Exploiting the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2006 Amsterdam
ISBN: 1-58603-682-3

An Enhanced Strategy for SLA Management in the Business Context of New Mobile Dynamic VO

Francesco D'ANDRIA[1], Josep MARTRAT[1], Giuseppe LARIA[2],
Pierluigi RITROVATO[2], Stefan WESNER[3]
[1]*Atos Origin - Research and Innovation, Diagonal 210-218, Barcelona, 08011, Spain*
Tel: +34 93 4861818, Fax: +34 93 4860766
Email: francesco.dandria@atosorigin.com, josep.martrat@atosorigin.com
[2]*Centro di Riverca in Matematica Pura ed Applicata, Università di Salerno,*
via Ponte Don Melillo, Fisciano (Sa), 84084, Italy
Tel: +39 089 964189, Fax: +39 089 964191, Email: {laria,ritrovato}@crmpa.unisa.it
[3]*High Performance Computing Centre Stuttgart, Allmandring 30, Stuttgart, 70550,*
Germany Tel: +49 711 685 4275, Fax: +49 711 685 5832, Email: wesner@hlrs.de

Abstract: This paper presents an approach for managing SLA contracts in new highly dynamic mobile Grid scenarios tackled in the Akogrimo project [1]. The problem of controlling context, negotiate and control QoS as well as undertake corrective actions is outlined. The main outcome is the validation of emerging specifications like WS-Agreement and WS-Agreement Negotiation. The purpose is to increase the technological experience of SLA contract management in Grids and propose some extensions of specifications. The paper also indicates the need to develop simplified and more flexible contract negotiation mechanisms in these environments, and the need for collocation protocols for negotiation. The lack of suitable and reliable SLA control mechanisms is slowing down the adoption of Grid technologies which provide numerous advantages such us optimal use of resources, higher flexibility and new business models in open and distributed e-business context for industry.

1. Introduction

Over the last few years, Grid technology has evolved from a technology designed largely for the needs of the High Performance Computing (HPC) community towards an open framework supporting the general business domain. This development from Science Grids to Business Grids comes with the promise to significantly change the way we do business today. Grid is not just thought as an infrastructure to embrace computing power geographically distributed but it has been becoming the enabling technology for creation of dynamic Virtual Organizations (dVO) – a future form for cooperative organisations.

Managing such environments is complex and still unsolved challenge due to the large number of potential participants, the heterogeneity of resources and the need for controlled and reliable sharing of them across organisational boundaries. Furthermore, with the spread of wireless devices and networks mobile or nomadic users will become fundamental actors of Virtual Organizations (VO) in the near future. In this scenario mobile devices (and potential services hosted on them) are full participants and resource providers within the virtual organization itself and they can be bought and sold as any other kind of resource offered by a service provider connected to a fixed network.

In this paper the particular results in the domain of electronic contracts expressed as Service Level Agreements (SLA) are discussed.

In order to support the envisioned environment many issues have to be addressed, in particular, we will focus on some of them that are more related to business requirements:

- Negotiation of resources (computational and network) has to be adapted depending on the run time conditions (pre-negotiated resources cannot be reachable yet due to change of the context of the mobile grid node such as location or bandwidth).
- Monitoring of agreements need to consider network related QoS and the network availability itself as a relevant component of the value chain for service provision.
- Management of resource context derived from its mobile nature and enforcement of suitable policies.

2. Objectives

In a business context, the management of Service Level Agreement with the agreed QoS parameters between Service Customer and Service Provider(s) is essential. The lack of appropriate mechanisms for SLA negotiation and monitoring is an important barrier for the Grid uptake by industry in distributed e-business environments. The goal of this paper is to describe the problem of managing SLA e-contracts in a mobile Grid scenario and to outline a built solution considering aspects such as contract negotiation, context management and application of suitable policies when QoS is not guaranteed. This is result of our findings in Akogrimo research project that is positioned in the domain of Collaborative Business Grids while adding as a new dimension: Mobility support and network layer integration for telecom operators.

