

Connecting Client Objectives with Resource Capabilities: An Essential Component for Grid Service Management Infrastructures

Asit Dan

IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532
asit@us.ibm.com

Catalin Dumitrescu

Computer Science Department,
The University of Chicago
1100 E 58th St., Chicago, IL 60637
cldumitr@cs.uchicago.edu

Matei Ripeanu

Computer Science Department,
The University of Chicago
1100 E 58th St., Chicago, IL 60637
matei@cs.uchicago.edu

ABSTRACT

In large-scale, distributed systems such as Grids, an agreement between a client and a service provider specifies service level objectives both as expressions of client requirements and as provider assurances. Ideally, these objectives are expressed in a high-level, service- or application-specific manner rather than requiring clients to detail the necessary resources. Resource providers on the other hand, expect low-level, resource specific performance criteria that are uniform across applications and can easily be interpreted and provisioned.

This paper presents a framework for Grid service management that addresses this gap between high-level specification of client performance objectives and existing resource management infrastructures. It identifies three levels of abstraction for resource requirements that a service provider needs to manage, namely: detailed specification of raw resources, virtualization of heterogeneous resources as abstract resources, and performance objectives at an application level. The paper also identifies three key functions for managing service level agreements, namely: *translation* of resource requirements across abstraction layers, *arbitration* in allocating resources to client requests, and *aggregation and allocation* of resources from multiple lower level resource managers. One or more of these key functions may be present at each abstraction layer of a service level manager. Thus, the composition of these functions across resource abstraction layers enables modeling of a wide array of management scenarios. We present a framework that supports these functions: it uses the service metadata and/or service performance models to map client requirements to resource capabilities, it uses business value associated with objectives in allocation decisions to arbitrate between competing requests, and it allocates resources based on previously negotiated agreements.