Chapter 5.9
SIP Protocol for Supporting Grid Computing

Aldo Campi
University of Bologna, Italy

Franco Callegati
University of Bologna, Italy

ABSTRACT

This chapter advocates migrating sufficient functionality into the network as to allow direct support to grid computing services. This approach is realized by making use of Session Initiation Protocol (SIP) as the medium to support the signaling among grid applications and network. The authors elaborate on the advantages of this approach, which benefits directly from the session management capabilities of sip and enables application-oriented functions.

INTRODUCTION

Grid computing is migrating from traditional high performance and distributed computing to pervasive and utility computing based on heterogeneous networks and clients. The success of this new range of services is directly linked to the effectiveness of the networking infrastructure used to deliver them.

Grid computing services aim at integrating and manage resources according to application requirements and within distributed, heterogeneous, dynamic environments. On the other hand, networks and, in particular their control planes, regardless of the specific implementation, deal with topology, capacity, connectivity and routing. In general, these two worlds have very different perspectives and talk very different languages. So far the direct link between the application and the network is still missing and the applications usually rely on virtual overlay topologies that provide connectivity according to a pre-defined scheme and are built a-priori in the network.

The present typical scenario is that, as proposed for instance by the Open Grid Forum (Open Grid Forum, n.d.), applications talk end-to-end and

DOI: 10.4018/978-1-4666-0879-5.ch5.9

resource management is provided by means of web-services, an approach that has the merit to provide a standard and flexible communication (Tuecke, Czajkowski, Foster, Graham, Kesselman, Maguire, et al, 2003)(Foster, Berry, Djaoui, Grimshaw, Horn, Kishimoto, et al, 2004). This is a typical overlay approach with associated merits and drawbacks. In particular it may result in inefficiencies in the use of the network resources and in limited functionalities with respect to what achievable with a more integrated approach. Moreover limitations exist for implementations targeting large populations of users in terms of scalability, fault tolerance, performance, quality of service provisioning etc.

An alternative approach may be that of migrating into the network the intelligence needed to support the Grid computing services, making the network node capable of interacting with the network in an application oriented language. The realization of this goal requires the disintegration of numerous barriers that normally separate application services, IT resources (computing systems) and networks. This chapter discusses this scenario and proposes a possible approach to solve the problem, based on an application layer signalling protocol able to work with a broad spectrum of existing and future IP protocols and services.

Up to now the application layer services (e.g. resource discovery, reservation, etc.) have been deployed with some form of centralized mechanisms based on web services, without bothering of exploiting any knowledge of the transport services. The result is that, in most cases, the Grid signaling is segregated on a separate infrastructure while the transport network is used as a best effort transport facility to carry the application data almost independently from the signaling. The reasons for this can be found in the fact that service provisioning is an operator task for the part concerning network connectivity and is a user "private" task for the part concerning IT resources. Currently, operator and user do not communicate and service provisioning is not available "on-the-fly" at the user premises.

Some solutions to provide users with an abstract view of the network resources (resource virtualization) have been proposed already, like the User-Controlled Light-paths (UCLP) by Bill St. Arnaud (St. Arnaud, et al, 2004)(St. Arnaud, et al, 2006). In UCLP users manage the optical network resources they own through a web service-based infrastructure. UCLP is based on the Grid services concept and builds on the Jini and JavaSpaces technologies. Similarly in the framework of the ITU-T there has been effort in the direction and trying to standardize high-level triggering of network services. The most significant example is the ITU-T Next Generation Network (NGN) architecture (Cochennec, 2002)(ITU-T Recommendation Y.2001, 2004)(ITU-T Recommendation Y.2011, 2004)(Recommendation Y.2012, Functional requirements and architecture of the NGN release 1, 2006). Through the logical partitioning of functions into a Service Stratum and Transport Stratum, NGN is capable of abstracting the connectivity services over a IP-based service-oriented network architecture.

Moreover there are three large research projects focus on developing mechanisms for end-to-end coordinated reservations across network domains of geographically distributed high-end computing resources:

- **Phosphorus (EU-IST-** http://www.ist-phosphorus.org**):** Focuses on enabling on-demand end-to-end network services across multiple heterogeneous domains, treating the underlying network as a first class Grid resource, thus developing integration between application middleware and transport networks.
- **Enlightened computing (NSF-** http://www.enlightenedcomputing.org**):** Develops Grid middleware that views the network as a Grid resource at the same level as the compute and storage resourc-

Related Content

Multi-Core Supported Deep Packet Inspection
Yang Xiangand Daxin Tian (2010). *Handbook of Research on Scalable Computing Technologies (pp. 858-873).*
www.irma-international.org/chapter/multi-core-supported-deep-packet/36437

Agent-Based Modeling: A Historical Perspective and a Review of Validation and Verification Efforts
Brian L. Heathand Raymond R. Hill (2010). *Handbook of Research on Discrete Event Simulation Environments: Technologies and Applications (pp. 28-57).*
www.irma-international.org/chapter/agent-based-modeling/38256

Efficient Management of Fragmented Replica in Data Grids
Chao-Chin Wu, Lien-Fu Laiand Jia-Xian Lai (2012). *International Journal of Grid and High Performance Computing (pp. 63-70).*
www.irma-international.org/article/efficient-management-fragmented-replica-data/66357

Visualization of Large-Scale Distributed Data
Jason Leigh, Andrew Johnson, Luc Renambot, Venkatram Vishwanath, Tom Peterkaand Nicholas Schwarz (2012). *Data Intensive Distributed Computing: Challenges and Solutions for Large-scale Information Management (pp. 242-274).*
www.irma-international.org/chapter/visualization-large-scale-distributed-data/62830

A Novel Solution for Scaling Video Shot Boundary Detection Based on Hadoop
Ahmed Diband Mokhtar Sellami (2018). *International Journal of Distributed Systems and Technologies (pp. 39-52).*
www.irma-international.org/article/a-novel-solution-for-scaling-video-shot-boundary-detection-based-on-hadoop/207691