3. Methodology Used

A Mobile Grid consists out of resources that are not subject to centralized control. Supporting various kinds of mobility is done using standardized, open, general-purpose protocols and interfaces in order to deliver nontrivial and optimized Quality-of-Service (QoS) depending on the current context of the resource or the user.[2]

In our vision a Mobile Grid Service refers to a service that considers mobility, context and session aspects in its functionality. The concept of “mobile” deals with the “nomadic” concept and not only supports mobile resources (hosts, mobile devices and even users) connected though standard networks and wireless links. SLA Management subsystem encompasses the SLA contract definition, SLA negotiation, SLA monitoring and SLA enforcement according to defined policies. The main point is to build a new layer upon the grid middleware able to create negotiation mechanism between providers and consumer of services. In addition, SLA Enforcement and monitoring subsystem have also the supervisor role in order to verify that the negotiated contract conditions of all running services are met. They use WS-Notification mechanism for alerting about abnormal situation so that SLA Management can undertake effective corrective decision according to defined policies.

An advanced feature of this infrastructure is the tight relationship that is established between the grid and the network layer in such a way that Grid middleware becomes aware of network capabilities and can be expressed as a cross layer co-operation. A clear example is shown in the management of the Quality of Service. The SLA contract and its negotiation considers QoS parameters that belong to both grid resources (CPU use, Memory disk, etc) and network capabilities (bandwidth, priorities for packet traffic, etc) by means of network bundles or profiles that telecom operators provide. Thus, the application QoS requests are mapped on these infrastructure QoS parameters.

This novelty is completed with the close interactions between network and grid at runtime. Thus, any changes on network performance are taken into account by the process

that is responsible for the monitoring of QoS parameters and corrective actions and penalties can be applied according to the defined policy in a per-case basis.

4. WS "Background" State-of-the-Art

A key aspect when forming a VO is to decide how the members should interact. What are the roles within this VO, who is responsible for what, when and how should a certain task be completed; rules that members are expected to live by and what penalties apply when these rules are broken. One way of regulating these interactions is by establishing a contract or agreement that formalizes the business relationship or other part of the relationship between two parties (SLA). In terms of VO there are a number of SLAs that must be agreed upon and there is much to be gained if the process of coming to an SLA could be fully or at least partially automated to follow the fast emerging business needs that a SLA expresses. Usually, for an agreement to be concluded, negotiations need to take place. In order to negotiate and set up agreements, but also monitor QoS, web services (and in particular) Grid Services need protocol that govern and structure interaction between them. In this section, we present the WS-Agreement and the WS-Agreement Negotiation specification.

4.1. WS-Agreement[3]

WS-Agreement specifies an XML-based language for creating assurances (contracts, agreements and guarantees) from offers between a service provider and a client. An agreement may involve multiple services and includes fields for the parties, references to prior agreements, service definitions and guarantee terms. The service definition is part of the terms of the agreement and is established prior to the agreement creation. An agreement is defined as being composed of:

1. *Name* identifies the agreement and is used for reference in other agreements.
2. *Context* includes parties to an agreement, reference to the service provided and possibly other related or prior agreements.
3. *Service Description Terms* provide information to instantiate or identify a service to which the agreement pertains.
4. *Guarantee Terms* specify the service levels that the parties are agreeing to and may be used to monitor and enforce the agreement. They consist of: 1) the list of services it applies to, 2) the list of variables representing domain-specific concepts (e.g. response time or bandwidth), 3) optional conditions that have to be met for the guarantee to be enforced, 4) conditions to satisfy the guarantee and 5) one or more business values (e.g. the penalty upon failure to meet the objective, the strength of a commitment by a service provider or the importance and confidence of meeting an objective).

The strength of WS-Agreement lies in a well-defined template for specifying agreements. The template or part of the template, such as service description and guarantee terms, can be used in the content of exchanged messages and used to establish long-lived contracts in various domains such as virtual organisations. Thus, this template is suitable in cases where interactions are concerned with reaching agreements and drawing up contracts.

