regard a changing $\mu^U(x)$ over time (which would result in an index t as in $\sigma_t^U(x)$).

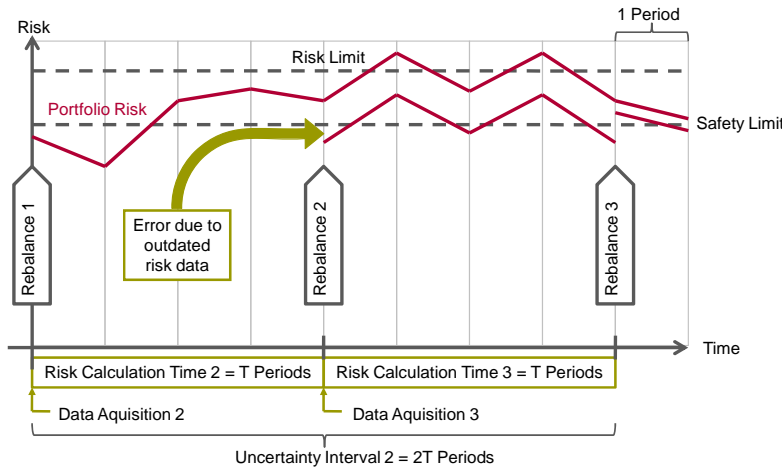


Figure 1. Period Model and Relevant Time Intervals.

As we see, with ascending x the overall returns as well as the overall risk of the enterprise are both increasing. On the one side the enterprise certainly strives for the highest possible return, on the other side the given risk limit constricts x .

It is common practice to use some variation of a random walk for the price movement on security and commodity markets. This approach goes ultimately back to Louis Bachelier (1900) who compared

the stock market with a “drunkards walk”. Although controversially discussed, it was picked up more than half a century later by Mandelbrot (1963; 1972) and Fama (1965) among others. Following this theory of random walks historical (e.g. daily or weekly) portfolio returns can be used for estimating mean μ and standard deviation σ_t of future portfolio returns. It is important to understand that in our model the standard deviation is possibly changing in each period (indicated by its index t).

Assumption 3 *The initial calculation of covariances starts in $t = 0$ and is finished after T periods. Each new covariance calculation begins in the finishing period of the previous covariance calculation.*

According to assumption 3, whenever a covariance matrix is completed, the input data used for its calculation are T periods old. We can immediately determine the portfolio risk by summing up the covariances in the matrix. This can then be used for rebalancing the portfolio. The moment before the next matrix is finished the input data used for risk calculations are already $2T$ periods outdated.

Therefore the uncertainty interval that has to be taken into account spans $2T$ periods: in the worst case the risk has been going up over $2T$ periods before the enterprise realizes that it is exceeding the maximum risk it is willing (or able) to take (see figure 1). Without loss of generality we will concentrate our analysis on the first covariance matrix calculation and the corresponding adjustment decision, therefore focussing on the time interval $[0; 2T[$. During this time the portfolio risk is fluctuating in a non-predictable way.

As originally introduced by Buhl et al. (2009), we now dwell on the portfolio risk in t , denoted as σ_t , modelling it as a random variable. This relates to a phenomenon known from the behaviour of stock market prices called *heteroscedasticity* (see e.g. Spremann, 2003, pp. 152). In analogy to returns we assume:

Assumption 4 *The σ_t are normally distributed.*

This distribution assumption can (and would in practice) be relaxed by approximating the distribution of the σ_t delivered by the calculated sequence of standard deviations. Arranged in increasing order one can easily deduce the quantiles needed in our model. Nevertheless, for reasons of simplicity and without significantly changing the general result we assume σ_t to be normally distributed here.

We will again focus on two distribution parameters: Our notation for the (strictly positive) mean will be μ_σ , for the standard deviation σ_σ (both tagged with an σ indicating that the distribution applies to the portfolio risk), thus $\sigma_t \sim N(\mu_\sigma; \sigma_\sigma)$ with N short for the normal distribution.

