Decision Support Systems 53 (2012) 234–244

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Workflow composition of service level agreements for web services ☆

M. Brian Blake [a,*], David J. Cummings [b], Ajay Bansal [c], Srividya Kona Bansal [c]

[a] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States
[b] Department of Computer Science, Stanford University, United States
[c] Department of Engineering, Arizona State University, United States

ARTICLE INFO

Article history:
Received 19 July 2010
Received in revised form 22 December 2011
Accepted 30 January 2012
Available online 8 February 2012

Keywords:
Service level agreements
Quality of service
Web services
Service-oriented computing

ABSTRACT

Service-oriented architecture enables an environment where businesses can expose services for use by their collaborators and their peer organizations. In this dynamic environment, organizations require the use of service level agreements (SLAs) to assure the quality of service (QoS) standards of services provided by their collaborators. In an ad-hoc workflow scenario, a business may need to perform real-time composition of existing services in response to consumer requests. In this work, we suggest that, in parallel to traditional web service composition, the business must also compose the existing SLAs in order to ensure the service levels that must be guaranteed to new consumers. Ultimately, this approach to SLA composition must align with the overarching principles of the provider and the priorities of the consumer. In this paper, we introduce a model and representations of service level agreement attributes appropriate for managing a service provider's expectations when adding new partners. Our evaluations suggest that the SLA composition can efficiently run concurrently with traditional service composition.

1. Introduction

A service level agreement, *SLA*, is a technical contract between two types of businesses, producers and consumers. A SLA captures the agreed-upon terms between organizations with respect to quality of service (QoS) and other related concerns. In simple cases, one consumer forms a SLA with a producer. In more complex cases, a consumer may form a SLA that defines a set of producer businesses. Considering a service-oriented computing environment, capabilities are shared via the implementation of web services exposed by a producer organization. The ultimate goal of service-oriented computing is for consumers to access these shared capabilities on-demand. As such, in cases where businesses have longstanding relationships, such as workflow and supply chain environments, peer companies that share services must be able to assure a level of service to their underlying customers [4,8].

New specifications, such as the Web Service Level Agreement (WSLA) and Web Service Agreement (WS-Agreement) [2] enable SLAs to be associated with an individual web service or even groups of web services. These specifications define an eXtensible Markup Language (XML)-based data model that can be used along with the Web Service Description Language (WSDL) documents that traditionally describe the web services. These specifications provide a significant opportunity. Organizations can specify QoS-related concerns in concert with the functionality concerns already captured in the WSDL files. As a result, when a new organization searches for a pertinent web service, the SLA-enhanced WSDL file can be used to determine the appropriateness of the service to meet the required business need. Furthermore organizations can use the SLA-enhanced WSDL file to negotiate the QoS terms.

Although these SLA technologies and specifications present new opportunities for service-oriented business processes, there are a number of significant barriers. When a consumer organization must create a new business capability that requires the workflow composition of multiple web services, then that organization will also need to understand the composite impact of the underlying SLAs. Consequently, in addition to composing web services that are functionally compatible, the organization will need to ensure that the web services are compatible with regard to their service levels. Also, the product of all the SLAs for a composition of web services must be within the required threshold of *feasibility* as defined by the end users. As the service-oriented computing paradigm increases in popularity, the consumer will have the option of many similar services that may meet a particular requirement. As such, the composition of web services that is most efficient for a particular business purpose will rest on the organization's ability to understand and optimize the corresponding composition of SLAs.

To deal with the aforementioned issues, we introduce the phrase, *workflow composition of SLAs*. Our approach suggests the multi-dimensional evaluation of existing agreed-upon QoS standards in order to predict the standards possible for the introduction of new agreements. While the notions of multi-dimensional analysis, optimization, dynamic programming are not new [9,10,12,17,35] in this

☆ This paper is a substantial extension of earlier work presented in [7] and [5].

* Corresponding author. Tel.: +1 574 631 1625; fax: +1 574 631 8007.
E-mail addresses: m.brian.blake@nd.edu (M.B. Blake), david.cummings@stanford.edu (D.J. Cummings), srividya.bansal@asu.edu, ajay.bansal@asu.edu (A. Bansal).

0167-9236/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2012.01.017

work, we identify the specific SLA-based attributes that allows for the introduction of new partners. Furthermore we develop a set of principles and the associated process that utilize the SLA information to estimate service levels. This approach favors services with clean request/response (RPC-type) communication, generally known as WSDL-based web services. Further investigation would be required to assess this approach as it relates to REST-based services [37].

In this work, we investigate several research issues relevant to the integration of web services-based workflow:

1. *What SLA measures and principles are appropriate to support QoS-based assessment of existing service level guarantees?*
2. *Given a group of SLAs and knowledge about current consumer service level needs, can an on-demand request be analyzed against existing SLAs to guarantee a certain service level for a new consumer?*

The paper proceeds in the following section with a discussion of related work. In Section 3, we discuss how the SLA-based QoS assessment values are derived from higher-level organizational principles. The formal details of the attributes are defined in Section 4, and how the attributes are physically captured in markup languages in shown in Section 5. Finally, in Section 6 we evaluate the performance SLA composition as it runs parallel with traditional service composition routines.