4.2. WS-Agreement Negotiation[4]

Negotiation is a decision process in with two o more parties make individual decision and interact with each other for mutual gain [8]. Proposals are sent to other parties, and a new proposal may be generated after receiving a counter-offer. The process continues till an agreement or a deadlock is reached.

WS-Negotiation is a domain independent language for generating agreements between a service provider and a service requestor. It can employ different types of agreement templates.

WS-Agreement negotiation describes the re/negotiation of the agreement between two parties. It's an XML language for Web Services providers and requestor. In general, WS-Negotiation contains three parts (see Figure ***1***):

- ***Negotiation Message***. This part describes the format of the messages exchanged. Some suggested message types are: Offer, Counter-Offer, Rejected, Accepted, etc. (The Schema and semantics of the messages are to be defined in the future). This part of WS-Negotiation shall tackle the "Initial Contact" and "Offer and Counter-Offer" tasks.
- ***Negotiation Protocol*** describes the mechanism and the rules the negotiation parties should follow to exchange messages. Messages contain the offers and counter-offers and can be exchanged between requestor and provider as well as a third-party negotiation service (Negotiation Support System-NSS) (see Figure 2). This part of WS-Negotiation shall also tackle the "Initial Contact" and "Offer and Counter-Offer" tasks.
- ***Negotiation Decision Making***. This is the component that takes the decisions. It is private and is based on the negotiation strategy each party has chosen (e.g. cost-benefit strategy) and the agreement template. This part of WS-Negotiation shall tackle the "Evaluation" task.

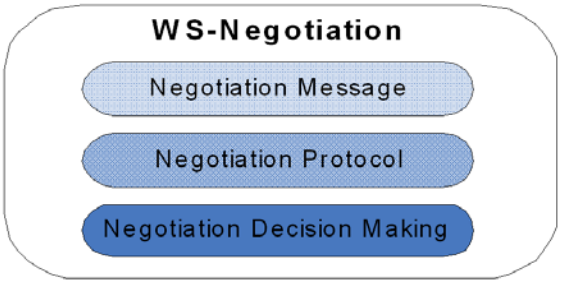

Figure 1 Three Parts of WS-Negotiation

5. How the SLA Subsystem Works in Akogrimo

Any SLA management strategy considers two well differentiated phases: the negotiation of the contract and the monitoring of its fulfillment in run-time. Thus, the Akogrimo SLA subsystem is split in two layers.

The upper layer is called "SLA High Level Services" and it is basically involved on the negotiation phase of all the services that belong to a workflow execution, as well as the establishment of a contract as a result of this negotiation. The participants of a negotiation are the service provider(s) (SP), who offers its services and the service consumer (SC) or better someone on behalf of him. Its goal is to check the availability of services able to carry out actions required by a user with the requested QoS level. According to a SOA approach, an application workflow is usually decomposed in a set of simple services.

Therefore, the negotiation of an application workflow requested by a SC implies the consequent negotiation of many simpler services that compose it. The "SLA High level services" must be able to cope with decomposition of SLAs. A SLA contract of a composed application is split in a set of individual SLAs that are negotiated separately. Only the positive answer of individual SLAs can be aggregated in a positive answer for full workflow negotiation.

When a partial negotiation is not able to offer the QoS requested, a counteroffer must be built by SP and shown to SC. While there are standards for simpler services negotiation, there is no specification that offers collocation protocols for this composed negotiation. In order to limit the number of iterations of this process, Akogrimo SLA subsystem makes a set of concrete offers to SC to choose. This strategy is aligned with the way that telecom

operators work, offering a set of network QoS profiles and the customer decides which to use. This approach has been also extended to the negotiation of the Grid resources. The low layer is called “SLA Enforcement” and it is in charge of controlling the correct evolution of on-going service execution from a QoS point of view, processing the information received from the monitoring subsystem and requesting the infrastructure of necessary corrective actions described in the policies. The negotiation protocol orchestrates the negotiation messages, which trigger and conduct the negotiation decision-making process.

