Optimizing Complex Service-Based Workflows for Stochastic QoS Parameters

Dieter Schuller, Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt, Germany
Ulrich Lampe, Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt, Germany
Julian Eckert, Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt, Germany
Ralf Steinmetz, Multimedia Communications Lab (KOM), Technische Universität Darmstadt, Darmstadt, Germany
Stefan Schulte, Distributed Systems Group, Vienna University of Technology, Vienna, Austria

ABSTRACT

The challenge of optimally selecting services from a set of functionally appropriate ones under Quality of Service (QoS) constraints – the Service Selection Problem – has been extensively addressed in the literature based on deterministic parameters. In practice, however, Quality of Service QoS parameters rather follow a stochastic distribution. In the work at hand, we present an integrated approach which addresses the Service Selection Problem for complex structured as well as unstructured workflows in conjunction with stochastic Quality of Service parameters. Accounting for penalty cost which accrue due to Quality of Service violations, we perform a worst-case analysis as opposed to an average-case analysis aiming at avoiding additional penalties. Although considering conservative computations, QoS violations due to stochastic QoS behavior still may occur resulting in potentially severe penalties. Our proposed approach reduces this impact of stochastic QoS behavior on total cost significantly.

Keywords: Optimization, Quality of Service, Service Selection, Simulation, Stochastic Quality of Service

INTRODUCTION

In today’s global economy, organizations are subject to high competitive pressure, which results in frequently changing business requirements, and thus, business processes. Because most business processes are nowadays enabled by Information Technology (IT), these changes primarily trigger the need for a tighter business/IT alignment. Thus, as business processes require to be flexible in order to allow for quick adaptations due to the mentioned changing requirements, the respective (supporting) IT requires to be quickly adaptable too. But as
current IT architecture landscapes often tend to be heterogeneous due to historically grown legacy systems and (business) applications, which are implemented with different programming languages and run on different middleware platforms and operating systems, the required level of flexibility is hard to achieve.

Recently, Service-oriented Architecture (SOA) has evolved as one important paradigm to support and enable agile business processes (Papazoglou, 2003). One of its key features is the ability to utilize and compose loosely coupled services in order to realize IT supported business processes and workflows, respectively. Depending on their granularity, these services provide a more or less coarse-/fine-grained functionality (Krafzig, Banke et al., 2004). Further, services are not necessarily located solely within the own enterprise but – following the vision of the Internet of Services – are offered by different service providers on multiple service market places. Thus, in case, services are available on such marketplaces, which are equally appropriate to realize certain tasks, enterprises have the possibility to choose among them based on the services’ cost and provided quality levels (Quality of Service – QoS). The selection of services from a set of appropriate ones that are able to provide the required functionality and thereby best meeting cost and Quality of Service (QoS) requirements – the Service Selection Problem (SSP) – is widely recognized in the literature (e.g., Ardagna & Pernici 2007, Menascé, Casalicchio et al., 2008, Huang, Lan et al., 2009, Strunk, 2010). The optimization of the SSP is thereby based on deterministic QoS values. A solution to the SSP describes an execution plan, i.e., an assignment of services to certain tasks of a workflow, which satisfy the mentioned cost and QoS constraints.

But QoS, e.g., the response time of a service or its availability, is not always deterministic in reality. Due to network latency or server load, response times of services may change dynamically. I.e., when the execution of the computed execution plan actually takes place, the perceived QoS might differ from the expected QoS which has previously been used for the calculation of the execution plan. Thus, although having computed an optimal solution to the SSP during design time which satisfies the constraints, it still is possible that these constraints are violated during runtime.

The work at hand addresses this issue. Based on a service broker scenario, which is presented in the following section, we describe how QoS violations due to stochastic QoS behavior negatively impacts total cost. In order to account for this impact of stochastic QoS parameters, we propose an integrated approach comprising an optimization, a simulation, and an adaptation step. In this respect, we present three different adaptation heuristics.

During the optimization step, we compute an optimal solution to the SSP, i.e., an optimal execution plan, satisfying the QoS constraints based on the deterministic QoS values denoted by the respective service providers. In the simulation step, we observe the expected runtime behavior of the computed execution plan in terms of QoS. This way, we can assess potentially occurring constraint violations. We thereby assume that violating QoS constraints is penalized, i.e., penalty fees become due in addition to service invocation cost. According to the results of the simulation, we apply greedy adaptation heuristics in order to reduce the impact of potentially occurring constraint violations.

In order to avoid developing algorithms that fit particularly to one dedicated service-based workflow, we use different, complex structured as well as unstructured workflows for evaluating our approach. In this respect, we are able to account for structured workflows comprising Sequences, AND-blocks (AND-split with AND-join), XOR-blocks (XOR-split with XOR-join), OR-blocks (OR-split with OR-join), and Repeat loops as well as for unstructured workflows such as Single-Entry-Multiple-Exit (SEME) loops, which distinguishes our work from related approaches in the according literature. While the former workflow patterns have been considered in our work in (Schuller, Polyvyanyy et al., 2011), the work at hand additionally presents and accounts for SEME loops (cf.
Related Content

Discovery of Web Services in a Multi-Ontology and Federated Registry Environment
www.igi-global.com/article/discovery-web-services-multi-ontology/3062?camid=4v1a

Ontology and Service Oriented Programming
www.igi-global.com/article/ontology-service-oriented-programming/3063?camid=4v1a
On Utilizing Web Service Equivalence for Supporting the Composition Life Cycle
[www.igi-global.com/article/utilizing-web-service-equivalence-supporting/50492?camid=4v1a](www.igi-global.com/article/utilizing-web-service-equivalence-supporting/50492?camid=4v1a)

Analytics in Public Policy Related to Service Sector
[www.igi-global.com/chapter/analytics-in-public-policy-related-to-service-sector/217831?camid=4v1a](www.igi-global.com/chapter/analytics-in-public-policy-related-to-service-sector/217831?camid=4v1a)