Bus Inf Syst Eng 67(4):495–510 (2025)
https://doi.org/10.1007/s12599-024-00890-7

RESEARCH PAPER

HySOR: A Simulation Model for the Sharing of Risk in a Service Level Agreement-Aware Hybrid Cloud

Michael Seifert · Stephan Kuehnel

Received: 3 July 2023 / Accepted: 17 April 2024 / Published online: 13 December 2024

Abstract More and more organizations are considering public cloud services for their business. Functional improvements, innovative features, and strategic factors are further driving demand. The adoption of public cloud services typically involves the integration into existing IT architectures and an established service structure that is ideally aligned with the non-functional requirements of the business processes to be supported. Hybrid cloud providers must be able to accommodate a variety of different public cloud providers while ensuring continuity of service or appropriate compensation prior to implementation. Existing literature focuses on the calculation and simulation of service availability, but less on service credit or business process outage costs of service compositions. In consequence, this paper presents a calculation and simulation model for the concept of “sharing of risk” in Service Level Agreement (SLA)-aware hybrid clouds (HySOR), focusing on the risk-sensitive simulation of the financial impact on hybrid cloud providers and customers. The model was implemented as an R-based application and evaluated with 12 leading experts in the field, yielding interesting implications for theory and practice.

Keywords Risk simulation · Cloud adoption · Hybrid cloud · Service level agreement

Accepted after two revisions by Óscar Pastor.

M. Seifert (✉) · S. Kuehnel (✉)
Chair for Information Systems, esp. Business Information Management, Martin Luther University Halle-Wittenberg, Universitaetsring 3, 06108 Halle, Germany
e-mail: michael.seifert@wiwi.uni-halle.de

S. Kuehnel
e-mail: stephan.kuehnel@wiwi.uni-halle.de

1 Introduction

Public cloud usage has grown steadily in recent years and is forecast to continue to grow even further, with Software as a Service (SaaS) remaining the largest segment (Gartner Inc. 2024). As public cloud SaaS adoptions increase, their integration into existing IT architectures towards hybrid clouds must be properly managed (Sun et al. 2008). However, managing this integration poses a number of challenges for organizations, e.g., changes to contractual commitments, different definitions and formulations of Service Level Agreements (SLA), or the economic evaluation of SLA breaches (Seifert et al. 2023), to name just a few. The various SLAs of different cloud providers offer customers a contractual basis for assessing the service commitment (Aljoumah et al. 2015; Yuan et al. 2015). For this purpose, however, the continuity-relevant components must first be formalized (Seifert 2021) before being aggregated across the various vertical and horizontal integration patterns (Breiter and Naik 2013) and the several cloud architecture levels (Comuzzi et al. 2009) up to the hybrid cloud composition (Theilmann et al. 2010).

An essential aspect of such hybrid cloud architectures is the one-to-many relationship between the (one) provider of public cloud SaaS and the (many) customers, including the provider of the hybrid cloud composition (Pan and Mitchell 2015; Seifert et al. 2023). The possibility of opportunistic behavior of several additional public cloud providers (Pan and Mitchell 2015) leads to increased risks in contractual commitment and partnership, which should be considered in the financial assessment of the desired hybrid cloud architecture (Seifert et al. 2023). Such an architecture often has to deal with different negotiability types of SLAs (Kritikos et al. 2016) and a resulting risk regarding the adequacy of the top-level SLA for the service composition

(Comuzzi et al. 2013). This may further lead to risks in the execution of the business processes being supported.

Assessing the financial risk of failure of external (cloud) services is already reflected in various risk calculations and simulations, as can be seen, for example, in Jiang et al. (2013), Yuan et al. (2015), and Wang and Franke (2020). However, existing models discussed in the literature lack the consideration of uncertainties that go beyond availability parameters, as described as a relevant issue by Franke et al. (2013) and Johnson et al. (2014). In addition, the consideration of business-critical SLA parameters in recent public cloud SLAs is considered a promising extension of risk calculation and simulation approaches (Seifert 2021).

The focus of this paper is on the assessment of hybrid cloud service compositions prior to implementation, which is motivated by a functional or strategic decision (Seifert et al. 2023). The provider – whether internal or external – has to deal with the risk of newly added ("incoming") public cloud services. The decisions on pricing or penalties for failing the top-level SLA must be made in a long-term partnership between the provider and the customer (Benlian et al. 2011; Goo et al. 2009), even if this leads to unacceptability or a "don't do it" decision. In this context, this paper presents a new risk calculation and simulation model and examines the following research questions (RQ):

- RQ1: What are the core elements and business-critical SLA parameters driving risk in hybrid cloud architectures?
- RQ2: How can the impact of simulated IT outages of participating components be taken into account in the risk calculation of hybrid cloud architectures and how does this affect the expected costs for providers and customers?
- RQ3: How can the financial impact for providers and customers be calculated and related to the availability from the service composition's top-level SLA to provide decision support for the sharing of risk?

The further structure of this paper is based on the Design Science Research (DSR) process of Peffers et al. (2007). Accordingly, we first present related work found within this multi-stage research project, addressing the "Objectives of a Solution" phase of DSR in Sect. 2. Next, the calculation model for the sharing of risk for SLA-aware hybrid clouds (HySOR) is presented in Sect. 3, which corresponds to the first part of the "Design & Development" phase. In addition, the simulation aspects are also presented here. The second part (with a stronger focus on development) is represented by the implementation of the HySOR model in an R-based application built on the "Shiny" library, along with the demonstration using a case study in Sect. 4. A survey of 12 leading experts in the field was conducted to obtain a summative evaluation, which is documented in Sect. 5 and represents the "Evaluation" phase of DSR. A second part of this phase can be found in the subsequent Sect. 6, which contains implications for practice condensed from the follow-up expert interviews. The limitations of this work, as well as opportunities for further research, are presented in Sect. 7. The paper closes with a conclusion, answers to the research questions, and a presentation of the contributions to theory and practice in Sect. 8.

2 Theoretical Background and Related Work

In this chapter, we draw on our own previous research (see Seifert et al. 2023), in which we conducted a comprehensive systematic literature review (SLR) of research on hybrid clouds resulting from the adoption of SaaS. The study identified six major challenges in dealing with hybrid clouds resulting from public cloud adoption, which we use as a basis for this conceptualization. It is assumed that further research is needed, especially in the area of modeling business-critical Qualities of Service (QoS), such as service commitment or service credit and their aggregation in hybrid cloud compositions (Seifert et al. 2023). Furthermore, the consideration of business and technological uncertainties as well as the changing commitment of the involved cloud providers serves as a research gap for the extension of the existing knowledge base (Seifert et al. 2023).

For consistent application of the existing literature to the identified research gap, we conducted a second short SLR in preparation for this design cycle with a special focus on risk calculation and simulation of cloud architectures and service compositions. We included research with a focus on the following topics (inclusion criteria): (I) "service downtime", "service outage", and "service failure" to address availability simulation and aggregation. (II) "Service credit" and "service penalty" to reflect business process outages and for calculating the financial impact on the customer side as well. We deliberately do not consider the contractual distinction between penalty and credit, as we see both as a payment from the provider to the (subscription) customer. And (III), "cloud", to ensure that both the characteristics of the different depths of negotiability (risking the business continuity), and multiple cloud component providers involved in the composition are addressed.

We excluded research with a focus on the following: (I) Optimization in terms of performance and cost (e.g., Guérout et al. 2014; Mateo-Fornes et al. 2019), (II) dynamics and SLA (e.g., Al-Ghuwairi et al. 2016; Faniyi et al. 2012), and (III) risk assessment based on transaction history (e.g., Hussain et al. 2010).

Based on the results of the SLRs, the following subsections define and derive both terminology and requirements that are used to conceptualize the HySOR risk calculation and simulation model. The requirements are highlighted in italics at the end of each sub-chapter.

2.1 Cloud Service Components, Cloud Service Compositions, and the Hybrid Cloud Architecture

When describing hybrid cloud architectures, a distinction must be made between individual components and different compositions of components involved in the provision of a cloud service. To this end, we build on the works of Comuzzi et al. (2009), Labidi et al. (2016), and Seifert and Kuehnel (2021), that use architecture modeling to design service compositions in order to formally map hybrid clouds and capture SLA-critical risk aspects (e.g., availability). The related business and technology perspectives are also promising for understanding the impact on economic risks (Comuzzi et al. 2013; Seifert and Kuehnel 2021).

A service component is a single IT service that fulfills a specific function, is required for the execution of a business process, and has an SLA. Service components can occur in any cloud service or cloud deployment model. In contrast, a service composition can be described as a combination of several service components, whereby these are divided into horizontal and vertical components depending on the type of integration (Seifert et al. 2023). Horizontal integration is the interaction of IT services of the same type from different sources (Breiter and Naik 2013), such as two SaaS components, where one is obtained from an existing private cloud and one from a public cloud. In contrast, vertical integration means that one service depends on another or consumes it (Breiter and Naik 2013), such as when SaaS is built on infrastructure as a service (IaaS).

