The Economics of Service Level Engineering

Axel Kieninger
Karlsruhe Service Research Institute (KSRI)
Karlsruhe Institute of Technology (KIT)
axel.kieninger@kit.edu

Jens Westernhagen
Karlsruhe Service Research Institute (KSRI)
Karlsruhe Institute of Technology (KIT)
jens.westernhagen@student.kit.edu

Gerhard Satzger
Karlsruhe Service Research Institute (KSRI)
Karlsruhe Institute of Technology (KIT)
gerhard.satzger@kit.edu

Abstract

Today, the management of service quality poses a major challenge for many service providers and their business customers. We argue that the trade-off between service cost and benefit incurred by both parties is not sufficiently considered when service quality is stipulated. Up to now, neither in practice nor in academia a commonly accepted engineering approach exists to determine business-relevant performance metrics and associated cost-efficient target values to precisely identify efficient service quality. It can be expected, though, that such a systematically developed and economically well-founded Service Level Engineering approach — based on the individual economic conditions of customer and provider — can enhance the value generation in service systems formed by both these parties. In this work we lay the foundation for the field of Service Level Engineering discussing its constitutional concepts and elements. After developing a generic framework for different Service Level Engineering scenarios, we analyze one scenario — the system view — in detail as a quantitative optimization problem: In particular, we aim at the derivation of cost-efficient target values given business-relevant performance metrics.

1. Introduction

Over the last 50 years the service sector has grown continuously and has become the major driver of value creation in most advanced economies\(^1\). Yet, there are still many challenges in service research [6]. Especially the management of service quality is a major challenge for many service providers and their business customers: service quality has to match service customers' business requirements. However, in practice and academia, up to now there is no commonly accepted engineering approach to determine business-relevant performance metrics (service level indicators) and associated cost-efficient target values (service level objectives) to precisely define service quality.

We illustrate this deficit using the example of IT services: In practice, service level indicators (SLIs) and service level objectives (SLOs) for IT services are often determined in an ad-hoc and heuristic way [26], which is sometimes even referred to as “pure guesswork” [19]. In addition, the current process used to define SLIs and SLOs often lacks the end-user business perspective [15].

In the field of IT outsourcing, technical metrics are defined instead of metrics that are relevant to the end-user. This entails that customer expectations are often not adequately considered, because they have to be “translated” into technical specifications. Due to the complexity of the customer’s organization, it is difficult to specify quality metrics that are meaningful to the business. Moreover, in most cases the providers do not have the necessary knowledge about the customer’s business processes to be able to provide a service that is aligned to business requirements in terms of service quality. Instead of delivering according to business requirements, the providers try to reach compliance with the technical SLOs that are stipulated in Service Level Agreements [27]. The lack of knowledge about the customer’s business processes also implies that the trade-off between service cost and benefit is not considered sufficiently. When a service is not delivering value, oftentimes the

\(^1\)https://www.cia.gov/library/publications/the-world-factbook/fields/2012.html
SLOs are raised instead of trying to really understand the business needs. As a result, SLOs are often set inappropriately, leading to overly high service cost compared to the benefit for business [26].

The situation exemplarily described for IT services calls for a systematic engineering approach to determine SLIs and SLOs that are ideal from a business point of view. It can be expected that such a systematically developed and economically well-founded Service Level Engineering (SLE) approach based on the individual economic conditions of customer and provider can enhance the value generation in service systems formed by both of these parties for services in general.

In this work, we want to lay the foundation for the field of SLE by discussing its constitutional concepts and elements in the following section. In particular, we aim at the derivation of cost-efficient SLOs given business-relevant SLIs. Works related to our approach are outlined in section 3, before we present an economic model implementing SLE in section 4. After presenting the various scenarios that are conceivable for SLE, we will model one of these — the system view scenario in which provider and customer collaborate to define service quality — as quantitative optimization problem. Although we use the example of IT outsourcing to demonstrate the application and the benefit of SLE, we characterize SLE as a generic approach that can be applied to various kinds of services. Hence, our work can be regarded as the basis for future work in the research area of SLE. Finally we conclude and outline next steps.