The standard deviation of the portfolio risk can be taken as an estimate for the expected portfolio risk μ_σ and thus as the starting point for the random walk of σ_t . In order to maximize x in equation (1) we have to consider the uncertainty interval $[0; 2T[$. Therefore, the standard deviation of the expected portfolio risk after $2T$ periods is again normally distributed with $N(\mu_\sigma; \sigma_\sigma \sqrt{2T})$. As a consequence, we have $\sigma_t^U(x) \sim N(x\mu_\sigma; x\sigma_\sigma \sqrt{2T})$ for the overall risk of the enterprise.

In order to rephrase the fuzzy formulation “the risk limit is probably not exceeded” we will follow an approach comparable to the VaR for quantifying portfolio risk. It is important to understand, that the VaR is not sufficient for this purpose, as is in our model the risk itself (and not the “value”) is risky. We speak of a Risk-at-Risk over a holding period and a confidence level α , $0 \ll \alpha < 1$ and think of it as the standard deviation $\bar{\sigma}$ which is exceeded within the holding period only with the probability of $(1 - \alpha)$. With $\Phi(x)$ denoting the standardized normal distribution function, we know for the distribution of $\sigma_t^U(x)$ over $2T$ periods that $P(\sigma_t^U(x) \leq \bar{\sigma}) = \Phi\left(\frac{\bar{\sigma} - x\mu_\sigma}{x\sigma_\sigma \sqrt{2T}}\right)$. At the same time we require the probability to be greater than or equal to the confidence level α :

$$P(\sigma_t^U(x) \leq \bar{\sigma}) \geq \alpha \text{ for } t = 2T \quad (2)$$

In the marginal case both sides of the equation are equal and we can therefore state—with q_α as the (onesided) α -quantile of the standardized normal distribution—that

$$\frac{\bar{\sigma} - x\mu_\sigma}{x\sigma_\sigma \sqrt{2T}} = q_\alpha \Leftrightarrow x = \frac{\bar{\sigma}}{q_\alpha \sigma_\sigma \sqrt{2T} + \mu_\sigma} \quad (3)$$

x gives us the portion of risk-free and risky investment objects in a way that equation (2) holds. We can calculate the overall expected benefits of the enterprise, given this capital allocation, as $B(x) = \mu^U(x) \cdot K = (i + x(\mu - i)) \cdot K$. By inserting x from equation (3) into $\mu^U(x)$ from equation

(1) we find for the benefits (with the number of periods needed for the completion of one covariance matrix as the independent variable)

$$B(T) = \left(i + \frac{\bar{\sigma}(\mu - i)}{q_\alpha \sigma_\sigma \sqrt{2T} + \mu_\sigma} \right) \cdot K \quad (4)$$

Considering equation (4) an enterprise could maximize its benefit by minimizing the time T that is needed to calculate a covariance matrix. Yet there is a trade-off between the benefits and the cost, i.e. cost caused by the infrastructure that is necessary to compute the calculations.

5 Determining the Required Computing Capacity on a SOI

From a SOI point of view the risk calculation can be regarded as a service providing its user with up-to-date information for the relevant investments. In this section we derive the relationship between the computing capacity (i.e. cost) allocated to risk quantification and the time needed for the computation. We denote with z the computing capacity necessary for estimating one covariance matrix in exactly T periods. We use CPUs as a measure for computing capacity and are aware of the fact that this means a one-dimensional view on matters as other determinants of system performance are ignored. We denote with w the workload—measured in CPU hours—for estimating one covariance. Applying a simple moving average technique with a rolling sample of historical data (Elton and Gruber, 1972, pp. 409) we get unbiased estimators of the expected value and (co)variances for every point in time.

Assumption 5 *The same workload w is necessary for estimating variances and covariances. Furthermore every covariance is estimated from scratch, i.e. no intermediate results are used.*

Variance estimation requires only one historical time series, so its intrinsic workload is smaller than the workload for covariance estimation. Nevertheless, this effect can be neglected since for n variances there are $n(n - 1)/2$ covariances in a given covariance matrix. With n sufficiently large the variances have merely no effect on the number of calculations. Using no intermediate results could be considered awkward for the simple moving average procedure but is a realistic approach for more sophisticated methods.