To put it simply, service compositions refer to the various possible combinations of different clouds and can describe, for example, multi-clouds or the hybrid clouds relevant to this work. In addition, the simple use of public cloud services alongside the existing IT landscape usually already leads to hybrid clouds, whereby the individual components differ in terms of their negotiability (e.g., availability). Since our risk model is applicable to both types of cloud deployments, we use the term cloud service composition as a generic term.

HySOR has to consider an architectural model consisting of (i) service components that may represent existing IT components used by an organization and (ii) incoming public cloud services that result in service compositions that support the business process.

2.2 Customer and Provider in Different Roles

There are also organizational aspects of hybrid cloud architectures that need to be taken into account and can become a challenge, e.g., with regard to the different roles. “These challenges become even more complex when cloud providers take on the dual role of provider and customer, for example, when their own cloud service offerings build on the services of external cloud providers” (Seifert and Kuehnel 2021). Regardless of whether it is an internal or external cloud service provider, someone needs to be responsible for the top-level SLA and discuss the risk of business process outages with their customers (Seifert and Kuehnel 2021).

Subscription is a very common pricing model for public cloud services (Mazrekaj et al. 2016). “With subscription pricing, users pay on a recurring basis to access software as an online service or to benefit from a service” (Mazrekaj et al. 2016). The role distinction in this model is crucial for determining who holds the subscription to the cloud (Zhang and Zhou 2009). Subscription in the open architecture of cloud computing consists, among other things, of a process and the roles (Zhang and Zhou 2009). Whoever subscribes to the cloud service is a contractual partner, which is also a decisive risk element for the calculation of penalties.

HySOR has to consider (i) the hybrid cloud customer and (ii) the hybrid cloud provider as the top-level contracting parties, with (iii) one of them subscribing to each of the cloud services involved in the service composition.

2.3 Service Level Agreement and Operational Level Agreement Formalization

A core facet of hybrid cloud architecture involves SLA and operational level agreement (OLA) aspects (Seifert et al. 2023). To quantify availability as a risk metric, we need to formalize SLA/OLA parameters, including their measurable and calculable parameters called QoS (Suakanto et al. 2012).

Availability is one of the most important attributes of cloud service quality, and most popular public cloud services claim their availability promise (Baset 2012; Gulia and Sood 2013; Seifert 2021). Availability is often described not only by a number but also by different parameters.

The categories of Yuan et al. (2015) describe this appropriately for our context. The first parameter is the measurement period, usually defined as one month. The service granularity defines which service scope is meant, and the time granularity is usually specified in the form of 1, 5, or 10 min. The coverage describes what must be running correctly and which services must be included. In addition, exclusions of unavailability are commonly defined.

Yuan et al. (2015) provide a useful formalization of the penalty function: a) "total charge ratio", b) "fixed value at different violation levels", or c) "downtime ratio". We can also see the public cloud penalty in the "service credit" category of Seifert (2021). Moreover, we can find the typical service penalty calculation for public cloud SaaS in Seifert (2021) as a combination of the methods from Yuan et al. (2015). An example is the software company SAP with the ratio of total charge and the ratio of downtime: "per 1% below availability (99.5) you get a credit of 2% of your monthly fee" (Seifert 2021). In addition, the maximum credit volume is given, which is also crucial for risk assessment.

Another important aspect of the SLA is its negotiability (Comuzzi et al. 2013; Kritikos et al. 2016), which only has a secondary effect on the formalization, i.e., as a formulation of fixed parameters in public cloud SLAs (Seifert 2021). The negotiability of the SLA plays a decisive role in the description of the "sharing of risk" in Sect. 2.6.

HySOR has to (i) simulate service continuity based on availability commitment parameters and (ii) enable integrated calculation of typical penalty functions from cloud SLAs so that these (iii) can be aggregated in the service composition.

2.4 Uncertainty as a Risk Driver

Uncertainty has to be taken into account in hybrid cloud architectures (Johnson et al. 2014). In the category "uncertainty in technology and business", Seifert et al. (2023) distinguish three dimensions of risk or uncertainty in this context. First, tangible risks, such as availability (Paquette et al. 2010), can be represented as a probability distribution, e.g., from the QoS history (Johnson et al. 2014). Second, there are intangible risks when the business process depends on an external cloud element (Paquette et al. 2010). This may lead to opportunistic behavior by the public cloud provider due to a one-to-many relationship in the hybrid cloud context (Pan and Mitchell 2015). Third, uncertainty regarding the knowledge of the architecture supporting the cloud compositions and business processes (Franke et al. 2013; Johnson et al. 2014; Rockmann et al. 2014) is an intangible risk driver and must also be considered.

HySOR has to consider (i) tangible risks, (ii) intangible risks arising from possible opportunistic behavior of the cloud component providers involved, and (iii) intangible risks arising from the business technology architecture.

2.5 Related Work on Risk Calculation and Simulation

Within the SLRs, we found three related papers on risk calculation or simulation models based on architectures with strong relevance to our context.

Yuan et al. (2015) motivate their approach with the lack of clarity in availability commitment and penalty for cloud consumers, and the business model for cloud providers to find the optimal penalty level. The tripartition of possible penalty functions, i.e., (I) ratio of total charge, (II) fixed value, and (III) ratio of downtime, seems promising for the development of risk calculation and simulation models. The financial impact of customer cost as provider price together with the impact of downtime appears reasonable. The important finding that the provider will reduce the penalty in order to compensate for higher availability requirements or claims leads us to the assumption made later in this paper that better risk sharing (e.g., more equally shared risks) is a gap in research to date. Due to the lack of consideration of the composition (and the composition provider), i.e., the combination of several components and their aggregation, this risk calculation is not adequate for our context with the above-mentioned concepts.

Jiang et al. (2013) motivate a QoS-based risk approach combined with business-oriented target monitoring. The five-stage procedure for finding and parameterizing a suitable failure probability distribution confirms the necessity of modeling based on distributions. In particular, parameterization is an interesting aspect of risk sensitivity of assumed failure probabilities in order to test decision support in variants. The lack of focus on the provider's revenue and loss of revenue (e.g., by focusing only on reducing usage) is an incentive for developing new methods.

Wang and Franke (2020) present a model for analyzing IT service outages for individual organizations and supply chains. An important part of their model is the cost of business process downtime (i.e., economic impact on the customer side) in terms of three function types: constant, linear, and quadratic. Another useful aspect is the frequency of breakdowns/downtime, represented by a Poisson arrival model – as found in the literature. On this basis, the downtime duration is modeled with a lognormal distribution, as is often used when modeling downtimes. The lack of consideration of uncertainties arising from business and technology (e.g., support of business processes by IT, see Sect. 2.4) is a point that requires adjustment. Furthermore, the consideration of today's public cloud penalty functions for compensation on the customer side and costs on the provider side can play an important role.

HySOR needs to (i) consider different business process downtime cost functions and demonstrate the financial *implications for (ii) the hybrid cloud composition customer and (iii) the provider.*

2.6 The Concepts “Sharing of Risk” and “Zone of Possible Agreements”

Finally, our approach also differs from related work as we do not aim at an economic optimization model for cloud providers. For example, Wang and Franke (2020) state that their “intuition behind the model is that capital K can buy better hardware, thereby reducing the frequency of downtime”. HySOR, however, does not primarily focus on investment opportunities to reduce downtime, as we assume a typical “take it or leave it” scenario for public cloud adoption, as described by Comuzzi et al. (2013). Combined with the assumption of a desired partnership between the hybrid cloud customer and provider, this leads to the need to appropriately judge the hybrid cloud architecture. The following model therefore focuses on the newly proposed concept “sharing of risk”. This concept opens up a negotiation corridor that allows hybrid cloud customers and providers to account for tangible and intangible decision factors by varying risk parameters. The sharing of risk reflects the range of risky financial consequences for the provider and the customer of a cloud service composition. To this end, we rely on one of the most well-known descriptive negotiation concepts – Raiffa (1982)’s Zone of Possible Agreements (ZOPA) (see Fig. 1). With this concept, Raiffa (1982) represents a two-person distributive bargaining problem that is bounded by the parties’ reservation prices (respectively their best cases/alternatives) (Ahlert and Sträter 2016).

If the buyer’s reservation price is greater than that of the seller, there is a ZOPA containing a possible agreement value (Ahlert and Sträter 2016; Raiffa 1982). If the reservation prices of both parties are the same, there is one possible point of agreement, whereby in this case the gains for both parties are zero. If the seller’s reservation price is higher than the buyer’s, there is no ZOPA (Ahlert and Sträter 2016).

Applied to the context of this study, we imagine a situation in which a customer and a provider of a hybrid cloud service composition enter into negotiations about risky financial consequences (instead of prices). The situation is more complex than in Raiffa (1982)’s original model of price negotiation, as the risky financial consequences for both parties are influenced by various aspects.

Two main aspects lead to increased risk and require sharing as part of a long-term partnership between hybrid cloud provider and customer. First, the non-negotiability of the incoming public cloud components involved in the composition increases the risk of an outage. Second, the underlying penalty functions of these components may be insufficient to compensate for the financial consequences in the worst case. For example, passing the subscription for a cloud component on to the provider could reduce the provider’s risk, while increasing the customer’s (financial) risk of process outages. However, this could be taken into account when negotiating the service credit or the pricing for the service composition between the hybrid cloud customer and the provider.