2. Related work

There are many related projects that investigate the general use of SLAs for web services [18]. Some projects characterize SLA approaches to specific domains, such as military, database management, or information systems [13,20,26,29]. There is also a large body of work that attempts to automate the management and negotiation of SLAs [11,16,23,27,33]. Other work attempts to use semantics to automate the negotiation of SLAs [15,25].

Our work leverages markup language (i.e. WS-Agreements) for providing SLA measures as in other studies [1,28,30]. All related work describes the importance of composing SLAs. In [28], their emphasis is on compatibility between user requirements and provider constraints. Their approach suggests a promising model-based approach to assuring the compatibility.

Our work is closely related to the comprehensive work performed by [9,10,35]. Each of these approaches investigates the QoS-based and constrained composition of web services. Although Canfora et al. [9] has an elegant approach that allows for the insertion and aggregation of any user-defined QoS attribute, our approach identifies the specific SLA measures that support the user-driven assessment of an environment where their SLAs dictate current system state. Table 1 shows a survey of SLA attributes and how they are exploited in related projects specifically in the service-oriented computing domain.

Our work can be loosely classified in the body of work that looks to automate the aggregation of QoS attributes [14,21,24,32]. The uniqueness of our approach is that we consider the impacts when new web service workflows must be added as they affect the existing operational SLAs. More specifically, if a composite capability overlaps multiple SLAs, then the characteristics of an early SLA can impact a later SLA in the composition routine. Canfora et al., Cardoso et al., Zeng et al., [9,10,35] concentrate on deriving a specific composition routine as constrained by QoS values. Canfora et al., Zhang et al., and Yu et al. focus on iterative multiattribute utility approaches [12] where to focus is on the overall optimization function and less on the details of each of the attribute. Our work attempts to consider both consumer and producer concerns when assessing the entry of a new workflow. As such, our work contains low-level details for each service level objective such that subsequent optimization approaches use them as a model for optimization that targets each attribute at a low-level. Although [22] has a similar approach where SLAs are aggregated formally, they do not consider consumer and producer services independently as in our work.

In summary, we define specific SLA measures and formally integrate measures across multiple SLAs. We also define a principled process for the composition of SLAs. This work extends related work [6] by concentrating on SLA measures in markup language files as opposed to Unified Modeling Language (UML) models. Unlike other work in QoS-based web service composition, we attempt to classify QoS attributes by those associated with the provider and those associated with the consumer. Another variation here is the introduction of several high-level criteria that can be used to characterize organizations. We believe that by aggregating all lower-level attributes into a smaller set of higher-level criteria then organizations can be quantitatively evaluated or *scored*. Further evaluation in this paper justifies that such on-demand assessment of SLAs performs feasibly in an operational environment where very large numbers of web service workflows exist.

3. Assessing an enterprise based on its SLAs

The typical SLA has a large number of measures and criteria. However, in this work, we attempt to choose the measures that are most closely aligned to aggregation of a group of SLAs and ultimately their assessment. In the operational notion of web service composition, a basic web services workflow system must ensure that the input information supplied by the consumer ultimately leads to the required actions and outputs required by that consumer. In parallel, the workflow management system must ensure that the predicates and requisites match (either by syntactical or semantic techniques) in each step of the workflow. It is the operational composition routines that motivate the set of SLA attributes relevant to our work.

In order to designate which attributes that are most relevant to our proposed innovation, we developed a set of principles important to managing the quality of an enterprise with many business processes. The three relevant principles are *Compliance (Suitability)*,

Table 1
Survey of research projects that consider SLA attributes for web services.

Author names	Run time	Reputation	Uptime (Avail)	Resp time ***	Negotiation (rebinding)	Cost (price)	Success rate/ reliability	Problem resolution	Maintenance
(Blake et al.) [7] and this paper	✓	✓	✓	✓	✓	✓	✓	✓	✓
(Canfora et al.) [9]	✓	✓	✓	✓	✓	✓	✓	✓	✓
(Yu et al.) [34] Zhang et al.) [36] **									
(Zeng et al.) [35]		✓	✓	✓		✓	✓		
(Cardoso et al.) [10]			✓	✓		✓	✓		
(Jin et al.)[18] *	✓	✓	✓	✓		✓	✓		
(Mohabey et al.) [22]		✓	✓	✓		✓	✓		

* [18] list attributes, but do not develop formal approaches for composition.
** Although Canfora et al., Yu et al., and Zhang et al. do not formally define each attribute, their work focuses on an approach that allows *any* attribute to be aggregated within the composition routine.
*** Response time and run time (and in other places Service Rate) have the same or different definitions. This table places those attributes into separate columns to show the different meanings across the survey of related literature.