Towards Compensable SLAs

Carlos Müller(✉), Antonio Manuel Gutierrez, Manuel Resinas, Pablo Fernandez, and Antonio Ruiz-Cortés

University of Sevilla, Seville, Spain
cmuller@us.es

Abstract. Service Level Agreements (SLA) describe the rights and obligations of parties involved (typically the service consumer and the service provider); amongst other information they could include the definition of compensations: penalties and/or rewards depending on the level of service provided. We coin the concept of Compensable SLAs to such that include compensation information inside. In such a context, in spite of important steps towards the automation of the management of SLAs have been given, the expression of compensations remains as an important challenge to be addressed. In this paper we aim to provide a characterization model to create Compensable SLAs; specifically, the main contributions include: (i) the conceptualization of the Compensation Function to express consistently penalties and rewards. (ii) a model for Compensable SLAs as a set of guarantees that associate Service Level Objectives with Compensation Functions. We provide some properties and aspects that have been used to analyse two real-world SLAs.

1 Introduction

The shift from product to services (software provided, human provided, or hybrids) in the industry is a general trend for developed countries. In such a context this evolution implies the creation of a network of dependable organizations that exchange services and, as a consequence, there is a craving for guarantees that support a reliable service consumption. In this challenge, Service Level Agreements (SLA) represents a first-class citizen to describe the rights and obligations of both service consumer and provider.

SLA-driven systems may be established with SLAs as pivotal element to define a set of Service Level Objectives (SLO) that should be enforced by one party (the guarantor) to another party (the beneficiary). In most cases the former correspond to the service provider, and the latter to the service consumer. Such information included in the SLAs may be the basis for the decision making of involved parties in SLA-driven systems along service lifecycle. In this sense, some SLAs may include a set of compensations that represent the consequences of underfullfiling (penalties) or overfullfiling (rewards) the SLOs. We coin the

This work was partially supported by the European Commission (FEDER), the Spanish and the Andalusian R&D&I programmes (grants P12-TIC-1867, TIN2012-32273, TIC-5906 and IPT-2013-0890-3).

© Springer International Publishing Switzerland 2015
G. Ortiz and C. Tran (Eds.): ESOCC 2014, CCIS 508, pp. 31–38, 2015.
DOI: 10.1007/978-3-319-14886-1_4

concept of Compensable SLAs referring to such SLAs that include at least a compensation action, either a penalty or a reward.

The expression of different kinds of compensations and the checking of some desiderable properties that compensations may fulfill, are still challenging tasks. In addition, the automated extraction of information from such compensations would bring important benefits for service consumers and providers in SLA-driven systems in general and in compensable SLAs in particular: (i) on the one hand, providers could automate the optimization of the provision of services based on the compensations involved, and (ii) service consumer could automate the analysis of guarantees in the SLA to understand its risk.

In this paper we go into details about the concept of Compensable SLAs by providing an appropriate specification model for compensation actions, either penalties or rewards. This model incorporates a catalog of properties that can be easily checked. The main element of our proposal is a compensation function inspired in a penalty function provided by Leitner et al. [4], but extended to support the rewards. Our proposal has been validated in two real world SLAs.

2 Motivating Scenarios

SLAs are widely used in the industry in situations where consumers and providers need or desire to explicitly express certain guarantees over the service transaction. These guarantees are typically tied to certain consequences in terms of penalties and rewards depending whether the guarantee is unfulfiled or overfulfilled; we commonly refer to these consequences as compensations.

In this section we motivate the need for formal compensable agreements with two real world scenarios that include both human-driven services and computing services. In both cases, there is a strong need to express compensations related to the guarantees defined.

GNWT Scenario. The Government of the Northwest Territories (GNWT) of Canada outsources the IT support. Specifically, the demanded services include issues related to: reporting, user support, problem correction, application enhancement, process and application improvement, and other services. They provide a template for establishing an SLA with an external vendor providing the mentioned kind of IT support with the desired service levels and compensations. Four examples of terms with compensations have been extracted from its SLA template[1] and they are depicted in Fig. 1.