2. Basics of service level engineering

During the phase of service development, provider and customer usually negotiate a Service Level Agreement (SLA) to stipulate the content, the quality, and the price of the service provided [4]. To specify the quality of service, SLIs and corresponding SLOs need to be defined. Because to date there is no commonly used engineering approach to determine business-relevant SLIs and cost-efficient SLOs, we want to propose a generic Service Level Engineering (SLE) approach. The objectives of SLE are:

1. For a service, to determine the SLIs that are relevant from a business point of view and that should hence be used to define the quality of this service in an SLA.
2. For the SLIs determined, to derive efficient SLOs by quantitatively considering the business impact of different SLOs.

To achieve the first objective one must start with the requirements of the business customer. This entails that business-relevant SLIs have to be measured end-to-end from the customer’s perspective [25] instead of being technical metrics for individual service components. For the second objective (on which we focus in this work) one can assume that the business-relevant SLIs are already known and that now efficient SLOs have to be derived. This has to be done by explicitly considering the monetary impact of different SLOs on the business. From these two objectives we derive our definition of SLE:

Service Level Engineering: A systematic engineering approach to determine business-relevant service level indicators (SLIs) and efficient service level objectives (SLOs) for Service Level Agreements (SLAs).

SLE is related to the research fields of Service Engineering, Service Level Management, Business Driven IT Management and Quality Management. Referring to the German Institute for Standardization (DIN), Service Engineering is a research discipline concerned with the systematic development of services (using suitable models, methods and tools) and the management of service development processes [8]. According to Bullinger et al. Service Engineering addresses three dimensions characterizing a service [5]: (i) a structure dimension concentrating on the allocation of resources, (ii) a process dimension defining service creation activities, and (iii) an outcome dimension determining the outcome of a service. Within this discipline, SLE addresses the outcome dimension supporting the systematic definition of service quality.

Service Level Management covers all tasks related to the definition and adaption of service quality parameters and their target values, the monitoring and reporting of service quality achieved and the control of corresponding management processes [27]. SLE supports this field offering enhanced methods to define SLIs and associated SLOs to precisely define service quality in Service Level Agreements.

In the field of IT, SLE can be categorized as an application of Business-Driven IT Management (BDIM). BDIM can be regarded as a continued development of IT Service Management (ITSM), which focuses particularly on the quantifiable business impact of IT [21].

According to the International Organization for Standardization [10] quality management “includes all the activities that organizations use to direct, control, and coordinate quality”. These activities include formulating a quality policy and setting quality objectives. Thus, SLE complements Quality Management as well.

Because services can be provided by either internal or external service providers, we will further differentiate between internal and external SLE. Internal SLE
SLE can further be subclassified. Increasing SLOs (i.e., increasing quality) have a business impact. For instance, if the availability of a service is increased, the probability for outages of this service decreases. Because one can assume that a service outage negatively impacts the affected business (e.g., lost revenues, increased cost, stakeholder dissatisfaction), higher availability should decrease the costs for the business.

Sauv´e et al. denote these costs as “business financial loss” [19]. We believe that this kind of cost can better be denominated as business opportunity cost. In economics, the “opportunity cost” of any decision is usually defined as “the value of the next best alternative that the decision forces the decision maker to forgo” [3]. For the context of SLE we will use the following definition:

**Business opportunity cost:** The cost for the customer that is incurred by an imperfect service compared to the cost of a perfect service.

A service is imperfect when it degrades from a perfect service that is provided at the theoretical maximum of each SLI. For instance, a perfect service would have an availability of 100%, a response time of 0, and a mean time to repair (MTTR) of 0. Therefore, the business opportunity cost of a specific SLO \(x^0\) is equivalent to the additional value that the perfect service would provide compared to the value that the service with SLO \(x^0\) provides. The decision for the SLO \(x^0\) forces the decision maker to forgo the value of the perfect service. Consequently, for every SLO that is optimized in an SLE context, the specific theoretical maximum value that constitutes a perfect service needs to be defined as a reference point. In practice, this maximum value may also be the highest SLO that can be supported by the provider (e.g., an availability of 99.999%). As for variable service cost, we can neglect fixed cost — we only need to consider the additional service cost that is variable with respect to the levels of the SLOs that are to be optimized. Therefore, we define variable service cost as follows:

**Variable service cost:** The cost that the provider incurs for providing the service. It includes all direct and indirect costs that are variable with respect to the SLOs, i.e., depend on the decision which quality of service to provide.