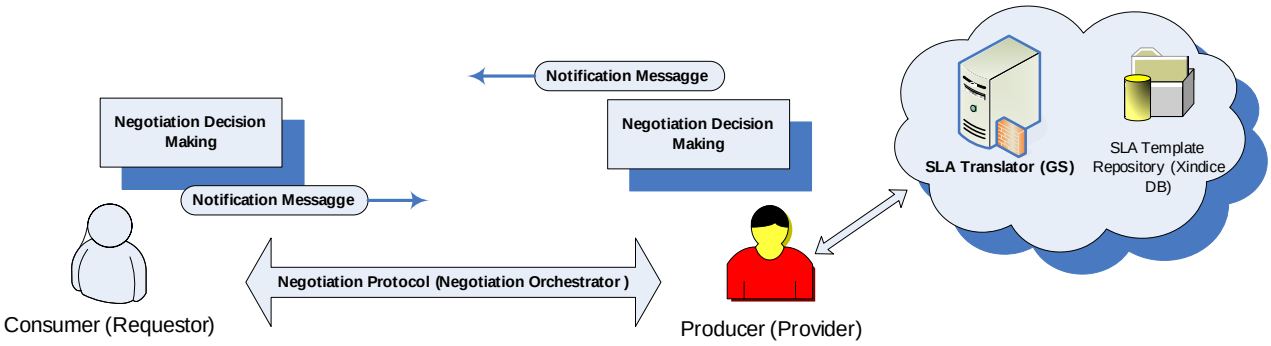

Figure 2 How the SLA Subsystem Works in Akogrimo

Negotiation issues change from one business domain to another but there are some issues that are common. Domain specific vocabularies are introduced for different types of business contexts. Our SLA management deals with this variety through SLA Translator module that is in charge of making the infrastructure QoS parameters independent of particularities of application domain.

According to the SLA Template Model (managed by the Translator Grid Service) domain specific vocabularies can directly or after negotiations (through WS-Negotiation) generate infrastructure SLA documents.

The interaction protocol as specified in the WS-Agreement Specification only allows for a single “request, accept” interaction, in which the requesting party receives either an acceptance or rejection message from the providing party as a response to an agreement request [5]. This is a quite limited interaction model. In the model proposed in Akogrimo, an additional accept/reject interaction sequence is introduced through the interaction with the Grid Execution Manager System (EMS). Using a sophisticated mechanism based upon the interaction with the Policy Manager, the Negotiator Service is able to build an offer based on different range of QoS low level parameters. This offer will be presented to the EMS that will be able to discover the most suitable service (Figure 3).

