Research on Layered Resource Discovery Model in Grid

Luo Zhong, Bo Zhu, Li Yang, Huazhu Song
Department of Computer Science and Technology Wuhan University of Technology Wuhan, 430070, China
zhubo0128@yahoo.com.cn

Abstract

Grid aims at sharing all kinds of geographically distributed and heterogeneous resources. The key to resource organization is how to accomplish the resource discovery efficiently. In this paper, a Layered Resource Discovery Model (LRDM) was proposed firstly according to different resource type in grid. Then the components of LRDM were analyzed and designed in detail with resource discovery module in Globus Tookit 4, mainly including resource description, resource information storage, resource information organization, resource request processing, and resource selection. Finally, the model was implemented and proved to have the characteristics high discovery efficiency and scalability because of combining the advantages of both centralized and distributed mechanisms.

Keywords: Layered resource discovery model; grid; virtual organization(VO); GT4

1. Introduction

Grid is trying to achieve full connectivity of all resources on the Internet, Therefore How to discover variety of dynamic, heterogeneous and distributed resources in the grid and provide transparent and consistent access interface to users becomes an important area of grid study. With the development of grid technology, resource discovery plays an increasingly important role. Resource discovery is the link between resource owner and resource requestor, and provide support for task scheduling. The effect of resource discovery directly determines the efficiency and friendly of a grid.

In terms of the grid resource features and the design goals of resource discovery mechanisms [1], a resource discovery mechanism which is distributed, able to adapt to the dynamic changes of resources, good performance positioning and scalable is required. It can find resource information quickly and accurately and improve the grid performance greatly. There are different advantages and disadvantages in variety single-mode, and now having a variety of domestic and international research, and has made some progress. In this paper, based on the research and analysis of the existing grid architecture [2,3]and resource discovery mechanisms and methods [4,5], a Layered Resource Discovery Model was proposed,
the components of the model were described and analyzed specifically with GT4 and the experimental example of the model was given.

2. Framework of Model

Grid provides shared platform for physically distributed resources. Virtual Organization (VO) is composed of multiple grid nodes or other VOs. VO collects information of each node in the organization, represents and provides resource information with unified perspective. The same VO members adopt a unified authentication mechanism and share all kinds of resources in the organization equally. Grid nodes can register to different VO and join or leave VO dynamically. In LRDM, the VO is organized based on the type of grid resources. Metadata is used to classify the resources from function or service in order to improve the efficiency of resource discovery. The structure of LRDM is shown in Fig. 1, the grid resources is classified into three parts in accordance with the logic function: courseware, multimedia, text.

![Figure 1 Structure of LRDM](image)

- Physical Resource Layer
  It includes a variety of heterogeneous resources information of ordinary nodes, which can be either Resource Provider (RP) or Resource Requester (RR). When they are RP, firstly, the underlying resource information is extracted, converted and managed. Then the corresponding resource description documents are generated based on resource metadata standard. And last their resource information is registered to the desired manager node in the VO, and resources updates is send to the manager node regularly. When they are RR, the query and search service can be provided directly.

- VO Layer
  As shown in Fig. 1, a manager node (MN) is configured to manage the VO uniformly in every VO, such as VO1, VO2, and VO3. Multiple manager nodes can be also configured as required. MN is responsible for accepting receiving the registration, updating and retrieval of ordinary nodes on physical resource layer, notifying the MN in other VOs of its own resources updates at regular time, accepting other MN’s request and transmitting if necessary, storing and maintaining homogeneous resource information in its VO and related resource information in other VOs on VO layer.

  In every VO, a centralized management mode of resource and service is adopted, so that the information can be managed and transmitted within the VO and not be transferred to the top level. Among the VO, distributed peer-to-peer networking is adopted. Peer-to-peer connections between nodes can overcome the deficiencies associated with centralized management mode.
3. LRDM

3.1 Experimental Environment

Globus Toolkit 4 [6] is as a service set based on open structure and open source code, called “GT4” for short. It is a basically grid development platform based on Web Services Resource Framework, provides API to build the stateful Web Services and its goal is to establish a distributed heterogeneous computing environment. MDS [7] is a module to provide information services in GT4. It provides a set of tools and application programming interface for discovering, publishing and accessing a variety of grid resource information.