Assumption 6 *The length of the calculation time frame T depends solely on the time needed for the computation, neglecting e.g. latency or transmission times. Correspondingly the only cost relevant is cost for computation which occurs in the form of a (internal or external) factor price per CPU hour over a given time.*

Note that for the calculation of covariance matrices on a SOI it is convenient that the computations can be distributed on several nodes and executed in parallel, as all pairwise covariances can be calculated independently from each other. Thus efficiency losses are considerably low. We can now deduce the computing capacity $z(T)$ (workload per time) that is required in every period over T periods. We already know that for n investment objects $n(n + 1)/2$ covariances have to be calculated. Multiplied with the workload per covariance this determines the total number of CPU hours needed. This in turn leads to the functional relationship

$$z(T) = \frac{n(n + 1)w}{2T} \Leftrightarrow T(z) = \frac{n(n + 1)w}{2z} \quad (5)$$

Given n investment objects and a computing capacity of z CPUs, the covariance matrix will be completed after $T(z)$ periods (In the following outcomes are assumed real-valued. In reality and in order to fit it to the discrete-time period model one has to check the neighboring integer values to obtain the discrete optimum). This constitutes an important parameter of the covariance estimation service since it describes the economic value that should be attributed to the consumption of capacity. By inserting (5) into equation (4) and neglecting the mean risk compared to the standard deviation of the risk (The exact quantification would be tedious to continue with in our analysis. In order to avoid

writing overhead we deliberately simplify our objective function by neglecting the mean risk, which is a numerically justifiable approximation in our context.) we quantify the benefits as

$$B(z) = a + 2b\sqrt{z} \text{ with } a = iK > 0, b = \frac{\bar{\sigma}(\mu-i)K}{2q_a\sigma_\sigma\sqrt{n(n+1)w}} > 0$$

$$B'(z) = \frac{b}{\sqrt{z}} > 0, B''(z) = -\frac{b}{2\sqrt{z}^3} < 0 \quad (6)$$

$B(z)$ is a strictly increasing and concave function.

In the following we will consider two alternatives for balancing the benefits described by equation (6) with the cost of computation: SOI vs. DS.

6 Service-Oriented Infrastructures vs. Dedicated Systems

As a base of the decision on SOI vs. DS, the enterprise faces the problem of determining the required computing capacity for its risk-/return management in advance. Especially in a DS setting, it is obviously necessary to acquire or reserve resources for a longer planning horizon (e.g. one year). Even in a SOI scenario, where one might expect capacity planning to be superfluous, under realistic conditions it is still necessary: In the case of internal provision the enterprise needs to decide on the overall capacity of its SOI thus affecting the availability and prices of the resources. When resources are provided externally it may be required to forecast and reserve capacity in advance. This is in accordance with the findings of Bhargava and Sundaresan (2004) who find that a pure pay-as-you-go model without reservation is not reasonable in most cases.

We will therefore determine suitable cost functions depending on the computing capacity used for risk quantification both on the DS and the SOI. In our notation of the model parameters, we will use the lower indices D for the DS and S for the SOI, respectively. For the DS, we apply a cost function $C_D(z)$ proportional to the computing capacity z provided per period which in fact assumes that there exist suppliers for DS with virtually any capacity. This leads us to a continuous cost function with a price of p_D per computing capacity used over one period. For the SOI we also specify a proportional cost function with slope p_S , but with a fixed minimum cost of $p_S z_m$, $z_m > 0$. Thus the SOI cost function has a slope of 0 to the point z_m and afterwards a constant slope of p_S , i.e. for all $z \geq z_m$ every unit of capacity over one period cost p_S . The break at point z_m accounts for the following situations that need to be considered in practice for an internal and external provisioning scenario respectively:

- The SOI capacity is obtained by an external service provider for the usage price p_S . The service consumer on the other side is obligated to a minimum purchase of z_m . Consuming less capacity nevertheless induces cost of $p_S z_m$.
- The SOI is instituted by an investment project of the enterprise. After initial investment cash-flows free capacity up to the amount of z_m is available for covariance estimations at variable cost of 0 because of pooling existing internal resources, until the free capacity generated by the SOI project is fully exploited. A capacity demand exceeding z_m induces a usage price of p_S – because of competing service consumers in the case of an internal SOI or because of additional external capacity priced with p_S .

