2013

Leveraging Service Incident Analytics to Determine Cost-Optimal Service Offers

Axel Kieninger
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, axel.kieninger@kit.edu

Detlef Straeten
IBM Deutschland GmbH, Walldorf, Germany, detlef.straeten@de.ibm.com

Steven Kimbrough
University of Pennsylvania, Philadelphia, USA, kimbrough@wharton.upenn.edu

Björn Schmitz
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, bjoern.schmitz@kit.edu

Gerhard Satzger
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, gerhard.satzger@kit.edu

Leveraging Service Incident Analytics to Determine Cost-Optimal Service Offers

Axel Kieninger[1], Detlef Straeten[2], Steven Kimbrough[3], Björn Schmitz[1], and Gerhard Satzger[1]

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{axel.kieninger,bjoern.schmitz,gerhard.satzger}@kit.edu
[2] IBM Deutschland GmbH, Walldorf, Germany
detlef.straeten@de.ibm.com
[3] University of Pennsylvania, Philadelphia, USA
kimbrough@wharton.upenn.edu

Abstract. In this work we address the challenge for an IT service customer to select the cost-optimal service among different offers by external providers. We describe the customer's optimization problem by considering the negative monetary impact of potential service incidents on its business. First, we demonstrate that the information currently used in service level agreements may lead to suboptimal customer decisions. Second, we discuss how providers' private information about the behavior of service delivery environments could be leveraged by the customer when selecting service offers. Third, we propose a procurement auction as a mechanism to optimize total cost for the customer – choosing from different service offers by risk-neutral providers. In introducing this approach, we suggest that customers and providers collaborate to define service performance measures, which allow providers to better tailor service offers to customers' business requirements.

Keywords: Service Level Management, Service Level Agreement, Adverse Business Impact, Service Incident, Procurement Auction

1 Introduction

Outsourcing of IT services is popular and growing. The worldwide revenue generated with outsourcing totaled $246.6 bn in 2011 – an increase of 7.8 percent compared to 2010 [1]. But in spite of the increasing significance of outsourced IT services, there are still opportunities for improvement (see, e.g., [2-3]).

In this work we address the business and IT alignment challenge (see, e.g., [4]) for an IT service customer to select the cost-optimal service among different offers by external providers. We describe the customer's optimization problem by considering the negative monetary impact of potential service incidents on its business.

Typically, the provision of outsourced services is governed by a *service level agreement* (SLA) – also including performance requirements for the vendor, such as

11th International Conference on Wirtschaftsinformatik,
27th February – 01st March 2013, Leipzig, Germany

throughput, response time, and service outages. We focus on service outages of end-to-end services, which directly affect the customer's business performance; however, our findings can in many cases be generalized to also cover other performance dimensions.

In practice, performance requirements for system outages are typically expressed by two specifications: the minimum availability of the service over a given period of time and the maximum duration of any single service outage incident (see, e.g., [5]). For example, the minimum availability might be expressed as 99% uptime every month and the maximum incident duration might be specified as 30 minutes. Let us call this requirements specification method the *(min, max)* regime, where the target values for *min* and *max* are called *service level objectives* (SLOs).

We address two primary research questions in this paper: First, we analyze why the *(min, max)* method of performance specification of service outages leaves purchasers under-informed and why, as a consequence (under reasonable assumptions), decisions based on this method may be significantly suboptimal. We discuss how providers' private information about the behavior of service delivery environments could be leveraged by the *customer* when selecting service offers. Second, we examine how the customer could optimize its total cost in choosing from different service offers by risk-neutral providers. We propose a procurement auction as a mechanism to support customers in determining the cost-optimal solution to their optimization problem considering the negative monetary impact of potential service incidents on their business.

The remainder of the paper proceeds as follows: Section 2 briefly reviews related literature. We introduce the base scenario, which we address later on, in Section 3. In Section 4, we analyze the monetarily quantified negative impact of selected incident duration distributions on the performance of a customer business process. We discuss how a provider could support the customer in solving its decision problem in Section 5. In Section 6, we propose a service procurement auction approach that enables the customer to select its cost-optimal service solution. Finally, Section 7 summarizes and discusses our approach, and outlines the next steps towards application of these ideas in industry.

2 Related Work

We conducted a literature review following the methodology proposed by Webster and Watson [6]. In an extensive forward and backward search without temporal restriction, Google Scholar and CiteSeer served as our main sources of search. Keywords used were 'service level objective', 'service level agreement', 'service level management', 'optimal choice', 'business impact', 'incident management', 'decision theory' and 'customer' as well as combinations thereof.

