Chapter 16
Accelerating Web Service Workflow Execution via Intelligent Allocation of Services to Servers

Konstantinos Stamkopoulos
University of Ioannina, Greece

Evaggelia Pitoura
University of Ioannina, Greece

Panos Vassiliadis
University of Ioannina, Greece

Apostolos Zarras
University of Ioannina, Greece

ABSTRACT

The appropriate deployment of web service operations at the service provider site plays a critical role in the efficient provision of services to clients. In this paper, the authors assume that a service provider has several servers over which web service operations can be deployed. Given a workflow of web services and the topology of the servers, the most efficient mapping of operations to servers must then be discovered. Efficiency is measured in terms of two cost functions that concern the execution time of the workflow and the fairness of the load distribution among the servers. The authors study different topologies for the workflow structure and the server connectivity and propose a suite of greedy algorithms for each combination.

DOI: 10.4018/978-1-61350-471-0.ch016

INTRODUCTION

A web service is typically defined in the literature –for example, see Alonso, Casati, Kuno and Machiraju (2004)—as an interface that describes a collection of operations provided through the internet and accessed through standard XML messages. The appropriate deployment of web service operations at a service provider site plays a critical role in the efficient provision of services to clients. To effectively provide solutions to users' tasks, web services are *composed* in *workflows* (see Chen, Zhou, & Zhang, 2006) that combine intermediate service results towards achieving a more complex goal. Such workflows are typically specified in appropriate languages such as BPEL (see Andrews, et al., 2003).

Motivating Example

Assume an electronic system that assigns rendezvous for patients that need to consult doctors. A workflow that arranges a meeting depending on the availability of a doctor is depicted in Figure 1. Once the meeting has been conducted, the system registers any prescribed medicines and communicates via operations with social security agencies to register the assignment of medicines to patients. The detailed description of these operations is not necessary for the purpose of the paper; still it is important to note that there are *operational services* that receive requests (in the form of XML messages) to which they react (by sending XML messages) and *decision operations* that regulate which operations are to be invoked depending on the state of the workflow.

The whole workflow is supported by web service operations, deployed by the ministry of health and social security. The ministry has 5 servers that can host any of the 15 operations of the workflow and *the problem is to decide which of the possible 5^{15} configurations of the deployment of operations to servers (a) provides the fastest closing of each patient case and (b) loads each server in a fair way, so that whenever additional workflows are deployed, or a server fails, a reasonable load scale-up is still possible*.

Background and Problem Statement

In the problem we are dealing with in this paper, *we assume that a service provider has several servers over which web service operations can be deployed. Then, given a workflow and the topology of the servers, the most efficient deployment of the operations must be discovered*. Different topologies refer to the possibility of different networking infrastructure for the servers; this might include particularities relating to the characteristics of the machinery data center,

Figure 1. Exemplary workflow