HySOR must consider risk-sensitive simulation parameters to predict, evaluate, and compare variants of the desired hybrid cloud composition architecture.

3 The Risk Calculation and Simulation Model HySOR

In the following, HySOR and its different elements are described concerning the calculation and the simulation model, which ensures the transparency and comprehensibility of our DSR project. Moreover, a detailed description of the artifact allows the fields of future research presented in Sect. 6 and 7 to be addressed in a targeted manner. For example, other scientists can exchange specific simulation components or integrate additional calculation modules.

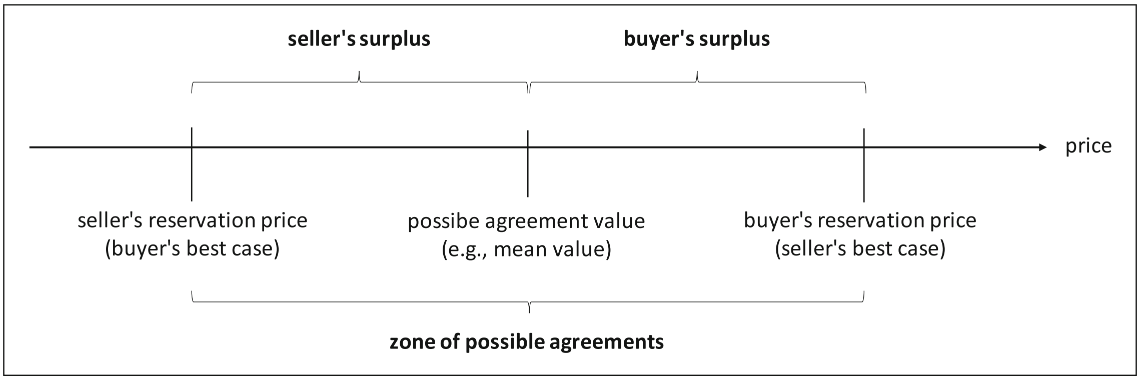

seller's surplus
buyer's surplus
price
seller's reservation price (buyer's best case)
possibe agreement value (e.g., mean value)
buyer's reservation price (seller's best case)
zone of possible agreements

Fig. 1 Raiffa’s descriptive negotiation concept (illustration adapted from Ahlert and Sträter (2016)

3.1 Core Elements and Characteristics

The core elements of HySOR are a service composition s consisting of service components i that are involved and functionally necessary to support the business process and its execution. The model refers to m months of a contract duration M. The QoS of the participating service components is then simulated for all hours (n) of this contract duration and then aggregated to the QoS of the service composition.

Core elements:

I: nonempty finite set of service components

$i \in I$: a service component from the set of I

$s \subseteq I$: a service composition representing a subset of I (i.e., consisting of i= 1,2,3,...,q service components)

$m \in \{1,2,3,\ldots,M\}$: month (m) of a contract duration (M) of the service composition (s)

$n \in \{1,2,3,\ldots,720 * M\}$: hour (n) of a contract duration (M) of the service composition (s)

The basic characteristics of the service components include the subscription ownership, the monthly cost of the service, the availability per month guaranteed in the SLA, and the penalty amount for missing the SLA (see the listed service component characteristics below). The tangible risks are modeled using the risk parameters lambda for the Poisson arrival of the downtime and the expected value and variance of the lognormal distribution of the downtime duration. The specific characteristics of these parameters can either be derived through benchmarking and from historical data (e.g., QoS history, as described in Sect. 2.4) or must be estimated by experts. As also described in Sect. 2.4, insufficient knowledge of the business technology architecture and its relevance for business process support is an intangible risk driver that needs to be considered. Ambiguities in the dependency of the business process on a component are therefore taken into account as a risk factor and are included in the model as a security perception parameter. The service components also feature downtime simulation and monthly availability aggregation capabilities. Finally, there are different types of penalty functions that can be modeled for each component.

Service component characteristics:

o_i: type of subscription ownership of the service component (i)

c_i: monthly costs for the service component (i)

s_i: in the SLA guaranteed availability of the service component (i) per month

pa_i: penalty amount for the service component (i) for missing the SLA

λ_i: expected value and variance for the Poisson arrival of the downtime of a service component (i)

μ_i: expected value of the downtime duration of service component (i)

σ_i: standard deviation of the downtime duration of service component (i)

$\mu = \ln\left(\frac{\mu_i^2}{\sqrt{\mu_i^2+\sigma_i^2}}\right)$: μ parameter for the lognormal distribution

$\sigma^2 = \ln\left(1+\frac{\sigma_i^2}{\mu_i^2}\right)$: σ^2 parameter for the lognormal distribution

$id_i = \begin{cases} 0, if\ i\ dispensable \\ 1, else \end{cases}$: indispensability of the service component (i) for the business process

$ri_i \in [0,1]$: risk factor regarding the indispensability of the service component (i) for the business process

$e_i = id_i * (1 - ri_i)$: security perception parameter for the indispensability of the service component (i) for the business process

$da_{i,n} = f(\lambda_i)$: simulated downtime arrival of the service component (i) per hour (n)

$d_{i,n} = f(da_i, \mu, \sigma)$: simulated downtime duration of the service component (i) per hour (n)

$a_{i,m} = f(d_{i,m})$: calculated availability of the service component (i) in percent per month (m)

$p_{i,m} = f(a_{i,m}, s_i, pa_i)$: penalty function for the service component (i) for missing the SLA per month (m)

The service composition has specific characteristics, such as the monthly cost for the customer, which is equivalent to the composition revenue for the provider. The service composition itself has a defined availability in the so-called top-level SLA, with which the aggregation of the availabilities of the components involved is later compared to measure SLA compliance. Analogous to the components, the composition has a penalty amount and a corresponding penalty function.

Another essential dimension of the service composition characteristics is the modeling of business risk (and related costs). According to our concept of "sharing of risk", the risk of penalties on the provider's side must be balanced with the risk of business process outages on the customer's side, both of which are connected via the service composition and corresponding penalty and cost functions. The business process outage cost function types for compositions are either constant, linear, or quadratic, as described in Sect. 2.5.

Service composition characteristics:

c_s: monthly costs for the service composition (s)

s_s: in the (top-level) SLA guaranteed availability of the service composition (s) per month

pa_s: penalty amount for the service composition (s) for missing the SLA

$d_{s,n} = f(d_i)$: aggregated downtime duration of the service composition (s) per hour (n)

$a_{s,m} = f(d_s)$: calculated availability of the service composition (s) in percent per month (m)

$p_{s,m} = f(a_{s,m}, s_s, pa_s)$: penalty amount for service composition (s) for missing the SLA per month (m)

$b_n^{type} = f(d_s, e_i)$: business process outage cost function type, depending on the unavailability of the service composition (s) per hour (n)

$b_m = f(b_n)$: business process outage costs caused by the unavailability of the service composition (s) per month (m)

3.2 Calculation and Simulation Procedure

As already mentioned above, we assume a Poisson arrival process as the basis for the availability simulation of service components, in line with Wang and Franke (2020). The downtime duration is simulated separately for each service component using a lognormal distribution.

Remark 1 *We assume that service components having a dedicated SLA also have independent outages, as they each have differentiated architectures and resilience procedures (corresponding to the SLA offered).*

Downtime arrival per service component per hour:

$$da_{i,n} = da_{\lambda_i} = f(n) = \frac{(\lambda)^n}{n!} e^{-\lambda},$$

$$\forall n \in \{1, 2, 3, \ldots, 720 * M\}$$

Downtime duration per service component per hour:

$$d_{i,n} = f\left(da_{i,n}\right) = \frac{1}{da_i * \sigma * \sqrt{2\pi}} \exp\left(-\frac{\left(\ln(da_i) - \mu\right)^2}{2\sigma^2}\right),$$

for $da_{i,n} > 0$,

$$\forall n \in \{1, 2, 3, \ldots, 720 * M\}$$

The aggregation of the service components into a service composition with regard to the downtime duration is determined by the component with the maximum downtime per hour (across all components). As part of the later availability aggregation, the values are then summed up for the month (with 720 hours per month).

Remark 2 *We assume that each month consists of 30 days of 24 hours each and that a standard contract year comprises 12 months. This is based on the assumption that business process outages have the same financial impact each month, as the associated costs are independent of the occurrence of downtime during the contract period.*

Downtime duration aggregation per service composition per hour:

$$d_{s,n} = f\left(d_{i,n}\right) = \max\ \left(d_{i,n}\right),$$

$$\forall i \in \{1, 2, 3, \ldots, q\},$$

$$\forall n \in \{1, 2, 3, \ldots, 720 * M\}$$

The aggregation of the downtime duration per hour is also crucial for determining the amount of the business process outage costs. This is because two or more components can fail simultaneously, resulting in the same service composition outage.

Remark 3 *We have a limitation in the case where a downtime arrival coincides with the downtime duration of an earlier downtime arrival. Due to a moderate arrival rate per hour, we assume this to be negligible.*

The business process outage costs are calculated per hour of the contract duration and subsequently aggregated for each month. The penalties are calculated for each component and for the composition based on the respective availability aggregation. This paper illustrates the calculation of the monthly penalty.

