LEGAL & OPEN MODEL TERMS FOR CLOUD SLA AND CONTRACTS

Initial position paper (technical)

D3.1

Dissemination level: Public

Work Package	**WP3, Technical Track**
Due Date:	M4 (01/05/2015)
Submission Date:	05/05/2015
Version:	0.2
Status	FINAL
Author(s):	Aimilia Bantouna (UPRC) Dimos Kyriazis (ICCS) Daniel Field (ATOS) Panagiotis Demestichas (UPRC) Teta Stamati (UPRC) Kostas Tsagkaris (UPRC) Panagiotis Vlacheas (UPRC)

The SLALOM Project is co-funded by the European Commission through the H2020 Programme under Grant Agreement 644720

Contents

Tables

Figures

1 Introduction

WP3 realizes the technical perspective of SLALOM activities. The overall goal is to provide the technical ground of the interactions with various stakeholders in the cloud ecosystem, and support the research projects with respect to SLA-related topics. In greater detail, the objectives of WP3 are the following:

- Gather input from the cloud research community relating to future cloud scenarios and their contractual challenges.
- Define the core SLA specification (and potential domain-specific extensions to it) through clear, common, provider-independent SLA terms in order to overcome today's great variability in SLA specifications.
- Deliver a common SLA reference model capturing the essential components across the SLA lifecycle, and their interactions. The reference model will also include potential extensions for added-value services.
- Support the cloud research community by demonstrating the relation of future cloud scenarios to the legal context and their enablement through SLALOM.
- Disseminate and communicate the technical SLALOM reference model to the Cloud and SLA sector, including recruitment of experts to the consensus groups.

To this end, Task 3.3 "Initial interaction with the cloud community" a) creates awareness within the community and through this recruiting key stakeholders to the consensus groups and b) identifies and selects experts, projects, working groups and coordination actions in order to distil the future scenarios which will be codified into the terms. Moreover, Task 3.5 "Support to SLA research projects", in the context of providing support to the SLA-related research performed in the framework of EC projects, collects the relevant requirements of the scenarios to be taken into account when defining the core SLA specification and the common SLA reference model (SLA lifecycle and interactions among its components).

This deliverable provides a position paper identifying the key scenarios being developed by the projects with an absolute emphasis on their bearing on, requirements of, and assumptions of cloud contracts and SLAs. Overall, it will be used as the starting point for the codification of these aspects into the SLALOM terms.

1.1 Document Structure

The document is structured as follows: Section 2 presents the approach that was followed. Section 3 summarizes the outcomes of report [1] to be exploited in SLALOM WP3 work. Sections 4-6 summarize the cloud scenarios and the respective requirements identified in cloud- and SLA-related projects. The feedback received from the cloud and SLA community is presented in Section 7. Finally, conclusions are drawn in Section 8.

1.2 Glossary of Acronyms

Acronym	*Definition*
CSP	Cloud Service Provider
SLA	Service level Agreement
SLO	Service level Objective
C-SIG	Cloud Select Industry Group
MSA	Master Service Agreement

2 Approach

SLA working group of the Unit E.2 “Software and Services, Clouds” of DG Connect has created report [1] which summarizes the outcomes of European and National research projects in the area (which have started working on clouds and SLAs before June 2013), makes recommendations that could support the on-going policy work and studies the impact that these recommendations could have if/when realized in different domains and areas. SLALOM, staying focussed on its aim to build on top of the existing past work and create a harmonized model with it, in such cases; it will exploit directly the elaborated results from the report.

Related SLA outcomes include specifications and term languages, such as the manifest in OPTIMIS [2], the blueprint in RESERVOIR [3] and 4CaaSt [4], the quality model in CONTRAIL [5], the QoS-oriented specification in Q-ImPrESS [6], the virtualised service network in IRMOS [7], and the service description in SLA@SOI [8]. Business aspects in the SLA lifecycle have also been considered - a representative example would be the business-enhanced template in ETICS, as well as frameworks supporting composite services as the cloud federations proposed in CONTRAIL [9], the eMarketplace in 4CaaSt [10] or the mechanisms in GEYSERS [11]. As the basis for the provision of QoS guarantees, interesting works regarding performance estimation and workload prediction have been developed in Cloud-TM [12], service network risk, uncertainty and dependability for critical infrastructures in SERSCIS [13], data reliability and safety in PrestoPRIME [14], while enhancements for trade-off analysis have been proposed in Q-ImPrESS [15]. The unified monitoring interface from Cloud4SOA [16], the adaptable monitoring tools from IRMOS [17], the SLA-driven monitoring from SLA@SOI [18], and the scalable and efficient monitoring from Stream [19] cover the monitoring aspects for SLAs. Novel negotiation approaches enabling dynamicity, automation, scalability and re-negotiation during runtime have been implemented by Cloud4SOA [16], OPTIMIS [20] and IRMOS [21] respectively. Regarding SLA enforcement, CloudScale tools for automatic root cause analysis [22], 4CaaSt developments for elasticity management [23], VISION cloud approaches for proactive SLA violation detection [24], as well as CumuloNimbo [25] and Cloud-TM outcomes [26] with respect to enforcement for transactional systems are worth mentioning.

Keeping up with the work already done, SLALOM consortium studied public documents from the EU-funded Cloud- and SLA-related projects after the release of the report to complement it with further results on the postulated changes to the value chain and service delivery foreseen in the key areas of hybrid clouds, mobile cloud, federated and heterogeneous infrastructures, service marketplaces, inclusion of network control (e.g. through SDN) and cloud brokerage. The use cases and the respective requirements considered in those projects are summarized in the next sections.

Furthermore, SLALOM consortium aiming to have an as soon as possible feedback from the cloud and SLA stakeholders, with respect to technical specifications as well, included some of the major areas of SLA research which are not currently (clearly) reflected in the ISO structure so as to also capture stakeholders' interest towards this direction. The results of the questionnaire are presented in Section 12.

Finally, SLALOM has already started interacting with SLA-Ready project [27], an Horizon 2020 project that promises to support SMEs with practical guides, tutorials and a social marketplace, so as to move from the typical "take it or leave it" approach to an informed, stepping-stone approach that will allow cloud and applications to grow with their business.

3 DG-Connect SLA working group Recommendations

Report [1] has concluded to the "technical" provided in Table 1. For each recommendation, the report also states the studied projects that have worked towards this direction and should be taken into account.

Table 1. Report [1] Summary

Recommendation	*Potential contributions*
Develop a Core SLA Specification and Differentiate SLAs and Contracts	▪ 4CaaSt Project: Blueprint Concept ▪ BREIN Project: Semantic Annotation in SLA templates ▪ CloudScale: Scalability Specification ▪ Cloud-TM Project: SLA Definition and Enforcement in Transactional Data Stores ▪ CONTRAIL Project: SLA Specification and Quality Model ▪ EGI Project: Service Catalogue in a Federated Environment ▪ IRMOS Project: SLAs at Different Levels ▪ OPTIMIS Project: Service Manifest ▪ Q-ImPrESS Project: QoS-oriented SLA Specification ▪ PrestoPRIME Project: SLA Specification for Preservation Services (risk of data loss) ▪ SLA@SOI Project: Service Description ▪ VISION Cloud: Content-related Terms in SLAs
Support composite and complex services	▪ 4CaaST Project: Blueprint Concept ▪ CONTRAIL Project: Quality Model and Multi-level SLA Interaction Model ▪ ETICS Project: SLAs for Composite Services

	▪ OPTIMIS Project: Service Manifest ▪ MCN Project: Distributed SLA Management ▪ Q-ImPrESS Project: QoS-oriented SLA Specification
Encapsulate legal terms and separate responsibilities and obligations	▪ CONTRAIL Project: Quality Model ▪ OPTIMIS Project: Service Manifest
Provide accurate runtime monitoring and reporting	▪ Cloud4SOA: Unified Monitoring Interface and Metrics ▪ IRMOS Project: Adaptable Monitoring and Evaluation ▪ MODAClouds Project: Unified Monitoring ▪ mPlane Project: Network Monitoring for SLAs ▪ SLA@SOI Project: Scalable SLA-driven Monitoring ▪ Stream Project: Scalable and Efficient Monitoring
Support runtime adaptability and dynamic SLA (re-)negotiation	▪ Cloud4SOA: Dynamic SLA Negotiation and Enforcement ▪ IRMOS Project: Dynamic SLA Re-negotiation ▪ MODAClouds Project: Runtime Re-negotiation ▪ OPTIMIS Project: Automated SLA Negotiation ▪ SLA@SOI Project: SLA Negotiation across Multiple Layers
Certify providers and enhance SLA enforcement for mission-critical applications	▪ 4CaaSt Project: Elasticity Management ▪ CloudScale: Automatic Root Cause Analysis ▪ Cloud-TM Project: Performance Estimation and Workload Prediction and SLA Definition and Enforcement in Transactional Data Stores ▪ CumuloNimbo Project: SLA Enforcement for Transactional Systems ▪ Q-ImPreSS Project: Trade-off Analysis and SLA Prediction ▪ VISION Cloud Project: Proactive SLA Violation Detection
Consider business models and objectives	▪ 4CaaSt Project: eMarketplace ▪ EGI Project: Federated Service Management ▪ ETICS Project: Business-enhanced SLA Template ▪ OPTIMIS Project: Service Manifest ▪ plugIT Project: Recommendation System ▪ SLA@SOI Project: Service Description
Adopt a reference baseline solution for SLA	▪ CONTRAIL Project: SLA Management for Cloud Federations ▪ GEYSERS Project: Converged SLA Management for Composed Virtual Infrastructures ▪ MCN Project: Distributed SLA Management

