Impact of Consumer Involvement in Grid SLA Creation

Tarirai Chani¹, Edgar Jembere², Salah K. Kabanda³, Murimo B. Mutanga² and Mathew O. Adigun²

¹Meraka Institute, Council for Scientific and Industrial Research (CSIR)
High Performance Computing (HPC) Research Group
P. Box 395, Pretoria, 0001
Tel: +27 12 841 2122, Fax: +27 12 841 4720
email: tchani@csir.co.za

²Department of Computer Science
University of Zululand, P. Bag X1001, KwaDlangezwa 3886
Tel: +27 35 902 6706, Fax: +27 35 902 6569
email: { ejembere, bethelmutanga, profmatthewo}@gmail.com

³Department of Information Systems
University of Capetown, Private Bag X3, Rondebosch 7701
Tel: +27 21 650 4253,
email: Salah.Kabanda@uct.ac.za

Abstract—Service Level Agreements (SLAs) have been identified as an important tool in guaranteeing Quality of Service (QoS) in Grid service provisioning and consumption in commercial Grids. SLA templates are commonly drawn and predefined by Grid service providers. However, due to the rigidity exhibited by this existing provider-initiated (PI) technique, service consumers are reluctant to take up the use of commercial Grid. We propose a consumer-initiated approach to SLA template creation (C-SLA) which makes use of the Weighted Euclidean Distance Function to select an SLA template appropriate to an individual consumer. Furthermore, we evaluate the impact of this approach on consumer satisfaction as compared to the classical PI approach to determine the possibility of our approach, C-SLA, being a viable solution to be implemented to strengthen the uptake of SLAs by Small Medium and Micro Enterprises (SMMEs) in the commercial Grid enterprise.

Index Terms—Grid Services, Service Consumer, Service Level Agreements, Service Provider, Web Services.

I. INTRODUCTION

With the advent of SLAs in commercial Grids, their templates have always been known to be drawn and predefined by the provider of the grid service. As a result, SLA templates almost always reflect the needs of the provider. This has proved to be unpopular with service consumers. In view of the foregoing, consumers are faced with unattainable business targets [1, 2]. As a result, this has led to the widespread resistance to commercial grid uptake especially by resource constrained Small, Medium and Micro Enterprises (SMMEs). SLAs are thus, still, far from being used as a regular day to day tool for establishing agreements between service providers and consumers [3]. It is essential therefore that Grid service providers be aware that SLAs are experiencing a shift from those sought in the past [1]. Traditionally, SLAs aimed at exceeding a particular threshold of QoS metrics. For instance, a service with an availability metric of 99% would be favored by a consumer but lately, consumers require QoS levels that are linked to tangible business productivity enhancements. Such a metric would probably be unnecessarily too high and thus expensive [1, 2, 7].

In this paper we investigate the impact of involving consumers to a larger extent in SLA creation particularly at the initial stages of the SLA life cycle [5] as we believe that it is at this stage that it is crucial to determine their SLA requirements thereby eliminating rigidity.

We therefore propose that service consumers initiate the SLA template creation phase, one of the phases of the SLA life cycle as described in [5]. By so doing, the consumer makes known their QoS metric and service level requirements as well as preferences to each of these metrics. It is therefore, highly likely that the resulting template to be used as a bases of negotiation would be more reflective of both parties’ needs as compared to existing techniques that are provider-initiated (PI).

An SLA template is selected for a particular consumer to suit their individual needs. To achieve this, the template selection process is based on a similarity distance measure, the Weighted Euclidean Distance Function [4]. In this paper we investigate the impact of using such an approach on
consumer satisfaction. We believe that if this approach impacts positively on consumer satisfaction then it may imply that uptake of commercial Grid may be enhanced especially by SMMEs.

The remainder of this paper is structured as follows: Section II discusses the related work. Section III outlines the proposed solution. Section IV discusses the simulation as well as the performance evaluation conducted. In Section V, we conclude the paper.