Availability aggregation per service component per month:

$$a_{i,m} = f\left(d_{i,n}\right) = \frac{720 - \sum d_{i,n}}{720},$$

$$\text{for } 0 < n < 721 \rightarrow m = 1, \text{ for } 720 < n < 1441 \rightarrow m = 2, \ldots, \rightarrow m = M$$

Availability aggregation per service composition per month:

$$a_{s,m} = f\left(d_{s,n}\right) = \frac{720 - \sum d_{s,n}}{720},$$

for $0 < n < 721 \rightarrow m = 1$, for $720 < n < 1441 \rightarrow m = 2, \ldots, \rightarrow m = M$

Business process outage costs per hour:

$$b_{da}^{constant} = f(d_s) = b_i^{a,e}, \quad \text{if} \quad d_{s,n} > 0,$$

$$b_{da}^{linear} = f(d_s) = d_{s,n} * b_i^{a,e},$$

$$b_{da}^{quadratic} = f(d_s) = {d_{s,n}}^2 * b_i^{a,e},$$

$$\forall n \in \{1, 2, 3, \ldots, 720 * M\}$$

Business process outage cost aggregation per month:

$$b_m = f(b_{da}) = \sum b_{da},$$

$$\text{for } 0 < n < 721 \rightarrow m = 1, \text{ for } 720 < n < 1441 \rightarrow m = 2, \ \ldots, \rightarrow \ m = M$$

Penalty calculations per month:

$$p_{k,m}^{\mathrm{fix}} = f(a_{k,m}, s_k) = a_{k,m} < s_k \rightarrow pa_i, \mathrm{else} \rightarrow 0,$$

$$p_{k,m}^{\mathrm{percentage}} = f(a_{k,m}, s_k, c_k) = \max(2c_k \lceil s_k - a_{k,m} \rceil, c_k),$$

$$p_{k,m}^{\mathrm{none}} = f(a_{k,m}, s_k, c_k) = 0,$$

$$k \in \{i, s\},$$

$$\forall m \in \{1,2,3, \ldots, M\}$$

Following the previous calculations at the levels of components, compositions, and business processes, the financial impact for the customer ($totalcosts_{cust}$) and the provider ($totalcosts_{prov}$) of the service composition can be determined. This is done by distinguishing between fixed costs, which are incurred on a monthly basis regardless of the simulated availability of the components, and variable costs, which are composed of business process costs and/or the penalties for the service components and composition.

Fixed costs per month:

$$fixcosts_{cust,m} = f(c_s, c_i) = c_s + \sum_{i=0}^{n} c_i,$$

$$\forall o_i \in \{customer\,public\,cloud\,subscription\},$$

$$fixcosts_{prov,m} = f(c_s, c_i) = -c_s + \sum\nolimits_{i=0}^{n} c_i,$$

$$\forall o_i \in \{provider\,public\,cloud\,subscription,\ provider\,private\,cloud\,component\},$$

$$\forall m \in \{1,2,3, \ldots, M\}$$

Remark 4 *For reasons of comprehensibility, the fixed costs of the composition* (c_s) *that a customer pays to the provider were modeled in the provider's calculation with the same variable, but with the mathematical sign reversed, i.e. as negative costs* ($-c_s$)

Variable costs per month:

$$varcosts_{cust,m} = f(p_{i,m}, p_{s,m}, b_m) = -p_{i,m} - p_{s,m} + b_m,$$

$$\forall o_i \in \{customer\,public\,cloud\,subscription,\ provider\,private\,cloud\,component\}$$

$$varcosts_{prov,m} = f(p_{i,m}, p_{s,m}) = -p_{i,m} + p_{s,m},$$

$$\forall o_i \in \{provider\,public\,cloud\,subscription\},$$

$$\forall m \in \{1,2,3, \ldots, M\}$$

Remark 5 *For reasons of comprehensibility, penalty payments for the service composition* ($p_{s,m}$) *that a provider pays to the customer for missing the SLA were modeled in the customer's calculation with the same variable, but with the mathematical sign reversed, i.e. as negative variable costs* ($-p_{s,m}$).

Financial impact calculations per month:

$$totalcosts_{cust,m} = fixcosts_{cust,m} + varcosts_{cust,m},$$

$$totalcosts_{prov,m} = fixcosts_{prov,m} + varcosts_{prov,m},$$

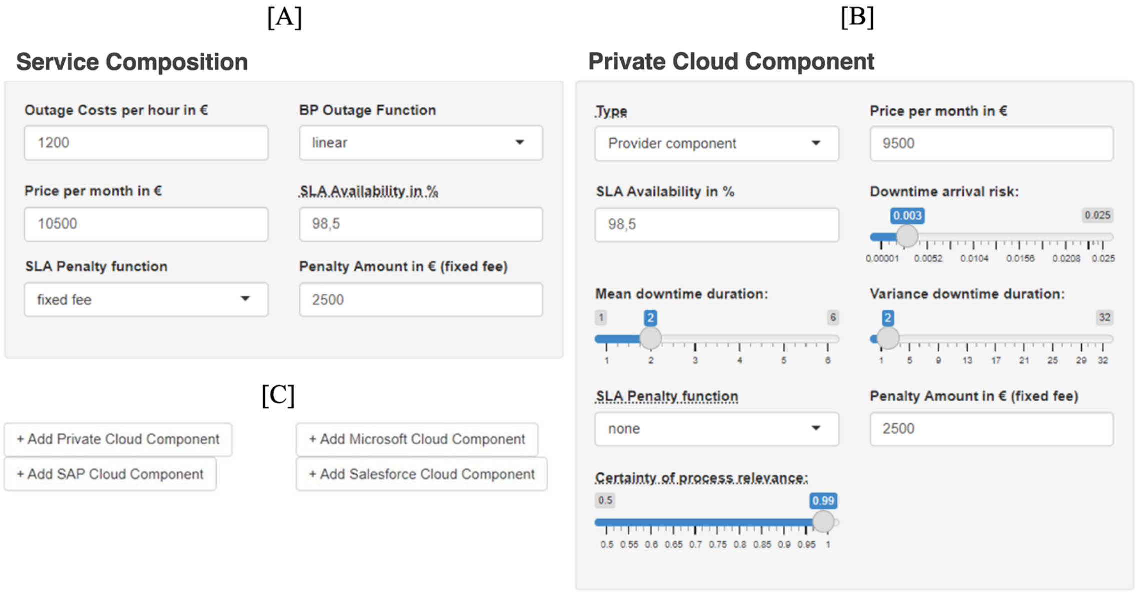
$$\forall m \in \{1,2,3, \ldots, M\}$$

The financial impact per month is the sum of the monthly variable and fixed costs. The implicit connection between the aggregated availability of the service composition and the financial impact on the customer and the provider serves as a basis for decision support regarding the sharing of risk. Using the fitted total cost graph for the customer and the provider, interesting considerations arise concerning the decision-making and negotiation options of both parties, as the following case study shows.

4 Development and Demonstration

HySOR was implemented using the integrated development environment RStudio/2023.03.1, with which an R-based application was developed that builds on the library "Shiny". The modeling of the Poisson arrival of service outages was realized by the R function rpois, the lognormal distribution of the downtime duration by rlnorm.

The input fields are distributed over three areas of the user interface, as shown in Fig. 2. In the first, the "Service Composition" panel (see Fig. 2 [A]), the characteristics of the service composition are represented by business process variables and SLA parameters at the top level. In the second area, the "<Type> Component" panel, the characteristics of each service component involved are parameterized, as shown in Fig. 2 [B] using the example of a private cloud component. The four "Add Component" buttons (see Fig. 2 [C]) are used to add an additional component to the composition as a third input area. The addition of a typical public cloud component was a requirement from an early test phase of the prototype and was implemented to strengthen the demonstrability of the model. SAP, Microsoft, and Salesforce were chosen as suitable examples for the expected case studies due to the



[A]

Service Composition

Outage Costs per hour in €: 1200

BP Outage Function: linear

Price per month in €: 10500

SLA Availability in %: 98,5

SLA Penalty function: fixed fee

Penalty Amount in € (fixed fee): 2500

[C]

+ Add Private Cloud Component

+ Add Microsoft Cloud Component

+ Add SAP Cloud Component

+ Add Salesforce Cloud Component

[B]