4 Helix Nebula [28]

4.1 Use cases

4.1.1 CERN-LHC Use Case

ATLAS and CMS are particle physics experiments on the Large Hadron Collider at CERN, Geneva. These detectors are searching for new discoveries in the collisions of protons of extraordinarily high energy. These experiments will learn about the basic forces that have shaped our Universe since the beginning of time and that will determine its fate.

The experiments are currently running a large scale distributed computing system to process the data acquired by the experiment. The ATLAS Distributed Computing environment (ADC) consists of several pieces: a Distributed Data Management component (DDM), a distributed production and analysis system (PANDA), and associated tools as well as the processing and analysis binaries. This use case is part of a wider project within ATLAS, to research the applicability of cloud computing to ATLAS computing.

Overall, the objectives of the use case are to

- evaluate the use of cloud technologies for LHC data processing,
- transparently integrate cloud computing resources with ADC software and services,
- implement the ATLAS cloud computing model within DDM, Panda, and related tools and services,
- identify the financial costs of network transfers of data into and out of the cloud resources, short and long-term data storage in the cloud, and CPU resources for running the various ATLAS use cases,
- define the appropriate Service Level Agreements, and
- Identify the policy and legal constraints in moving scientific data across academic networks into commercial resources and back again.

4.1.2 European Molecular Biology Laboratory (EMBL) use case

Next Generation DNA Sequencing technologies have had a huge impact on how biological and medical research is performed today. It has helped large-scale sequencing to become affordable and revolutionized the sequencing of complete genomes. Scientists today easily generate huge amounts of quality sequence data within a few days. The analysis and management of these vast data sets, however, require high performance computing and fast data storage infrastructures as well as bioinformatics expertise, which are often challenging for many labs.

The European Molecular Biology Laboratory (EMBL) is developing a portal for cloud-supported analysis of large and complex genomes. This will facilitate genomic assembly and annotation, allowing a deeper insight into evolution and biodiversity across a range of organisms.

Overall, the objectives of the use case are to

- Open up new possibilities for scientists to perform large-scale genomic analysis without making large capital investments in computing infrastructure, thereby making de novo assembly and genome annotation affordable to many more laboratories,
- Provide a leading bioinformatics pipeline to perform fast and on-demand genomic data analysis,
- Provide a basis for future extension of genomic research using cloud computing infrastructures.

4.1.3 European Space Agency (ESA) SuperSites Exploitation Platform (SSEP) Use Case

The European Space Agency (ESA), in partnership with the Centre National d'Etudes Spatiales (CNES) in France and the German Aerospace Center (DLR) is collaborating with the National Research Council (CNR) in Italy, to create an Earth observation platform focusing on earthquake and volcano research.

Geohazard Supersites is an international partnership of projects, organizations and scientists, involved in monitoring the dynamic and complex solid-Earth system and the assessment of geohazards. Its mission is to support the Group of Earth Observation (GEO) in the effort to reduce loss of life and property caused by the natural disasters of earthquakes and volcano eruptions. The Geohazard Supersites project will advance scientific understanding of the physical processes which control earthquakes and volcanic eruptions as well as those driving tectonics and Earth surface dynamics.

Overall, the objectives of the use case are to

- provide an open source, unified e-infrastructure for solid Earth data,
- improve data products for solid Earth monitoring,
- ensure secure solid Earth data sharing on the cloud,
- enable international collaboration, and
- Enlarge the science user base.

4.1.4 Port d' Informació Cientifica (PIC) Use Case

Medical research in neurodegenerative diseases, both academic and industrial (pharmaceutical), greatly benefits from the results of MRI and PET data processing. Doctor-Researchers are not interested in computing – they just want someone to turn the raw image datasets into feature-rich analysis datasets as quickly as possible. There are very few "Technician-Researchers" to do this job and there are many different ways to process datasets, so a friendly interface is needed. A key factor in accelerating research is turnaround time. Clusters, Grids and Clouds can help here, each with their pluses and minuses.

The port d'informació científica (PIC), has developed a web-based platform running over a cluster that allows researchers to easily store, organize and process data with a vast improvement of turnaround time (from several months to a few days), the so called PICNIC (PIC Neuroimaging Centre).

The PICNIC use case aims to implement a commercial cloud back-end for PICNIC so as to

- improve the speed and quality of research for finding surrogate biomarkers based on brain images,
- achieve early detection of neurodegenerative diseases, and
- Study of pharmaceutical effects.

4.2 Requirements

The following attributes/requirements were identified to match most (if not all) the use cases of the project and may need to be dealt in SLA level as well:

- Scale of Resources Used
- Federation/Aggregation of Datasets
- Long-Term Archiving of Data
- On-Demand Processing
- Potential Increase of Users
- Interoperability
- Data Security
- Access to License-Controlled Software

5 CELAR [29]

5.1 Use Cases

5.1.1 Translational Cancer Detection pipeline (SCAN)

The identification of genes that are mutated and hence drive oncogenesis has been a central aim of cancer research since the advent of recombinant DNA technology. Institute of Cancer Research (ICR) has developed several pipelines to capture and analyse genomic, proteomic and clinical information by using several biology tools such as BWA, GATK, The Global Proteome Machine, MaxQuant, CellProfiler and Cytoscape.

The SCAN pipeline comprises four processes: a) NGS data process with Linux system; b) Mass Spectrometry sample data process with Windows system; c) Cell Image data process with Linux Web server; and d) Integrative network analysis with Linux system. The resources required for the four processes are very different, such as Linux, Windows, Web services, etc. Currently, ICR has to prepare the different systems (hardware and software) with various run-time environments in advance in order to run the whole pipeline in five steps. However, this approach is unproductive and highly inefficient. For example, a dedicated windows system for the protein discovery application is required but only used for about 30 minutes in a total of 90 h of the whole pipeline execution. Furthermore, over-provisioning the required systems with maximum hardware capability is currently mandatory for few special cases (e.g., the analysis of some very complicate and very large patient data sets), although they are not necessary for most of the cases.

CELAR can provide system resources automatically to the heterogeneous applications of the SCAN analytic pipeline in a “just enough, just on time” manner and can thus enable the SCAN pipeline to terminate smoothly without interruption. In particular, CELAR is expected to monitor the state of execution of the various SCAN steps, so that it can dynamically allocate the required resources for each step of the pipeline when needed.

5.1.2 EverythingHere Application

EverythingHere is a web based policy game, developed by Playgen, which will utilize Big-Data from the government website http://data.gov.uk, pulling in historical and real time information in order to model and simulate the infrastructure of London as games challenges. Players will be put in the shoes of policy makers and progress through the game by slicing and analysing data. In order to complete game challenges they must discover emerging stories within the data sets (Correspondence analysis). The game will encourage players to use its inbuilt data tools to do sentiment analysis and deep data mining to complete their objectives and progress further. The game will allow multiple concurrent users to access it ubiquitously. The game will be deployed and run over the CELAR system in order to demonstrate the capability of CELAR for elastic processing of huge volumes of highly volatile social data produced and updated during a cloud social game as well as accommodating a large varying number of concurrent user accesses.

In this use case, CELAR is envisaged to provide the necessary resources so that CPU- and I/O-intensive applications which comprise the application, e.g., Data Analysis, can operate adequately and execute queries at the shortest time possible. CELAR is also expected to monitor the application executed on the CELAR platform and keep appropriate metrics so that overloading is avoided and dynamic addition or removal of resources will help achieving a scalable application performance.