We assume that always $0 < p_S < p_D$ reflecting the typical advantages of a SOI. Firstly, an intrinsic characteristic of a SOI is that a potentially high number of services share a common infrastructure. Capacity must not be determined by summing up the requirements of each single service independently. Instead multiplexing gains resulting from less than perfectly correlated demand structures need to be considered (Chandra et al., 2003). Therefore, total capacity may be significantly smaller compared to independent dedicated systems. Additionally, a SOI may comprise standardized components which can be assumed to be substantially cheaper. Finally, on a SOI capacity can be varied by adding or removing resources. Thus, it is possible to react to varying environment conditions which is a crucial part of the results of our optimization model.

Consequently, we define the cost functions, C_D and C_S , as well as the objective functions, Z_D and Z_S , for the DS and the SOI with recourse to equations (5) and (6) for $z > 0$ as

$$\begin{aligned} C_D(z) &= p_D z, & Z_D(z) &= B(z) - C_D(z) \\ C_S(z) &= \max(p_S z_m; p_S z), & Z_S(z) &= B(z) - C_S(z) \end{aligned}$$

$Z_D(z)$ and $Z_S(z)$ are again strictly concave functions. $C_D(z)$, $C_S(z)$ and $Z_D(z)$, $Z_S(z)$ are intersecting at $z_1 = \frac{p_S z_m}{p_D} < z_m$. We set $c = p_S z_m = p_D z_1$ and obtain a cost setting as depicted in figure 2.

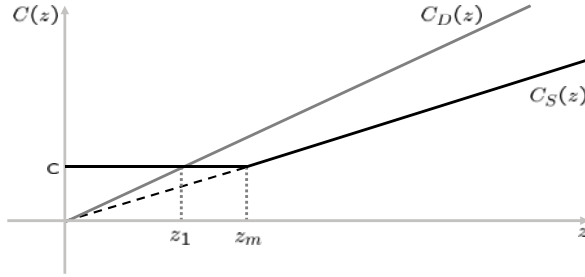


Figure 2. Cost Functions for DS and SOI.

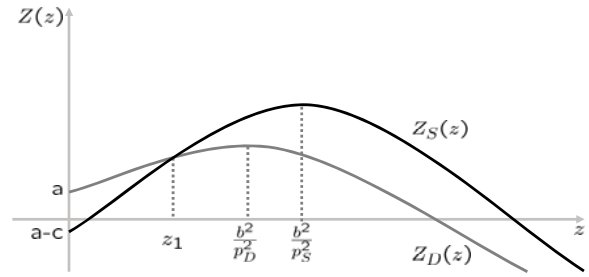


Figure 3. Objective Functions for DS and SOI.

In practice one would insert realistic numbers (e.g. expected values for the planning interval) into the model parameters. In the following we will deduce analytically the conditions under which a long-term investment in a DS and a SOI respectively is optimal. We will achieve this in three steps.

Step 1: Obviously we have, as illustrated in figure 3, $Z_D(z) \geq Z_S(z) \Leftrightarrow 0 < z \leq z_1$, with equality exactly for $z = z_1$. The first derivative $Z'_D(z) = \frac{b}{\sqrt{z}} - p_D$ features its only null at $\frac{b^2}{p_D^2} > 0$. Due to the strict concavity of $Z_D(z)$ the only maximum of $Z_D(z)$ on $]0; z_1]$ is at $z_D^* = \min\left(\frac{b^2}{p_D^2}; z_1\right)$. The associated value of the objective function is

$$Z_D(z_D^*) = \begin{cases} Z_D\left(\frac{b^2}{p_D^2}\right) = a + \frac{b^2}{p_D} & \text{for } \frac{b^2}{p_D^2} < z_1 \text{ and thus } z_D^* = \frac{b^2}{p_D^2} \\ Z_D(z_1) = a - c + 2b\sqrt{z_1} & \text{for } \frac{b^2}{p_D^2} \geq z_1 \text{ and thus } z_D^* = z_1 \end{cases} \quad (7)$$

Step 2: Furthermore we have $Z_S(z) \geq Z_D(z) \Leftrightarrow z \geq z_1$, with equality exactly for $z = z_1$. Since $Z_S(z)$ is strictly increasing on $]0; z_m[$ the maximum can only occur on $[z_m; \infty[$. The first derivative $Z'_S(z) = \frac{b}{\sqrt{z}} - p_S$ for $z > z_m$ features its only null at $\frac{b^2}{p_S^2} > 0$.

