A Semantic Service Discovery Framework for Intergrid

Mahamat Issa Hassan and Azween Abdullah

Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia
(mahamat.hassan,azweenabdullah)@petronas.com.my

Abstract. Resource/service discovery is a very critical issue in intergrid (grid of grids) system, which is the recent advancement in grid technology that uses multi-middleware. In this paper we present an intergrid service discovery framework that integrates semantic technology, peer-to-peer network and intelligent agents. The framework has two main components which are service description, and service registration and discovery models. The earlier consists of a set of ontologies that are used as a data model for service description and services to accomplish the description process. The service registration is also based on ontology, where nodes of the services (service providers) are classified to some classes according to the ontology concepts, which means each class represents a concept in the ontology. Each class will have an elected head that plays the role of a registry in its class and handles the interclass communication. We further introduce intelligent agents to automate the discovery process. The framework is evaluated via simulation experiments, and the result of which confirms the effectiveness of the framework in satisfying the required RD features (interoperability, scalability, decentralization and dynamism).

Keywords: Grid, semantic technology, intelligent agent, and peer-to-peer network.

1 Introduction

Intergrid/Global grids (in some literatures is also called multi-grid [1]) are known as a grid of grids, since they are a collection of small grids that cross organizational boundaries to create very large virtual systems that can be accessed from anywhere in the world. A very basic and first step in sharing resources over intergrid is the detection of suitable resource for a given task/application which is commonly known as Resource Discovery (RD). The RD process entails description of the resource through its properties, registration/indexing of the described resource in common registry(s), and discovering the registered resources that match with resource request specifications. These steps correspond to the main components of the RD system, which are Description, Registration and Discovery (which is composed of search and selection). RD is very important as resource reservation and task scheduling are based on it. Unfortunately, intergrids are normally associated with some complexities such as resources/services and users are distributed across different locations; resources are
heterogeneous in their platforms; status of the resources is dynamic (resources can join or leave the system without any prior notice); and use of multi-middleware. These complexities pose a challenge to the development of an efficient RD system to discover the resources and services. In fact, these complexities also yield some requirements that should be fulfilled by any developed RD. These requirements include high searchability (interoperability) to retrieve the relevant and precise resources and services, and high performance (scalability, decentralization, and dynamism) to make the RD system sustainable with the scale of the intergrid.

Currently, there is a wealth of work on grid RD (e.g. Globus\(^1\), Condor\(^2\), [2], [3], and [4]) which can be classified into two classes based on the description component, which are keyword-based RD systems and semantic-based RD systems. Keyword-based system uses syntactic information and data models such as directories and special databases to describe and discover the resources and services. Unfortunately, syntactic information and data models are not efficient in describing resources at intergrid level. This is because resources and services are initially described by using multi information services that belong to different grid middlewares. As a matter of fact, much of the efforts in keyword-based RD systems have been focused on achieving the high performance requirement; staring from introducing centralized registration models such as Globus MDS-1\(^3\) [6], R-GMA\(^3\) [7] and Hawkeye [8]; then followed by hierarchical registration models [9], [10] and [11], and lastly peer-to-peer (P2P) registration models [12], [13], [4] and [14]. Keyword-based RD systems that are based on P2P registration models have achieved high performance compared to the centralized and hierarchical models, but we cannot go far as to say that they have achieved full scalability. Moreover, their use of syntactic description, especially at the intergrid level, prevents them from fulfilling the high searchability requirement. Semantic-based RD systems, on the other hand, use semantic information and data models (ontology and ontology languages)\(^1\) to describe and discover the resources and services. Although, there is a considerable amount of work on semantic-based RD systems (e.g. [16], [17]), most of the existing approaches fail to achieve high searchability. This is due to the lack of a proper use of semantic description mechanism as the semantic technology is initially imported from the semantic web [18]. Actually, we have argued in an earlier study [19] that the main obstacle that leads to the continuous existence of this issue is the ad hoc research nature of these semantic-based RD studies (different research communities doing the same thing by different ways).

In this paper, we introduce a new intergrid RD framework that can overcome the shortcoming of the current studies and meeting the above mentioned requirements. The framework contains two main components which are service description, and service registration and discovery models. The earlier consists of a set of ontologies and services. Ontologies are used as a data model for service description, whereas the services are to accomplish the description process, we detail that in section 2.

The service registration is also based on ontology, where nodes of the services (service providers) are classified to some classes according to the ontology concepts, which means each class represents a concept in the ontology. Each class has an

\(^1\)http://www.globus.org/.

\(^2\)http://www.cs.wisc.edu/condor/.

\(^3\)Relational Grid Monitoring Architecture: http://www.r-gma.org/index.html