In addition, SLOs also have a business impact. More specifically, higher SLOs should entail lower costs for the business customer. For example, if the availability of a service is increased, the probability for outages of this service decreases. Because one can assume that a service outage negatively impacts the affected business (e.g., lost revenues, increased cost, stakeholder dissatisfaction), higher availability should decrease the costs for the business.

We believe that efficient SLOs can be derived from a model-based optimization approach that minimizes total cost. The basic idea is that service levels and consequently SLOs affect two kinds of cost - variable service cost and business opportunity cost. In the following, we discuss these terms in detail.

Intuitively, variable service cost should rise with increasing SLOs (i.e., increasing quality). For instance, if the availability of a service is to be increased, cost will increase accordingly, because additional or more expensive resources will be required. In our optimization context we can neglect fixed cost — we only need to consider the additional service cost that is variable with respect to the levels of the SLOs that are to be optimized. Therefore, we define variable service cost as follows:

**Variable service cost:** The cost that the provider incurs for providing the service. It includes all direct and indirect costs that are variable with respect to the SLOs, i.e., depend on the decision which quality of service to provide.

As stated by the definition above, SLE is not restricted to IT services but basically applicable to various kinds of services. We argue that SLE is easiest to apply when the service in question is directly used by a business process or a human end-user. In this instance there is a direct business impact of the service, which can be measured. In this work, we will focus on the application of SLE to IT services. Because the context of this work is IT outsourcing, we will only consider external SLE in the following. Furthermore, we will assume that the business-relevant SLIs are already known, and therefore only address the second objective of SLE.

As we have shown, SLE is more complicated, because the service system involves two different organizations.

When services are provided externally, we can distinguish between bilateral and unilateral SLE regarding the second objective of SLE. Bilateral SLE denotes that provider and customer collaborate to derive optimal SLOs. Unilateral SLE in contrast means that one of the parties optimizes SLOs on its own. Again, it can be expected that bilateral SLE is more complex than unilateral SLE, because two parties are involved. However, it can also be expected that the potential benefit of bilateral SLE is higher than that of unilateral SLE. The different forms of SLE are depicted in figure 1.

![Figure 1: Classification of Service Level Engineering](image-url)
because a financial metric is easier to use in a business context compared to utility.

Since variable service cost and business opportunity cost are the cost types that are affected by a change in SLOs, these are exactly the factors that should be considered for the optimization of SLOs. Ideally, SLE models have to derive efficient SLOs by quantifying the trade-off between these two cost types.

Exemplary cost curves for variable service cost and business opportunity cost are shown in figure 2. With an increasing SLO, variable service cost rises, while business opportunity cost declines. Summing up both cost functions leads to a total cost function, which can be minimized. An SLO that minimizes total cost will be denoted as system-efficient SLO. When multiple SLIs are to be optimized simultaneously, the solution is a combination (vector) of multiple SLOs (one SLO for every SLI).

As we will show in section 4, both parties can always be better off by choosing system-efficient SLOs instead of SLOs that are not system-efficient. In practice, however, it may not be possible to derive the SLOs that minimize total cost, because provider and customer may not be willing to share their private cost information. In this case, the question arises how SLOs can be optimized alternatively and how the results of such optimizations will differ from the system-efficient solution. We will discuss this in detail in section 4 as well.

3. Related work

Scientific support on how to define business-relevant SLIs and efficient SLOs is scarce. Within the field of IT, there is a very limited number of papers that derive optimal SLOs for IT services (e.g. [9, 13, 16, 19, 20]). We will discuss some of them in the second half of this section. For services in general, we could not identify any works with this objective after extensive literature research. However, there are two areas in operations research and one area in marketing research that are related to our work: First, there is a stream of research that uses queuing theory to model service systems and that considers the opportunity cost of a service to answer various research questions. Second, there are inventory management approaches that are very similar to SLE. Third, customer value management, which belongs to the field of marketing, provides approaches to determine business opportunity cost.

