

SLAs A Key Commercial Tool

Bryce Mitchell¹, Paul Mckee²

¹*BT, Orion Bldg G PP13 Adastral Park Martlesham Heath, Ipswich, IP5 3RE, UK
Tel: +44 (0)1473 606172, Email: bryce.mitchell@bt.com*

²*BT, Orion Bldg G PP13 Adastral Park Martlesham Heath, Ipswich, IP5 3RE, UK
Tel: +44 (0) 1473 606185, Email: paul.mckee@bt.com*

Abstract: SLAs are likely to become a key tool for the service provider in offering commercial services in Service Oriented Architectures. A well chosen SLA helps in the establishment of credibility for the service provider; it is a means to attract customers in a competitive environment and establishes the customer-provider business relationship. It may help to retain customers and will be used as a mechanism for service differentiation. SLAs have many complex issues associated with them and are difficult to use and understand, particularly in a value chain, so a common approach to the representation of SLAs is desirable. Most of the current SLA representations focus almost exclusively on the functional aspects of the service being offered such as performance levels and ignore the non-functional factors that have a considerable influence on the successful establishment of a business relationship. Automation of this process is further inhibited by the confidentiality requirements that often surround commercial SLAs.

1. Introduction

The development and use of modern commercial software is driven by a number of trends, businesses are always striving to reduce cost, be it of operational support staff or the hardware upon which their systems run. At the same time there is a growing need to be able to introduce new applications quicker and with lower costs to be able to rapidly respond to changing demands or market requirements. It is therefore highly desirable that new applications be assembled quickly and easily from existing high level building blocks, by application domain specialists rather than laboriously coded from scratch. There is an increasing need for business to business collaboration and increasing dynamism and agility in the application stack.

In order to address these goals a number of technological innovations are taking place in the enterprise. Service oriented architectures (SOA) are increasing in importance as the preferred direction for the implementation of new enterprise applications. The goal of SOA's is to move away from monolithic inflexible applications towards flexible services that allow reuse and on demand access. This represents the latest attempt to reap the benefits of modularity and reuse promised by all component based technologies and is seen as a major step towards business agility. This increased flexibility and agility does of course pose a major challenge in the resource layer of enterprises. Management of the hardware assets of a company is seen as an area where cost reductions are also desirable and the management task is addressed by the introduction of virtualisation to hide the complexities of the hardware layer from the software. But when viewed in conjunction with the service oriented architecture the execution and management of modularised service components requires a much finer grained view of the underlying resources.

This represents an ideal opportunity for the use of Grid technology. As Grid is intended to effectively manage collections of resources it provides an ideal framework for SOAs to produce a platform with the potential for flexible process execution and business

agility. With the convergence of Web Services and Grid standardisation efforts the potential industrial application of Grid technology to achieve the required reductions in cost with the added flexibility looks a foregone conclusion.

There are however a few remaining problems that need to be resolved. All of the current Grid systems deployments do not offer true guaranteed qualities of service, and if automated deployment of services on demand is to become a reality then the performance requirements of end users need to be considered at service deployment time. This shortcoming is well recognised and there are a number of, sometimes competing, initiatives underway to address the issues of guaranteed quality of service and hence service level agreements (SLA) in Grid environments.

In the standards arena the Global Grid Forum (GGF) has a proposed standard, WS-Agreement for specifying agreements between service providers and consumers [1] which may be evaluated as part of the CREMONA architecture [2] developed by IBM. Previous work by Dan et al of IBM focussed on WSLA driven automated management [3]. Other vendors have similar activities Hartig and Reedy of SUN [4] report on a project designed to associate SLAs to applications in a dynamic environment while Sahai et al from HP consider automated SLA management for Web Services [5,6,7]. An interesting model driven approach to create precise SLAs is under development by Skene et al at UCL in the UK [8] as part of the TAPAS project.