5.2 Requirements

The following requirements have been identified (among those defined within the project) that may add value to SLALOM work:

- elasticity directives and optimization policies will be provided by the users and will need to be respected
- reservation of the right type and amount of cloud resources
- the monitoring system should be able to collect metrics from multiple infrastructure layers, combine and evaluate them against costs and benefits
- ability to perform elastic resource allocation in a completely transparent manner: Users and owners of the application should be able to perceive the performance and corresponding costs of their applications to vary, at all times, within the limits they defined when submitting the application
- ability to adaptively add and remove resources in real time, so that the perceived behaviour range is always bound by the user's requirements
- scalability regarding both the number of managed applications as well as to the load/data that each actor of the application produces
- taking elastic resize actions over multiple applications and huge resource pools from the IaaS provider
- High Availability: Users should be able to manage their applications consistently and reliably via CELAR. Consequently, the system should exhibit very high levels of robustness both against

hardware and software failures, utilize redundancy, load-balancing, etc., in order to ensure that CELAR components and services will be highly available
- CELAR has the potential of being offered by different IaaS providers as a service that will dynamically manage application resource allocation

6 PaaSage [30]

6.1 Use Cases

6.1.1 Flight Scheduling

NetLine/Sched is an application from the NetLine product suite, which is used for airline schedule planning. Today's airlines need to permanently revise their schedule plans in response to competitor actions, or to follow updated sales and marketing plans, while constantly maintaining operational integrity. This makes schedule management a very complex process. These challenges call for a state-of-the-art scheduling system which optimally supports the development, management and implementation of alternative network strategies. NetLine/Sched supports all aspects of schedule development and schedule management. It offers powerful and easy to use schedule visualization and modification, supports alternative network strategies and schedule scenarios and measures the profitability impacts of alternative scheduling scenarios. The system is used every day by more than 30 airlines around the globe, ranging from small to large carriers and using different business models.

In order to enhance the performance of the application (in terms of reduced complexity, improved quality and reduced development time / reduced cost) flexible infrastructure and application software is required. The used application software must be designed to scale and to use the given resources very efficiently. Cloud computing is one of the key factors to realize this flexibility but in order the application to run in a cloud environment, it needs abstraction from specific cloud service providers to prevent a vendor lock-in, to allow shorter development cycles for new products and to gain additional benefit for the application user by providing additional features.

PaaSage is expected to benefit this use case by a) enabling the realization of the application also across different cloud infrastructures and b) supporting deployment into hybrid clouds easily (build up on customer and provider cloud infrastructures).

6.1.2 Industrial ERP

BEWAN is an IT Service company located in Belgium, delivering products and services in the domain of IT infrastructure, software development and consultancy. All of BEWANs applications have been developed in-house and can be easily adapted in order to fulfil specific requirements from customers. However, those (licensed-) applications are not cloud-ready, not SaaS-ready and therefore run on private machines. BEWAN is in a process of redeveloping its standard applications and the objective is to propose SaaS – multi tenant software solutions in the cloud. Depending on the usage and load, BEWANs

objective is to be able to deploy applications to its private cloud and also to be able to scale out to high performance public clouds when needed.

6.1.3 Electronic portal for citizen-city

Most municipalities host their own applications locally, or in cooperation with directly neighbouring municipalities. To enhance the services they are responsible for, e.g., in terms of human resources needed, with more efficient and Citizen-centric solutions there is a need to renew and rethink how current applications are used. In a typical case such as managing a request for building a house, municipalities also need to take advantage of external services such as databases on housing regulations or detailed maps of the area, and integrate these within their processes. The integration should be between the processes as well as archives. Software provided to municipalities can be remote or in premises, or ASP vs. SaaS, thus several delivery models need to work together.

6.1.4 Resource intensive simulations

One of the major research fields of HLRS is computationally intensive science that is carried out in HPC environments (eScience) including molecular dynamics simulation and biomechanical simulation e.g. blood flow, bones and bone-implant-systems. HLRS is particularly pushing the aspect of convergence between high-end and low-end programming, to enable common developers to exploit new resource infrastructures that scale both vertically (HPC) and horizontally (cloud). Furthermore HLRS is cooperating with ASCS to perform scientific research on numerical simulation in the field of automotive engineering. However scientific computing requires an ever-increasing number of heterogeneous resources to deliver results for growing problem sizes in a reasonable timeframe and with the current business procedure of HPC, it is difficult for users to access and manage the execution of such applications, in particular applications that involve parameter sweeps, as elaborated in detail in next section. With the recent cloud hype, there has been a growing interest from the eScience and HPC community to exploit cloud infrastructure, as they seem to offer just the capabilities required by the researchers because of its well-known advantages, i.e.,

1. Strong computing resources (Scalability)
2. “on-demand resources” (Elasticity)
3. High availability
4. High reliability,
5. Large data scope
6. Reduced capital expenditure (cheap).

The main objectives targeted to be achieved through the PaaSage project for facilitating the execution of large scale and heterogeneous-resource demanding simulation workflows, are listed as below:

- Deploy large-scale simulation applications cross HPC and cloud:
- Better resources utilization
- Expose the simulation application as services

6.2 Requirements

The requirements that may pose challenges in clouds and the respective SLAs and have been identified within PaaSage are as follows:

- The appropriate public, private or hybrid cloud platform should be identified for each module
- The preferences for location of deployment units can be specified
- Different sizes of data at different locations
- The appropriate VM instance size and type must be selected
- The different latencies required for different modules should be known
- When a user moves e.g. from the desktop browser to a mobile client, he/she expects to see the same data after login to the same application.
- It should be possible to increased interoperability between applications, at least for applications of the same application suite
- Deployments (Services) must therefore be able to communicate seamlessly across different cloud-based applications
- It should be possible for example to support a process where a workflow and User Interface is run in a private cloud, but it reuses public data/Open Data-databases, and integrated with locally installed archiving and accounting systems for a municipality.
- It should be possible for existing applications with a large local installed base to integrate these with a cloud offering delivering standardized processes where the process is run in the cloud, but closely integrated with the business applications installed at each individual customer
- Support methods to ensure access control across data sources
- Security concerns must be covered at all time, moving from a private cloud e.g. into a public cloud (even for parts of the system) must be possible in a secure and reliable way. It must provide trustworthy services.
- The integrity and authenticity of data should be guaranteed end-to-end
- Deployed services should be available globally.
- Access to external interfaces is a vital part for such deployments.
- Data can be exported and imported using a standard file format and data can be sent to other departments or to partners. High availability is important.
- Data must be constantly updated but operational integrity must be maintained.
- It should be specified which modules need to scale, how much and when to ensure the e.g. availability and response time requirements
- Scalability other than elasticity must also be defined, e.g. defining how much memory could be allocated to an application.
- Elasticity and scalability across datacentres, and across business processes over the year should be specified
- The expected application load for an application should be defined
- Deployed systems should be available every day by many customers around the globe, ranging from small to large companies and using different business models.

- The sizing of the servers must be done up front and elasticity of the database servers must be anticipated (e.g. transform an Oracle single node database server into a cluster (RAC) database server)
- Hybrid cloud models should be supported with some services in private clouds and some services in public clouds
- Gradually shift from locally installed software to gradually more cloud based models
- Full portability of the cross cloud deployments must be guaranteed
- Availability of deployed components must be monitored.
- Performance depends on the network connection into the cloud: cloud specific network optimizations are needed
- The response time of deployed components must be monitored
- Relocation of deployed services and data based on user experience
- Relocation of services and data based on network experience
- Cloud bursting should be possible
- During execution, there is real-time checking whether the performance is as you expect (via updates to the MD-DB). Several options prioritise alternative resources when an SLA is violated. If performance drops below acceptable levels in the SLA
- Framework for «SOA/Cloud» management should keep control on dependencies
- The reputation of available cloud providers should be managed: it should be based on past performance
- The application load should be distributed across the cloud resources.
- Take into account the urgency and priority of the request.
- It should be possible to easily access other IT-systems within a company and outside of the company
- Time zones should be taken into account when deploying an application in multiple clouds, especially when the application must be accessible globally from anywhere in the world.
- It should be possible for existing applications with a large local installed base to integrate these with a cloud offering delivering standardized processes where the process is run in the cloud, but closely integrated with the business applications installed at each individual customer
- Elasticity and scalability across datacentres, and across business processes over the year should be specified
- Data can be exported and imported using a standard file format and data can be sent to other departments or to partners. High availability is important.

7 Handout and Questionnaire Feedback

Further to the above survey that has been done, SLALOM consortium also requested feedback from cloud stakeholders through the questionnaire that was conducted [31]. The questionnaire was well based on the SLALOM Handout with respect to the way that SLALOM considers the cloud- and SLA-related issues [32].