- Case 1: $\frac{b^2}{p_S^2} \leq z_m \Rightarrow Z'_S(z) < 0$ for $z > z_m \Rightarrow Z_S(z) = Z_S(z_m) + \int_{z_m}^z Z'_S(t) dt < Z_S(z_m)$ for $z > z_m$. Therefore in this case $z_S^* = z_m$ is the only maximum.
- Case 2: $\frac{b^2}{p_S^2} > z_m \Rightarrow \exists z_0 > z_m$ with $Z'_S(z_0) > 0$ and hence $Z'_S(z) > 0$ for $z_m < z \leq z_0 \Rightarrow Z_S$ has its only maximum on the right side of z_m at the point $\frac{b^2}{p_S^2}$, i.e. $z_S^* = \frac{b^2}{p_S^2}$ in this case.

Due to the strict concavity of $Z_S(z)$ for $z > z_m$ the only maximum of $Z_S(z)$ on $[z_1; \infty[$ is at $z_S^* = \max\left(\frac{b^2}{p_S^2}; z_m\right)$. The associated value of the objective function is

$$Z_S(z_S^*) = \begin{cases} Z_S\left(\frac{b^2}{p_S^2}\right) = a + \frac{b^2}{p_S} \text{ for } \frac{b^2}{p_S^2} > z_m \text{ and thus } z_S^* = \frac{b^2}{p_S^2} \\ Z_S(z_m) = a - c + 2b\sqrt{z_m} \text{ for } \frac{b^2}{p_S^2} \leq z_m \text{ and thus } z_S^* = z_m \end{cases} \quad (8)$$

Step 3: Finally the optimal values of the objective functions have to be compared. We have

$$Z_D(z_D^*) >^{[=]} Z_S(z_S^*) \Leftrightarrow a + \frac{b^2}{p_D} >^{[=]} Z_S(z_S^*),$$

since $Z_D(z_1) = Z_S(z_1) < Z_S(z_S^*)$. This is equivalent to

$$a + \frac{b^2}{p_D} >^{[=]} Z_S(z_m) = a - c + 2b\sqrt{z_m},$$

since $Z_S\left(\frac{b^2}{p_S^2}\right) = a + \frac{b^2}{p_S} > a + \frac{b^2}{p_D}$, which again can only be satisfied (with $c = p_S z_m$) for

$$2b\sqrt{z_m} - \frac{b^2}{p_D} <^{[=]} p_S z_m \Leftrightarrow p_S >^{[=]} \frac{2b\sqrt{z_m} - \frac{b^2}{p_D}}{z_m}.$$

The overall optimum z^* and the associated type of system can then be expressed without loss of generality depending on the value of the parameter p_S relative to the other relevant model parameters

p_D , b and z_m . Substituting for simplicity $\frac{b}{\sqrt{z_m}} = p_1$ and $\frac{2b\sqrt{z_m} - \frac{b^2}{p_D}}{z_m} = p_2$ we can state the overall optimum (in the case $p_S = p_2$ the DS as well as the SOI are optimal).

$$z^* = \begin{cases} \frac{b^2}{p_D^2} \Leftrightarrow p_2 \leq p_S \text{ and } p_1 \leq p_S \Rightarrow \text{DS optimal} \\ z_m \Leftrightarrow p_1 \leq p_S \leq p_2 \Rightarrow \text{SOI optimal} \\ \frac{b^2}{p_S^2} \Leftrightarrow p_S < p_1 \Rightarrow \text{SOI optimal} \end{cases} \quad (9)$$

This can be interpreted as a decision rule determining whether or not the investment into a SOI or into a DS is ex ante economically rational. Clearly, the investment into a DS is optimal when the variable cost of the SOI exceeds a certain threshold relative to the other relevant model parameters. For p_S at this threshold the enterprise would be indifferent on the DS and the SOI (with different capacities each). Below the threshold the SOI is optimal with a capacity – depending on z_m – of either z_m or $\frac{b^2}{p_S^2}$.