Within the NextGrid FP6 project our goal is to define an architecture for Grid that meets the requirements for business users, obviously quality of service is a prime consideration and will be the focus of our attentions. In addition to the technical challenges of SLA negotiation, enforcement and management this project will address the business issues surrounding SLAs in Grid and ensure that SLAs provide a proven business benefit when viewed against the costs and consequences they embody.

2. Role of an SLA

Given all of the research activity surrounding SLAs in service oriented architectures and Grids it is perhaps easy to forget that there is a large knowledge base to draw upon in the area of SLAs, their content and the business implications. The Information Technology Association of America (ITAA) has published a document detailing the key issues that application service providers (ASP) need to consider when negotiating an SLA [9]. The TeleManagement Forum (TMF) publishes a 4 volume SLA management handbook [10] that examines in great detail the role and content of an SLA. The following discussion draws on both of these sources.

The key initial observation regarding SLAs is that they are a commercial tool and the resources required to monitor, enforce and administer the SLA must be justified in terms of business benefits for both the provider and consumer. If we consider the simplest case: a customer agrees a level of delivered service with a supplier, at a price, and with a suitable penalty for failure to deliver. In order for the system to work the price and penalty must have real meaning. This means the price and penalty values must be rooted in financial terms and must be placed in the context of a marketplace for the SLA to have relevance. The consumer must have the choice of an alternative supplier and the penalty must impose some cost on the supplier. A commonly used guarantee "Complete Satisfaction or your money back" is actually worthless in a service context as it allows the provider to provide a limited service with payment from those that successfully manage to use it. This imposes no real obligation on the service provider. Similarly in a situation found in many academic and commercial environments where there is an in-house mandatory supplier any SLA has little impact, the customer cannot exercise choice and use

another supplier, and the supplier will continue to be financed until a catastrophic failure occurs.

Given our assertion that an SLA is primarily a commercial tool what roles does an SLA fill:

1. The SLA manages customer expectations; by supplying a clear and unambiguous description of the service problems may be avoided. The customer has a clear view of costs, the availability of reports, and the scale of any associated penalties.
2. SLAs with a standard format support product differentiation and aid consumer choice, this is particularly relevant in a dynamic market place.
3. SLAs aid in the attraction of new customers and the retention of existing customers
4. As an advert: an SLA is indicative of a supplier's confidence in their ability to deliver, and may be used as part of the discovery and selection method together as a mechanism for establishing trust between providers and consumers.
5. Helps consumers evaluate their contracted service.
6. The SLA may aggregate risk for the consumer, service providers may incorporate the services of other providers in their product offerings. The principle supplier will in that case have subordinate SLAs with the subcontractors, and will have to make appropriate assumptions regarding the degree of risk they are happy to assume in guaranteeing the composite service.

SLAs are not a universal panacea and a number of negative issues need to be considered. Many of these are discussed by Thomas Davis and Fitzgerald [11] and focus on the problems associated with overzealous application of SLAs, and the reliance on them instead of normal commercial due diligence. Attempting to specify an overambitious SLA with heavy penalties may increase the cost to a level that the customer finds unacceptable. The penalty may itself be unenforceable. An overly large penalty may cause financial difficulties in the provider and cause a loss of service, particularly problematic if the service is not available from another provider.

As a final observation neither party wants the SLA to be violated, the customer wants a high level of service for their key business processes, not a payment for SLA violation, which will never compensate for the loss of business. The provider similarly does not want to suffer the loss of market trust and credibility, which may affect many more accounts than that affected by the SLA violation. The SLA must be in place to foster a co-operative business approach to common goals not an adversarial relationship, and the content of any SLA should reflect this partnership.

3. Structure of an SLA

This section aims to illustrate what we believe is the structure of a SLA that can be used within a commercial environment. This section discusses current representations of SLAs and the emerging specifications from the Grid community. The section will then describe the contents and structure of our proposed specification, comparing and contrasting with other proposals as appropriate.

Existing SLA Agreements

SLAs have been a key tool for a number of industries, such as telecommunication and application providers, for a long period of time. These industries now have considerable expertise and experience in the use of SLAs to maintain and develop successful business relationships.

