QoS-Driven Web Services Selection in Autonomic Grid Environments

Danilo Ardagna¹, Gabriele Giunta², Nunzio Ingraffia², Raffaela Mirandola¹, and Barbara Pernici¹

¹ Dipartimento Elettronica e Informazione, Politecnico di Milano, Italy
² Engineering Ingegneria Informatica S.p.A, R&D Lab, Italy

Abstract. In the Service Oriented Architecture (SOA) complex applications can be described as business processes from independently developed services that can be selected at run time on the basis of the provided Quality of Service (QoS). However, QoS requirements are difficult to satisfy especially for the high variability of Internet application workloads. Autonomic grid architectures, which provide basic mechanisms to dynamically re-configure service center infrastructures, can be be exploited to fullfill varying QoS requirements. We tackle the problem of selection of Web services that assure the optimum mapping between each abstract Web service of a business process and a Web service which implements the abstract description, such that the overall quality of service perceived by the user is maximized. The proposed solution guarantees the fulfillment of global constraints, considers variable quality of service profile of component Web services and the long term process execution. The soundness of the proposed solution is shown through the results obtained on an industrial application example. Furthermore, preliminary computational experiments show that the identified solution has a gap of few percentage units to the global optimum of the problem.

1 Introduction

The recent trend towards shorter business cycles encourages the use of the emerging grid technologies and of the flexible service oriented development paradigm to meet the goal of construction and management of reliable and adaptable e-business applications at low cost. Our work aims to meet the requirements and the challenges posed by e-business applications by using the new emerging technological trends. Specifically, we intend to pursue the QoS-driven selection and composition of Web services for e-business applications in autonomic grid environments.

SOA paradigm foresees the creation of complex applications, described as business processes, from independently developed services that can be selected at run time. Usually, a set of functionally equivalent services exist, that is, services which implement the same functionality, but differ for non-functional characteristics, i.e., QoS properties. In this context, our goal is to discover the optimum mapping between each abstract Web service of a business process and a Web service which implements the abstract description.
However, QoS requirements are difficult to satisfy especially due to the high variability of Internet application workloads. Internet workloads can vary by orders of magnitude within the same business day [12]. Such variations cannot be accommodated with traditional allocation practices, but require autonomic computing self-managing techniques [15], which dynamically allocate resources among different services on the basis of short-term demand estimates.

This dynamic fulfillment of varying QoS requirements can be enhanced by grid computing. Grid middleware provides basic mechanisms to manage the overall infrastructure of a service center, implementing service differentiation and performance isolation for multiple Web services sharing the same physical resources and simplifying the re-configuration of the physical infrastructure.

In the literature, resource allocation in scientific grid workflows has been analyzed in depth [23,28]. However scientific workflows and composed e-business processes have different computing requirements, which pose diverse constraints on the re-configuration of the infrastructure. Indeed, in scientific grid, each task is computation intensive and is executed by a single computing resource. Vice versa, in e-business applications, the execution of a single Web service operation requires few CPU seconds and the high computation requirements are due to the high incoming workload. Furthermore, grid workflow and composed Web services have different characteristics: scientific grids introduce many tasks, generally in the order of thousands (the higher is the number the higher is the parallelism level of the workflow). Vice versa, e-business composed processes introduce a lower number of tasks since they correspond to high level application operations [9,25]; also the number of candidate resources for each task is very high (since correspond to available computing and storage resources) for scientific grids and moderate for e-business grids where resources are Web services candidate for the execution of high level operations. For this reasons resource allocation in scientific and e-business grids requires different approaches.

In this paper we present a reference framework to support the execution of Web services based e-business applications in autonomic grid environments. The problem of selection of Web services in composed services such that the QoS for the end user is maximized is formulated as a Mixed Integer Linear Programming (MILP) model. The formulation guarantees the fulfillment of global constraints and extends the work in [5] allowing the execution of stateful Web services. The devised solution is then applied to a case study derived from an industrial settings and the results obtained so far show the effectiveness of the proposed approach.

The paper is organized as follows. Section 2 briefly reviews related works. Section 3 describes the adopted case study. In Section 4 we describe the composed Web service specification and the quality model adopted. Section 5 is devoted to the description of the problem of resource allocation in grid environments, while Section 6 presents the numerical results obtained in the considered example application. Finally, conclusions are drawn in Section 7.