The major areas of SLA research which are not (clearly) reflected in the ISO structure and for which feedback was requested are the following:

- SLAs at different levels
- Multi-level SLA interaction model
- SLA negotiation across multiple layers
- Automated SLA re-negotiation
- Proactive SLA violation detection

7.1 SLAs at different levels

The example provided for clarifications to the respondents of the questionnaire for this research area was as follows:

Anastasia - focused on the media domain - would like to obtain a cloud service but as a non-technical user, she can only specify terms at a "high-level" (i.e. not related to resource attributes such as CPU or memory). She wants to specify terms at her own "understandable language" level (e.g. frames per second) so that these terms will be translated to the corresponding technical ones.

The IRMOS EU project has proposed two types of SLAs, namely application and technical along with a mechanism for performing the required translation between these agreements.

As depicted in Figure 1, Figure 2 and Figure 3, "SLAs at different levels" research area was overall evaluated as a core requirement (48%) with the percentage of Cloud Adopters reaching the value of 63% while the percentage of CSPs remained at the level of 36%. Contrarily to the cloud adopters, most of the CSPs (43%) consider this research area 'somewhat important' but not a 'core requirement'.

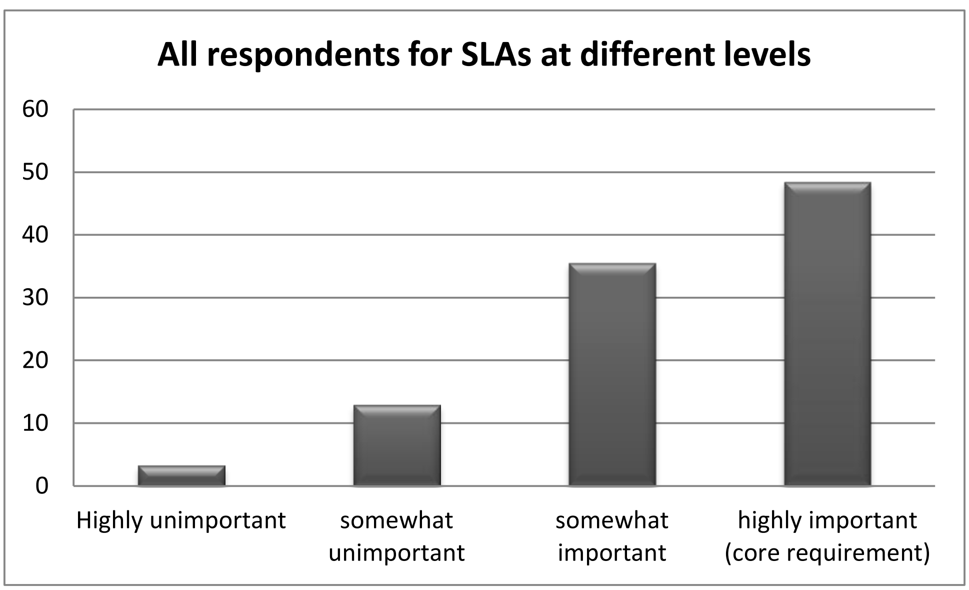

Figure 1. Overall feedback on the importance of the "SLAs at different levels" research area

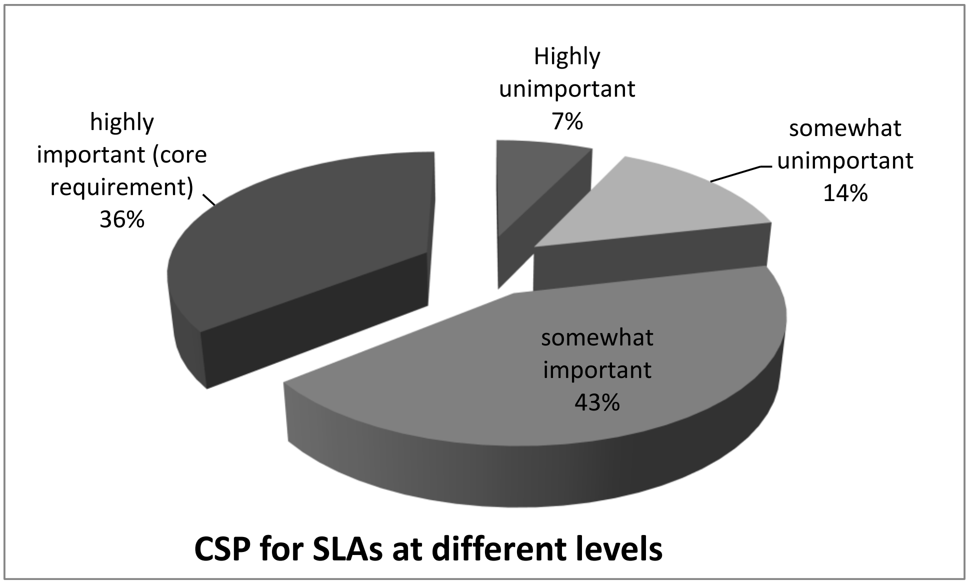

Figure 2. CSPs' feedback on the importance of the "SLAs at different levels" research area

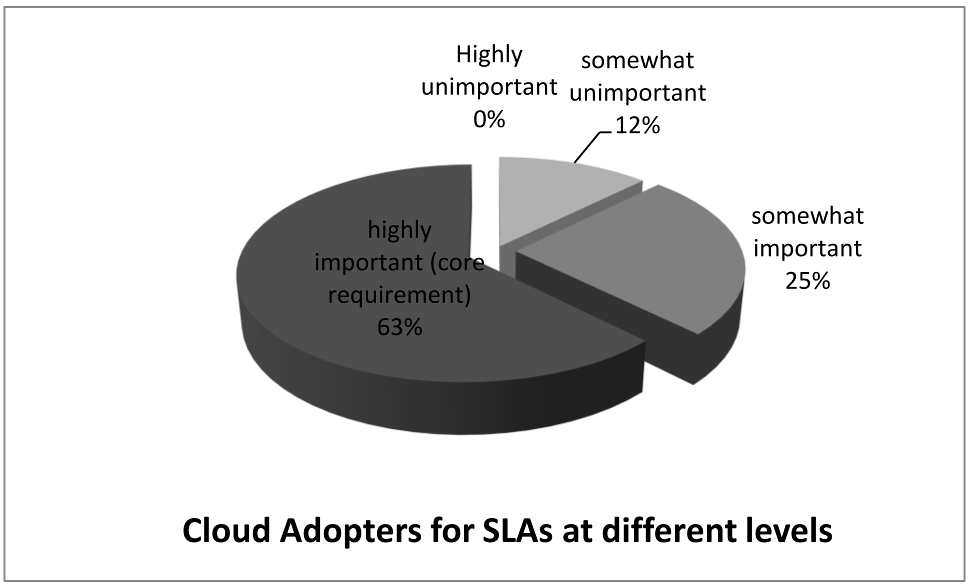

Figure 3. Cloud Adopters' feedback on the importance of the "SLAs at different levels" research area

7.2 Multi-level SLA interaction model

The example provided for clarifications to the respondents of the questionnaire for this research area was as follows:

Irene is a data analyst in a financial trading company. She aims at utilizing resources / services offered from two different providers, as her service is both computational- and data-intensive and different providers offer the corresponding "best" services. To this end, a model for cloud federations is required. The model is based on automated SLA offer generation and enables the user to negotiate an SLA with

the federation and the federation looks for the best way to satisfy it by negotiating SLAs with one or more providers (on behalf of the user).

The CONTRAIL EU project has proposed an approach for SLAs in multi-provider environments, including a scheme for SLA splitting, which allows for service-, resource-, or performance-based SLA splitting and revenue sharing / compensation provision.

As depicted in Figure 4, Figure 5 and Figure 6, "Multi-level SLA interaction model" research area was overall evaluated as 'somewhat important' (39%) although most of the cloud adopters (50%) treat it more like a 'core requirement'.

7.3 SLA Negotiation across multiple layers

The example provided for clarifications to the respondents of the questionnaire for this research area was as follows:

Yannis is a doctor and would like to deploy his eHealth application in a cloud environment. The application requires workflow management that highlights the need for software-layer negotiation on top of the infrastructure-layer negotiation. What is more, the negotiation process across these two layers should be transparent, while domain-specific (i.e. medical) knowledge should also be incorporated in the SLA lifecycle without additional human intervention.

