Ensuring Resource-Level Quality for Services on Grids

André Lage Freitas, Jean-Louis Pazat, Nikos Parlavantzas

2010 IEEE 6th World Congress on Services

Ensuring Resource-Level Quality for Services on Grids

André Lage Freitas, Jean-Louis Pazat, Nikos Parlavantzas
INRIA / IRISA – MYRIADS Team
Campus de Beaulieu – 35 042 Rennes, France
Emails: {Andre.Lage,Jean-Louis.Pazat,Nikos.Parlavantzas}@irisa.fr

***Abstract*—Service-Oriented Architectures address the development of distributed and dynamic service-based applications. Due to the dynamics of their environments, services should be self-adaptable in order to maintain agreed resource-level qualities. To support building such services, this work proposes the Self-Adaptable Service Execution Manager (SASEM), responsible for monitoring and controlling the service execution on grid resources in order to prevent SLA violations. SASEM builds on the Dynaco adaptation model and on an enhanced grid monitoring and actuation system, implemented using the XtreemOS grid operating system.**

I. Introduction

Service-Oriented Computing (SOC) [7] enables the development of enterprise applications by flexibly integrating services available across diverse networks and devices. Due to the intrinsic dynamism of their operating environment, service-based applications must be able to dynamically detect and react to unforeseen changes. Dynamic adaptation techniques are thus necessary in order to ensure that the applications are operating properly. Most current approaches [9] for dynamic adaptation in SOC focus on modifying service compositions, treating individual services as black-boxes. We focus instead on ensuring the correct operation of individual, complex services running over heterogeneous dynamic environments, such as Grids. Grid computing [5] proposes a distributed infrastructure that addresses large-scale sharing of heterogeneous and low-cost resources, aiming at supporting high performance computing and storing huge amounts of data. Thereby, grids present themselves as a worthwhile basis capable of providing a dependable and effective support for service execution.

We aim at ensuring quality-compliant service execution grids. In order to address this problem, we rely on the Dynaco [2] component adaptation model by providing self-adaptable support for service execution. As the underlying service infrastructure, we leverage the XtreemOS [11] grid platform. This paper is structured as follows. Section II discusses Grid Computing and presents XtreemOS. In Section III, we propose the design of a *Self-Adaptable Service Execution Manager* (SASEM) followed by its prototype implementation. Ultimately, we discuss related approaches and we conclude in Section IV.

II. The XtreemOS Grid Operating System

Grid computing [5] proposes a standardized way of sharing distributed computing resources in order to better take advantage of their computational capabilities. It provides an infrastructure that enables the exploitation of the computation power of low-cost and heterogeneous resources for providing high performance processing in a transparent fashion. Grid capabilities may be provided by middleware platforms, such as the Globus Toolkit (version 4) [6], or by operating systems, such as the XtreemOS grid operating system [4]. Thanks to its kernel module implementations, XtreemOS transparently addresses hardware heterogeneity and data storage. Its native support for VO management provides safe and flexible control of grid resources, while its application interface (XOSAGA) supports a simple grid usage. Furthermore, the XtreemOS Virtual Node feature supports replicating services to improve reliability.

III. A Self-Adaptable Service Execution Manager

Our goal is to ensure quality-compliant service execution, that is, for qualities that depend on the infrastructure on which an atomic service runs[1]. Resource-level qualities are usually specified in Service-Level Agreements (SLAs) with high-level terms, such as request response time and throughput. Therefore, it is necessary to firstly translate such terms to low-level resource requirements which are directly understood by the infrastructure – e.g., the number and type of machines needed to execute the service as well as their characteristics. Ultimately, self-adaptable mechanisms should be used in order to ensure the quality compliance by means of reacting to both infrastructure and quality changes.

In order to address the quality assurance for service execution, we use the grid to propose a self-adaptable support for services as depicted in Figure 1. The Service Provider uses the *Self-Adaptable Service Execution Manager* (SASEM) to avoid that infrastructure disruptions or minor changes impact on the service provision. The self-adaptable capability of SASEM is based on Dynaco (Dynamic Adaptation for Components) [2]. The Dynaco model decomposes adaptation into four distinct components: *monitor, decider, planner* and *executor*. Dynaco uses the Grid Monitoring and Actuation system to obtain information about and control the service implementation units and their underlying resources.

To demonstrate SASEM, let us suppose that SLA negotiation is initiated between a service requester and the provider (A1). The resource-level qualities are translated to resource requirements by the *Quality Translator* (A2). Then the SLA Negotiator asks the monitoring system if the such resources are available (A3) and it finally agrees on a contract (A4).

[1]We refer to *service local adaptation* ([8], p. 6).

978-0-7695-4129-7/10 $26.00 © 2010 IEEE
DOI 10.1109/SERVICES.2010.75

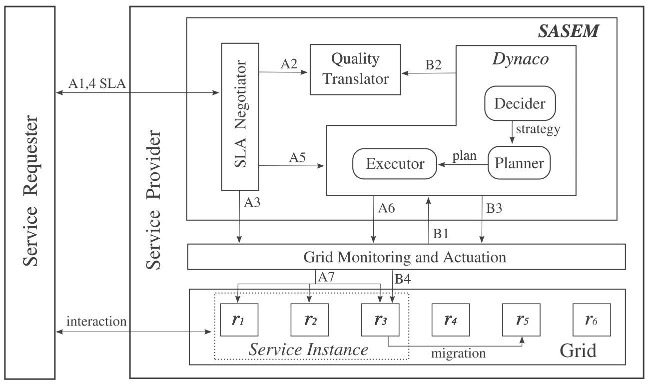

Fig. 1. The *Self-Adaptable Service Execution Manager* (SASEM) automatically manages service execution on the grid by keeping compliant to resource-level qualities.