Focusing on the research field of IT (outsourcing) we found six papers – by Jin et al. [7], Sauvé et al. [8], Cheng et al. [9], Taylor and Tofts [10], Barroero et al. [11] and Franke [12] – which are related to our approach that aims to support the consideration of adverse business impact in service offers. None of these works, however, closely matches our scenario and approach. Our own prior work contains discussions of these papers and is more directly relevant:

In [13] we discuss the customer's optimization problem to select the cost-optimal SLA having concluded a long-term outsourcing contract with a single external provider. We propose a methodology to solve the optimization problem in a scenario without penalties. In [14] we extend that approach and discuss the customer's optimization problem in a setting where a provider has to compensate the customer for every single service incident to occur. We propose a single-attribute procurement auction approach to identify and select the cost-optimal service offer out of different providers' bids. In contrast to this work, our prior work does not consider the cases of providers bidding tuples of target values for service level indicators (i.e., performance measures) and a price (multi-attribute procurement auction) or of receiving incentive bonus payments for 'better-than-promised' performance. In [15] we propose a simulation-based procedure to monetarily quantify the negative impact of single service incidents on the customer business.

Furthermore, there are several areas of research complementary to our work: Taking a more technical perspective, Wittgreffe et al. [16] provide concepts aiming at the design of end-to-end SLAs that are directly targeted at business applications of the service customer. Breitgand et al. [17] present an SLA design approach which enables service providers to derive optimal response time SLOs for a service by analyzing historical performance data of the applied IT infrastructure. Blau et al. [18] use a multidimensional procurement auction to determine the socially efficient service composition in a service value network aiming at the maximization of welfare.

Summing up, none of the related works closely matches our scenario and approach, namely: (i) We monetarily measure the negative business impact of service offers by (ii) quantifying the incurred business cost resulting from the occurrence of combinations of service incidents and (iii) explicitly considering business cost functions which do not develop linearly with service performance attributes. We do this in a setting in which providers have private information they do not want to disclose (iv).

3 Base Scenario

In our base scenario, which we address later on, a customer company wants to purchase an end-to-end IT service from an external service provider to support a particular business process. This non-mission-critical, yet non-commodity service directly affects the customer's business performance, i.e., it contributes measurably to the business value created. Competing providers are able to offer the particular service with different SLAs, i.e., at different service levels and different prices.

We assume that all functional and non-functional properties (including base sizing) of the service, except for performance and price, have already been documented in a 'Request for Proposal' (RfP). For this purpose, requirements of the business process to be supported have been considered. The RfP also states a fixed delivery duration, which may be split into a number of reference periods.

To define service performance the customer has specified a set of service level indicators (performance measures), and associated service level objectives (target values). An SLA is represented by a tuple of service level objectives and a service price, which the customer will have to pay to the selected provider on a periodic basis for service provisioning. In the RfP, the customer has also defined penalty functions for each service level indicator. These specify the amount of money a provider will have to pay in case of service level breaches. Further, the RfP grants the customer the right of constant access to the monitoring information that the provider collects.

To compare service providers' offers, we assume the customer will estimate and monetarily quantify the negative impact of imperfect service (at specific service levels) on its business process. Also, the customer is assumed to consider its expectations of providers to achieve, to under- or even outperform the stated SLOs (cf. [13]). Using these 'business cost' estimates, the service price, as well as the expected penalty payments, the customer can assess the financial implications resulting from the selection of a specific SLA, i.e., it can determine total customer cost. We suppose that the rational customer in our setting aims to identify the cost-optimal service offer.

Since customer and provider have agreed on a fixed delivery duration, capital budgeting approaches, such as the net present value method, can be applied to discount cash flows (i.e., total customer cost) of several future reference periods. This allows the comparison of SLAs which state service prices and business cost changing over time, i.e., varying in different reference periods. For reasons of clarity, however, we will focus on a single reference period.

4 The Adverse Business Impact of Service Incidents

The *(min, max)* performance measures normally found in SLAs for specifying requirements for service outages may be called *aggregating* and *limiting* service level indicators. An example: *A logistics company C runs a high-bay warehouse with a management system that is operated and managed by the provider P as a service. This end-to-end service is defined to be available 99% of the time of a reference period [this is the SLO of the aggregating service level indicator min], whereas no single service incident may last longer than 30 minutes [this is the SLO of the limiting service level indicator max].*

The information lost in the aggregation may, however, have undesirable consequences for making decisions – as will be illustrated in the following. Using these established service level indicators the customer has no information about the specific probabilities of a service to achieve, to under- or even outperform the stipulated SLOs. Consequently, the customer is not able to determine with any precision the business cost resulting from the selection of an SLA. Even if the service level objective combinations of aggregating and limiting service level indicators are exactly met, customers may not be able to determine the adverse business impact. Let us see why.

Adverse business impact, resulting in business cost, is caused by each service incident (SI), i.e., by each 'unplanned interruption' to a service or 'reduction in the quality' of a service, which is observable from a business point of view (adapted from the definition of ‘incident’ in [19]). We use business cost functions to describe the business cost induced by a single service incident of a certain type with regard to its incident attribute values. These business cost functions may develop non-proportionally to service incidents’ attribute values, i.e., they do not need to increase or decrease linearly.