II. RELATED WORK

The frameworks to be discussed have been categorized into either third party-based negotiation frameworks or direct service-interaction negotiation frameworks. Within either of these two categories, each framework has also been classified as one that deals with or is specific to single service offerings or composite service offerings (Table I.)

<table>
<thead>
<tr>
<th>Framework</th>
<th>Direct Service Interaction SLA Frameworks</th>
<th>Third-Party Based SLA Frameworks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hasselmeyer et al., 2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaminski and Perry, 2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yan et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ludwig et al., 2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andrieux et al., 2005</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1) Third party-based negotiation frameworks - This refers to frameworks whose interaction and communication of offers and counter offers between parties is conducted through either agents or brokers.

2) Direct service-interaction - A framework that allows for direct communication between the consumer and provider without any agents or brokers involved, communicating on their behalf.

3) Single service offering - This refers to a framework whose SLA creation is based on simple tasks that require invocation of only a single service for them to be successfully executed.

4) Composite service offerings - This refers to a framework whose SLA creation is based on complex tasks that require invocation of more than one service in order to ensure successful completion of that task.

Hasselmeyer [11] proposed outsourcing of negotiation functionalities to a third party Negotiation Broker. This framework requires that consumers have some negotiation components on their side and also risks using two brokers for a single negotiation. This further increases the negotiation costs. The authors argued that the solution guarantees autonomy of individual business entities and allows for gradual migration towards brokered negotiation.

Consumer requirements are specified with the exception of desired preferences and the broker makes the decision. There is no interactive negotiation in the framework, only a third party that makes the decision on behalf of the parties involved. Once the broker makes the decision, there are no further negotiations hence the framework is rigid and might not work well where current SMME have different needs and capabilities.

Kaminski and Perry [7] proposed the automation of SLA creation from a set of Service Level Objectives (SLOs), employing software agents and adopting a social order function by incorporating it into the decision process. Employing these agents has a tendency to impact the cost to the consumer. Their SLA Negotiation Manager makes use of templates proposed in Web Service Level Agreement (WSLA)[6]. This framework offers predefined templates that have a rigid and very limited degree of customization. The customization process is very complex and tedious as the consumer is loaded with the burden of defining data collection algorithms as well as arbitrary input parameters. Such unnecessary activities affect the usability of the system resulting in lack of consumer uptake. The flexibility of this approach is limited and only suitable for a small set of variants of the same type of service using the same QoS parameters and a service offering that is not likely to undergo changes over time.

Yan et al., [12] proposed a framework aimed at supporting autonomous establishment and maintenance of service level agreements in order to guarantee end-to-end quality of service requirements in the provisioning of composite services. To enable this, an innovative framework is proposed in which intelligent agents on behalf of the service requestor and the service providers negotiate SLAs in a coordinated way. The use of intelligent agents can be costly as it will require powerful technology with high memory capacity. These costs could be propagated to the consumer through high price for service consumption, a situation not desirable especially with infrastructure and budget constrained SMMEs in mind. Hasselmeyer et al.[11] acknowledge that high maintenance costs is one of the many reasons SMMEs do not take up proposed Grid SLA solutions in e-Business. No clarification as to what extent consumers’ preferences are captured and whether customization or flexibility, is considered at all during SLA creation. Their work encompasses re-negotiation this implies that consumers are presented with a proposal to which they make counter offers for negotiation and re-negotiation until an agreement is reached. This process proves to be time-consuming and hence costly for the consumer considering that time is money in the business arena.

WS-Agreement [8] specifies a very simple negotiation protocol. It is more of a “take it or leave it” protocol as it does not allow offer refinement. This is a very limited interaction model as a result it limits consumer expressiveness immensely. Consumers can only fill in some customizable parts of the template that are already prepared for by the provider. As QoS parameters are now playing a key role in selecting Grid resources in order to ensure alignment to consumer needs as well as realizing optimized resource usage efficiently [10], an SLA language like WS-
Agreement should allow the specification of these QoS needs in order to derive more satisfactory SLA templates and subsequently make use of the template to select the satisfactory Grid service. This is essential in order to specify tradeoffs between the parameters. For instance, a consumer could be interested in making tradeoffs between response time and availability, or between size and quality of the data. Another interesting possibility for relating QoS parameters is for consumers to give them a weight or level of importance, allowing consumers to establish their priorities. This could prove useful to the SLA provisioning system as it asserts the priority or level of importance attached to that QoS parameter by the service consumer.