In the emerging world of Grid a number of efforts have been made to produce a standard specification for the representation of a SLA. The two lead efforts have been made

by IBM with the WSLA specification [3] and more recently the Global Grid Forum (GGF)'s WS-Agreement specification [1]. Both of these specifications describe an XML language for representing the agreement between the two parties.

The proposed structure and content of the emerging specifications from the Grid community differ considerably from the more mature SLA's currently in use by the industry sectors mentioned above. In our opinion the proposed specifications place too much focus on describing the monitoring of the service and not enough attention to the business issues, key for a commercial relationship to be successful.

In general a service provider will be using a SLA in order to manage the risk they are exposed to by offering the service at the specified levels and the service consumer will be using the SLA in an attempt to gauge the confidence of the provider in being able to achieve the stated quality of service. It is not possible to make these judgements by focusing solely on the low-level measurable metrics of the service, as is the aim of WSLA and WS-Agreement; instead a more rounded view of the service offered and the requirements of the consumer must be presented.

It is expected this rounded view will include some elements that are not requirements of the service itself but are important elements in the creation of an environment for service execution that the consumer feels confident is as at least as secure and resilient as one they could provide themselves. For example a typical consumer of a service will most likely want to know details of the disaster recovery and backup provisions put in place by the provider to ensure service continuity. To this end the SLA may include a clause to express the mean time to recover from a failure. There are increasing legislative requirements in many consumer oriented industries that create new requirements for SLAs. For example the length of time log data must be available for auditing, or the geographical location data is processed. All of these terms need to be assured by an SLA. These requirements may exist long after the primary provider – consumer interaction ceases and provide a major challenge for SLAs, particularly in the financial industries. .

In other cases the service consumer may wish to gain a view of the policies in place within the service provider that influence the operational environment. For example, the company's help desk and support provision may be referenced in such a policy. Issues such as a companies support policy are more general and less likely to change than the quality of service levels offered for particular services; but they need to be included in a SLA, if it is to achieve the aim of facilitating the commercial relationship between the service provider and consumer.

Proposed NextGRID SLA

Within NextGRID we have attempted to exploit the partners' knowledge of existing SLA structures and the requirements partners have identified in the role of both providers and consumers to build a representation of a SLA that can truly be used to define a business relationship between parties. We have also concentrated effort on producing a management framework to evaluate the performance of the SLA over the service lifecycle. Our proposed representation is expressed as an XML Schema that is composed of three sections. The first section details the parties involved in the SLA, the second section describes the negotiable terms that correspond to the elements to be monitored of the service and lastly the final section contains the general more static terms to represent the environment the service will operate in. Each section is illustrated in Figure 1 and discussed in more detail below.