Example (cont’d): *The warehouse management system is used to locate goods in the high-bay warehouse that have to be taken out of stock and shipped by truck. Every five minutes, one truck arrives at the warehouse in order to be loaded. Outages of the warehouse management system (i.e., service incidents of the type ‘outage incidents’, which are described by the single attribute ‘outage duration’) interrupt this workflow. For each minute a truck has to wait to be loaded, C incurs business cost of € 1. Figure 1 shows how C’s aggregated business cost increases with longer outage durations.*

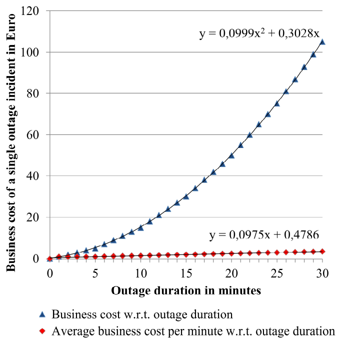

Fig. 1. Business cost of a single outage incident w.r.t. outage duration

The blue triangles in Figure 1 represent the (discrete) business cost values of single outage incidents with respect to their duration.[1] The red diamonds in Figure 1 describe the average business cost per minute an outage incident causes with respect to its duration.[2] For instance, an outage incident that lasts 25 minutes induces business cost of € 75, i.e., average business cost of € 3 per minute of outage.

When an aggregating service level indicator is used, the achieved service quality is calculated considering the values of a specific attribute of all service incidents of the same type (service incident class) that occur within a certain reference period. In order to quantify business cost in case of moderately non-linear business cost functions[3], however, it is essential for the customer to understand which combination of incident attribute values actually realizes a specific *achieved* value (of an aggregating

[1] The discrete business cost values (blue triangles) can be approximated by the continuous business cost function $y=0.0999x^2+0.3028x$.

[2] The average business cost per minute values (red diamonds) can be approximated by the continuous function $y=0.0975x+0.4786$.

[3] We obtained our results below assuming a moderately non-linear business cost function. Due to space limitations we cannot report a full sensitivity analysis. The business cost function we employ, however, is realistic and reasonably robust.

service level indicator). The distribution of single incidents' attribute values, e.g., of their outage durations, has to be considered.

To illustrate the potential consequences of different incident attribute value distributions on business cost, we continue our example:

We assume the provider to exactly achieve the target availability of 99% (i.e., 432 minutes of total outage time within a month) and to avoid single service incidents lasting longer than 30 minutes.

Let O_T be the total outage time permitted and x_{max} be the maximum length of a single outage incident. Furthermore, let x_i be the length of outage i during the reference period T and let $b(x_i) = 0.0999x_i^2 + 0.3028x_i$ be the non-linear business cost function. Now assume $x_i \leq x_{max}$, $\forall i$, and, $x_i = x_j$, $\forall i,j$.

If all outages occurring in T have the same duration x_{fix}, this leads to significantly different total business cost for different values of x_{fix} (see Table 1):

Table 1. Different outage durations lead to significantly different business cost

x_{fix}	$b(x_{fix})$	number of incidents within T (rounded)	total business cost caused within T
10	*€ 13.02*	*43*	*€ 559.77*
20	*€ 46.02*	*21*	*€ 996.34*
30	*€ 98.99*	*14*	*€ 1,385.92*

Having shown the impact of different outage durations using this simple calculation we can now discuss a more complex setting: *Figure 2 shows the relative frequencies of different outage incident durations for selected beta-distributions*[4].

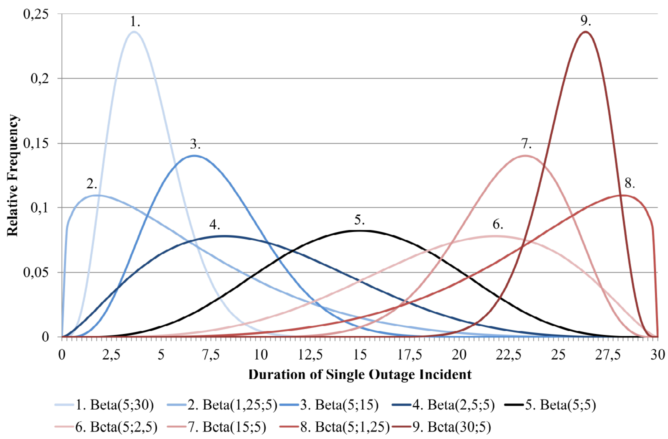

Fig. 2. Relative frequencies of beta-distributed outage incident durations on the interval (0;30]

[4] We chose the beta-distribution because it can be specified within definite limits, it is very flexible and is analytically tractable. The beta-distribution allows to model a large variety of incident behaviors of service delivery environments, e.g., that of an environment which is characterized by many short outages, some outages of medium length and few long outages.