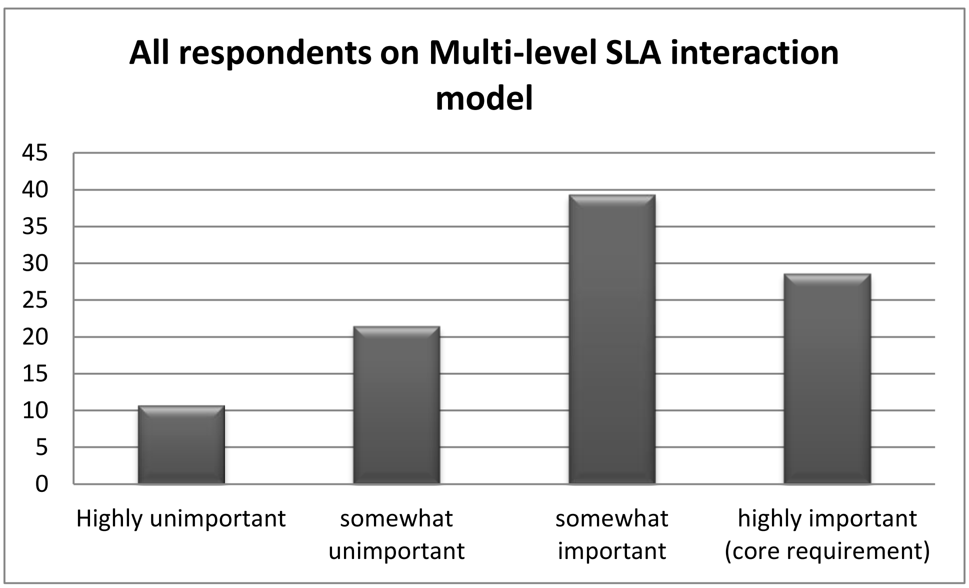

Figure 4. Overall feedback on the importance of the "Multi-level SLA interaction model" research area

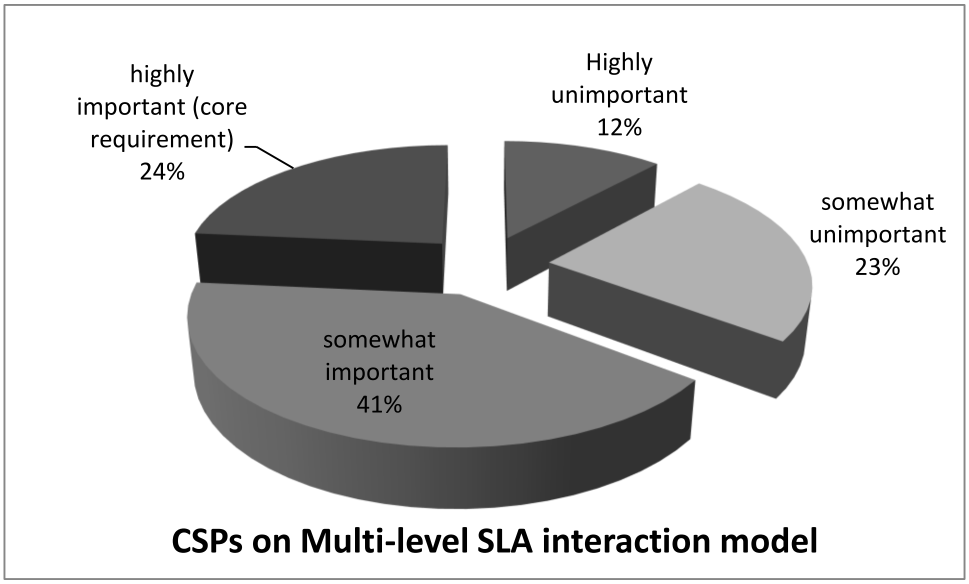

Figure 5. CSPs feedback on the importance of the "Multi-level SLA interaction model" research area

The SLA@SOI EU project implemented a framework that enables different protocols to be injected through an engine, so as to facilitate the interaction between the different layers and entities. Since the protocol will be used for specific interaction, it may include domain specific content.

As depicted in Figure 7, Figure 8 and Figure 9, opinions among CSPs and cloud adopters differ enough regarding the “SLA negotiation across multiple layers” research area. In particular, although most cloud adopters (50%) consider this aspect as a ‘core requirement’, most CSPs (35%) tend to consider it as an aspect with limited interest, i.e., ‘somewhat unimportant’.

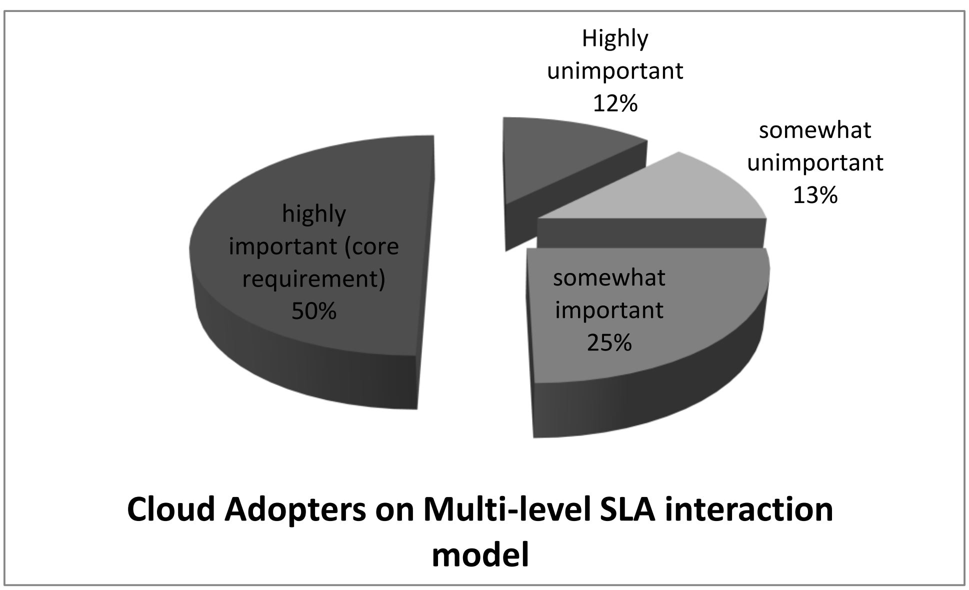

Figure 6. Cloud Adopters feedback on the importance of the "Multi-level SLA interaction model" research area

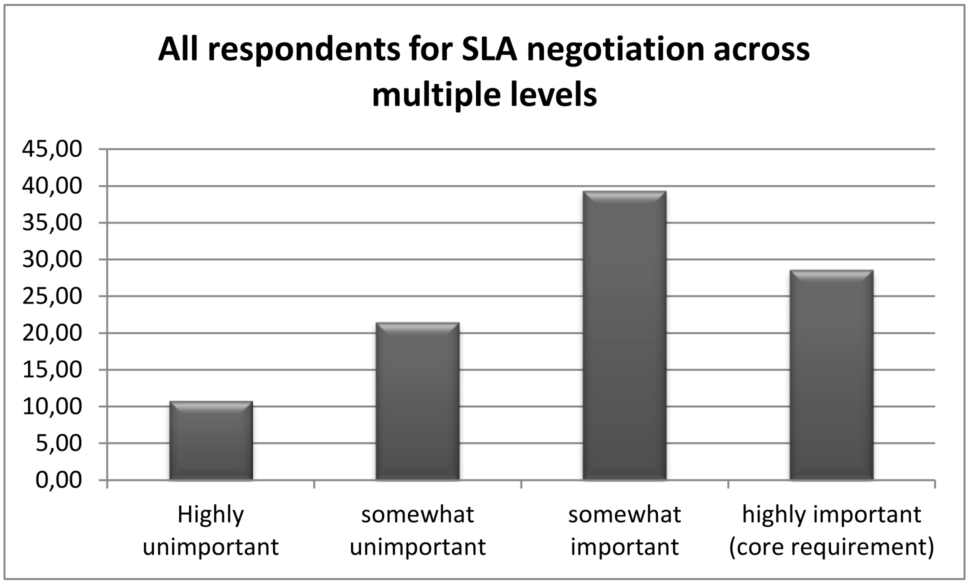

Figure 7. Overall feedback on the importance of the "SLA negotiation across multiple levels" research area

7.4 Automated SLA re-negotiation

The example provided for clarifications to the respondents of the questionnaire for this research area was as follows:

Michael is an entrepreneur who has developed a new tourist service that mixes virtual and augmented reality. As an interactive real-time application, there are specific requirements (towards the cloud infrastructure that hosts it) which however change during runtime based on the number of end-users. His goal is to sign a contract with an SLA provider, which will allow automated re-negotiation during runtime without human intervention. The re-negotiation should also address cases where the causes and origin of an agreement violation could be addressed by establishing again a process of negotiation.