AWS EC2 Scenario. Amazon Web Services (AWS) is a service catalogue that has boosted the idea of cloud computing in the industry; amongst them, the Elastic Computing Cloud (EC2) represents a widely used Infrastructure as a Service providing a set of virtualized resources that can be consumed and paid on an on-demand basis.

The aim of this service is to provide an escalable infrastructure to organizations that have variable needs or they need to grow seamlessly without the

[1] Available at http://www.fin.gov.nt.ca/ocio/sim/sdlc/3/resources/sla.htm.

Type	Measurement	Penalty
Quarterly Status Report	Delivered at quarterly intervals and not less than five business days before scheduled review meeting	5% of monthly invoice

Sample 1 (page 5 of the template)

Severity Code	Initial Response	Estimation Response	Subsequent Responses	Resolution
1	15 minutes	2 hours	Every 30 min.	4 hours

Type	Measurement	Reward	Penalty
Severity 1 Resolution	All Severity 1 problems are resolved in less than 2 hours.	10% of monthly fees	NA
	One or more Severity 1 problems are resolved in over 4 hours.	NA	10% of monthly fees

Sample 2 (page 7 of the template)

Type	Measurement	Reward	Penalty
Maximum Problem Aging	No problem is older than 60 days.	5% of monthly fees	NA

Sample 3 (page 8 of the template)

Type	Measurement	Reward*	Penalty
Project Delivery	Total elapsed days until delivery is more than 20% greater than planned.	NA	10% of the amount invoiced for the project.
	Total elapsed days until delivery is 20% less than planned.	5% of the amount invoiced for the project.	NA

Sample 4 (page 9 of the template)

Fig. 1. Compensations actions extracted from the SLA of GNWT

investment for an internal data center. In this context, the reliability of a virtualized infrastructure represents a key point for customers (i.e. IaaS consumers) in order to choose a service like AWS EC2.

As a consequence, Amazon has explicitly published an SLA for EC2[2] that is based on the idea of Monthly Uptime Percentage (MUP); this element, characterizes a guarantee over the availability of the virtual resources requested. Specifically, in case the MUP drops below 99.95 percent and in case the MUP drops below 99 percent. In this scenario, the actual compensation is defined as a 10 and 30 percent of discount in the next billing cycle a.k.a Service Credit Percentage (SCP), respectively.

3 Compensable SLAs

In this section we identify the concept of Compensation Function as the key element to express both, the penalties and rewards in a consistent structure that allows a formal definition of compensable SLAs.

A compensation function represents the relation between the level of fulfillment of a guarantee and the possible compensations for the involved parties in the service consumption (the service consumer and the provider). Complementary, in the context of a guarantee two roles can be played: the guarantor and the beneficiary; in most cases, these roles are respectively covered by consumer

[2] Available at http://aws.amazon.com/es/ec2/sla/.

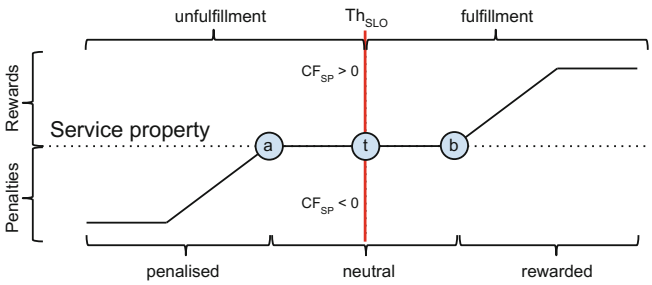

Fig. 2. A generic example of compensation function

and provider. However, there could exist scenarios where service consumers act as guarantor developing some sort of guarantee to the provider and vice versa.

A compensation function for a given service property sp, denoted by CF_{SP}, is a function from SP to $\mathbb{R}$, where SP denotes the set of all possible values of the service property ($SP = \{v_1, ..., v_n\}$). Thus, each v_i maps to a given compensation value and we identify a partition in three sets for the service properties values: Rewarded values that are positive values (e.g. $v_i \mapsto 7$) and represent situations where the beneficiary must compensate the guarantor ($CF_{SP}(v_i) > 0$); Penalised values that are negative values (e.g. $v_i \mapsto -7$) and represent situations where the guarantor must compensate the beneficiary ($CF_{SP}(v_i) < 0$); and Neutral values (e.g. $v_i \mapsto 0$) which represent the lack of compensation ($CF_{SP}(v_i) = 0$).