Since no single service incident may last longer than 30 minutes (SLO of the limiting service level indicator) the distributions are limited to the interval (0;30]. By assuming the provider exactly meets the SLOs promised, we are able to calculate the business cost a customer has to expect if these targets are realized through different beta-distributions of single outage durations.

We compute the following integral (see equation 1) for the given beta-distributions, where $f_{Beta}(x, \alpha, \beta, 0, 30)$ is the probability density function of a beta-distribution which is limited to the interval (0;30] and $g_{AvgBC/Min}(x)$ describes the average business cost per minute (depicted as red diamonds in Figure 1):

$$\int_0^{30} f_{Beta}(x, \alpha, \beta, 0{,}30) * 432 * g_{AvgBC/Min}(x)\, dx \qquad (1)$$

The probability density function is multiplied with the total outage duration, i.e., 432 minutes, in order to obtain the time which is spent on a certain incident attribute value. Figure 3 shows the resulting business cost values.

In case of non-constant marginal business cost functions, the customer has to consider the distribution of incident attribute values (a target value of an aggregating service level indicator is realized through), since different distributions will usually lead to different business cost.

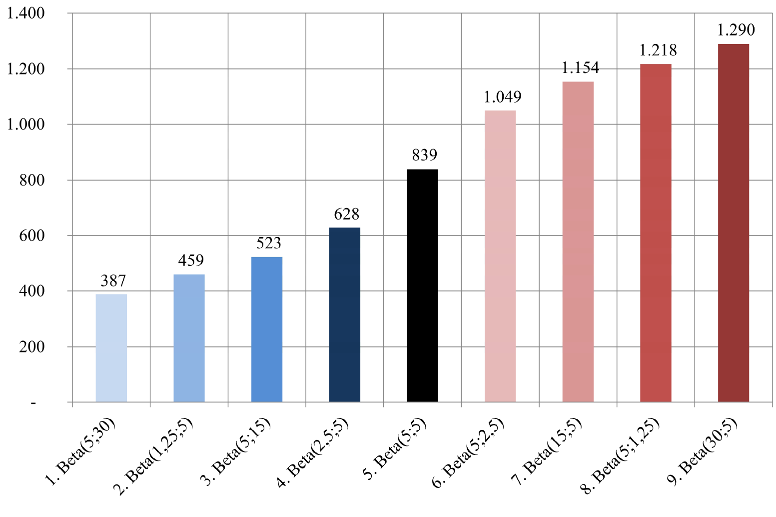

Fig. 3. Expected business cost given different beta-distributions for single outage durations

Even if the customer in our example knew that incident durations are beta-distributed and assumed a specific beta-distribution in order to estimate its expected business cost, this could lead to a significant misjudgment in case the customer selects inaccurate beta-distribution parameter values.

Table 2 shows the errors with regard to expected business cost (in percent) resulting from the assumption of a specific beta-distribution, while the actual outage duration distribution follows a different beta-distribution.

Table 2. Error with regard to expected business cost in percent if a 'wrong' beta-distribution is assumed

		Assumed beta-distribution				
		B(5;30)	**B(5;15)**	**B(5;5)**	**B(15;5)**	**B(30;5)**
Actual beta-distribution	**B(5;30)**	0.0 %	35.0 %	116.5 %	198.1 %	233.1 %
	B(5;15)	-25.9 %	0.0 %	60.4 %	120.9 %	146.7 %
	B(5;5)	-53.8 %	-37.7 %	0.0 %	37.7 %	53.8 %
	B(15;5)	-66.5 %	-54.7 %	-27.4 %	0.0 %	11.7 %
	B(30;5)	-70.0 %	-59.5 %	-35.0 %	-10.5 %	0.0 %

Consequently, the definition of SLOs for aggregating and limiting service level indicators is not sufficient to compare different service offers regarding the total business cost (and, thus, total customer cost) induced in the case of non-linear business cost functions.[5] Performance measures have to consider the monetarily quantified negative business impact of service incidents instead of incident attribute values only.

Furthermore, the example shows that it would be negligent from a customer point of view not to reflect distributions of incident attribute values in service level agreements in case of moderately non-linear business cost functions. This idea, however, hinges on the providers' ability to ascertain incident attribute value distributions at reasonable (additional) cost and, thus, price increases.

5 Service Incident Patterns

Providers usually possess private information about the service delivery environments (SDEs), i.e., combinations of organizational and technical service components they use to provide a service. Since outsourcing providers have many customers who purchase similar services, they are able to collect data about the frequency of service incidents that occur when using a particular type of SDE, i.e., to establish service incident histories for each SDE. By analyzing these data, providers can predict the frequency distributions of service incidents of a specific type having certain incident attribute values, i.e., determine the SDEs' characteristic 'service incident patterns'. If the customer had the deep insight a provider has, it could use this information to better address its optimization problem. In the following we will denote a tuple of attribute values of a single service incident as its service incident level (SIL), with the incident type determining the tuple structure.