With respect to queuing theory, Luski [11] for example models two service providers that compete for customers. The customers choose the provider based on the net gain, which is made up of the reward from the service, the price paid to the provider, and the waiting cost for queuing. In an internal service context, Dewan and Mendelson [7] determine the optimal internal pricing and capacity for an internal service facility that is used by multiple user departments. To do this, they maximize the total benefit for the organization, which equals the sum of the value generated by the service, the delay cost incurred by additional service users, and the cost of system capacity. The work by Rao and Petersen [18] aims at deriving a pricing structure for a service with different priority classes, so that an optimization by the individual customers leads to the “system optimal” solution. Thereby, the customers choose one out of different priority classes (i.e., quality classes) by maximizing their sum of revenues, service price, and waiting cost.

The three papers discussed above are related to our work in that the decisions are also targeting a quantitative maximization of the sum of multiple monetary variables. In particular, the waiting or delay cost, which is similar to business opportunity cost, is incorporated into the decision. In contrast to our work, none of these papers (and of others that build on queuing theory) directly addresses the derivation of optimal SLOs. They also only consider one single SLI (waiting time or delay), and not other potential SLIs.

In the field of inventory management, approaches similar to SLE are already used to determine optimal stocking policies [2, 23]. For example, the optimal stocking level or the optimal availability (probability of having an item on stock) can be derived by minimizing the sum of the expected holding cost and the expected shortage cost. Holding cost (cf. service cost) mainly includes the cost for storing and the cost of capital, and shortage cost (cf. opportunity cost) measures the cost of lost customer demand and of short notice reorders. Shy-con and Sprague [22], for example, propose to maintain an optimal service level that balances losses from stockouts (shortage cost) against the cost of carrying larger inventories (holding cost).

Customer value management aims at the demonstra-
To our knowledge, there are only a few papers that address the derivation of efficient SLOs based on service cost and business opportunity cost in an IT context. The first work that proposes that the trade-off between these two cost types should be considered to determine the optimal capacity (and thereby the response time) of an IT system is the one by Doherty and Kelisky [9]. Based on empirical data, the authors note that the optimal system capacity minimizes the total cost of hardware operating cost and user cost (waiting cost). However, they only consider the response time element of opportunity cost. Furthermore, the objective of the paper is rather to discuss best practice approaches for IT management than to derive efficient SLOs.

During the last years a few papers written by a group of authors including Sauvé, Marques, and Moura directly aimed at deriving efficient SLOs for web services. In their first work [19], they use a BDIM approach to find optimal SLOs for availability and response time by minimizing the sum of IT infrastructure cost and business financial loss, whereas business financial loss is defined as the adverse financial impact of an imperfect IT infrastructure on business.

Building on this work, further papers by the authors extend the model by elaborating the derivation of business financial loss based on the four Balanced Scorecard perspectives and on the criticality of different business processes [16]. Only one recent paper [13] adapts the model to IT outsourcing relationships. Here, the objective of the optimization becomes to maximize the “profit margins” of provider and customer.

Compared to the papers by Sauvé et al., our work is much more universal. The cited papers mainly focus on web services within e-commerce scenarios, whereas our SLE approach pertains to virtually all kinds of services. SLE, as we define it, comprises internal, external, unilateral, and bilateral SLE, whereas the cited papers mainly focus on internal SLE. External SLE, which we consider a main area of application, is only briefly discussed in [13]. Contrary to this paper, our model takes a value co-creation perspective within service systems [12] to evaluate the optimality of SLOs. Moreover, our model uses the concept of Pareto optimality to assess combinations of SLOs and service prices. In addition, we differentiate and analyze different scenarios of information sharing between provider and customer.

4. A Model for service level engineering

To demonstrate the application and the benefit of SLE, we will model a simple service setting. This setting can be analyzed for different assumptions that we will make about the sharing of information between provider and customer. We will also discuss under which conditions SLE is beneficial to the service system of provider and customer.