Figure 2 shows a generic example of an increasing compensation function. This kind of increasing function would model the compensation for service properties such as the MUP in which the higher values the more interesting for beneficiary. On the contrary, decreasing compensation functions would model the compensation of service properties such as the resolution hours in which the lesser values the more interesting for beneficiary. In this sense, to denote that a value v_1 for a service property is less interesting for the beneficiary than other value v_2 by $v_1 \preceq v_2$. Since omission means a lack of compensation in natural language, when there is not explicit definition of a penalty or a reward, we consider $CF_{SP}(v_i) = 0$. Figures 3 shows the different compensation functions from the GNWT scenario. In the Figure, while a dark point denotes the inclusion of the value in the interval, gray points means the value exclusion.

In order to analyse and characterize the possible compensation functions, we identify the following list of interesting properties:

Property P_1 *(Consistent)* Let CF_{SP} a compensation function, it is said to be consistent if the compensation for a lesser degree of fulfillment of service property is lesser or equal than the compensation for a greater degree of fullfillment of such property. This property assures the monotonicity of compensation functions. Samples 1 and 2 of Fig. 3 depict a decreasing and an increasing consistent compensation function, respectively.

$$\text{consistent}(CF_{SP}) \iff \forall v_1, v_2 \in SP \cdot v_1 \preceq v_2 \Rightarrow CF_{SP}(v_1) \leq CF_{SP}(v_2)$$

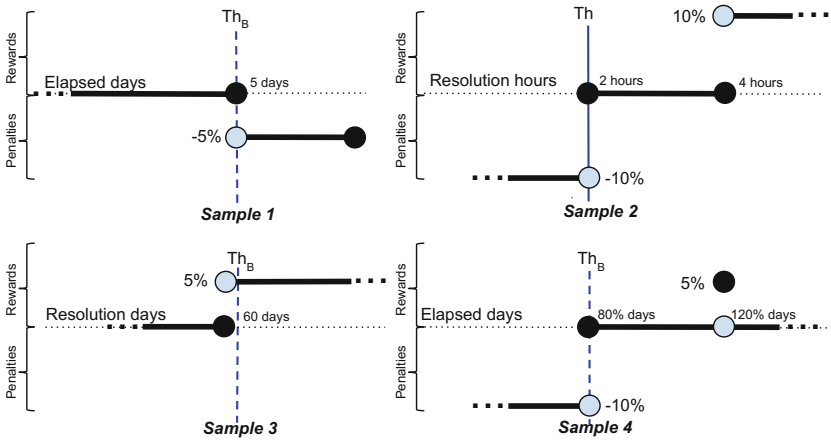

Fig. 3. Samples from GNWT SLA

Property P_2 *(Saturated)* Let CF_{SP} a compensation function, it is said to be saturated if there exist a minimum (v_{min}) and a maximun value (v_{max}) for the service property, that delimit the higher compensation, either penalty or reward.The compensation functions of Fig. 3 are all saturated.

$$saturated(CF_{SP}) \iff \forall v_i \in SP, \exists v_{max}, v_{min} \in SP \cdot \\ CF_{SP}(v_i) \leq CF_{SP}(v_{max}) \wedge CF_{SP}(v_i) \geq CF_{SP}(v_{min})$$

Property P_3 *(Valid)* Let CF_{SP} a compensation function, it is said to be valid if it is consistent and saturated. We consider this because of the following two aspects that, in our opinion a compensation function should assure: (1) on the one hand, consistency assures that the compensation for a more interesting service property value for the beneficiary has more reward or less penalty than a less interesting service property value; (2) on the other hand, saturability assures that penalties and rewards are bounded. Samples 1–3 of Fig. 3 are valid compensation functions, but sample 4 is not valid due to a lack of consistency because the service property value 120 % has more reward than higher values.

$$\text{valid}(CF_{SP}) \iff \text{consistent}(CF_{SP}) \wedge \text{saturated}(CF_{SP})$$

The formalization of Service Level Agreements (SLA) have been a research topic during last decade. Amongst other results, WS-Agreement [1] currently represents the most prominent specification for formal agreements for both academia and industry. This specification formalizes the SLA as a set of terms for describing the service and the associated guarantees. Specifically, a guarantee term, (guarantees from now on) is defined around the idea of a Service Level Objective (SLO) that should be guaranteed (by the guarantor) to the‘beneficiary (e.g. *Response Time* < 100 ms or *Monthly Uptime Percentage* $\geq 99.95\,\%$) Complementary, this specification also includes a placeholder to express penalties and rewards that can be extended with domain-specific languages [6,7].