Considering one service incident type[6] only, Table 3 illustrates a provider's private knowledge of an exemplary service delivery environment *e*. It lists in schematic form

[5] … even if we assume the service level objective combinations of aggregating and limiting service level indicators to be met. In practice, the stipulated SLOs might be even outperformed to certain degrees in different reference periods.

[6] Service incidents sharing the same set of (one or more) attributes are regarded to be of the same service incident type (e.g. outage incidents, reduced throughput incidents, etc.). Incidents of the same service incident type affect the performance of a customer business process in a similar manner, but may differ in their attribute values (e.g., in the duration of outage (outage incident) or the degree of throughput reduction (reduced throughput incident)).

the data records about the combinations of sequential service incidents (CoSIs) the provider has collected about this type of service delivery environment when using it (i) during past reference periods (ii) for one or several other customers. The function $f_t(SIL_{a, b, ..., n})$ describes the absolute frequency with which service incidents at the levels $SIL_{a, b, ..., n}$ occurred during a specific reference period, whereas the index t labels a data record. We denote this table as CoSI/SIL frequency history.

To continue with our example: *A row $CoSI_2$ in the CoSI/SIL frequency history could, for instance, document that there were five outage incidents with a duration in the interval (0;5) minutes ($f_2((0;5))=5$), three incidents with a duration in the interval [5;10) minutes ($f_2([5;10))=3$), … in the data record '2'.* [7]

Based on the information contained in the CoSI/SIL frequency histories, the provider determines the SDEs' characteristic incident patterns (one pattern for each service incident type), which state the expected (absolute) frequencies of service incidents at the different service incident levels to occur within a reference period.[8]

Table 3. A provider's private knowledge of the incident characteristics of a specific service delivery environment e (CoSI/SIL frequency history)

		Absolute frequencies of SIs at a certain SIL (historical data)				
		$SIL_{1,1,...,1}$	**…**	**$SIL_{1,1,...,n}$**	**…**	**$SIL_{a,b,...,n}$**
Service Delivery Environment e	**$CoSI_1$**	$f_1(SIL_{1,1,...,1})$	…	$f_1(SIL_{1,1,...,n})$	…	$f_1(SIL_{a,b,...,n})$
	$CoSI_2$	$f_2(SIL_{1,1,...,1})$	…	$f_2(SIL_{1,1,...,n})$	…	$f_2(SIL_{a,b,...,n})$
	…	…	…	…	…	…
	$CoSI_t$	$f_t(SIL_{1,1,...,1})$	…	$f_t(SIL_{1,1,...,n})$	…	$f_t(SIL_{a,b,...,n})$

Providers have only little understanding of the business cost the customer incurs when a certain CoSI is realized. In our work we assume that the customer is able to describe business cost functions, which define the business cost caused by single service incidents with respect to their specific service incident levels (SILs). That is, the customer can state (discrete) BusinessCost/SIL vectors for each service incident type.

Example (cont'd): *A customer's BusinessCost/SIL vector for the service incident type 'outage' states that outages with a duration in the interval (0;5) minutes induce business cost of € 1.81, outages with a duration in the interval [5;10) minutes induce business cost of € 8.82, …*

Furthermore, in the following, we assume the customer discloses this information to selected providers, treating them as 'trusted suppliers'. Knowing the customer's BusinessCost/SIL vectors (one vector for each service incident type), providers can combine the information contained therein with their private knowledge of service delivery environments' service incident patterns. In doing so, they are able to calcu-

---

[7] The classification of attribute values into attribute intervals reduces the number of service incident levels to be considered and, thus, improves the applicability of our approach.

[8] In future works, we will focus on the determination of service incident patterns from CoSI/SIL frequency histories.

late (provider-internal) indicators describing their service delivery environments' impact on the customer business process in terms of business cost.

In order to determine the monetary adverse business impact – for a single service incident type and a certain service delivery environment – a provider multiplies the characteristic service incident pattern *p* with the BusinessCost/SIL vector *b*. Equation 2 formally describes this operation.

$$p * b = r, \text{ where } p \in \mathbb{R}^s, b \in \mathbb{R}^s, r \in \mathbb{R} \tag{2}$$

The resulting value *r* represents the business cost value the customer would incur in a specific reference period for a specific type of incident. Assuming historical data to allow the prediction of service incident patterns to occur in future reference periods and service incident patterns to be stable – i.e., to have low variances with regard to frequencies of service incidents at certain SILs to occur – providers can arrive at accurate business cost estimates. These measures reflect the adverse business impact a customer incurs in case of constant *and* non-constant marginal business cost functions. Using this private information, providers can better understand the impact of their services on customers' business and, thus, define their service offers accordingly.