Private Cloud Component

Type: Provider component

Price per month in €: 9500

SLA Availability in %: 98,5

Downtime arrival risk: 0.003

Mean downtime duration: 2

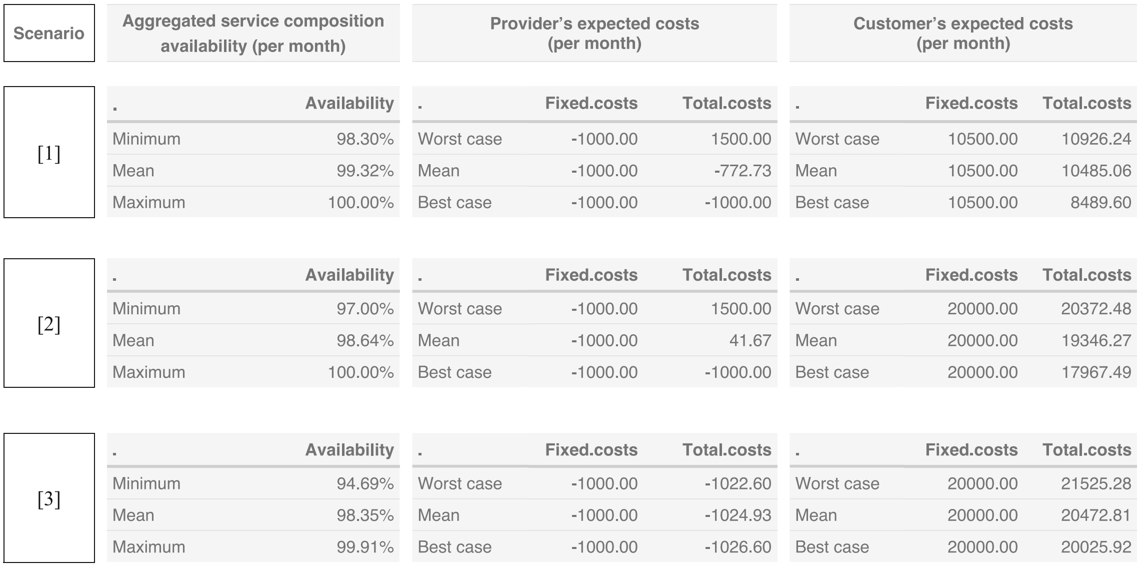
Variance downtime duration: 2

SLA Penalty function: none

Penalty Amount in € (fixed fee): 2500

Certainty of process relevance: 0.99

Fig. 2 Screenshot of the input fields of the HySOR prototype: Service Composition [**A**], < Type > Component [**B**], and Add Component [**C**]



Scenario	Aggregated service composition availability (per month)		Provider's expected costs (per month)			Customer's expected costs (per month)		
	.	Availability	.	Fixed.costs	Total.costs	.	Fixed.costs	Total.costs
[1]	Minimum	98.30%	Worst case	-1000.00	1500.00	Worst case	10500.00	10926.24
	Mean	99.32%	Mean	-1000.00	-772.73	Mean	10500.00	10485.06
	Maximum	100.00%	Best case	-1000.00	-1000.00	Best case	10500.00	8489.60
	.	Availability	.	Fixed.costs	Total.costs	.	Fixed.costs	Total.costs
[2]	Minimum	97.00%	Worst case	-1000.00	1500.00	Worst case	20000.00	20372.48
	Mean	98.64%	Mean	-1000.00	41.67	Mean	20000.00	19346.27
	Maximum	100.00%	Best case	-1000.00	-1000.00	Best case	20000.00	17967.49
	.	Availability	.	Fixed.costs	Total.costs	.	Fixed.costs	Total.costs
[3]	Minimum	94.69%	Worst case	-1000.00	-1022.60	Worst case	20000.00	21525.28
	Mean	98.35%	Mean	-1000.00	-1024.93	Mean	20000.00	20472.81
	Maximum	99.91%	Best case	-1000.00	-1026.60	Best case	20000.00	20025.92

Fig. 3 Screenshot showing simulation results of the HySOR prototype: monthly aggregated service composition availability and expected costs for provider and customer in three different scenarios

varying nature of their SLAs in terms of service commitment and service credit. The fourth button can be used to add a private cloud component that may already be part of the existing IT architecture.

Remark 6 *An extension with any pre-parameterized buttons is conceivable. Furthermore, the added components – regardless of their name – can be parameterized to describe the targeted hybrid cloud scenario as accurately as possible.*

The prototypical demonstration of HySOR involves three scenarios, each of which is based on a contract term of one year, i.e., 12 months. For the calculations, one run per hour was simulated, which resulted in a total of 8640 simulation runs over the course of a year (24 hours/day * 30 days/month * 12 month/year) for each scenario. The simulation results are likewise displayed in three areas (columns) of the user interface, as shown in Fig. 3. The first column shows the minimum, mean, and maximum monthly availability of the composition for the contract year and is referred to as "Aggregated service composition

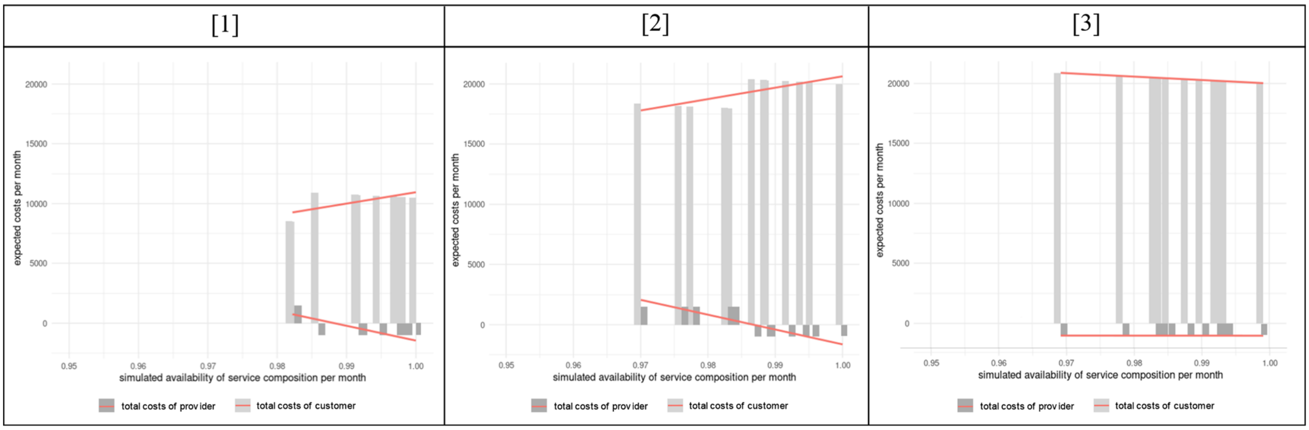


Fig. 4 Screenshot showing the output diagrams of the HySOR prototype: comparison of the provider's and customer's monthly financial impact depending on availability in the case study's three scenarios

availability (per month)". The second column shows the financial impact of the provider and is called "Provider‘s expected costs (per month)". There, a distinction is made between fixed costs and total costs to enable the effects of the different cost types to be seen at a glance. The tabular display of the best (in Fig. 3 called "Best case"), average (in Fig. 3 called "Mean"), and worst months (in Fig. 3 called "Worst case") gives the user a quick overview of the provider's simulation results. The third column, "Customer's expected costs (per month)", is used to show the financial impact on the customer and is structured in the same way as column 2. Thus, the changes in the parameterization and its effects on both parties can be seen in a nuanced way, which serves to ensure transparency and provides a basis for negotiating the sharing of risk.

Remark 7 *Figure 3 should not be read row by row, but column by column within the respective scenarios. In other words, the minima, mean values, and maxima of the availabilities do not necessarily lead to the best cases, mean values, or worst cases of providers' and customers' costs.*

The output diagrams displayed in the third area of the user interface can also be used for this purpose (see Fig. 4), allowing different scenarios with different parameterizations to be compared visually. Each output diagram shows the provider's and customer's financial impact depending on availability given a specific scenario.

A deliberately straightforward case study was carried out to illustrate the application of the HySOR approach and also as an example for the subsequent evaluation. In the initial situation (scenario [1]), a service composition is modeled with a linear business process outage cost function (see Fig. 2 [A], "BP Outage Function" = linear) and €1200 outage costs per hour (see Fig. 2 [A], "Outage Costs per hour in €" = 1200). The price for the service composition is €10500 per month (see Fig. 2 [A], "Price per month in €" = 10500). The penalty is a fixed fee (see Fig. 2 [A], "SLA Penalty function" = fixed fee) of €2500 (see Fig. 2 [A], "Penalty Amount in €" = 2500) if the availability QoS of 98.5% of the top-level SLA (see Fig. 2 [A], "SLA Availability in %" = 98.5) is not met. In this initial scenario [1], the service composition consists of exactly one service component (private cloud component), which is also modeled with an availability SLA of 98.5% (see Fig. 2 [B], "SLA Availability in %" = 98.5). The costs for the component are €9500 per month (see Fig. 2 [B], "Price per month in €" = 9500). The security perception parameter for the indispensability of the private cloud component for the business process is set to a default value of 99% (see Fig. 2 [B], "Certainty of process relevance" = 0.99).

The simulation results for scenario [1] can be seen in tabular form in Fig. 3, line [1], and visually in Fig. 4, diagram [1]. If the simulated availability of the private cloud component is 98.5% or higher, the top-level availability SLA is met, which in the best case generates a monthly profit of €1000 for the provider (€9500 – €10500 = – €1000, see also Fig. 3, scenario [1], column 2: Total.costs (Best case) = –1000).