Taking this conceptualization as a starting point, we coin the concept of Compensable SLAs referring to such SLAs that include at least a compensation action, either a penalty or a reward to be considered when at least one of the comprised guarantees is underfulfilled or overfulfilled, respectively. Thus, at least one of the guarantees included within a compensable SLA is composed of an SLO and a compensation function that defines the penalties and rewards derived from underfullfilling or overfullfilling the SLO. For instance, Sample 2 of GNWT scenario (Fig. 1) includes the SLO in the first table and the compensation function in the second table. The other Samples of Fig. 1 just provide a compensation function but not the SLO. Figure 2 shows a typical compensation function that depicts the relationships between the fulfillment regions delimited by the SLO and the compensation regions defined by the compensation function. Moreover, as shown in this figure, it is important to highlight that fulfillment regions are not necessary coupled with compensation regions; specifically, figure exemplifies a case having neutral service properties values between a and b without any compensation: service properties values between t and b are unfulfilled but not penalized, and similarly service properties values between a and t are fulfilled but not rewarded. In addition, figure shows how threshold Th_{SLO} delimits the fulfillment from the unfulfillment values.

Figure 3 presents our modeling of guarantees identified in the GNWT scenario. Each example, corresponds with a guarantee showing the specific compensation function (as a black line) along with the Th_{SLO} (as a solid vertical line) derived from the SLO; in case there is no SLO explicit, we have inferred a threshold (Th_G or Th_B) depending on whether the SLA was specified by the guarantor (AWS EC2 scenario) or the beneficiary (GNWT and Telecomm. SLA scenarios); these inferred thresholds are depicted as discontinuous lines.

In a guarantee with both, an SLO and a compensation function, we must study the coherence between them. Then we identify the following preconditions that could be checked by an automated tool to assure the validity of guarantees in compensable SLAs: (i) The compensation function should be valid (i.e. Satisfy P_3) (ii) SLO definition must imply a partition of two sets (fulfilled and unfulfilled) in the service property domain (iii) Fulfillment set of values defined by SLO should be coherent with compensation function model: Th should be such fulfilled value that is less interesting for the beneficiary ($Th \preceq v_i, \forall v_i \in SP \cdot CF_{SP}(v_i) \geq 0$) (iv) SLO must define at least a neutral service property value without compensation.

3.1 Analysis of Real Compensable SLAs and Related Work

Table 1 presents a comparative study of the properties fulfilled by the compensations found in both, the real world SLAs of Sect. 2, and examples found in relevant research papers that include the idea of SLAs with penalties and/or rewards. Regarding the properties fulfilled by the compensations of the real word SLAs, we highlight that GNWT Sample 4 of Fig. 3 depicts an inconsistent compensation function; this may represent a mistake in the SLA derived from the usage of natural language.

Table 1. Comparative study of the properties fulfilled by the examples found.

		P_1	P_2	P_3
GNWT Sample 1		✓	✓	✓
GNWT Sample 2		✓	✓	✓
GNWT Sample 3		✓	✓	✓
GNWT Sample 4		✗	✓	✗
AWS EC2 Sample		✓	✓	✓
Letiner et al. [5]	TimeToOffer<=2 (Pag.2)	✓	✓	✓
	OrderFulfillment<=5 (Pag.2)	✓	✓	✓
	ProcessLeadTime<=6 (Pag.2)	✓	✓	✓
	CostCompliance<=5 (Pag.2)	✓	✓	✓
Leitner et al. [4]	SLA Cost = Penalties (Pag.3)	✓	✓	✓
Buco et al. [2]	Penalties (Pag.12)	✓	✓	✓
	Penalties & Reward (Pag.18)	✗	✗	✗
Grabarnik et al. [3]	Penalties & Reward (Pag.7)	✓	✓	✓
Rana et al. [6]	ExecutionTime (Pages 6-7)	✓	✗	✗