As SLAs are basic building blocks for Grid resource orchestration and distributed resource management Pichot et al., [13] showed how a bilateral WS-Agreement based negotiation process can be used to dynamically negotiate SLA templates. For this, they propose a simple extension of the WS-Agreement [8] protocol in order to support a simple offer or counter-offer model. In order to use their model in the WS-Agreement protocol, they proposed a new function called negotiateTemplate. Thus, the negotiation becomes an iterative process before an SLA can be reached. It can, however, be very costly or slow. More so, iterative negotiation indicates lack of knowledge of the initiator’s QoS requirements. The cost of negotiation not only entails monetary costs but also resources usage (i.e. resources needed to facilitate negotiation between disparate administrative domains e.g. network), time spent during negotiation.

Match making techniques using the Euclidean space have been applied in work such as [9] and [14] in an effort to aid in service selection in the grid.

III. THE CONSUMER INITIATED SLA FRAMEWORK

A. Interaction of Components in the Framework

Having identified the need to encompass consumers in the creation of their SLA template, the Consumer-initiated SLA framework (Fig. 1) is proposed: The service consumer application submits QoS requirements of the desired service to the service selector component contained within the generic commercial Grid middleware. This component will select the appropriate service for that given consumer. The consumer requirements information specification consists of the QoS metric levels and the weights attached to these QoS metrics. These weights are an indication of the consumer’s individual preferences on each of these metrics. Once the appropriate service has been identified and selected, its attribute values and other information are captured and inserted into a generic SLA template (gSLA) to form one that is specific to that consumer. The gSLA template will contain all the necessary components [6] which include names of the signatory parties, guarantees, obligations, except for the Service Level Objectives (SLOs) which would be obtained from the appropriate service selected by the service selector for that given consumer. The resulting SLA template is then displayed to the consumer for approval.

If the consumer chooses to reject the SLA template, they would be requested to change its initial QoS requirement submissions and the selection process is redone or it can choose to discontinue by terminating the process. The consumer can also terminate the selection process if it takes longer than desired. However, if the consumer is satisfied with the selected SLA template, it would have to accept it. By accepting the SLA template, an SLA is established between the two parties and hence service provisioning commences.

During service deployment, conformance to the SLA is monitored and policy enforcers are employed to avert any predicted violations of the SLA. Given that a violation has occurred, compensation procedures are taken as stipulated by that particular SLA. After service provisioning, the SLA is stored for record purposes as well as the audit trail for the whole transaction.

B. The Service Selection Algorithm

For simplicity purposes, we assumed 4 metrics including availability, reliability, response time, and Cost respectively. Let $S = \{\text{all services in a grid system}\} = \{S_i\}$ where $i = 1,...,l$ is the service index, $j = 1,...,m$ is the variant index within a service.

A consumer requirement information specification $(R)$ is given by,

$$R = f(X,Y)$$

where, $X = (X_1, X_2, X_3, X_4)$ is a tuple representing consumers preference values for a requested service.

$Y = (Y_1, Y_2, Y_3, Y_4)$ is a tuple representing the weights attached to service attributes.

Service attributes refer to the QoS metrics associated with each service variant. We use the terms service attribute and QoS metric interchangeably.

Due to the difference in the measurement units of each of the QoS metric values, there is a need to normalize [9] them to be in the range $[0, 1]$. As suggested in [9], we use the following equations:

$$\text{Normalize } (X) = \frac{X_{max} - X}{X_{max} - X_{min}} \quad (2)$$
Normalize \( (X) \) = \( \frac{X_v - X_{v_{\min}}}{X_{v_{\max}} - X_{v_{\min}}} \)  \( (3) \)

where \( X_v \) = the QoS metric to be normalized,

\( X_{v_{\max}} \) = the maximum value for the same QoS metric for each relevant grid service variant returned by the system, and

\( X_{v_{\min}} \) = the minimum value for the same QoS metric for each relevant grid service variant returned by the system.