6 A Multi-Attribute Procurement Auction to Select Cost-optimal Service Offers

Based on the constructs introduced in the previous sections, we develop a procedure enabling a customer to solve its optimization problem – given different service offers by risk-neutral providers.

Today, a typical contract negotiation in IT outsourcing follows a structured approach [20]. First, by sending potential service providers a request for information, the customer identifies suppliers that offer the service in question. Afterwards, the customer defines a request for proposal specifying additional information about technical, economic and legal characteristics (sizing, mission-criticality, service partnership models, etc.) and further reduces the number of potential partners before the actual negotiation process begins.

Our proposal is to use a service procurement auction to conduct the negotiation. Auctions constitute a way to elicit bidders' private information [21]. In our scenario we use an auction to allow the customer to draw conclusions regarding the providers' reservation prices for delivering their services at characteristic service incident patterns. Thus, we enable the customer to better compare different service offers. In procurement auctions providers 'compete for the right to sell' their services [21].

For reasons of clarity, the method presented focuses on a single reference period only. As discussed in Section 3, capital budgeting approaches could be applied to compare multi-reference period service contracts, which state service prices and business cost changing over time.

6.1 Definition of the Procurement Auction

The objective of our approach is to compare service offers in order to identify and select the one that is minimizing the total customer cost. Therefore, we define an 'optimal mechanism' [21] for the procurement of the service.

In the following, we assume the customer reveals its business cost functions to the selected providers participating in the procurement auction. Providers taking part in the procurement auction have to bid a tuple of service incident patterns (one pattern for each service incident type with regard to service incident levels predefined by the customer) and service price.

A provider in our example, for instance, would bid a tuple consisting of an 'outage incident pattern' p (stating the absolute frequencies for outage incidents at the different service incident levels to occur) and the service price it demands with regard to this pattern.

We define the following two rules in order to make sure that providers have no incentive to state service incident patterns of SDEs untruthfully in their bids:

If the amount of service incidents at a specific service incident level exceeds the number stated in the service incident pattern (i.e., a provider's bid) we define that the future contractual partner will have to compensate the customer for every additional service incident to occur in the amount of business cost defined in the corresponding customer business cost function. Due to this '1:1'-penalty rate, providers would directly add the additional business cost to be expected (for all types of service incidents) to their service prices.

On the other hand, we define that the future contractual partner (provider) will be rewarded for every single incident at a specific SIL it avoids (with regard to the number specified in the service incident pattern) in the amount of business cost defined in the corresponding customer business cost function. Due to this '1:1'-bonus rate, providers in a competitive situation would give the customer a discount for business cost they expect to avoid considering future rewards.

Since the customer has defined all types of service incidents in a way that these are independent of one another, providers will add up all expected penalties (positive) and rewards (negative) and the net service price to a single monetary value (the gross service price). As mentioned above, providers are supposed to bid a tuple of service incident patterns and gross service price.

Therefore, we can use a multi-attribute auction to determine the cost-optimal service offer for the customer. In such a 'multi-attribute context' where the customer reveals the negative impact of service incidents on its business and the providers have different 'market expectations', first-price auctions seem to achieve the best results [22].

Based on these findings, we define a multi-attribute, first-price, sealed-bid auction for our setting (see e.g., [21]). In our first-price auction, each provider is asked to submit one sealed bid, which states a tuple of service incident patterns and service price it offers. The provider who submits the bid leading to lowest total customer cost wins the auction and has to be prepared to deliver the service as described in the RfP, with the promised service incident patterns and at the offered gross service price.

6.2 Conducting the Procurement Auction

We now briefly describe the procurement auction.

1. **Definition of service incident types** (customer)**:** For the service in question, the customer formally describes the different business-relevant types of service incidents including their corresponding sets of attributes. Further, the specific value range of each attribute is defined.
2. **Definition of service incident levels** (customer)**:** The customer divides the value range of each attribute (characterizing a single service incident type) into a set of disjoint intervals. Therefore, the endpoints of each interval are defined in a way that attribute values leading to significantly different adverse business impacts are elements of different intervals. Then, the customer calculates the Cartesian product of the different sets of attribute intervals for each service incident type, i.e., all service incident levels to be considered are defined.
   Example: The service incident type 'reduced throughput incident' is characterized by the two attributes 'duration of throughput reduction' (measured in minutes) and 'degree of throughput reduction' (measured in percent) with the value ranges (0;c] and (0;z]. Considering the negative business impact of different attribute value combinations the customer now divides the range of the duration interval into three parts – (0;a), [a;b) and [b;c] – and the range of the degree interval into four parts – (0;w), [w;x), [x;y) and [y;z]. Calculating the Cartesian product of these two sets of attribute intervals, twelve service incident levels (3 x 4) are defined.
   This step is repeated for all types of service incidents of the service in question.
3. **Business cost assessment** (customer)**:** The customer considers the negative business impacts, which service incidents at the different service incident levels (SIL) might cause, and assesses these monetarily. Consequently, for each service incident type the customer specifies a BusinessCost/SIL vector stating the business cost for each SIL defined in step 2. That is, the (discrete) business cost function regarding each service incident type is defined.
4. **Invitation of providers** (customer)**:** The customer invites a number of selected providers to take part in the procurement auction and sends the RfP including the business cost functions to these trusted suppliers.
5. **Definition of bids** (providers)**:** According to the procedure introduced in the previous section, each provider determines the characteristic incident patterns (one pattern for each service incident type) for the SDE it wants to offer. Furthermore, it calculates the expected penalty and benefit payments it assumes to have to pay / receive. Due to the competitive auction setting the providers will add / subtract these expected values to / from the net service price. That is, the resulting gross service price reflects all expected cash flows.
6. **Bidding** (providers)**:** The providers submit their sealed bids, i.e., tuples of service incident patterns and service price, to the customer.
7. **Winner determination** (customer)**:** Among all service offers the customer identifies the bid resulting in the lowest total customer cost. Therefore, the customer multiplies the absolute frequencies for service incidents at certain SILs to occur as stated in the service incident patterns (first part of a provider bid) with its BusinessCost/SIL vectors (one vector for each service incident type). It adds the re-