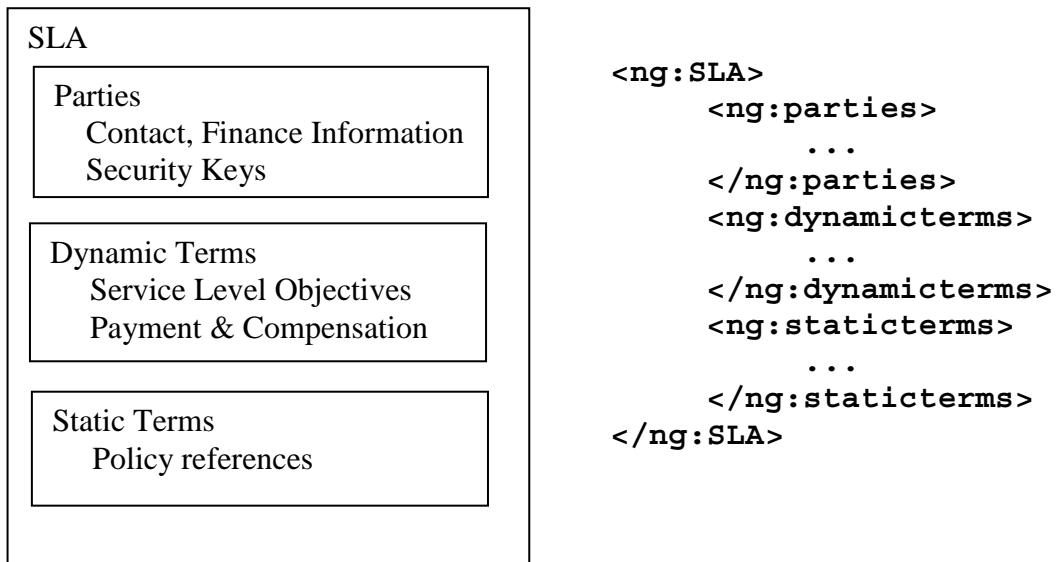


Figure 1- Structure of NextGRID SLA

Parties

This section uniquely identifies the SLA and will contain references to the service involved and agreement time scales. Similar to the definitions pre-amble in a contract that define the involved parties this section explicitly declares the provider and consumer of the service, with reference to a named individual who is accepting the SLA on behalf of the company. Details of the bank account to credit and debit with payments from either party will be provided in this section. As the SLA contains sensitive information such as bank accounts it is obvious the document must be encrypted in some form, as will messages that flow between the parties on the state of the SLA and the service. In order to provide this security keys may be provided or referenced within this section.

This section also allows for the declaration of one or more supporting parties. This could be a third party who is trusted by both the supplier and consumer of the service who will also receive the monitoring events and will be the agreed point for arbitration over any disputes over SLA violation.

WS-Agreement's *context* section is very similar in purpose to this section. Although WS-Agreement does not go to the lengths of defining named individuals or financial and security information.

The section will be specific to the individual agreement but will only require completion of the details: no negotiation over these terms will take place.

Dynamic Terms

The details in this section will outline what is to be measured, how it is to be measured, the threshold for violation and the statistical method for combining the measurements to determine if violation has occurred. This section has been heavily influenced by the notion of *metrics* in WSLA.

The details of this section will be determined by the service covered by the SLA and will be limited to the elements that it is possible to measure. These measurements will either have to be able to be captured from sensors or probes operating within the network or from some other proxy component. A negotiation process will most likely occur between the two parties before agreeing on the SLA. Due to the complexity, cost and time involved in individual negotiation of each value it is thought more likely that a package of offerings will be presented with the option to choose different pre-quantised quality levels. For this

process to be economical it should be largely automated and our proposed XML format allows for this.

Static Terms

This is the section that is most alien to the existing specifications proposed by the Grid community and contains the elements we consider to be key in order to use SLAs in a commercial world. The processes and policies that a provider has in place can aid the provider in the differentiation of their service from competitors.

The contents of this section will be originally based around the guidelines produced by the IT Association of America for Application Service Providers [9] and will describe not only the behaviour but the overall environment the application will run in. It is not thought that any of these terms will be negotiable, as they will either be offered by the service provider or not. For example the provider may be ISO 9001 compliant or not. This section may be thought of as similar in intent to the due diligence investigation that is often undertaken at the formative stages of any normal business relationship. As mentioned below possession of these attributes may be sufficient to allow providers to move to an approved list and then be capable of joining an automated market place. These attributes are those that contribute more directly to the business oriented goals of an SLA and help establish a trust relationship between provider and consumer of services, coupled with a community generating references a good picture of a suppliers capability may be inferred prior to any contract negotiations

4. Automated Processing and Confidentiality

In many situations SLAs are standard and not negotiated, this may be typified by the gold, silver, bronze service levels offered in many commercial environments, and will of course lead to reductions in cost. They may also increase customer confidence as they are buying a standard product with known performance characteristics and greater confidence in the ability to deliver. If the customer requires a specialised SLA heavy negotiation may be required particularly if additional monitoring is required and costs will inevitably rise. There is therefore considerable commercial motivation to minimise variation and cost when offering SLAs.