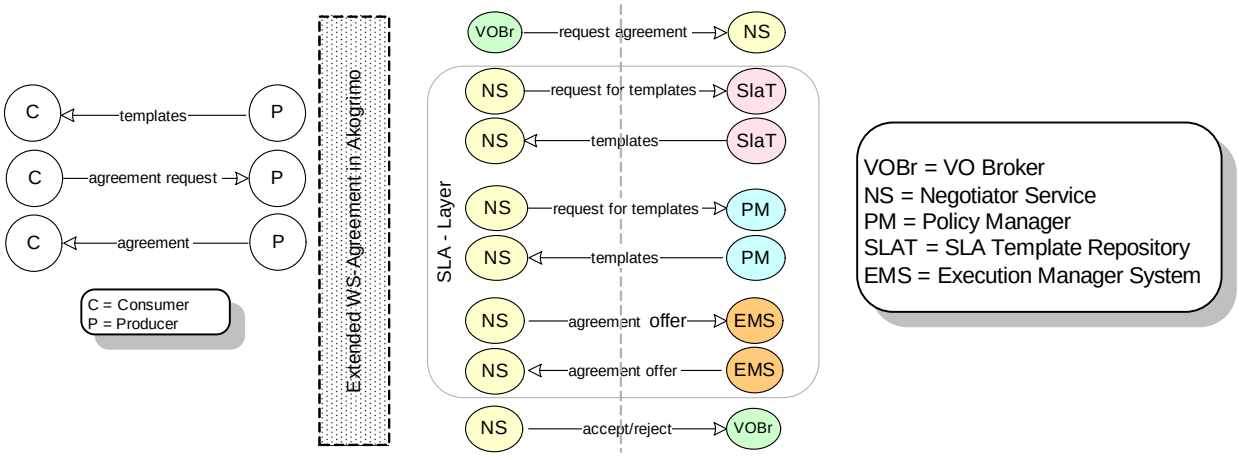

Figure 3 Extended WS-Agreement Protocol

After the negotiation phase, once an application service instances have been created, it is necessary to ensure that the SP respects the contractual terms. The EMS notifies to

Monitoring that a new service is being used (see the step 3 in the *Figure 4*). At this moment the monitoring component (Monitoring and Metering Services) is in charge to measuring and sending QoS value if this service to the SLA-Controller component. While the metering is able, for each host, to record and send QoS measurements related to the services (like CPU load, bandwidth, memory used, service invocation), the Monitoring is able to collect the information sent by all Metering in the VO and distribute these measurements to the interested modules. The SLA Controller is responsible for the verification that these QoS measurements are within the thresholds established in the contract.

Whenever a service does not meet these conditions this agent notifies it (through WS notification mechanism) to the SLA Decisor module, which is in charge to starting the appropriate recovery action.

The SLA Decisor asks for the policy to be applied to the service interacting with a Policy Manager, taking into account overall execution context. For example, if the SLA Decisor catches a "light" violation interacting with the EMS, it could request a higher service priority, but if it catches a strong violation the SLA Decisor could ask interacting with a Workflow Manager service to stop the service and re-instantiate it in another host.