sulting business cost value to the gross service price (second part of a provider bid) and, thus, determines the total customer cost.

The provider bid winning the procurement auction determines the total customer cost (fixed value), since the winning provider has to compensate the customer for negative deviations from the service incident pattern offered but is rewarded for positive deviations as well. If the provider is rewarded, on the customer's side business cost are avoided in the same amount. Thus, the winning provider is rewarded if the service in question is delivered at a higher quality than stated in the service incident patterns, and is penalized if realized service quality is lower.

7 Conclusion and Outlook

To summarize, we have addressed a service customer's optimization problem of identifying and selecting a cost-optimal service offer meeting the required level of service criticality and assuming typically non-linear business impact behavior.

First, we analyzed the information that SLA decisions are typically based upon and illustrated deficiencies in the service level indicators usually used today. With regard to our first research question, we showed that the traditional *(min, max)* indicator regime for measuring service performance is not sufficient to support proper decision making by the customer in case of quantified negative business impact to develop non-linearly with service quality. Furthermore, we discussed how providers' private information about the behavior of service delivery environments could be leveraged by the customer when selecting service offers.

In order to address our second research question, we developed a procedure enabling customers to solve their optimization problem – given different service offers – using a procurement auction. Our proposed seven-step auction approach enables providers to calculate business cost measures and, thus, to predict their services' quantified negative business impact using the customer's business cost estimates. Thus, providers can tailor service offers according to customers' business requirements and serve their clients' needs more profoundly. Using the information contained in the customer's business cost functions as the basis for penalty and bonus rules, providers are free to balance resources they use to deliver a service. Moreover, the application of our procedure enables the customer to compare service offers regarding their expected adverse business impact and, thus, to select the total customer cost-optimal solution of its optimization problem – in case of linear and non-linear business cost functions.

Having put forward our approach, we are well aware of a number of limitations and challenges. First, we assumed providers are able to determine the discrete incident attribute value distributions (e.g., regarding outage lengths) when using specific service delivery environments. In order to demonstrate this capability, we are currently working on a data mining approach using monitoring data. Second, we assumed the customer is able to determine its business cost functions for all types of service incidents. We realize that a precise determination of business cost functions may be difficult to achieve. Nevertheless, we argue that an economically well-founded investment decision about the service offer to purchase has to be made based on information about business cost. Even a rough estimate of business cost functions might significantly improve a decision with regard to total customer cost. Third, we assumed providers to be risk-neutral. This allows us the simplification of adding expected penalties to and subtracting expected bonus payments from the net service price when defining service offers, i.e., without adding risk premiums. This should be a realistic assumption at least for large providers supporting numerous, heterogeneous customers as they can spread and diversify their risk. In addition, we assumed that for non-mission-critical, yet non-commodity services, providers will accept a '1:1 penalty rate' as well as a '1:1 bonus rate'. These rates ensure that business cost information contained in providers' offers is not distorted when service prices are defined and that providers have no incentive to state service incident patterns untruthfully. The penalty rate will be accepted by providers, since they can add all expected penalties to the service price. In future work we plan to analyze the influence of different provider risk preferences on the solution of the customer optimization problem. Finally, we assumed the customer to reveal its business cost functions, which indicate the criticality of the service for its business. The competition among providers, which participate in the procurement auction, will prevent them from taking advantage of this information and claiming the complete consumer surplus, which results from different service quality and service price combinations, leading to specific total customer cost.

We are convinced that our approach, enabling customers to compare different service offers and to select the cost-optimal service solution, provides valuable insights for both, service providers and their customers. It will help both parties to address the challenge of business and IT alignment in a different way – exploiting analytical insights gained for service delivery environments. At the same time, the limitations mentioned above leave a broad field for further research.