From a managerial perspective it is an important aspect that IT investment decisions in this specific context can be based on cost as well as benefits. In general cost for IT infrastructures can be quantified quite well, whereas benefits most often are only roughly estimated. With our model we provide an overall quantification (under certain assumptions). We can moreover examine how input parameters like K , μ or i affect investment decisions. For example the larger the capital K , the higher is b and p_1 respectively. Accordingly, the price p_S is likely to be lower than p_1 and thus investing into SOI becomes more attractive. The same argumentation holds for the risk premium ($\mu - i$).

7 Limitations of the Model and Conclusion

In this paper we demonstrated how the economic value that can be derived from risk/return management calculations can be measured considering an enterprise that has to decide on the amount of capital to be reserved to cover potential losses resulting from a risky investment portfolio. Several assumptions were necessary to achieve an analytical solution. Using the covariance approach as an

example we moreover developed an optimization model that delivers the optimal amount of computing capacity to be allocated to risk calculations at a time. Although covariances are fundamental and widely used in financial applications we covered only one element of numerous risk/return management methods and algorithms. Other approaches and applications for SOI concepts have to be examined as well. In fact, most of the basic principles introduced in this paper can be adapted to more sophisticated and complex scenarios.

From a managerial view, a SOI is especially advantageous when market parameters determining the benefits of risk calculations are highly volatile as could be observed during the crisis since July 2007 resulting in varying demand for computing capacity. With a SOI, resources can be reallocated at any time to reach an economic optimum. As discussed, not only benefits but also (opportunity) cost may vary depending on the total demand for capacity. E.g., during “quiet times” risk calculations may be computed more frequently generating added value out of readily available excess capacity even if benefits are comparably small. On a DS on the other hand an economical allocation cannot be guaranteed. When expected values are applied for capacity planning of a DS the economic optimum is systematically missed when parameters are deviating from expectations. Accordingly, while a DS for risk quantification might be less expensive in “quiet times”, a SOI might deliver the required flexibility during a crisis.

After all, this paper is a contribution to understand the application of service-oriented infrastructures in the specific domain of risk/return management. Although a validation of our findings based on real-world data is still subject to further research, in our point of view, such a systematic and economic analysis is a requirement as a first step for the further development of the new concepts like service-oriented computing or utility computing.

8 References

- Alexander, C. (2008). *Market Risk Analysis Volume II: Practical Financial Econometrics*. Wiley, Chichester.
- Bachelier, L. (1900). *Théorie de la spéculation*. *Annales Scientifiques de l'École Normale Supérieure*, 17 21-86.
- Basel Committee on Banking Supervision. (2004). *International Convergence of Capital Measurement and Capital Standards*. 2009, 12/20, <http://www.bis.org/publ/bcbs107.pdf>.
- Berman, F. (2005). *Grid Computing: Making the Global Infrastructure a Reality*. Wiley, Chichester.
- Bhargava, H.K., Sundaresan, S. (2004). *Computing as Utility: Managing Availability, Commitment, and Pricing Through Contingent Bid Auctions*. *Journal of Management Information Systems*, 21 (2), 201-227.
- Brownlees, C.T., Contini, S., Di Meo, R. and Sullo, V. (2006). *Financial Risk Management Via Multi Mode Inference GRID Applications*. In *Proceedings of Science, 1st International Workshop on Grid Technology for Financial Modeling and Simulation*, Palermo, Italy.
- Buhl, H.U., Fridgen, G. and Hackenbroch, W. (2009). *An Economic Analysis of Service-Oriented Infrastructures for Risk/Return Management*. In *Proceedings of the 17th European Conference on Information Systems (ECIS)*, p. 2060-2072, Verona, Italy.
- Buyya, R., Chapin, S. and DiNucci, D. (2000). *Architectural Models for Resource Management in the Grid*. In *Proceedings of the First IEEE/ACM International Workshop: Grid Computing — GRID 2000*, p. 18-35, Bangalore.
- Chandra, A., Goyal, P. and Shenoy, P. (2003). *Quantifying the Benefits of Resource Multiplexing in On-demand Data Centers*. In *Proceedings of the First Workshop on Algorithms and Architectures for Self-Managing Systems*, San Diego, CA.
- Copeland, T.E., Weston, J.F. (1988). *Financial Theory and Corporate Policy*. 3rd Edition, Addison-Wesley Publishing Company, West Sussex.
- Crespo, D., Gil, J., Gómez, A. and Mota, E. (2006). *Grid Solution for market and counterparty risk calculation. Real life problems from the point of view of the system developer*. In *Proceedings of*