4.1. Scenario description

The service setting is assumed to consist of one external provider that is to provide a certain service to one customer for a specified length of time. For this service, the customer will pay a fixed price to the provider each time period (e.g., every month). The price paid is assumed to be subject to negotiation between customer and provider. This entails that there exists no uniform market price for the provided service and that price differentiation is possible.

For a uniform market price to exist, there are certain economic requirements: the homogeneity of the product, the absence of personal, temporal, or spatial preferences, as well as perfect information. Especially perfect information and product homogeneity are necessary for a uniform market price [3]. However, both are usually not existent in real markets.

Price differentiation denotes that a supplier sells identical products at different prices. In our service setting this means that the provider can individually negotiate the price for the service with the customer. Services are ideal for price differentiation because they cannot be stored or resold. Moreover, price differences for services are more willingly tolerated than for products [24].

With respect to the characteristics of the service provided we assume that the functional properties and all non-functional properties [17] except for quality have already been determined by the context of the business process to be supported. Furthermore, we assume that the SLA for the service specifies one or more SLIs to measure the quality of service. For each SLI exactly one SLO has to be specified. For reasons of simplicity, we will only consider one SLI with one associated SLO. However, the model can easily be extended to a larger number of SLIs. In our model, an SLA will be represented by a combination of an SLO and a price. As discussed, all other elements that are usually part of an SLA (basically functional properties and non-functional properties except for quality) are assumed to be fix [4].

The model variants that we will present in this work address the second objective of SLE; they model the optimization problems in order to derive efficient SLOs.
As discussed in section 2, the cost functions for variable service cost and business opportunity cost are required to do this. When both cost functions are available, different assumptions about the distribution of cost information are possible: it can be differentiated which cost information each of the parties has. We will assume that both the provider and the customer always have complete information about their own cost function. What we will distinguish is whether this cost information is disclosed to the other party. This distinction leads to four settings, which are shown in figure 3. For each setting it can be further differentiated how and by whom an optimization is performed. In the figure we have depicted eight different scenarios. In this work, we will concentrate on the first scenario: the system view.

In scenario A, both cost functions are disclosed: The customer discloses his business opportunity cost function to the provider and the provider reveals his variable service cost function to the customer. This means that both parties are in possession of the same information, which is the simplest setting for SLE. It is a bilateral SLE scenario, because both parties collaborate in order to apply SLE. They co-create value for the service system, which motivates the term “system view”. Although our model is catered to external SLE, note that the system view scenario could also apply to internal SLE.

In setting B, the world is no longer ideal but becomes more realistic. Here, the provider does not disclose his variable service cost function, whereas the client still reveals his business opportunity cost function. Generally, a provider or producer will be very hesitant to reveal his private cost information, because this information may impair his position in price negotiations. A customer disclosing opportunity cost, however, would only reveal the criticality of the service for his business, which makes this cost information less valuable than the one of the provider. Because the provider still possesses all required information to apply SLE, he is the natural choice to optimize the service system (scenario B.1). In practice, the provider could offer to apply SLE as a service, thereby differentiating himself from competitors. Because the customer cooperates by disclosing his cost function, this is again a scenario of bilateral SLE.

It is also conceivable that the customer optimizes SLOs without knowledge of the variable service cost function (scenario B.2). In this case, he will minimize the sum of business opportunity cost and price, whereas the latter could be assumed to be a function of the SLO. This function could be a price list offered by the provider or the customer’s estimation of the provider’s price proposal. The two scenarios of setting B are, for example, addressed by customer value management approaches.

Setting C is based on the assumption that the customer does not want to reveal business opportunity cost, while the provider is willing to disclose his variable service cost function. This setting is highly unlikely in practice, because as discussed the provider will probably not disclose his cost information. Although this setting is not realistic, we would differentiate two scenarios: Similarly to the provider in B.1, the customer could use all necessary information to optimize SLOs from a system perspective (scenario C.1). Alternatively, the provider could maximize the difference of the price paid by the customer and his variable service cost (scenario C.2). In this case, the price could be the provider’s estimation of the customer’s willingness to pay, defined as a function for every SLO.