References

1. Gartner: Gartner Says Worldwide IT Outsourcing Market Grew 7.8 Percent in 2011, http://www.gartner.com/it/section.jsp
2. Unterharnscheidt, P., Kieninger, A.: Service Level Management: Challenges and their Relevance from the Customer's Point of View. In: Proceedings of the 16th Americas Conference on Information Systems, Paper 540. Association for Information Systems, Lima, Peru (2010)
3. Lacity, M.C., Kahn, S.A., Willcocks, L.P.: A review of the IT outsourcing literature: Insights for practice. The Journal of Strategic Information Systems 18 (3), 130–146 (2009)
4. Luftman, J., Ben-Zvi, T.: Key Issues for IT Executives 2010: Judicious IT Investments Continue Post-Recession. MIS Quarterly Executive 9 (4), 263–273 (2010)
5. Sturm, R., Morris, W., Jander, M.: Foundations of Service Level Management. SAMS, Indianapolis, Indiana (2000)
6. Webster, J., Watson, T.: Analyzing the past to prepare for the future: Writing a literature review. MIS Quarterly 26 (2), xiii–xxiii (2002)
7. Jin, L., Machiraju, V., Sahai, A.: Analysis on Service Level Agreement of Web Services. Technical report HPL-2002-180, Hewlett-Packard Laboratories, Palo Alto, California (2002)

8. Sauvé, J., Marques, F., Moura, A., Sampaio, M., Jornada, J., Radziuk, E.: SLA design from a business perspective. In: Schönwälder, J., Serrat, J. (eds.): Proceedings of the 16th International Workshop on Distributed Systems, Operations and Management. LNCS, Vol. 3775, pp. 72–83. Springer, Barcelona (2005)
9. Cheng, F., Gamarnik, D., Jengte, N., Min, W., Ramachandran, B.: Modeling Operational Risks in Business Processes. IBM Research Report RC23672, Thomas J. Watson Research Center, Yorktown Heights, New York (2005)
10. Taylor, R., Tofts, C.: Death by a thousand SLAs: a short study of commercial suicide pacts. Technical report, Hewlett-Packard Laboratories, Bristol (2005)
11. Barroero, T., Motta, G., Pignatelli, G., Bochicchio, M., Longo, A.: Aligning IT service levels and business performance: A case study. In: Proceedings of the International Conference on Services Computing, pp. 570–577. IEEE Computer Society, Miami, Florida (2010)
12. Franke, U.: Optimal IT Service Availability: Shorter Outages, or Fewer?. IEEE Transactions on Network and Service Management 9 (1), 22–33 (2012)
13. Kieninger, A., Satzger, G., Straeten, D., Schmitz, B., Baltadzhiev, D.: Business Cost Budgets: A Methodology to Incorporate Business Impact into Service Level Agreements. International Journal of Service Science, Management, Engineering, and Technology 3 (3), 49–64 (2012)
14. Kieninger, A., Schmitz, B., Straeten, D., Satzger, G.: Incorporating Business Impact into Service Offers: A Procedure to Select Cost-Optimal Service Contracts. In: Proceedings of the 18th Americas Conference on Information Systems, Paper 19. Association for Information Systems, Seattle, Washington (2012)
15. Kieninger, A., Berghoff, F., Fromm, H., Satzger, G.: Simulation-based Quantification of Business Impacts Caused by Service Incidents. In: Cunha, J.F., Snene, M., Nóvoa, H. (eds.): IESS 2013 Porto. LNBIP, Vol. 143, pp. 170-185. Springer, Berlin et al. (2013)
16. Wittgreffe, J., Dames, M., Clark, J., McDonald, J.: End-to-end service level agreements for complex ICT solutions. BT technology journal 24 (4), 31–46 (2006)
17. Breitgand, D., Henis, E., Shehory, O., Lake, J.M.: Derivation of Response Time Service Level Objectives for Business Services. In: 2nd International Workshop on Business-Driven IT Management, pp. 29–38. IEEE Press, Munich (2007)
18. Blau, B., Conte, T., van Dinther, C.: A Multidimensional Procurement Auction for Trading Composite Services. Electronic Commerce Research and Applications 9 (5), 460–472 (2009)
19. OGC, Office of Government Commerce: IT Infrastructure Library (ITIL): Service Operation. The Stationary Office (TSO), London, UK (2007)
20. Bräutigam, P.: IT-Outsourcing: Eine Darstellung aus rechtlicher, technischer, wirtschaftlicher und vertraglicher Sicht. 2nd edition, Schmidt, Berlin (2009)
21. Krishna, V.: Auction Theory. 2nd edition, Academic Press, Amsterdam (2010)
22. Bichler, M.: An experimental analysis of multi-attribute auctions. Decision Support Systems 29 (3), 249–268 (2000)