An SLA Support System for Cloud Computing

Guilherme Sperb Machado, Burkhard Stiller

An SLA Support System for Cloud Computing

Guilherme Sperb Machado, Burkhard Stiller

Department of Informatics IFI, University of Zurich
Binzmühlestrasse 14, CH—8050 Zürich, Switzerland
[machado|stiller]@ifi.uzh.ch

Abstract. Nowadays, even with the existence of many Cloud Providers (CP) in the market, it is still impossible to see CPs who guarantee, or at least offer, an SLA specification to Cloud Users (CU) interests: not just offering percentage of availability, but also guaranteeing specific performance parameters for a certain Cloud application. Due to (1) the huge size of CPs' IT infrastructures and (2) the high complexity with multiple inter-dependencies of resources (physical or virtual), the estimation of specific SLA parameters to compose Service Level Objectives (SLOs) with trustful Key Performance Indicators (KPIs) tends to be inaccurate. This paper proposes the initial design and preliminary approach for an SLA Support System for CC (SLACC) in order to estimate in a formalized methodology — based on available CC infrastructure parameters — what CPs will be able to offer/accept as SLOs or KPIs and, as a consequence, which increasing levels of SLA specificity for their customers can be reached.

1 Introduction

Within CC environments a contract or a Service Level Agreement (SLA) needs to exist between two parties: Cloud Providers (CP) and Cloud Users (CU), *e.g.*, organizations or individuals. These two parties need to agree on a set of parameters expressed through the SLA. However, even with the existence of many CPs in the market (*e.g.*, Amazon, SalesForce, Rackspace, or Google), it is still impossible today to see CPs, who guarantee or at least offer an SLA specification tailored to CU's interests; however, tailoring interests have a great importance for tomorrows CC, since very general requirements (such as the "availability needs of a given service" [1], [8], [3]) do not match commercial needs for guaranteed CC services. Thus, CPs need accurate definitions of objective values that can be derived automatically and offered to their customers. An example of a specific SLA parameter is the Return to Operation (RTO) time, in case of virtual machine failures. If the RTO is estimated beforehand (and continuously), CPs can compose a Service Level Objective (SLO) offering guarantees of Key Performance Indicators (KPI) with a high precision (*e.g.*, RTO under 3 minutes, measured by the bootstrap time of virtual machines). Nevertheless, due to (1) the huge size of CPs' IT infrastructures and (2) the high complexity with multiple inter-dependencies of resources (physical or virtual), the estimation of specific SLA parameters to compose SLOs with trustful KPIs tends to be inaccurate. This inaccuracy can result in penalties for a CP, if an unrealistic set of values was proposed and consequently agreed upon in an SLA. Therefore, the lack of an automated system that maps and aggregates low-level measures into SLOs is the key barrier for (a) less risky and (b) customer-specific SLA-based CC service provisioning.

As far as known today, there is no past or current work that addresses this problem of mapping low-level measures of interdependent resources into SLOs inherent to typical Cloud services. Moreover, solutions like SLA assessments [5] and SLA monitoring [2] that provide an approach of SLA assessment, do not take into consideration the CC infrastructure as whole, but just very specific network parameters.

Therefore, this paper proposes the *SLA Support System for Cloud Computing* (SLACC) which aims to design, build, and evaluate a Decision Support System (DSS) for CC in order to estimate in a formalized methodology (*e.g.*, statistical analysis, machine learning) — based on available CC infrastructure parameters — what CPs will be able to offer/accept as SLOs or KPIs and, as a consequence, which increasing levels of SLA specificity for their customers can be reached. Furthermore, SLACC will handle specific knowledge about the CC infrastructure in support of the negotiation of dedicated SLA contracts. Thus, SLACC's main objectives include: (1) CPs will benefit from SLACC to propose accurate SLA parameters and SLOs/KPIs beforehand and (2) once CPs receive CU requests for dedicated SLOs/KPIs, the CP can evaluate, if such values can be guaranteed in his CC infrastructure. In both cases SLACC will take into consideration inter-dependencies of resources inside the CC infrastructure, since SLA parameters of high-level Cloud applications are composed by the sum of multiple low-level factors.