Remark 8 *Positive financial impacts, i.e., profits, are currently represented as negative costs in the model and prototype.*

If the availability drops below 98.5%, the fixed fee penalty of €2500 applies, resulting in a deficit of €1500 for the provider (€9500 + €2500 – €10500 = €1500, see also Fig. 3, scenario [1], column 2: Total.costs (Worst case) = 1500). If the composition does not achieve 100% availability, the business process outage cost function for the customer applies. In scenario [1], the worst case for the customer is an availability of 98.5%, since the provider does not yet have to pay the fixed fee penalty, but the customer has business process outage costs due to 1.5% unavailability that remain uncompensated (see Fig. 3, scenario [1], column 3: Total.costs (Worst case) = 10926.24). The customer's best case is an interesting combination of the payment of the fixed fee penalty (which the customer receives from the provider) minus the customer's business process outage costs. Interestingly, in this case the customer's business process outage costs are overcompensated by the provider's penalty payment, so that the total costs for the customer are even lower than its actual fixed costs. Thus, in scenario [1], the customer could plan with the average total costs (see Fig. 3, scenario [1], column 3: Total.costs (Mean) = 10485.06) being slightly below the fixed costs (see Fig. 3, scenario [1], column 3: Fixed.costs (Mean) = 10500).

Figure 4, diagram [1] illustrates the monthly financial impact depending on the composition's monthly availability for the provider (dark gray bars) and for the customer (light gray bars). In addition, the lower linear trend line shows the fitted financial impact for the provider and the upper linear trend line that of the customer.

Remark 9 *In the output diagrams of Fig. 4, the bars represent the financial impact for one month (both the dark gray bars for the provider and the light gray bars for the customer). Due to the reduced scaling in the user interface, the bars partially overlap, which is why thicker bars represent several months.*

In the diagram of scenario [1], the linearly fitted trend line shows that as the availability of the service composition decreases, the customer's total costs tend to decrease (despite business process outage costs) and the provider's total costs tend to increase (which may be okay given the nature of its business). Nevertheless, the provider is likely to be motivated to reduce its average cost. This could be done by adjusting the technical design of the composition or by changing financial parameters such as price or penalties in a service composition negotiation.

During the development of the prototype, it proved useful to compare the results of an initial state, as illustrated in Fig. 4, diagram [1], with the simulation results from one or more possible target states (see Fig. 4, diagrams [2] and [3]).

In the further course of the case study, i.e. in scenario [2], a change was needed due to the customer's functional request to add a public cloud service to the service composition. This was done ad hoc in the application with the button "Add SAP Cloud Component". The result of this adjustment can be seen in Fig. 3 [2] and Fig. 4 [2]. The SAP Cloud Component is added as a "customer subscription", i.e., the cost of €9500 is added to the customer's monthly fixed costs, and any service credit for this component (based on a new, independent SLA for this service) is paid to the customer. At the same time, no changes are made to the parameterization of the service composition in Fig. 2 [A]. Nevertheless, the provider is now responsible for a service composition consisting of two components, one of which it cannot control and for which it also receives no financial compensation. It follows that the provider in scenario [2] is on average worse off (with average total costs of €41.67) than in scenario [1] (with average negative total costs of –€772.73, i.e. profit) (see also Fig. 3, scenario [2], column 2: Total.costs (Mean) = 41.67 compared to Fig. 3, scenario [1], column 2: Total.costs (Mean) = –772.73). The situation of the customer in scenario [2] must be considered in more detail. Although his fixed costs for the new SAP cloud component increase by €9500 to a total of €20000 (€10500 for the existing composition + €9500 for the SAP cloud component), this is done at the customer's own request for functional reasons. Nevertheless, the customer has an average total cost of €19346.27 (see Fig. 3, scenario [2], column 3: Total.costs (Mean) = 19346.27), which is due to the lower average availability and the corresponding higher frequency of fixed fee penalties paid by the provider. The result is that the customer is better off (considering the SAP cloud component he wants) since his average total costs are lower than his fixed costs.

In scenario [3], we now show the variant in which the provider takes the subscription and thus increases the service composition price to a total of €20000. In addition, the penalty function of the service composition is changed from a fixed fee to a percentage-based approach, which may be an improvement for the partnership mentioned in several expert interviews. For scenario [3], a clear change can be seen in the linearly fitted trend lines in Fig. 4, diagram [3], which flatten out for both parties. On average, the provider in scenario [3] would improve its financial situation (see Fig. 3, scenario [3], column 2: Total.costs (Mean) = –1024.93, i.e. profit) compared to scenario [2] (see Fig. 3, scenario [2], column 2: Total.costs (Mean) = 41.67) and the customer would have to accept average total costs that are above his fixed costs (see Fig. 3, scenario [3], column 3: Total.costs (Mean) = 20472.81).

With knowledge of the two possible target states (Fig. 4 [2] and Fig. 4 [3]) and given that the customer has additional functional requirements compared to the initial scenario [1], the sharing of risk can be negotiated, taking into account the financial impact on both parties. Compared to the initial scenario [1] (the provider makes a profit on average; the customer has average total costs that are below the fixed costs), the provider in scenario [2] and the customer in scenario [3] are on average worse off. One possibility for a partnership agreement in the course of the desired functional extension could be a price reduction of the composition. Taking scenario [3] as a starting point, a reduction in the composition price to, for example, €19750 would bring the average financial impact for the provider and customer closer to a ratio similar to that in scenario [1]. However, there are other factors to consider. For example, the provider may want to include the additional administrative effort for the added component in the calculation. In addition, the customer must consider the possibility of business process outages as a result of the functional extension. All of this can be implemented ad hoc via the parameterization in HySOR, providing both parties with transparent decision support.

5 Evaluation Method and Survey Results

We evaluated HySOR using the implemented demonstrator and a survey of 12 leading experts in the field. The number of potential users of the model and the prototype is relatively small compared to the number of end users of other DSR artifacts. This is because such top-level contracts are usually signed by an IT decision-maker and there are usually very few of these per company. Therefore, we have a small target audience of specialized practitioners. In addition, not all companies have an established cloud strategy, and in many companies, the topic has not yet emerged or has only been discussed to a limited extent.

With the aim of a more "naturalistic" evaluation (Venable et al. 2016), the survey was conducted with selected leading experts who have already been confronted with such cloud adoption scenarios. The naturalistic evaluation design was substantially supported by our exemplary case study (presented in chapter 4) and the subsequent possibility of implementing and applying individual cloud scenarios in the prototype. The respondents were previously categorized into two dimensions – position and provider type – each with two variants, in line with the recommendations of Iivari et al. (2021) on dealing with "a small sample of the target community" and realistic subgroups. In the first dimension (position), the leading experts were distinguished between "Decision makers", who sign and are responsible for such service compositions, and "Architects", who are responsible for obtaining the necessary information for a decision maker. In the other dimension (provider type), a distinction is made between internal service composition providers ("Internal SCP"), including those who provide a service composition for their own company (internally), and those who provide them as an external service for other companies ("External SCP"). The two dimensions were each represented by 50% of the total number of respondents, thus avoiding the risk of over- or under-representing the views of a particular subgroup. Table 1 provides an overview of the number of experts surveyed in the two dimensions, additionally broken down into small (s), midsize (m), and large (l) companies, as we were also interested in a suitable cross-section of different company sizes. In total, respondents from three small companies (≤ €100 million revenue), four from midsize companies (> €100 million and ≤€1 billion revenue), and five large from companies (> €1 billion revenue) were surveyed.

The naturalistic evaluation aimed to survey experts with sufficient experience. While this can generally be assumed for decision-makers (most decision-makers were C-level executives; average professional experience of all decision-makers > 25 years), we particularly paid attention to sufficient professional experience when selecting the architects (average professional experience of the architects > 12 years).

Since the questionnaire by Iivari et al. (2021) proved to be suitable in a pre-test (high response rate from practitioners due to short and concise questions) and fits our scenario, we adapted it for our context and used it for surveying. Our focus was on the use of the model and prototype in a meaningful and relevant practical scenario (importance) and on the improved capability to achieve desired results through its use (effectiveness) (Iivari et al. 2021). The questions on importance adapted to the context of HySOR are represented in Table 2 by "IMP1" and "IMP2", and on effectiveness by "EFF1", "EFF2", and "EFF3". For data collection, we used verbal-numerical 7-point Likert-style scales ranging from 1 (strongly disagree) to 7 (strongly agree).

Table 1 Number of experts surveyed, categorized by position, provider type, and company size

	Internal SCP			External SCP			$\sum$
	s	m	l	s	m	l	
Decision makers (> 25 years of experience)	1	1	1	1	2	0	6
Architects (> 12 years of experience)	0	0	3	1	1	1	6
$\sum$	6			6			12

SCP = service composition provider; s = small company; m = midsize company; l = large company; $\sum$ *= sum*

Table 2 Adapted questions on importance and effectiveness, based on Iivari et al. (2021)

Item	Question (adapted to our context)
IMP1	In my view, HySOR addresses a real problem in my professional practice.
IMP2	In my view, HySOR addresses an important – acute or foreseeable – problem in my professional practice.
EFF1	Compared to my current situation, I believe that HySOR would enhance my effectiveness in my job.
EFF2	Compared to my current situation, I believe that HySOR would improve the reputation of excellence of my organization/company.
EFF3	Compared to the current situation, I believe that HySOR would improve the innovativeness of my organization/company.