For setting D we will differentiate three scenarios. Because both parties do not disclose their cost information in this setting, neither provider nor customer obtain the necessary information to apply SLE from a system perspective (by maximizing total cost). In scenario D.1 the customer unilaterally optimizes the SLO based on business opportunity cost and the price paid to the provider. In contrast to B.2, in this scenario the provider does not know the opportunity cost function. It will be of particular interest how this optimization and its results differ from the system view scenario. In scenario D.2 the provider optimizes himself by maximizing the difference between variable service cost and the price paid by the customer. This scenario is equivalent to C.2. Since business opportunity cost is not considered and since the optimization is performed completely unilaterally, this is a unilateral SLE scenario. Scenario D.3 describes a negotiation between the provider and the customer about the quality and the price of the service. This scenario is probably the status quo and can be regarded as a special case of bilateral SLE. Although the parties don’t disclose their cost information, the parties interact intensively in finding an efficient SLO. For each of these scenarios, the following questions should be answered:

1. What is the optimization problem that should be solved to apply SLE?
2. Which conclusions can be drawn about the solution to this optimization problem?
3. Can the SLO that solves the optimization problem constitute a Pareto efficient SLA?

The concept of Pareto efficiency will be used to assess whether an SLA (combination of SLO and price) is efficient in such way that there is no other SLA that better serves the needs of the service system of provider and customer. In economics, an allocation or outcome is said to be Pareto efficient (or Pareto optimal) “if it is impossible to make some individuals better off without making some other individuals worse off” [14]. Applied to our SLE model, this means that an SLA is Pareto efficient, if there is no other feasible combination of SLO and price that improves the situation of one party without worsening the situation of the other one.

4.2. Formal model representation

As we have described before, the basic setting for this model comprises a provider $P$, who provides a certain service to a customer $C$. For this service there is one SLI with an associated SLO $x$, which is subject to optimization.

The SLO $x$ will be assumed to be in a pre-defined closed interval $X := [x_{min}, x_{max}]$. There are two reasons for this assumption. One the one hand, technical constraints may limit the range of an SLO. Response time for example can usually not be 0. On the other hand there could be business requirements which limit an SLO to a certain interval. It may not make sense for a business user to have the response time SLO set to values larger than ten seconds, if the service is designed to be interactive. When determining the interval for one or more SLOs, one should beware of defining the interval too narrowly. A model-based derivation of efficient SLOs is only possible, when intervals for all SLOs are reasonably set. Restricting intervals too much in advance may exclude efficient solutions.

We assume that the cost functions for variable service cost $f(x) : X \rightarrow \mathbb{R}$ and business opportunity cost $g(x) : X \rightarrow \mathbb{R}$ are given. Both cost functions are assumed to be continuous or lower semi-continuous and only dependent on $x$. All other functional and non-functional properties that affect costs are assumed to be fix (cf. 4.1) and therefore do not need to be considered. Furthermore, we assume that the variable service cost and business opportunity cost functions are both constant over all time periods, which entails that cost can be measured per time unit (e.g., per hour or month). Variable service cost $f(x)$ is assumed to be monotonically increasing with $x$, whereas business opportunity cost $g(x)$ is assumed to be monotonically decreasing with $x$. The rationale for the basic shape of both functions was discussed in section 2. As we have also mentioned, fixed cost does not have to be considered for the cost functions. For $f(x)$ this means that only the service cost that is generated by an SLO $x$ compared to the service cost that is generated by the SLO $x_{min}$ needs to be measured. Hence, $f(x_{min})$ can be set to 0. For $g(x)$ it entails that $g(x_{max})$ can be defined as 0, and that only the additional opportunity cost that is induced by lower SLOs than $x_{max}$ has to be considered. We further assume both variable service cost and business opportunity cost to be cash items, which directly affect the cash flows of provider and customer.

The next assumption is that for the provision of the service a price per time unit $p$ is paid by the customer to the provider. Because price differentiation is viable, this price can be subject to negotiation between provider and customer. A tuple of an SLO $x$ and a price $p$ constitutes an SLA $(x, p)$. The last assumption is that all variables and parameters are deterministic.