2 Approach

The SLACC decision support system will estimate SLA parameters (*e.g.*, KPIs based on SLOs) to enable the design of more specific SLA documents. The system will map high-level requirements into low-level factors that, combined together in a balanced manner, form an estimation. Thus, the following key steps should be undertaken and are described in the remainder of this section: an integrated architecture, a well-defined Cloud IT Infrastructure Model, and an estimation algorithm.

The SLACC must be based on a scalable and fully interoperable architecture. Fig. 1 shows the abstract view of this architecture, which will serve as the starting point for SLACC development. The SLACC interacts with the Accounting Records Repository, the SLAs Repository, and the Infrastructure Model. The Infrastructure Model component enables an updated view of all inter-dependencies of the Cloud IT Infrastructure. It is important to reflect exactly the organization of the physical IT environment, otherwise the SLA DSS will be based on erroneous and not up-to-date data. The CP Operator interacts with the SLA Designer in order to build a well-defined SLA, using an SLA model/language. The SLA DSS can be split in many sub components such as the estimation engine (implementing an estimation algorithm). Such sub components' interface should be defined using an API (Application Programming Interface) to interact, in a standardized manner, with other components of a common SLA management architecture. This API will serve as the CPs openness factor as well as the supporting interface for any inter-domain interactions.

The key mechanism within the SLACC decision support systems is the design and development of the algorithm estimating with a defined level of confidence — may be in a configured manner — SLA parameters, such as the "minimum database query time" for a given application. Based on an example, the principle operation of the estimation algorithm is described as follows. The CU proposes an SLA with a specific SLO, which is the

"RTO of Virtual Machines under 3 minutes". It is known that the Return Time to Operation can be measured in different ways, but the KPI associated to this SLO is measured by a composition of low-level values inherent to the bootstrap of virtual machines. The SLACC decision support system will consult the CC IT Infrastructure knowledge base to check "what are the factors (time-wise) that matter for a successful bootstrap of a virtual machine?". Based on relations defined in an CC IT Infrastructure Model, a set of factors will be determined. In this case, as an example, it can be assumed that the following factors were mapped: (1) network bandwidth from the virtual machine's template repository to the physical server, which the virtual machine will be hosted on — assuming a transfer from the repository to the assigned physical server; (2) processing capacity from the physical server, which hosts the virtual server; (3) average workload of the physical server in an interval period of time; (4) time to deploy and configure the specific requested virtual machine template in the virtual server; and (5) time to (re)configure the deployed virtual machine in the load balancing front-end of the CP. The estimation algorithm will consider a viable distribution to compose and balance these factors to estimate the final result. Statistical methods such as non-linear Regression Analysis can be employed. At the last step, the CP can evaluate based on known facts, if the SLO "RTO of Virtual Machines under 3 minutes" proposed by the CU can be guaranteed by the CP, or if the CP has to negotiate, in this case, this parameter's value to a higher value, or if the CP has to offer different parameter(s).