This action is followed by the service deployment request (A5) which is treated by Dynaco by deciding to execute the service on the grid (A6). The grid then instantiates the service by using the resources *r1, r2, r3* (A7) thus the service provider and requester can finally interact with each other. Let us now suppose that an event is sent to Dynaco informing that the resource *r3* is faulty (B1). Based on the quality translation (B2), Dynaco decides to use the resource *r5* as a replacement. Through the executor, Dynaco sends the right grid commands (B3) responsible for migrating the task from *r3* to *r5* (B4).

A prototype that implements SASEM is under development. The prototype relies on Web Service (WS) technologies for exposing the service implementation in order to guarantee interoperability and to allow leveraging earlier research results from the field of SLA negotiation. Moreover, the prototype relies on the XtreemOS grid operating system as the underlying grid platform. XtreemOS provides enhanced management capabilities that greatly facilitate building the Grid Monitoring and Actuation system. In terms of monitoring, XtreemOS supports extensible monitoring metrics, which can be used to implement sensors for resource-level qualities such as request throughput. In terms of actuation, XtreemOS provides powerful job management features, such as job check-pointing, restarting and migration.

IV. Related Work and Conclusions

Several approaches tackle the problem of dealing with resource-level qualities for services, such as [1], [13], [12]. These approaches have only dealt with one or two types of specific qualities in distinct SOA scopes. In contrast, we propose to focus on the service execution scope but ensuring further resource-level qualities. In [3], a framework design is proposed that leverages WS, Autonomic Computing and Application-Oriented Network technologies to tackle issues related to the next generation of networks; however, unlike our work, the design does not define how resource-level qualities can be guaranteed. This issue is actually discussed in [10], but their approach is not suitable for larger scale and more heterogeneous resource infrastructures. Such a context would increase the complexity of their service model and consequently compromise the benefits of their approach.

This work has addressed the quality assurance of service execution. We propose a self-adaptable support for service execution (SASEM) that is able to negotiate resource-level qualities and to automatically manage the service execution. Such resource-level qualities are translated to resource requirements and finally used to manage the service execution in the underlying infrastructure. In addition, SASEM uses the Dynaco adaptation model to perform automatic service execution management on grids. Finally, we are implementing a prototype of SASEM which uses the XtreemOS grid operating system.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement 215483 (S-CUBE).

References

[1] D. Battre, M. Hovestadt, O. Kao, A. Keller, and K. Voss. Virtual Execution Environments for ensuring SLA-compliant Job Migration in Grids. In *SCC '08: Proceedings of the 2008 IEEE International Conference on Services Computing*, pages 571–572, Washington, DC, USA, 2008.

[2] J. Buisson, F. André, and J.-L. Pazat. Dynamic adaptation for Grid computing. In *EGC '05: Proceedings of The European Grid Conference*, pages 538–547, Amsterdam, June 2005.

[3] Y. Cheng, A. Leon-Garcia, and I. Foster. Towards an autonomic service management framework: A holistic vision of SOA, AON, and autonomic computing. In *IEEE Communications Magazine*, volume 46, pages 138–146. IEEE, May 2008.

[4] T. Cortes, C. Franke, Y. Jgou, T. Kielmann, D. Laforenza, B. Matthews, C. Morin, L. P. Prieto, and A. Reinefeld. XtreemOS: a Vision for a Grid Operating System. Technical report, XtreemOS Consortium, May 2008.

[5] I. Foster. What is the Grid? - A Three Point Checklist. *GRIDtoday*, 1:22–25, 2002.

[6] I. Foster. Globus Toolkit Version 4: Software for Service-Oriented Systems. *Journal of Computer Science and Technology*, 21:513–520, 2006.

[7] M. P. Papazoglou and D. Georgakopoulos. Service-Oriented Computing, Introduction. *Commun. ACM*, 46(10):24–28, 2003.

[8] S-CUBE Consortium. Basic Requirements for Self-healing Services and decision support for local adaptation. Deliverable 2.3.2, March 2009.

[9] S-CUBE Consortium. Taxonomy of Adaptation Principles and Mechanisms. Deliverable 1.2.2, May 2009.

[10] G. Tesauro, R. Das, W. E. Walsh, and J. O. Kephart. Utility Functions in Autonomic Systems. In *ICAC '04: Proceedings of the First International Conference on Autonomic Computing*, pages 70–77, Washington, DC, USA, 2004. IEEE Computer Society.

[11] XtreemOS Consortium. XtreemOS - Revised System Architecture. Deliverable 3.1.7, January 2009.

[12] C. Zhang, R. N. Chang, C.-s. Perng, E. So, C. Tang, and T. Tao. An Optimal Capacity Planning Algorithm for Provisioning Cluster-Based Failure-Resilient Composite Services. In *SCC '09: Proceedings of the 2009 IEEE International Conference on Services Computing*, pages 112–119, Washington, DC, USA, 2009. IEEE Computer Society.

[13] H. Zhang, K. Keahey, and W. Allcock. Providing Data Transfer with QoS as Agreement-Based Service. In *SCC '04: Proceedings of the 2004 IEEE International Conference on Services Computing*, pages 344–353, Washington, DC, USA, 2004. IEEE Computer Society.