QoS metrics are normalized either by minimization using equation \( (2) \) or by maximization using equation \( (3) \). A metric, which optimizes user utility when its value is minimal, like \textit{response time}, would be normalized by minimization using equation \( (2) \) while those, which optimizes user utility when its value is maximal, like \textit{reliability} would be normalized by maximization using equation \( (3) \).

We defined an array \( A = \{1, 1, 0, 0\} \). A 0 indicates that the QoS metric in question should be normalized using equation \( (2) \) and a 1 indicates that QoS metric in question should be normalized using equation \( (3) \) \( [9] \).

Let \( Q_j = \{\text{QoS values of a service variant } j\} \) and \( Q_{jk} = \) the QoS attribute value for the \( k \)-th attribute of variant \( j \) where \( j = 1, ..., m \), \( k = 1, ..., n \), \( m \) = number of variants of a particular service, and \( n \) = number of attributes used to differentiate variants of a particular service.

If \( d_{\text{min}} \) = the Euclidean distance for a particular variant an SLA template is selected by considering the preference values \( (X) \) and weights \( (Y) \) expressed on the service by the consumer.

Therefore, the weighted Euclidean distance function is used to measure the closeness of the consumer-defined preference values to the available variant values \( Q \). This is given by:

\[
    d_{\text{min}}(X, Q_j) = \sqrt{\sum_{i=1}^{n} Y_i (X_i - Q_{i,k})^2}
\]

The SLA template corresponding to the minimum value of \( d_{\text{min}}(X, Q) \) represents a selection that is close to consumer preference and is therefore recommended.

IV. SIMULATION RESULTS AND ANALYSIS

A simulation was conducted using Java with Development Kit Version 6 (JDK6). We simulated both the Consumer-Initiated and the Provider-Initiated SLA template creation framework. Fig. 2 illustrates the simulation setup.

The simulation monitor component acts as an entry point to pass parameters to the two schemes as well as to collect the results. The system generates service requests and both schemes would have to select and recommend the most appropriate service variant.

A. Impact of C-SLA on Consumer Satisfaction

To evaluate the impact of the proposed approach, C-SLA template selection framework against the classical Provider-initiated SLA (PI) creation framework, the following performance metric was used:

\[
    \text{Consumer Satisfaction} = \begin{cases} 
    \text{distortion in consumer satisfaction} & \text{if } d_{\text{min}}(X, Q_j) < P \text{ and consumer's Satisfaction Threshold utility} (d_u) \text{ given by } d_u = \sqrt{2(1-P)}, \text{where P is the consumer defined satisfaction threshold. Given that } d_{\text{min}} \text{ falls above the consumer’s Satisfaction Threshold utility this implies that the select SLA template derives decreased consumer utility.} \\
    \end{cases}
\]

B. Experimental Results Analysis

This section describes in detail the experiments carried out. It also analyzes and discuss the results that were obtained from the experiments.

1) Experiment I: Satisfactory level of PI and C-SLA

AIM: Conducted to determine the consumer satisfaction that each approach derives.

The experiment was set up as follows: 900 consumer requests were randomly generated with specified QoS parameters. These requests were made to a database with 502 simulated services. The selection of the appropriate services was then done by the two schemes. Both the requests and the corresponding results generated by each scheme were observed and analyzed. To test the performance of the two approaches, a z-test of difference of two proportions was conducted upon the observed results.

The hypotheses were as follows:

i) Null hypothesis \( (H_0) \): There is no difference in consumer satisfaction of the C-SLA and PI SLA template selection process.

ii) Alternative Hypothesis \( (H_1) \): Consumer initiated SLA template creation better satisfies the consumer than the Provider Initiated one.
TABLE II shows the results obtained from the experiment.