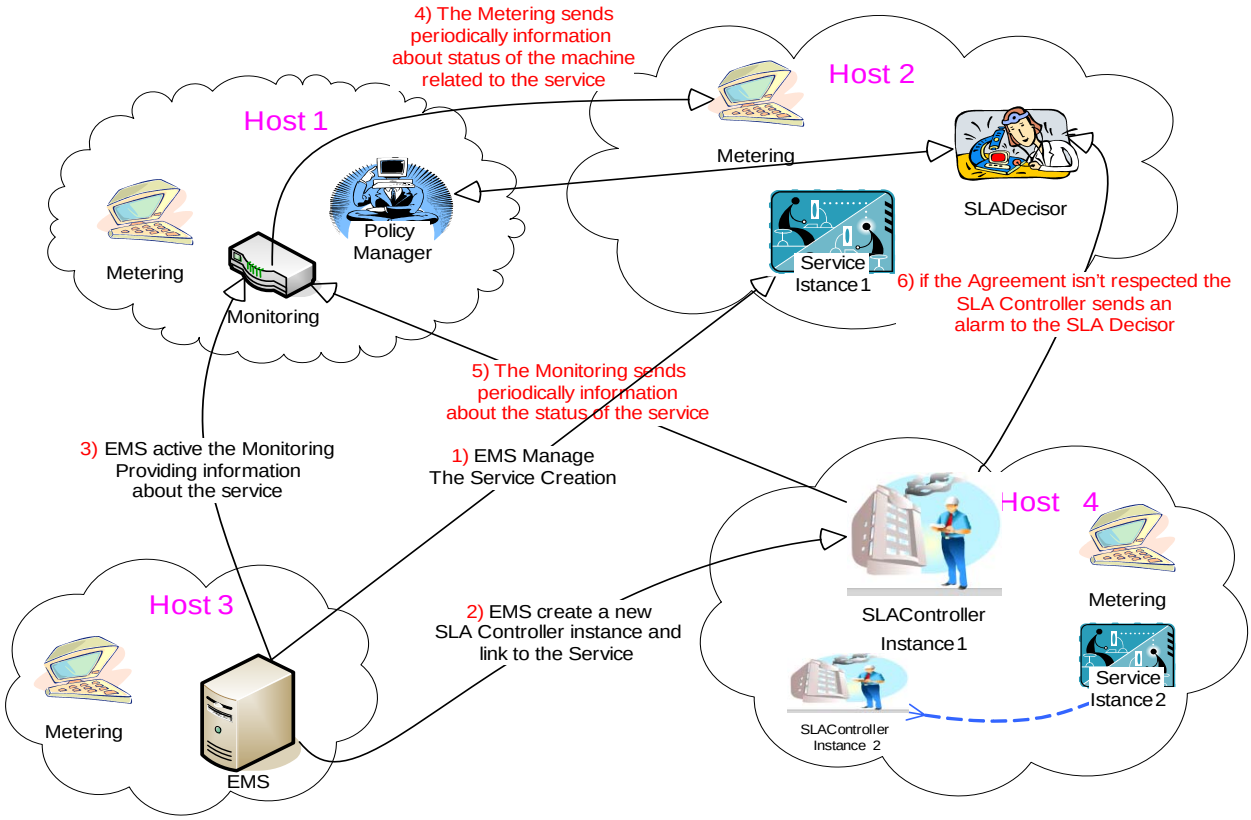

Figure 4 SLA Enforcement (run-time)

6. Business Case Description

The Akogrimo goal is to deliver a framework for the provision of IT added value services combining Grid technologies and the functionality developed for beyond 3G infrastructures following the "all-IP" connectivity.

Challenging mobile grid scenarios become even more interesting if implemented in business cases that take advantage of potential benefits such as faster market response, better utilization and service level performance and lower IT operating costs. Derived business models are attractive for both general service providers and telecom operators.

The Akogrimo results are validated through 2 scenarios (a third scenario is under consideration) from different application domains: i) eLearning and ii) eHealth. The chosen

scenarios are complementary in their imposed requirements and economic sectors. We are explaining how they can benefit from the described SLA management framework.

6.1. eLearning Scenario

A group of students, all equipped with a 4th generation PDA, are working on the Field Trip project in a Greek archaeological site. During their activity they store information, experience, emotion, in terms of photos, video clips, text notes, audio comments, etc. The PDA host a data source service that can be invoked in order to retrieve the gathered information through the appropriate network (the PDA is available via the network service provider operative in that zone that provides network communication according to the bandwidth necessary, the price, etc.).

During the daily work in the field trip some collaborative sessions with other students in the school are required to share their experiences. To this purpose the students collect with their PDA some photo of the field trip and the objects discovered and send them to the Virtual Collaborative Grid service asking to make them in 3D.

The VC Grid service invokes the high performance 3D modelling and rendering tools (provide by the High Performance Computing Centre of the University) to make the 3D reconstruction and render them

This is an example of next generation learning. The attractiveness of this learning vision is the capability of harnessing and sharing the almost universal availability of mobile devices to education and training. This widespread availability can be harnessed and exploited to provide access to training opportunities for those who otherwise might be at a disadvantage for geographic (wireless networks span the rural-urban divide), economic (mobile handsets are relatively inexpensive) or social reasons, but mainly to take in place innovative contextualised learning approach, where the learner "achieves" knowledge and skills in an active way instead of simply storing information.

Many economic implications result from such vision, in fact, it implies an eLearning delivery process that involves several parties, making existing resources and services available across organizations and domains.

A such environment needs for a strength of the flexible accounting capabilities, the support of service level and dynamic negotiation management, that is, the topics about Akogrimo SLA subsystem provides possible solutions.

6.2. eHealth Scenario

The Heart Monitoring and Emergency Management scenario to be prototyped within an Akogrimo testbed has the objective of early recognizing a heart attacks and properly treating the patients as fast as possible.