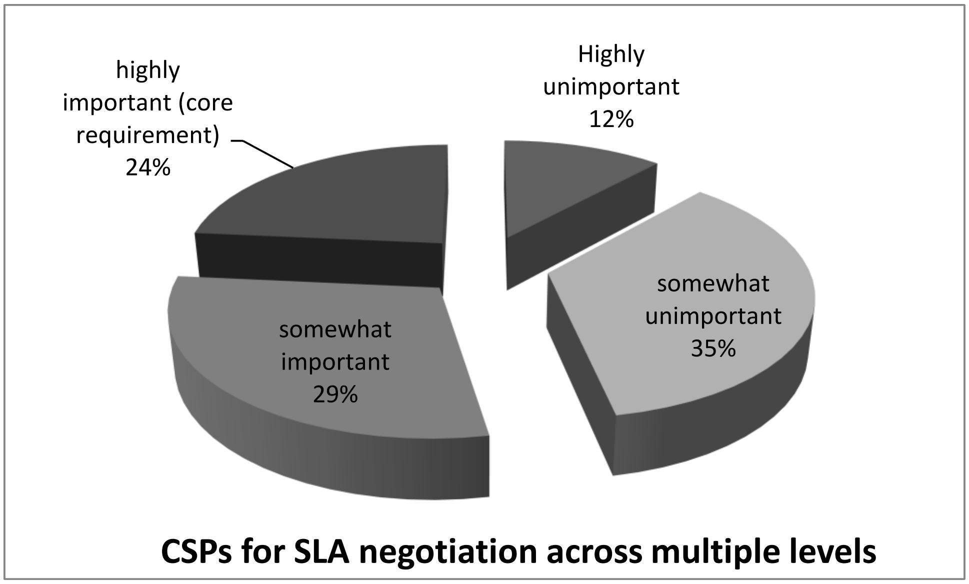

Figure 8. CSPs feedback on the importance of the "SLA negotiation across multiple levels" research area

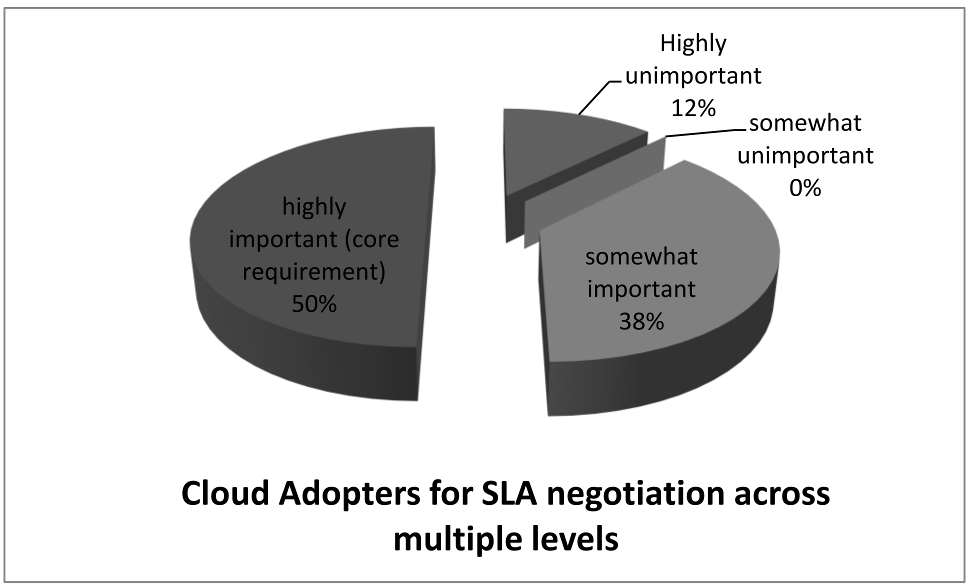

Figure 9. Cloud adopters' feedback on the importance of the "SLA negotiation across multiple levels" research area

The IRMOS EU project has developed an SLA re-negotiation mechanism that can be triggered either by the user (e.g. change in application parameters), by one of the providers (e.g. detection of potential SLA violation) or by the application (e.g. elasticity rules). The mechanism allows for automated re-negotiation during runtime without human intervention.

As depicted in Figure 10, Figure 11 and Figure 12, contradictory are the opinions among CSPs and cloud adopters in this research area as well. In particular, half of the cloud adopters that answered SLALOM questionnaire suggest that "Automated SLA re-negotiation" is highly importance and thus a core requirement while 53% of the respondent CSPs oppose supporting that this research area is 'somewhat

unimportant'. The two contradicting opinions result in an overall characterization of the research area as a 'somewhat important' one supported by ~39% of the overall respondents.

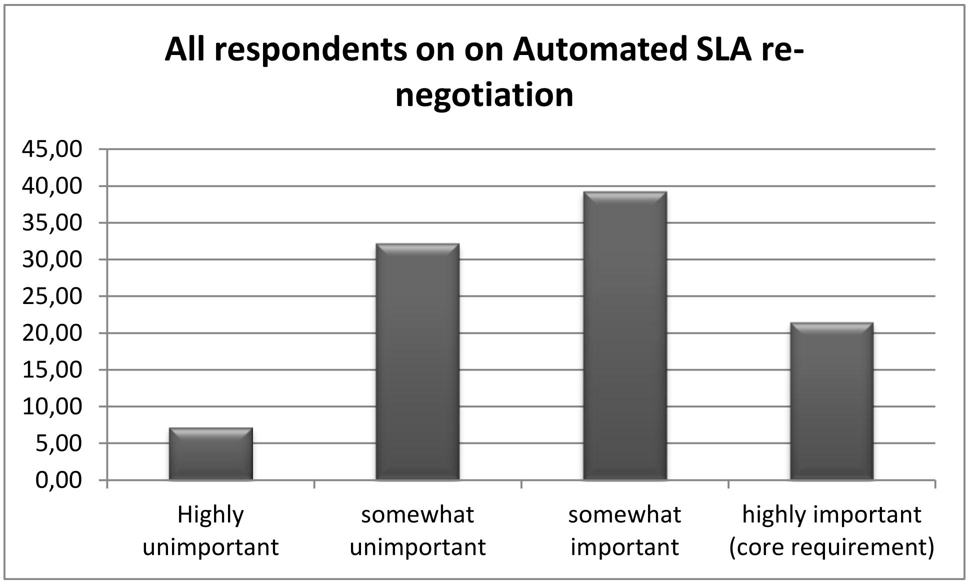

Figure 10. Overall feedback on the importance of the "Automated SLA re-negotiation" research area

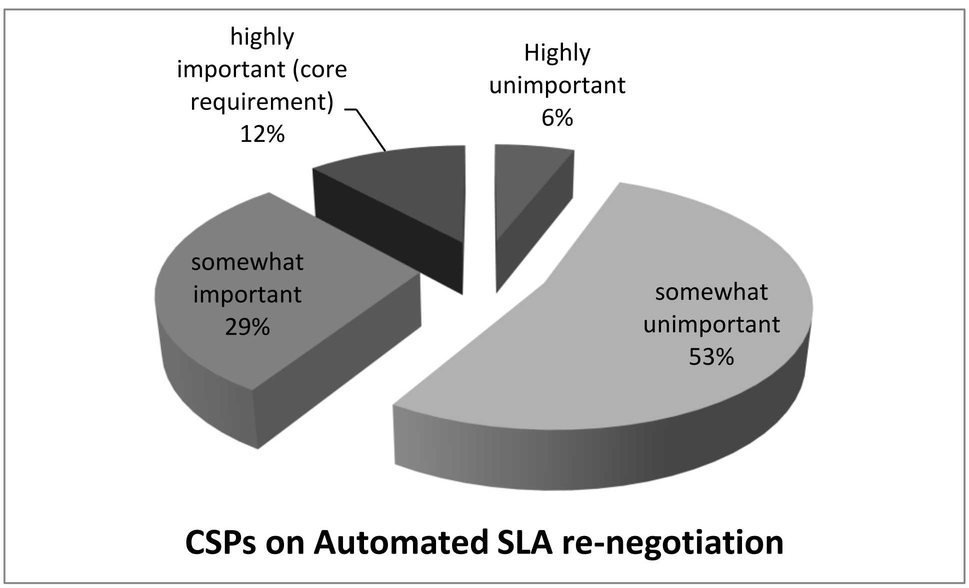

Figure 11. CSPs feedback on the importance of the "Automated SLA re-negotiation" research area