<table>
<thead>
<tr>
<th>PI</th>
<th>C-SLA</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>327</td>
<td>752</td>
<td>0.5994</td>
</tr>
</tbody>
</table>

The observation (p) and sample size (n) are shown in the table above. The z-test statistic for the difference between two proportion was calculated from the data in TABLE II using the equation (5) below.

\[
z_{act} = \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}
\]

where \( \hat{p} = \frac{p_1 + p_2}{n_1 + n_2} \)

Since the calculated z-test statistic is greater than the statistic obtained from statistical tables (Z(0.05) = 1.449), we rejected the Null hypothesis (with 95% confidence) and concluded that the Consumer-Initiated SLA template creation process performs better than the Provider Initiated one. This is due to the fact that the C-SLA framework has knowledge of the consumer requirements before it can select an SLA template. It is based on this knowledge that the framework recommends a satisfactory SLA template.

3) Experiment III: Average RRT Vs. Number of Service Variants

This experiment was carried out in order to determine the effect of an increase in the numbers of service variants on the Request Response Time (RRT) (i.e. the time it takes to retrieve an SLA template after a request has been made). A particular service variant was searched for as numbers of service variants were being increased. The average time taken for that particular service variant to be retrieved was recorded. The scatter plot presented in Fig. 4 shows a linear relationship between RRT and Number of service variants. This implies that the C-SLA framework scales well with increases in the number of service variants.

It was observed that the increase of SLA template requests does not in any way affect the C-SLA framework’s ability to recommend a satisfactory SLA template to the consumer. For instance, an increase in the number of requests from 500 to 800 maintained a steady template acceptance ratio of 0.8 for C-SLA whilst that of PI fluctuated from 0.33 to 0.34 to 0.36 and then dropped to 0.35. The huge gap between the performances of the two frameworks is attributed to the C-SLA’s ability to select and recommend an SLA template based on prior knowledge of the consumers QoS preferences. The C-SLA, therefore, scales well with an increase in number of requests and maintains high levels of Template Acceptance Ratio (TAR).

The possibility of service consumers initiating or taking
part in SLA template creation in commercial Grid markets has not been fully explored in literature but, can impact positively on consumer satisfaction as revealed by C-SLA framework. This factor would most definitely aid in enhancing commercial Grid uptake by SMMEs.

Although C-SLA has been proved to be an applicable approach to dynamic selection of the appropriate SLA template, it has some limitations which could be recommended for future enhancements. For instance, in practice network factors would have a direct effect on many QoS properties such as response time, negotiation time etc. In this work, these factors were not considered. It would be interesting to see how these network-metrics influence the behavior of C-SLA or vice-versa. Also in our work, only limited QoS metrics were used for the experiments and evaluation. In future, we intend to investigate the impact of other QoS metrics in the SLA template selection.

ACKNOWLEDGMENT

Gratitude goes to the University of Zululand’s Centre of Excellence - Centre for Mobile Computing, for providing the facilities from which this work was undertaken. This work was done while the principle author was still affiliated there. The authors would like to acknowledge Telkom South Africa and Huawei South Africa for funding this work.

REFERENCES


BIOGRAPHY

T. Chani is currently with the Meraka Institute of the CSIR where she is pursuing her PhD studies (e-mail:tchani@csir.co.za, tarraiachani@gmail.com). She studied her Msc. At the University of Zululand and obtained her Bsc. Hons. Degree at the Midlands State University, Zimbabwe. Her research interests include, grid security, Service Level Agreements and web services.

E. Jembere is with the University of Zululand where he obtained his MSc. He is currently doing his PhD with the Department of Computer Science.

S. Kabanda is with the University of Cape Town, where she is lecturing and pursuing her PhD with the Department of Information Systems. Her Msc, she obtained from the University of Zululand

M. Mutanga is with the University of Zululand, where he obtained his MSc. He is currently doing his PhD with the Department of Computer Science.

M.O. Adigun is the Professor and Head of the Centre of Excellence, Computer Science Department, University of Zululand.