The patient's Electrocardiogram (ECG) data are measured by a wearable ECG device, forwarded to the patient's mobile phone. The Mobile Grid based Heart Monitoring and Emergency Service (HMES) recognizes the availability of the ECG data, automatically requests the data and calls an ECG analysis service to check for aberrations. When an aberration occurred, concurrently, the emergency handling service starts the following actions: determine the location of the patient, identify the responsible emergency dispatch centre and the general practitioner or cardiologist currently attending the patient, and a locally available emergency ambulance.

The ambulance dynamic navigation system receives the patient location and shortest route to get there. A comprehensive patient history is compiled from the available patient records of the hospitals and a medical data analysis service prepares the data with respect to the current emergency and the visualization according to device terminal capabilities.

A finder service is looking for an appropriate hospital that is informed about the emergency. Thus, the admittance and emergency staff gets immediate access to all patient data. By calculating the arrival time the scheduling of the emergency staff and necessary resources (room, monitoring devices) can be optimized.

This scenario involves several parties and the charging of the HMES consists of several parts. For using the service the patient has to pay an annual flat rate to the HMES Provider. The data transferred over the network is charged by the network operator and is a part of the patient's monthly telephone bill. The services provided during the emergency handling process are paid by the patient's health insurance. In order to manage this complex charging requirements the Akogrimo SLA subsystem provides the basilar foundations necessary in such dynamic environment.

7. Conclusions and Summary Recommendations

This paper has introduced the approach for managing SLA contracts in dynamic mobile Grid scenarios explored in the scope of Akogrimo project. It shows the proposed mechanism during the two phases: SLA negotiation and SLA monitoring in run-time.

The implemented solution considers QoS parameters for both Grid resources (CPU, memory, etc) and mobile networks (Bandwidth, packet priorities) which is suitable to integrate Grid technologies in telecom operator networks. The paper shows how the approach undertakes corrective actions considering the context changes such as different bandwidth profiles. The solution takes into account particularities of business application domain in the SLA contract definition but also makes it independent in terms of infrastructure management by means of a parameter translator.

An important outcome is the use and validation of emerging specifications like WS-Agreement and WS-Agreement Negotiation with the purpose of increasing the technological experience of SLA contract management in Grids. The lack of reliable SLA management mechanisms is slowing down the adoption of Grid technologies which provide numerous benefits such as optimised user of resources, higher flexibility, cost reduction and enabling new business models round dynamic virtual organizations.

The paper indicates the need of working in more flexible but simplified contract negotiation mechanisms in these environments, concretely in the collocation negotiation protocols in the composed negotiation of workflows. The work on enriched semantic definition of policies is also an important challenge for having flexible and self-adaptive SLA management methods that distributed business contexts are requesting. In addition, only through suitable SLA monitoring subsystem is possible to derive a fair accounting subsystem in these kinds of scenarios.

References

[1].Access to Knowledge through the Grid in a mobile World (Akogrimo) http://www.mobilegrids.org/

[2].Antonios Litke, Dimitrios Skoutas, Theodora Varvarigou, Mobile Grid Computing: Changes and Challenges of Resource Management in a Mobile Grid Environment http://mobilesummit2006.org/

[3].A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, Web services agreement specification (WSAgreement) (draft). https://forge.gridforum.org/projects/graap-wg, 2006.

[4].A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, Web services agreement negotiation specification (WSAgreementNegotiation) (draft). https://forge.gridforum.org/projects/graap-wg, 2004.

[5].d.g.a. mobach, b.j. overeinder, and f.m.t. brazier, A WS-Agreement based resource negotiation framework for mobile agents.

[6].A Critical Analysis of the WS-Agreement Specification http://www.csc.liv.ac.uk/~sha/WS-Agreement/

[7].Web Services Notification http://www-128.ibm.com/developerworks/library/specification/ws-notification/

[8].L. Thompson "The Mind and Heart of the Negotiation" Prentice-Hall Inc., 1998