INTRODUCTION

Web services and their orchestrations are now considered an infrastructure of choice for managing business processes and workflow activities over the Web infrastructure (van der Aalst, 2002). BPEL (WSBPEL, 2007) has become the industrial standard for specifying orchestrations. Besides BPEL, the Orc formalism has been proposed to specify orchestrations, by W. Cook and J. Misra at Austin (Misra, 2007) (Kitchin, 2009). Orc is a simple and clean academic language for orchestrations with a rigorous mathematical semantics. For this reason, our study in this paper relies on Orc. Its conclusions and approaches, however, are also applicable to BPEL.

When dealing with the management of QoS, the commitments of each subcontractor with regard to the orchestration are specified via contracts in the form of Service Level Agreements, SLA (Bohj, 2001). Most SLAs commonly tend to have QoS parameters which are mild variations of the following: response time (latency); availability; maximum allowed query rate (throughput); and
Flexible Probabilistic QoS Management of Orchestrations

security (Keller, 2003). From QoS contracts with sub-contractors, the overall QoS contract between orchestration and its clients can be established. This process is called contract composition. Then, since contracts cannot only rely on trusting the sub-contractors, monitoring techniques must be developed for the orchestrator to be able to detect possible violation of a contract, by a sub-contractor. Finally, upon contract violation, the orchestrator may consider reconfiguration, i.e., replacing some called services by alternative, “equivalent” ones — we do not address this last task here.

To the best of our knowledge, with the noticeable exception of (Liu, 2001) and (Hwang, 2007), all composition studies consider performance related QoS parameters of contracts in the form of hard bounds. For instance, response times and query throughput are required to be less than a certain fixed value and validity of answers to queries must be guaranteed at all times. When composing contracts, hard composition rules are used. Typical examples are addition or maximum (for response times), or conjunction (for validity of answers to queries). Whereas this results in elegant and simple composition rules, this general approach by using hard bounds does not fit the reality well and may lead to over pessimistic promises. Indeed, real measurements of response times for existing Web services reveal that they vary a lot and are better represented through their histogram. Thus we have proposed in (Rosario, 2008) using soft probabilistic contracts instead. In such contracts, hard bounds are replaced by probabilistic obligations, i.e., a QoS parameter is considered probabilistic and a distribution probability is agreed for it. The obligation is that the called service should behave “no worse” than this agreed distribution regarding this QoS parameter, in a sense that we formalize in this paper.

Adopting a probabilistic approach for QoS has many advantages, but also raises some issues when performing contract composition and contract monitoring. Analytical solutions for deriving the distribution of the composition from the distribution of its components exist for simple cases where the control flow of the composition is not affected by the data values of the queries and their responses, and other timing issues. Queuing network techniques can be used in simple cases like this. More sophisticated stochastic Petri nets can also be used, but they require restricting to exponential distributions. These elegant analytical approaches, however, are not applicable in general to services orchestrations where responses to queries and timing (via timeouts) interfere with the control flow – The CarOnLine example in the next section is an instance of this. We thus need to develop new techniques to perform contract composition and contract monitoring, adapted to probabilistic contracts.

Contributions: In this paper we extend and systematize the approach of (Rosario, 2008) and (Bouillard, 2009) by extending it beyond the case of soft probabilistic contracts for Response Time.

Our first contribution consists in proposing a comprehensive approach for Soft Probabilistic QoS Contracts encompassing a large class of QoS parameters taking values in partially ordered domains, together with means to build composite QoS parameters and contracts and reason about them.

A second contribution consists in a procedure to perform flexible contract composition, which consists in relating the obligations binding the pair {client, orchestration}, to the obligations binding the different pairs {orchestration, called service}.

A third (minor) contribution consists in the extension of the technique proposed in (Rosario, 2008) for contract monitoring to our generalized case. This extension turns out to be straightforward, as we shall see.

Last but not least, we discuss languages features that are useful in making our approach effective. Not surprisingly, QoS domains must be declared along with their characteristics allowing to perform contract composition. We also found it very useful to introduce a language feature that is