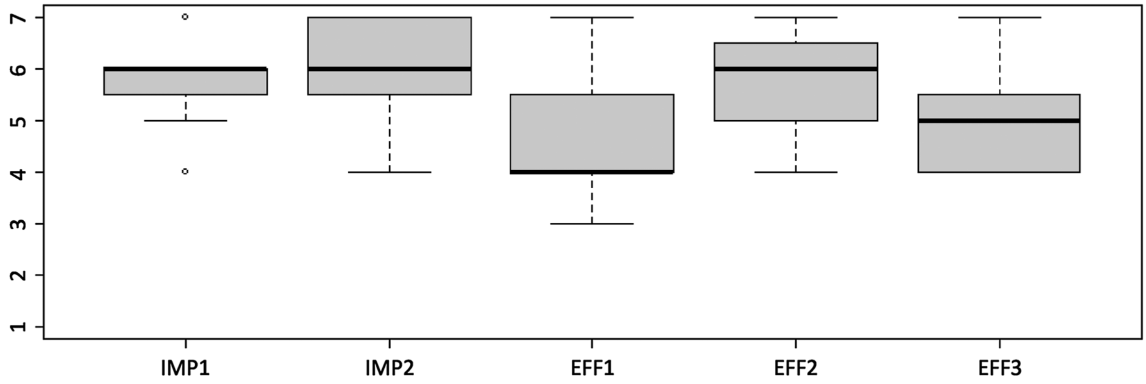


Fig. 5 Evaluation results regarding importance and effectiveness of HySOR (n = 12)

The survey procedure comprised the following seven parts: (1) an introductory text about the research project and the status quo, (2) some socio-economic, contextual, and demographic questions (job role, area of responsibility, practical experience, provider type, company size), (3) an explanation of the HySOR risk simulation model, (4) a link to the HySOR prototype with the presentation of the case study, (5) the possibility to enter one's own case study, (6) the statements on IMP1, IMP2, EFF1, EFF2, EFF3, and (7) a text field for comments. The text field for comments was included to allow for a brief initial qualitative evaluation. This was intended to identify aspects for the further development of the prototype regarding better practical applicability on the one hand and the need for further research in future design cycles on the other.

Overall, the questions received mostly positive feedback, as can be seen in Fig. 5. The questions with the highest median scores (6 out of 7) are IMP1, IMP2, and EFF2, suggesting that our research addresses both a real and an acute/foreseeable problem in business practice and that the use of HySOR has the potential to improve the reputation of companies/organizations. EFF1 and EFF3 received quite different scores. While the enhancement of job effectiveness (EFF1) received the lowest median rating (4 out of 7), the improvement of the innovativeness of the company/organization (EFF3) was rated slightly higher by the respondents (5 out of 7).

In addition, three aspects could be derived from the comments in the free text fields. First, several respondents from large enterprises indicated that they are strongly interested in decision support for business risks arising from cloud integration. Second, it was commented that the complexity of compositions increases in the presence of highly heterogeneous IT infrastructures. Third, it was mentioned that the high relevance of the topic for companies/organizations is due to great risks (very high costs or even the loss of approval), especially for larger enterprises. Overall, the comments we received are further proof of the relevance of our research, especially for large companies/organizations.

6 Discussion with Experts

Based on the information provided in the free text fields at the end of the questionnaire, the commentators were invited to expert interviews for further discussion. By analyzing these results qualitatively, some interesting implications for theory and especially for practice could be derived.

First, there were useful hints about the type and nature of the users of the application. For example, it has to be considered that, although the resulting information is interesting for the decision-makers, architects are always needed to correctly model the underlying input data and interpret the results due to their complexity. This was noted by both architects and decision-makers. Furthermore, it was mentioned that the applicability in practice needs to be well guided and requires a certain amount of prior knowledge, both in relation to the subject matter (i.e., sharing of risk in hybrid clouds) and to the use of the prototype. In this context, it was also found that the inclusion of negative costs and negative penalties (see remarks 4, 5, and 8) is useful for transparency, however, without knowledge of the underlying formulas, its interpretation is limited. Consequently, we conclude that there is still considerable potential for improving usability and comprehensibility to be addressed in a subsequent design cycle.

Another topic of discussion was the challenge of appropriate process integration in typical IT service management. The following questions were raised by the experts in this context and are seen as suggestions for further developments: At what point in the IT service management process should HySOR be used? What could a management process look like that takes risk aspects into account in procurement or in the case of an upcoming cloud migration? Who should be responsible for using HySOR? Is the use of HySOR by IT procurement staff sufficient? Or should the assessment be done by experts or IT decision-makers? We have already answered some of these questions in this paper, such as those relating to required competences or responsibilities. Other questions are to be addressed as part of a future case study on IT procurement.

Further aspects that were raised were the scope and possible applications of HySOR. The majority of interviewees stated that HySOR can provide an impetus for decision-making. However, there were indications that the decision had often already been made at other levels (e.g., due to the corporate strategy), elsewhere (e.g., by functional requesting departments), or at earlier stages (e.g., through preliminary cloud adoption studies). In these cases, the tool would nevertheless provide useful transparency about the risks involved.

The ability to visualize decisions in an objective and quantified way (especially using the linear trend lines shown in Fig. 4) and/or to point out possible inconsistencies was identified as a potential added value by most of the leading experts interviewed. It was confirmed that the concept of sharing of risk can be used as a corridor for negotiation as part of a long-term partnership. Not least for this reason, we consider the results of the qualitative analysis as evidence and a positive conclusion of our research efforts.

7 Limitations and Future Research Directions

To assess the impact of our results appropriately, the limitations of our study should be taken into account. Some of these have already been explained and justified as part of the assumptions and remarks when describing the HySOR model. The remaining limitations are discussed below.

First, a full cost analysis is not performed, i.e., possible additional costs are not considered. For example, the composition provider could incur additional costs for monitoring the added cloud services, identifying errors, documenting non-compliance, and notifying credit (i.e., reporting to the public cloud provider to get the service credit). There could also be additional costs for the customer, at least if the customer owns the subscription. However, for a fair comparison, in addition to the missing cost components, the benefits for providers and customers also need to be considered adequately. For example, the benefits of the customer's business processes should also be taken into account. In contrast, our model focuses on the cost side and partly works with negative costs to represent profits, which has also been identified as a practical limitation as part of the expert discussion (see Sect. 6). We therefore consider the further development of our model towards an even better risk assessment with a dedicated focus on both the economic cost and benefit effects to be a suitable field for further research.

A second limitation is the focus on a two-party bargaining situation, in line with Raiffa (1982)'s ZOPA concept. Consequently, a promising starting point for future research would be the consideration of third parties, which could offer new opportunities, such as the involvement of insurers (see, e.g., Wang and Franke (2020)), if two parties cannot agree on risk sharing without risk protection.

A third limitation is that only experts from German-speaking countries were surveyed and interviewed. Socio-cultural differences in the perception of risk in hybrid cloud architectures cannot be ruled out. However, there are two reasons why this limitation should not be given too much weight. First, the survey of experts from large companies included people from an international environment. Second, it cannot necessarily be assumed that assessments of financial impacts vary fundamentally depending on socio-cultural attributes. Nevertheless, we see a need for larger-scale studies that not only include more participants, but also those with different socio-cultural values.

Irrespective of the limitations, we were able to identify two further starting points for future research. First, there are SLA aspects that have not yet been considered in HySOR and that could potentially influence the financial impact and thus the sharing of risk. For example, the SLA category of service maintenance specified by Seifert (2021) could be taken into account. This refers to issues that describe the planned unavailability of cloud services. In addition, the expert interviews revealed – contrary to remark 2 – that business process outage costs in practice vary greatly depending on the time frame. This is immediately apparent at the weekends, during which many companies' business processes do not incur any outage costs. A mapping of different outage cost functions in relation to time frames can lead to further interesting results, especially when considering scheduled maintenance periods.

A second interesting direction for further research is the development of suitable estimation methods and, in connection with this, the examination of historical data for modeling network, interface, or other unknown risks which could influence the availability of service compositions (e.g., depending on the number of service providers and service components involved). This could be factored into the simulation model without the experts having to use their intuition to make estimates.

8 Conclusion

This paper presents a risk calculation and simulation model that was developed considering artifacts of the existing knowledge base in a rigorous DSR process iteration.

The core elements of service components, service compositions, associated SLAs/OLAs as well as their owners (in different roles), supported business processes, and various risks were derived from the literature, defined, and conceptualized. The areas of service commitment and service credit, which were identified as financially relevant in connection with the SLAs, complement this to answer the first research question (RQ1). Using the service component and service composition characteristics of HySOR, as well as the corresponding functions, availability simulation and aggregation, business process outage calculation and aggregation, and penalty calculation and aggregation, the impact of IT downtimes of components was implemented in the simulation of customer and provider costs to answer the second research question (RQ2). The third research question (RQ3) is answered on the basis of the results of simulating and calculating the aggregated availability of the service composition and the resulting financial impact for the customer and the provider in the best, mean, and worst case. Finally, the visualization of the financial impact depending on the top-level availability in the form of bar charts, supplemented by linear trend lines, was considered helpful by the leading experts surveyed.

The results of this work illustrate the relevance of the topic and confirm the different perceptions of transparency in hybrid cloud architectures. HySOR provides useful impulses for the negotiation of hybrid cloud service compositions in the context of the sharing of risk.