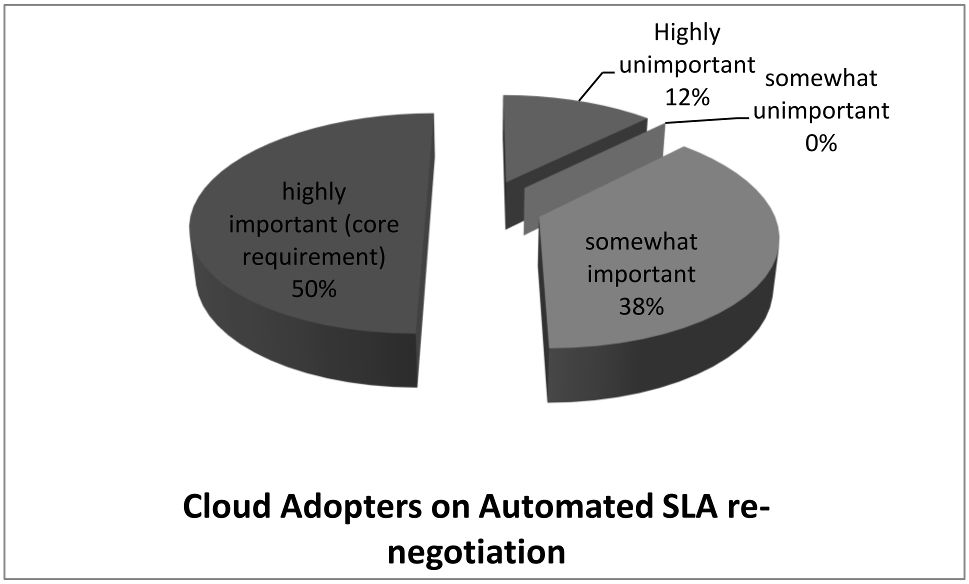

Figure 12. Cloud adopters' feedback on the importance of the "Automated SLA re-negotiation" research area

7.5 Proactive SLA violation detection

The example provided for clarifications to the respondents of the questionnaire for this research area was as follows:

Gabriel is a computer engineer at the operation centre of the publication transportation company. The operation centre software has been deployed and is running on a cloud infrastructure. There are specific requirements for downtime (near zero) and availability of these mission critical services. Therefore, he sets explicit values for the aforementioned terms and requires "proactive" mechanisms that will protect his application from any performance degradation or failure.

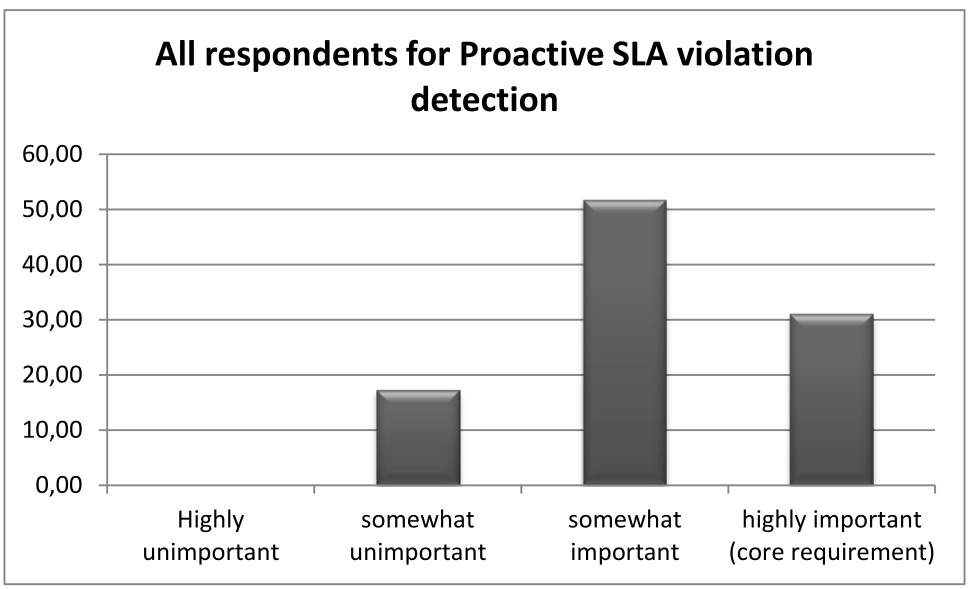

Figure 13. Overall feedback on the importance of the "Proactive SLA violation detection" research area

The VISION Cloud EU project has proposed a proactive SLA violation detection mechanism that bases estimations on the analysis of monitoring data. The analysis enables discovering of repetitive patterns and identification of potential relationships between the different parameters to identify dependencies that may affect the evolution of the parameter values during runtime.

As depicted in Figure 13, Figure 14 and Figure 15, both CSPs (53%) and cloud adopters (45%) agree that "Proactive SLA violation detection" research area is 'somewhat important'. However, it should also be noticed that there is an almost equal percentage of cloud adopters (44%) that consider this research area 'highly important' and thus a 'core requirement'.

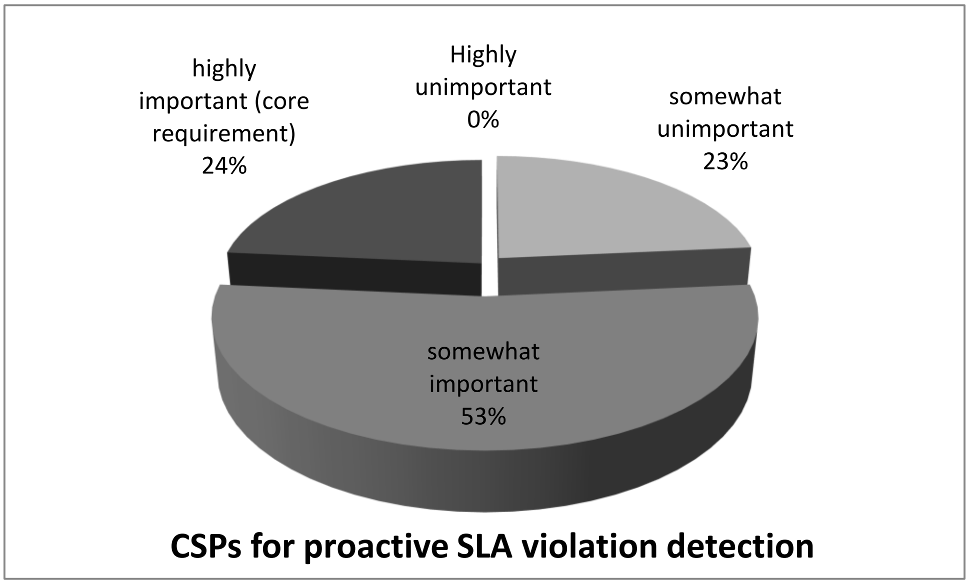

Figure 14. CSPs feedback on the importance of the "Proactive SLA violation detection" research area

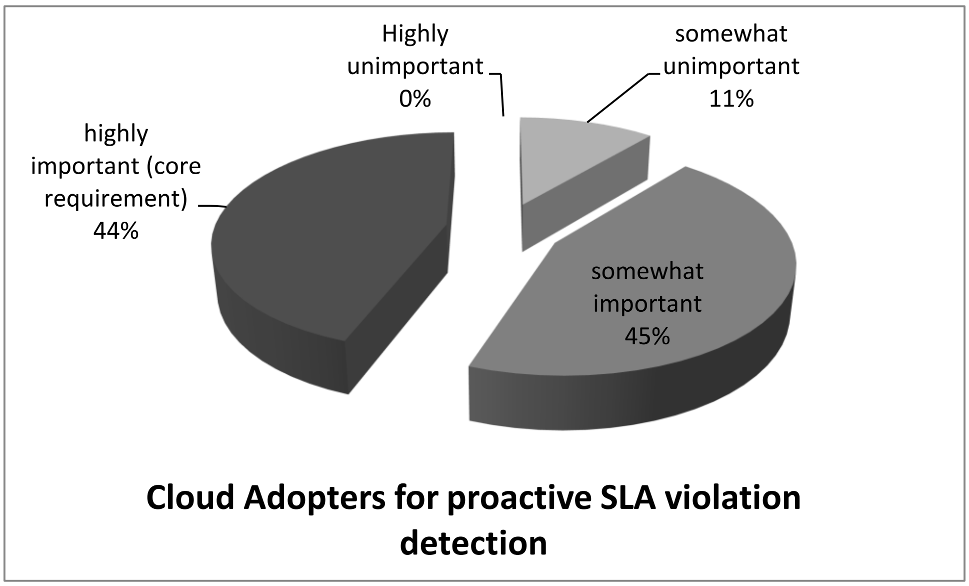

Figure 15. Cloud Adopters feedback on the importance of the "Proactive SLA violation detection" research area

7.6 Other

Two more aspects that SLALOM consortium was advised by CSPs to take into consideration were:

- Expansion of SLA at different levels - should be split out by geographic location due to high variability in user experience based on country. For example, Russia & China users have up to 10x worse performance than US users. This is a much larger impact on performance than anything else in the cloud provider's infrastructure.
- Where the client's cloud service requirement is jurisdiction-critical (e.g. banking/ investment management; legal; health), regular audit reports for the Supplier to the Client to enable the Client to comply with its Quality Management/ Licencing audit requirements

Cloud adopters agreed with CSPs comment, also suggesting that "combined effects of location of data centres, local laws and security" should be considered.