4.3. The system view scenario

In this bilateral SLE scenario, provider and customer collaborate to optimize the service system. To find the SLO that is optimal from a system perspective, we have to minimize total cost $c(x) = f(x) + g(x)$. This leads to the optimization problem:

<table>
<thead>
<tr>
<th>Find: $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Minimizing: $c(x) = f(x) + g(x)$</td>
</tr>
<tr>
<td>Subject to: $x \in [x_{min}, x_{max}]$</td>
</tr>
</tbody>
</table>

Table 1: Optimization problem for system view scenario

Because $c(x)$ is at least lower semi-continuous, it is also bounded and will have a (not necessarily unique) minimum in the interval $[x_{min}, x_{max}]$. We will assume that a solution of the optimization problem is $x^*$ with $c(x^*) \leq c(x)$ for all $x \in X$. This SLO $x^*$, which minimizes total cost $c(x)$, will be denoted as system-efficient SLO or system-efficient point. Exemplary cost functions for this scenario have already been shown in figure 2. In this example, all cost functions are strictly convex on $X$, and thus, the point $x^*$ is a unique global minimum.

To check whether an SLO $x$ can constitute a Pareto efficient SLA $(x, p)$, we have to look at the cash flows of provider and customer. Because we assume that $f(x)$ and $g(x)$ are cash items that directly affect payment, we can define the cash flows (per time unit) of customer and provider, which are generated by the outsourcing agree-
ment, as

Provider: \( CF_P(x, p) = p - f(x) \) \hspace{1cm} (1)

Customer: \( CF_C(x, p) = -p - g(x) \). \hspace{1cm} (2)

The cash flow of the provider (per time unit) \( CF_P(x, p) \) is the price received for the service less the service cost \( f(x) \). The cash flow of the customer (per time unit) \( CF_C(x, p) \) is the negative price, which is paid to the provider, minus the business opportunity cost \( g(x) \). As stated in section 4.2, fixed cost is not to be considered in this model.

Assume that the provider and the customer stipulated an SLA \( (x^0, p^0) \) with \( x^0 \neq x^* \) and \( c(x^0) > c(x^*) \). This SLA is not Pareto efficient, because both parties can be made better off by moving from \( x^0 \) to an SLO with lower total cost while simultaneously adjusting the price. The provider would be better off, if an SLA \( (x, p) \) would be chosen with

\[ CF_P(x, p) > CF_P(x^0, p^0) \]
\[ \iff \quad p > p^0 + f(x) - f(x^0). \] \hspace{1cm} (3)

The customer would be better off with

\[ CF_C(x, p) > CF_C(x^0, p^0) \]
\[ \iff \quad p < p^0 + g(x^0) - g(x). \] \hspace{1cm} (4)

If provider and customer move from the system-inefficient SLO \( x^0 \) to the system-efficient SLO \( x^* \), there exists an interval \( P^* \) for the new price \( p^* \), so that both provider and customer are better off than with the old SLA \( (x^0, p^0) \). This interval can be defined as

\[ P^* := \bigg( p^0 - (f(x^0) - f(x^*)) , p^0 - (g(x^*) - g(x^0)) \bigg). \] \hspace{1cm} (5)

The definition of the interval implies that

\[ g(x^*) - g(x^0) < f(x^0) - f(x^*), \] \hspace{1cm} (6)

which will hold true if \( x^0 \neq x^* \) and \( c(x^0) > c(x^*) \). Condition (6) simply states that \( x^0 \) is not system-efficient and that therefore \( c(x^0) \) is not a minimum of \( c(x) \) on \( X \).

Now let \( x^0 \neq x^* \) with \( f(x^0) > f(x^*) \) and \( g(x^0) \leq g(x^*) \). This means that \( x^0 \) is greater than \( x^* \), i.e. the stipulated SLO is set too high to be system-efficient. This case is shown in figure 4. By moving from \( x^0 \) to the smaller \( x^* \), service cost \( f \) is reduced while opportunity cost \( g \) is increased. Because the reduction in service cost overcompensates the increase in opportunity cost, total cost decreases. This cost reduction can be distributed to provider and customer by adjusting the price. When a relatively small \( p^* \) within \( P^* \) is chosen, most of the cost advantage is given to the customer, whereas when a relatively large \( p^* \) within \( P^* \) is chosen, most of the cost advantage is given to the provider. The same argument can be made for an SLO \( x \) which is set lower than \( x^* \). In this case it will hold true that \( f(x^0) < f(x^*) \) and \( g(x^0) \geq g(x^*) \). The interval for \( p^* \) for both parties to be better off will again be \( P^* \).