- Science, 1st International Workshop on Grid Technology for Financial Modeling and Simulation, Palermo, Italy.
- Elton, E.J., Gruber, M.J. (1972). Earnings Estimates and the Accuracy of Expectational Data. *Management Science*, 18 (8), B-409-B-424.
- Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50 (4), 987-1007.
- Faisst, U., Buhl, H.U. (2005). Integrated Enterprise Balancing mit integrierten Ertrags- und Risikodatenbanken. *Wirtschaftsinformatik*, 47 (6), 403-412.
- Fama, E.F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38 (1), 34-105.
- Fill, H., Gericke, A., Karagiannis, D., and Winter, R. (2007). Modellierung für Integrated Enterprise Balancing. *Wirtschaftsinformatik*, 49 (6), 419-429.
- Foster, I., Kesselman, C. (1998). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco, CA.
- Gericke, A., Fill, H., Karagiannis, D. and Winter, R. (2009). Situational method engineering for governance, risk and compliance information systems. In *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, DESRIST '09*, (Vaishanvi, V. and Puro, S. Eds.), paper 24, Philadelphia, PA.
- Hull, J., White, A. (1998). Incorporating Volatility Updating into the historical Simulation Method for Value-at-Risk. *Journal of Risk*, 1 (1), 5-19.
- Huther, A. (2003). *Integriertes Chancen- und Risikomanagement - Zur ertrags- und risikoorientierten Steuerung von Real- und Finanzinvestitionen in der Industrieunternehmung*. Deutscher Universitäts-Verlag, Wiesbaden.
- Jackson, P., Maude, D.J., and Perraudin, W. (1998). Back capital and Value at Risk. Bank of England working papers, 79
- Kundisch, D., Löhner, F.M., Rudolph, D., Steudner, M. and Weiss, C. (2007). Bank management using Basel II-data: Is the collection, storage and evaluation of data calculated with internal approaches dispensable? In *Enterprise Risk Management Symposium Monograph*, Chicago, IL.
- Kupiec, P.H. (2007). Financial stability and Basel II. *Annals of Finance*, 3 (1), 107-130.
- Lintner, J. (1965). Security Prices, Risk and Maximal Gains from Diversification. *Journal of Finance*, 20 (4), 587-615.
- Mandelbrot, B. (1972). Correction of an Error in "The Variation of certain speculative Prices". *The Journal of Business*, 45 (4), 542-543.
- Mandelbrot, B. (1963). The Variation of certain speculative Prices. *The Journal of Business*, 36 (4), 394-419.
- Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, 34 768-783.
- Nabrzyski, J., Schopf, J.M., and Weglarz, J. (2003). *Grid Resource Management: State of the Art and Future Trends*. Kluwer, Dordrecht.
- Schröck, G. (2001). *Risk Management and Value Creation in Banks*. Wiley, Hoboken.
- Schumacher, J., Jaeckel, U. and Zimmermann, F. (2006). Grid Services for Derivatives Pricing. In *Proceedings of Science, 1st International Workshop on Grid Technology for Financial Modeling and Simulation*, Palermo, Italy.
- Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19 (3), 425-442.
- Silva, V. (2006). *Grid computing for developers*. Charles River, Hingham, MA.
- Singh, M., Huhns, M.N. (2004). *Service-Oriented Computing. Semantics, Processes, Agents*. Wiley, West Sussex.
- Spremann, K. (2003). *Portfoliomanagement*. 2nd Edition, Oldenbourg, München.
- Taylor, J.W. (2005). Generating Volatility Forecasts from Value at Risk Estimates. *Management Science*, 51 (5), 712-726.
- Wilson, T.C. (1996). Calculating Risk Capital. In *The Handbook of Risk Management and Analysis*, (Alexander, C. Ed.), p. 193-232.
- Wolski, R., Brevik, J., Plank, J.S., and Bryan, T. (2004). *Grid resource allocation and control using computational economies*. Wiley, West Sussex.