Finally, respondents that belong neither to the CSPs nor to the cloud adopters groups complemented their feedback by proposing the following aspects to be investigated:

- Cloud orchestration standards
- Technical Cloud Brokerage standards (interfacing, API's...)
- Cloud Management Portals standards
- Cloud Services Monitoring & Reporting Standards
- SLA auditing and monitoring process for public Clouds SLAs: 3alib SLA auditing component, http://www.artist-project.eu/tools-of-toolbox/209"
- Translating SLA terms to necessary resource management actions: http://users.ntua.gr/gkousiou/publications/MOCS2011.pdf"

All the above comments will be considered to be included in the work taking place within the SLALOM project.

8 Conclusions

There is a huge amount of work done within the cloud and SLA areas covered by both working groups and research initiatives. SLALOM builds on top of the existing work in order to expand it (e.g., in the case of report [1] and ISO) and support it (e.g., in case of national and European research projects). This document summarizes existing work with respect to technical specifications and provides stakeholders' feedback so as to enable the delivery of a common SLA reference model capturing the essential components across the SLA lifecycle, and their interactions. The reference model will also include potential extensions for added-value services.

9 References

[1] D. Kyriazis (editor), "Cloud Computing Service Level Agreements: Exploitation of Research Results", June 2013, available at http://ec.europa.eu/digital-agenda/en/news/cloud-computing-service-level-agreements-exploitation-research-results [last accessed: Feb. 2015]

[2] OPTIMIS Service Manifest, Scientific Report, http://www.optimis-project.eu/sites/default/files/content-files/document/service-manifest-scientific-report.pdf

[3] RESERVOIR Project, http://www.reservoir-fp7.eu/

[4] M. Papazoglou, W. van den Heuvel, "Blueprinting the Cloud," IEEE Internet Computing, 2011

[5] R. Cascella, L. Blasi, Y. Jegou, M. Coppola, C. Morin, "Contrail: Distributed Application Deployment under SLA in Federated Heterogeneous Clouds", Springer, Lecture Notes in Computer Science, 2013

[6] S. Becker, L. Bulej, T. Bureš, P. Hnetynka, L. Kapová, J. Kofron, H. Koziolek, J. Kraft, R. Mirandola, J. Stammel, G. Tamburrelli, M. Trifu, Service Architecture Meta-Model (SAMM), http://www.q-impress.eu/wordpress/wp-content/uploads/2009/05/d21-service_architecture_meta-model.pdf

[7] A. Menychtas, D. Kyriazis, S. Gogouvitis, K. Oberle, T. Voith, G. Galizo, S. Berger, E. Oliveros, M. Boniface, "A cloud platform for real-time interactive applications ", 1st International Conference on Cloud Computing and Services Science (CLOSER), Noordwijkerhout, The Netherlands, 2011

[8] K. Kearney, F. Torelli, C. Kotsokalis, "SLA*: An Abstract Syntax for Service Level Agreements", Grid Computing (GRID) 2010, 11th IEEE/ACM International Conference on Grid Computing, Brussels, Belgium, 2010.

[9] G. Carrozzo, N. Ciulli, P. Donadio, A. Cimmino, "The Path Computation Element for the Network Service and Business Plane – Computation of route offers and price modelling for inter-carrier services", 17th European Conference on Network and Optical Communications (NOC 2012)

[10] A. Menychtas, J. Vogel, A. Giessmann, A. Gatzioura, S. Garcia Gomez, V. Moulos, F. Junker, M. Müller, D. Kyriazis, K. Stanoevska, T. Varvarigou, "4CaaSt marketplace: An advanced business environment for trading cloud services", Future Generation Computer Systems, 2014

[11] GEYSERS Project, http://www.geysers.eu/

[12] D. Didona, P. Di Sanzo, R. Palmieri, S. Peluso, F. Quaglia, P. Romano, "Automated Workload Characterization in Cloud-based Transactional Data Grids", 17th IEEE Workshop on Dependable Parallel, Distributed and Network-Centric Systems (DPDNS), 2012

[13] SERSCIS Project, http://www.serscis.eu/

[14] PrestoPRIME Project, http://www.prestoprime.org/

[15] R. Calinescu, L. Grunske, M. Kwiatkowska, R. Mirandola, G. Tamburrelli, "Dynamic QoS Management and Optimisation in Service-Based Systems", IEEE Transactions on Software Engineering 37(3), 2011

[16] Cloud4SOA Project, http://www.cloud4soa.eu/

[17] G. Katsaros, G. Kousiouris, S. Gogouvitis, D. Kyriazis, A. Menychtas, T. Varvarigou, “A Self-adaptive hierarchical monitoring mechanism for Clouds”, Elsevier Journal of Systems and Software, 2012
[18] W. Theilmann, J. Lambea, F. Brosch, S. Guinea, P. Chronz, F. Torelli, J. Kennedy, M. Nolan, G. Zacco, G. Spanoudakis, M. Stopar, G. Armellin, “SLA@SOI Final Report”, September 2011.
[19] V. Gulisano, R. Jimenez-Peris, M. Patiño-Martínez, P. Valduriez, “A Large Scale Data Streaming System”, 30th IEEE Int. Conf. on Distributed Systems (ICDCS), Genoa, Italy, 2010
[20] A. Lawrence, K. Djemame, O. Wäldrich, W. Ziegler, C. Zsigri, “Using service level agreements for optimising cloud infrastructure”, International Conference ServiceWave , Ghent, 2010
[21] G. Gallizo, R. Kübert, G. Katsaros, K. Oberle, K. Satzke, S. V. Gogouvitis, E. Oliveros, “A Service Level Agreement Management Framework for Real-time Applications in Cloud Computing Environments”, CloudComp Conference, 2010
[22] G. Brataas, E. Stav, S. Lehrig, S. Becker, G. Kopcak, D. Huljenic, “CloudScale: scalability management for cloud systems”, 4th ACM/SPEC International Conference on Performance Engineering, New York, USA, 2013
[23] P. Kranas, V. Anagnostopoulos, A. Menychtas, T. Varvarigou, “ElaaS: An innovative Elasticity as a Service framework for dynamic management across the cloud stack layers”, 6th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS), Palermo, Italy, 2012
[24] N. Mavrogeorgi, S. Gogouvitis, A. Voulodimos, G. Katsaros, S. Koutsoutos, D. Kyriazis, T. Varvarigou, E. Kolodner, “Content Based SLAs in Cloud Computing Environments”, IEEE International Conference on Cloud Computing (CLOUD), 2012
[25] F. Perez-Sorrosal, R. Jimenez-Peris, M. Patiño-Martinez, B. Kemme, “Elastic SI-Cache: Consistent and Scalable Caching in Multi-Tier Architectures”, VLDB Journal, 2011
[26] D. Didona, Paolo Romano, S. Peluso, F. Quaglia, “Transactional Auto Scaler: Elastic Scaling of In-Memory Transactional Data Grids”, 9th International Conference on Autonomic Computing (ICAC 2012), San Jose, CA, USA, 2012
[27] SLA-Ready project website, http://www.sla-ready.eu/ [last accessed: May 2015]
[28] Helix Nebula project website, http://www.helix-nebula.eu/ [last accessed: May 2015]
[29] CELAR project website, www.celarcloud.eu [last accessed: May 2015]
[30] PaaSage project website, www.paasage.eu [last accessed: May 2015]
[31]SLALOM Questionnaire, available at bit.ly/SLALOMWantsToKnow [last accessed: May 2015]
[32] SLALOM Handout, “SLA Model Terms and Specifications: SLALOM Project Overview and Request for Feedback”, available at bit.ly/SLALOMeHandout [last accessed: May 2015]