In conclusion, an SLA is Pareto efficient if and only if it is constituted by a system-efficient SLO \( x^* \). When an SLA contains a system-efficient SLO \( x^* \) in combination with an arbitrary price, there exists no other feasible SLA that makes one party better off without making the other one worse off.

Proposition 1. Service Level Engineering under a system view will generate a cost advantage, which can be distributed to provider and customer in such a way that both parties are better off than without Service Level Engineering — the only exception being that provider and customer stipulate a system-efficient SLO by chance.

Proposition 2. An SLA \( (x,p) \) is Pareto efficient if and only if it is constituted by a system-efficient SLO \( x^* \), which minimizes total cost \( c(x) \).

At this point we want to present an example to illustrate the results above.

Example: Assume that the SLI availability shall be optimized for an IT service. This availability is measured end-to-end from the user’s perspective. This is how the SLI availability should be defined to make sure that it is meaningful and comprehensible to the customer [25]. We will assume that provider and customer agreed to derive the optimal SLO \( x \) for availability on an interval between 10\% and 90\%. This means that the domain for \( x \) is \( X = [0.1, 0.9] \). We will further assume that the cost functions for service cost and business opportunity cost (in thousand EUR per month) are given as

\[ c(x) = f(x) \]
\[ g(x) \]

Figure 4: Cost reduction by moving from a system-inefficient SLO to a system-efficient SLO.
We have described a service setting and presented different scenarios based on the private cost information provider and customer disclose to each other.

Our third contribution is an extensive discussion of the system view scenario, which is based on the assumption that both, provider and customer, disclose their cost information. Using the example of IT outsourcing, the scenario shows that SLE will generally generate a cost advantage, which can be distributed to provider and customer so that both parties are better off than without SLE.

With this work we have laid the foundation for the field of SLE by discussing its constitutional concepts and elements. Building on these foundations, we focus our future research on three areas: i) the determination of business-relevant SLIs, ii) the evaluation of the SLE approach in practice, and iii) the extension of the model presented. All areas will be sketched in the following.

In this paper we have addressed the second objective of SLE, whereas the first objective (the determination of business-relevant SLIs), has not been discussed. Currently, we determine business-relevant SLIs for IT services in the context of IT outsourcing, which are measured end-to-end from the customers’ business process perspective. Based on these SLIs, we aim to develop a method to derive business opportunity cost and, thus, allow for the evaluation of the SLE approach in practice. Today, business opportunity cost may not be readily available yet, because providers and/or customers do not possess the know-how to determine these.

Regarding our SLE model, we will approach various possibilities for extension. First, we will consider multiple SLIs instead of only one. Second, instead of assuming all factors as deterministic, we will study the determination of SLOs under uncertainty. Third, we will analyze how penalties for non-compliance with SLOs or other incentives change the results. Fourth, we will incorporate transaction costs of negotiating and determining efficient SLOs into our model, which are particularly relevant in practice.

In addition to these extensions of the SLE model, there are issues we will address in further works. Since we could only present the system view scenario in detail in this paper, we will discuss the remaining SLE scenarios. Moreover, we will deal with the distribution of the cost advantage that is generated by SLE. Besides that, for this work we have assumed that SLOs can be chosen unrestrictedly on a continuum between a minimum and a maximum value. However in practice, it may be more realistic to choose from a finite set of possible SLOs for each SLI. Last but not least, it would be interesting to analyze how external SLE models can benefit from risk-reward sharing between provider and customer.
6. Acknowledgements

This work results from a cooperation with IBM Germany in the context of an SLE project. We thank our colleagues at IBM, in particular Justin Taylor, Claudia Dahl, and Martin Winter, for many fruitful discussions.

References