In a future Grid environment it will be highly desirable that negotiation of the SLA parameters be automated and that the resulting SLA be used as input to an automated resource allocation system in order to meet the required level of performance. This raises a number of problems associated with the confidentiality requirements of the process and the size of the negotiation task. In our outline structure for a SLAs contents we have produced a 3 section SLA, only the dynamically variable properties of the system need to be negotiated at or near run time. There are a large number of parameters associated with the SLA that are either not negotiable or are largely invariant and these may be used as some sort of pre-qualification assessment for service suppliers and consumers.

We believe that commercial adoption of grid will not result in a completely open and dynamic electronic market place instantaneously. Companies require trust relationships with their key suppliers and in many cases have lists of pre-approved suppliers for many commodities that they routinely purchase. We foresee a similar structure in early commercial grids supported by the rich data specified in a complete SLA. A secondary benefit of the pre-selection is to manage access to the SLA negotiation process. A competitor masquerading as a potential customer might obtain valuable market sensitive

information by going through an SLA negotiation and obtaining details of prices and performance and this can be prevented with the pre-qualification procedure.

5. Conclusions and Further work

There are fundamental questions that need to be answered regarding the successful use of SLAs in a Grid context, particularly concerning their exact role in any business relationship. It is essential that a business benefit be demonstrated for both service providers and service consumers and that realistic expectations are established for their use. In addition it is likely that any proposed system will be required to operate at large scale requiring both simplicity and low cost, making automation a key requirement. Any proposed SLA management system must be fit for use by business or it will not be adopted.

The NextGRID project will continue to evaluate the proposed schema for SLA by using it in a number of experiments ranging from negotiation of a SLA, monitoring and evaluation of the dynamic terms and the provisioning and enactment of the service covered in the SLA. By using the SLA schema in these experiments and through feedback from other NextGRID partners it is hoped to gain a fuller understanding of which elements of the schema are useful and which may benefit from further refinement. It may also be possible that the need for other elements may become apparent.

The project also hopes to provide feedback to the Grid community on their specifications based on the experience of using our schema in the project.

References

- [1] WS-Agreement <http://www.gridforum.org/Meetings/GGF11/Documents/draft-ggf-graap-agreement.pdf>
- [2] Cremona: An Architecture and Library for Creation and Monitoring of WS-Agreements H Ludwig, A Dan, R Kearney. ICSSOC04 Nov 2004
- [3] Web services on demand: WSLA-driven automated management, A Dan et al , IBM Systems Journal Vol 43, No 1, 2004
- [4] Associating Service Level Agreements to Applications in a Dynamic Environment. K Hartig, D Reedy <http://rio.jini.org/>
- [5] Towards Automated SLA Management for Web Services, A Sahai A Durante and V Machiraju HP Laboratories Technical report July 2002 <http://www.hpl.hp.com/techreports/2001/HPL-2001-310R1.html>
- [6] Specifying and Monitoring Guarantees in Commercial Grids through SLA, A Sahai, S Graupner, V Machiraju and A van Moorsel, 2002, <http://www.hpl.hp.com/techreports/2002/HPL-2002-324.pdf>
- [7] Automated SLA Monitoring for Web Services, A Sahai, V Machiraju, M Sayal, L J Jin and F Casati <http://www.hpl.hp.com/techreports/2002/HPL-2002-191.pdf>
- [8] SLAng: A Language for Defining Service Level Agreements D Lamana, J Skene, W Emmerich, FTDCS'03
- [9] ITAA's Application Service Provider (ASP) Service Level Agreement (SLA) Guidelines, ITAA www.itaasla.org/asp/itaasla.pdf
- [10] SLA Management Handbook, TeleManagement Forum Member Evaluation Version 2.0 July 2004
- [11] Deconstructing Service Level Agreements, H Thomas Davis Jr, Patrick K Fitzgerald New York Law Journal March 4th 2002