Summary

Service Level Agreements (SLA), as well as all its concerning facets such as SLA definition language, negotiation, monitoring, etc., have been subject of research for years, but the advent of Cloud computing, and the need of means for defining and ensuring Quality of Service (QoS) levels have greatly increased the interest on these developments. As the size and complexity of Cloud systems increases, the manual management of these platforms becomes a challenging issue. Therefore, the automation of large scale systems management is another promising feature of the SLA-driven autonomic Cloud solutions. Additionally, SLA-driven Clouds need mechanisms to represent, store and retrieve the information related to their particular domain. Usually the domain changes between different platforms, and henceforth custom models are built to capture this information and ad-hoc implementations are used to store and retrieve it.

This Ph.D. Thesis contributes to these topics by proposing a generic methodology for the representation of the domain in Cloud solutions. This methodology uses the WS-Agreement specification for capturing and manipulation arbitrary domain information using SLA fragments. SLA fragments are parts of SLA documents that describe a single computational element, and are composed on the fly in response to user request to generate a complete SLA document. This methodology provides the generality, extensibility and flexibility to unify the modeling of the domain in arbitrary Cloud services. A SLA composition algorithm enables a prototype implementation of the methodology in Cloudcompaas, a SLA-driven Cloud framework that manages the complete resource (e.g. Virtual Machines, software, services) lifecycle. This framework features an extension of the WS-Agreement SLA specification, tailored to the specific needs of Cloud computing. In particular, Cloudcompaas enables Cloud providers with a generic SLA model to deal with higher-level metrics, closer to end-user perception, and with flexible composition of the requirements of multiple actors in the computational scene. Moreover, Cloudcompaas provides a framework for general Cloud computing applications that dynami-

cally reacts to changes on Cloud infrastructures to correct QoS level guarantee violations.

The two major contributions of this Thesis are a generic methodology for the description of Cloud services, and the architecture, design and implementation of a SLA-driven Cloud framework. A use case provides a quantitative measure of the utility provided by the methodology from a Cloud user and Cloud provider point of view, reducing the price and increasing the ratio of users served. The effectiveness of the framework is demonstrated through the simulation of several realistic workload profiles, where Cloudcompaas achieves minimum price and maximum utility under highly heterogeneous utilization patterns.