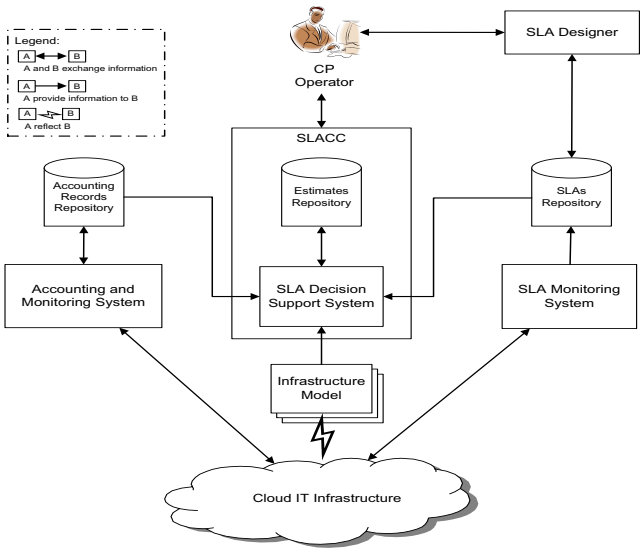

Fig. 1. SLACC architecture

In the process to estimate a certain SLA parameter, the CP operator can intervene within the SLACC system if there is a lack of available information provided by the Accounting Records Repository. Therefore, the operator can manually point some other factors that he judges that has an influence. For example, if the "processing capacity from the physical server (holding the virtual server)" is not measured/specified, the operator can leave it or point an other variable with the same unit (*e.g.*, processing capacity from an other physical server with the same characteristics) to be included in the estimation process.

Existent SLA management solutions like [6], [4], and [7], take into consideration the use of historical data that is collected through SLA monitoring processes or by accounting means. However, SLACC estimation algorithm will consider a wider range of parameters inside the CC IT infrastructure, balancing historical information, current IT infrastructure status (*e.g.*, server's load, network bandwidth at a given moment), and how the Cloud is organized internally, including its IT inter-dependencies.

In order to evaluate SLACC's benefits and advantages, it must be shown that the system can provide accurate estimates to CPs in order to better enhance its SLAs. For this to be proven, it is planned to test the functionality of SLACC by using different CC IT infrastructure's scenarios in a test-bed. Based on the estimates for some SLA parameters, these will be monitored in order to evaluate the confidence level of such generated values. Moreover, comparisons between estimates generated by humans and by SLACC should be observed.

3 Conclusions

This paper sketches a new approach for estimating accurate SLA parameters in order to evaluate what CPs will be able to offer/accept as SLOs or KPIs. The solution proposed and partially outlined in terms of key aspects will increase the level of SLA specificity, not just handling service's availability but also aiming to embrace a wider range of specific performance parameters. The respective and general architecture of the new system termed SLACC was described. An example concerning the estimation algorithm were discussed, also presenting key differences from other approaches in the area of SLA management. Furthermore, for the upcoming fine design, refined solution, and implementation of SLACC (future work), this paper also briefly presented a possible evaluation.

References

1. Amazon.com Web Services: *Amazon Elastic Compute Cloud (EC2) Service Level Agreement*. Available at: http://aws.amazon.com/ec2-sla. Last visited on February 2011.
2. M. Comuzzi, C. Kotsokalis, G. Spanoudakis, R. Yahyapour: *Establishing and Monitoring SLAs in Complex Service Based Systems*, IEEE International Conference on Web Services (ICWS2009), IEEE Computer Society, Washington, DC, USA, 6-10 July 2009, pp 783-790. doi:10.1109/ICWS.2009.47
3. Google.com Apps: *Google App Service Level Agreement*. Available at: http://www.google.com/apps/intl/en/terms/sla.html. Last visited on February 2011.
4. J. Padgett, I. Gourlay, K. Djemame (eds), AssessGrid Deliverable 1.3: *System Architecture Specification and Developed Scenarios*, Version 0.30, December 2006.
5. R. Serral-Gracià, Y. Labit, J. Domingo-Pascual, P. Owezarski: *Towards an Efficient Service Level Agreement Assessment*, IEEE Infocom, Rio de Janeiro, Brazil, 19-25 April 2009.
6. RESERVOIR Project Website: *Service Manager Scientific Report*. Available at: http://www.reservoir-fp7.eu/fileadmin/reservoir/delivarables/A4_ServiceManager_ScientificReport_V1.0.pdf. Last visited on February 2011.
7. SLA@SOI Project Website: *Empowering the service industry with SLA-aware infrastructures*. Available at: http://sla-at-soi.eu. Last visited on February 2011.
8. SalesForce.com Website: *The Leader of Customer Relationship Management (CRM) and Cloud Computing*. Available at: http://www.salesforce.com. Last visited on February 2011.