The contribution of this paper is twofold. On the empirical side, we address a practically important topic with an effective artifact implemented. The evidence from the evaluation of leading experts provides promising directions for further research. On the conceptual side, the description of the calculation and simulation elements of the HySOR model/artifact can serve as an evaluated foundation for other researchers.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Ahlert M, Sträter KF (2016) Refining Raiffa – aspiration adaptation within the zone of possible ag. Ger Econ Rev 17:298–315. https://doi.org/10.1111/geer.12096

Al-Ghuwairi A-R, Khalaf MN, Al-Yasen L, Salah Z, Alsarhan A, Baarah AH (2016) A dynamic model for automatic updating cloud computing SLA (DSLA). In: Proceedings of the international conference on internet of things and cloud computing. ACM, New York, pp 1–7. https://doi.org/10.1145/2896387.2896442

Aljoumah E, Al-Mousawi F, Ahmad I, Al-Shammri M, Al-Jady Z (2015) SLA in cloud computing architectures: a comprehensive study. Int J Grid Distrib Comput 8:7–32. https://doi.org/10.14257/ijgdc.2015.8.5.02

Baset SA (2012) Cloud SLAs: present and future. SIGOPS Oper Syst Rev 46:57–66. https://doi.org/10.1145/2331576.2331586

Benlian A, Koufaris M, Hess T (2011) Service quality in software-as-a-service: developing the SaaS-Qual measure and examining its role in usage continuance. J Manag Inf Syst 28:85–126. https://doi.org/10.2753/MIS0742-1222280303

Breiter G, Naik VK (2013) A framework for controlling and managing hybrid cloud service integration. In: 2013 IEEE international conference on cloud engineering, pp 217–224. https://doi.org/10.1109/IC2E.2013.48

Comuzzi M, Kotsokalis C, Rathfelder C, Theilmann W, Winkler U, Zacco G (2009) A framework for multi-level SLA management. In: Dan A, et al (eds): Service-oriented computing. Springer, Heidelberg, pp 187–196. https://doi.org/10.1007/978-3-642-16132-2_18

Comuzzi M, Jacobs G, Grefen P (2013) Clearing the sky: understanding SLA elements in cloud computing. BETA publicatie: working papers vol. 412, Eindhoven University of Technology, Eindhoven, pp 1–25. https://research.tue.nl/en/publications/clearing-the-sky-understanding-sla-elements-in-cloud-computing. Accessed 20 Jul 2024

Faniyi F, Bahsoon R, Theodoropoulos G (2012) A dynamic data-driven simulation approach for preventing service level agreement violations in cloud federation. Procedia Comput Sci 9:1167–1176. https://doi.org/10.1016/j.procs.2012.04.126

Franke U, Johnson P, König J (2013) An architecture framework for enterprise IT service availability analysis. Softw Syst Model 13:1417–1445. https://doi.org/10.1007/s10270-012-0307-3

Goo J, Kishore R, Rao HR, Nam K (2009) The role of service level agreements in relational management of information technology outsourcing: an empirical study. MIS Q 33:119–145. https://doi.org/10.2307/20650281

Guérout T, Medjiah S, Da Costa G, Monteil T (2014) Quality of service modeling for green scheduling in clouds. Sustain Comput Inform Syst 4:225–240. https://doi.org/10.1016/j.suscom.2014.08.006

Gulia P, Sood S (2013) Comparative analysis of present day clouds using service level agreements. Int J Comput Appl 71:1–8. https://doi.org/10.5120/12335-8603

Hussain O, Dong H, Singh J (2010) Semantic similarity model for risk assessment in forming cloud computing SLAs. In: Meersman R et al (eds) On the move to meaningful internet systems, OTM 2010. Springer, Heidelberg, pp 843–860. https://doi.org/10.1007/978-3-642-16949-6_12

Iivari J, Hansen M, Haj-Bolouri A (2021) A proposal for minimum reusability evaluation of design principles. Eur J Inf Syst 30:286–303. https://doi.org/10.1080/0960085X.2020.1793697

Gartner Inc. (2024) Gartner Forecasts Worldwide Public Cloud End-User Spending to Surpass $675 Billion in 2024. In: Gartner Newsroom, Information Technology, Press Release, https://www.gartner.com/en/newsroom/press-releases/2024-05-20-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-surpass-675-billion-in-2024. Accessed July 20, 2024.

Jiang M, Byrne J, Molka K, Armstrong D, Djemame K, Kirkham T (2013) Cost and risk aware support for cloud SLAs. 2184–5042. https://doi.org/10.5220/0004377302070212

Johnson P, Ullberg J, Buschle M, Franke U, Shahzad K (2014) An architecture modeling framework for probabilistic prediction. Inf Syst E-Bus Manag 12:595–622. https://doi.org/10.1007/s10257-014-0241-8

Kritikos K, Plexousakis D, Plebani P (2016) Semantic SLAs for services with Q-SLA. Procedia Comput Sci 97:24–33. https://doi.org/10.1016/j.procs.2016.08.277

Labidi T, Mtibaa A, Brabra H (2016) CSLAOnto: a comprehensive ontological SLA model in cloud computing. J Data Semant 5:179–193. https://doi.org/10.1007/s13740-016-0070-7

Mateo-Fornes J, Solsona-Tehas F, Vilaplana-Mayoral J, Teixido-Torrelles I, Rius-Torrento J (2019) CART, a decision SLA model for SaaS providers to keep QoS regarding availability and performance. IEEE Access 7:38195–38204. https://doi.org/10.1109/ACCESS.2019.2905870

Mazrekaj A, Shabani I, Sejdiu B (2016) Pricing schemes in cloud computing: an overview. International Journal Advance Computer Science Applications 7. https://doi.org/10.14569/IJACSA.2016.070211

Pan W, Mitchell G (2015) Software as a service (SaaS) quality management and service level agreement. Infuture 26:225–234. https://doi.org/10.17234/INFUTURE.2015.26

Paquette S, Jaeger PT, Wilson SC (2010) Identifying the security risks associated with governmental use of cloud computing. Gov Inf Q 27:245–253. https://doi.org/10.1016/j.giq.2010.01.002

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manag Inf Syst 24:45–77. https://doi.org/10.2753/MIS0742-1222240302

Raiffa H (1982) The Art and Science of Negotiation: How to resolve conflicts and get the best out of bargaining. Harvard Univ. Press, Cambridge

Rockmann R, Weeger A, Gewald H (2014) Identifying organizational capabilities for the enterprise-wide usage of cloud computing. In: PACIS 2014 Proceedings. http://aisel.aisnet.org/pacis2014/355

Seifert M, Kuehnel S, Sackmann S (2023) Hybrid clouds arising from software as a service adoption: challenges, solutions, and future research directions. ACM Comput. Surv. Vol. 55, No. 11. Article 228:1–35. https://doi.org/10.1145/3570156

Seifert M (2021) Analysis of public cloud service level agreements - an evaluation of leading software as a service provider. In: Kühnel S, Sackmann S, Trang S (eds): Proceedings of the First International Workshop on Current Compliance Issues in Information Systems Research (CIISR'21), Co-located with the 16th International Conference on Wirtschaftsinformatik (WI'21), Online (initially located in Duisburg-Essen, Germany), March 9th, 2021. CEUR Workshop Proceedings 2966, pp 22–35. https://ceur-ws.org/Vol-2966/paper2.pdf

Seifert M, Kuehnel S (2021) "HySLAC" - a conceptual model for service level agreement compliance in hybrid cloud architectures. In: Reussner RH, Koziolek A, Heinrich R (eds): INFORMATIK 2020, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021 205–218. https://doi.org/10.18420/inf2020_19

Suakanto S, Supangkat SH, Suhardi, Saragih R (2012) Performance measurement of cloud computing services. International Journal Cloud Computing Services Architecture 2 2 9 20. https://doi.org/10.5121/ijccsa.2012.2202

Sun W, Zhang X, Guo CJ, Sun P, Su H (2008) Software as a service: configuration and customization perspectives. In: 2008 IEEE congress on services part II, pp 18–25. https://doi.org/10.1109/SERVICES-2.2008.29

Theilmann W, Happe J, Kotsokalis C, Edmonds A, Kearney K, Lambea J (2010) A reference architecture for multi-level SLA management. J Internet Eng 4:289–298. https://doi.org/10.21256/zhaw-1757

Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a framework for evaluation in design science research. Eur J Inf Syst 25:77–89. https://doi.org/10.1057/ejis.2014.36

Wang SS, Franke U (2020) Enterprise IT service downtime cost and risk transfer in a supply chain. Oper Manag Res 13:94–108. https://doi.org/10.1007/s12063-020-00148-x

Yuan X, Li Y, Jia T, Liu T, Wu Z (2015) An analysis on availability commitment and penalty in cloud SLA. In: 2015 IEEE 39th annual computer software and applications conference, pp 914–919. https://doi.org/10.1109/COMPSAC.2015.39

Zhang L-J, Zhou Q (2009) CCOA: cloud computing open architecture. In: 2009 IEEE international conference on web services, pp 607–616. https://doi.org/10.1109/ICWS.2009.144