With regards to the examples found in the related work, it is remarkable that the proposal of Leitner et al. in [5] formalizes the problem of finding the optimal set of adaptations, which minimizes the total costs arising from SLA violations and the adaptations to prevent them. In this work, a model for penalty functions is presented; this formalization has been the starting point of our motivational scenario description presented in Sect. 2 and consequently, our approach represents an extension to this model in order to develop a complete formalization for Compensable Guarantees and SLAs. Based on our model, we have studied 4 examples (page 2) included presented in [5] relating to the cost of violations of one service property, namely: time to offer, order fulfillment time, process lead time, and cost compliance. In [4] the same authors present an approach for optimally scheduling incoming requests to virtual computing resources in the cloud, so that the sum of payments for resources and loss incurred by SLA violations is minimized. The studied example (page 3) includes a linear penalty function with two point of discontinuities. The example relates the penalty with a service property representing the duration of requests to virtual computing resources in the cloud. Other examples taken from relevant related works are the following: Buco et al. propose in [2] an SLA management system, called SAM that provides penalties in a Service Level Management process. In the first studied example (page 12) the compensation function relates some penalties with a service property denoting the alert time for SLA managers. In the second example some penalties and rewards are specified depending on the service level fulfillment of the SLAs; these last examples define overlapping values for the rewards (i.e. same value can have different compensations) and therefore, the compensation function cannot be defined. Grabarnik et al. propose in [3] a model that can be used to reduce total service costs of IT service providers using alternative delivery teams and external service providers. The studied example (page 7) includes penalties and rewards for a service property that represents the process execution time. Rana et al. identifies in [8] how SLOs may be impacted by the choice of specific penalty clauses. From such a work we have studied an example

(pages 6–7) that relates the penalty of different levels of service execution time. It is important to note that the compensation function does not meet the saturation property and therefore it does not fulfill the validity property.

4 Conclusions

In this paper we characterise the concept of Compensable SLA, that include a definition of guarantees with penalties and rewards for involved parties the Service Level Objective (SLO) is underfulfilled or overfulfilled. We motivate our proposal upon the study of two real world SLAs over human-driven or computing services, and define a model to formally express Compensable SLAs by conceptualizing the Compensation Function as the appropriate artifact to consistently combine penalties and rewards in the context of an SLO. In order to support the modelling and analysis of Compensable Functions and SLAs, we define several properties and aspects about the compensation functions and SLOs of guarantees. This list of properties is used to develop a comparative study over several samples taken from the real SLAs joint with a set of samples found in related work from the literature.

References

1. Andrieux, A., et al.: Web Services Agreement Specification (WS-Agreement) Version 1.1 draft 20, September 2006
2. Buco, M.J., et al.: Utility computing sla management based upon business objectives. IBM Syst. J. **43**(1), 159–178 (2004)
3. Grabarnik, G., et al.: Management of service process qos in a service provider-service supplier environment. In: The 9th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE), pp. 543–550, July 2007
4. Leitner, P., et al.: Cost-efficient and application sla-aware client side request scheduling in an infrastructure-as-a-service cloud. In: 2012 IEEE 5th International Conference on Cloud Computing (CLOUD), pp. 213–220, June 2012
5. Leitner, P., et al.: Cost-based optimization of service compositions. IEEE Trans. Serv. Comput. **6**(2), 239–251 (2013)
6. Müller, C., et al.: Automated analysis of conflicts in ws-agreement. IEEE Trans. Serv. Comput. **7**(4), 530–544 (2014)
7. Müller, C., et al.: Comprehensive explanation of SLA violations at runtime. IEEE Trans. Serv. Comput. **7**(2), 168–183 (2014)
8. Rana, O.F., et al.: Managing violations in service level agreements. In: Grid Middleware and Services Chapter Title - Managing Violations in Service Level Agreements, pp. 349–358 (2008)