Five PCs are simulated to represent five grid resource node (one node represents a class of resource). Every PC’s basic configuration is the Pentium4, 2.8GHz CPU, 512MB RAM, Redhat Linux Advanced Server 4 operating system, some components are installed such as JavaBean, Globus Toolkit 4.2.1 middleware tools, MDS4 and so on.

3.2 Model components

The components of LRDM, which mainly includes resource description, resource information storage, resource information organization, resource request processing and resource selection, are analyzed and designed in detailed in combination with MDS4 module of GT4.

1) Resource Description

In order to achieve the resource sharing goal, firstly the attributes of all distributed and heterogeneous resources are described and packaged. Thus the specific realization will be concealed, foreign interface will be presented uniformly and On-demand invoking services will be satisfied. Resource Specification Language (RSL) [8] is adopted to describe resources in LRDM. The resource description includes two parts, class name and specification, which determine the user needs together. The RSL resource description mechanism is shown in TABLE 1.

Table 1 RSL resource description mechanism

| <resource_description>::=<class_name>[“?”<specification>] |
| <class_name>::=<string> |
| <specification>::=(<attribute><bop><value>){<cop>(<attribute><bop> <value>)} |
| <attribute>::=<string> |
| <value>::=<string>|<resource_description> |
| <bop>::=“>”|“>”|“<”|“<”|“=”|“!”|“!”|include|exclude|satisfy |
| <cop>::=AND|OR|NOT |

Class name which is a string is defined to specify the request resource type. It can be a name which is user-friendly, easy to remember and semantic, also can be a number for performance sake which is easy for machines to process. Either way, class name is necessary. Users can also further describe their demands in specification. Specification which is a logical expression is composed of pairs of <attribute, value> and connectors. This field is optional, for example, the field is empty means that users only require a certain class of resources and do not care the specific resources attributes. TABLE 2 is a resource specification example of multimedia resources.
As shown in TABLE II, MultimediaResource class has the following attributes: size, time, Audio, Video and so on. The term Audio is an object of the class AUDIO which attributes are samplerates, bitdepth, etc. The term Video is an object of the class VIDEO which attributes are datarates, samplesize, etc. The attribute value is an object of a class and the operator "satisfy" expresses that the attributes should satisfy the following constraints. After simple supplement, any quantity of resources or any of the various types of resources can also be described like this.

2) Resource Information Storage

In LDRM, as the unified interface of grid information access and storage, the Lightweight Directory Access Protocol (LDAP) [9] stores the grid nodes resource information. In the LDAP directory information is stored in a tree structure, commonly referred to as Directory Information Tree (DIT). DIT consists of many of entry which represents the information in the LDAP. Each entry has a unique identity ‘dn’ (distinguish name) and contains zero or more pairs of <attribute, value> to specific the resources attributes. A multi-layered DIT structure is adopted to store the resource information in this paper, the benefits are to discover resource information easily and provide the resources discovery ways efficiently.

3) Resource Information Organization

Resource information should be organized effectively based on the LDAP definition and description of resources information. Resource information organization is offline and has nothing to do with the execution of the request. It can be seen as a pretreatment and can enhance the searching performance. According to the classificatory features in LRDM model, hierarchical structure is adopted to organize the resource information and is easy to discover and add resource information. The two-layered structure combined with P2P and hierarchy is shown as Fig. 2.

The two-layered structure combined with P2P and hierarchy compromises the central and distributed merits, manages resources registration and discovery services by using information nodes with good ability, and decreases the information node scale. The model becomes a structural and ordered resource information space in logical based on classification, enhances forwarded objectivity of resource discovery request among MN, reduces network traffic and improves the performance of resource discovery.

4) Resource Request Processing

The whole resource request processing is divided into two parts: local and global. Local processing refers to searching is made in VO, local policies are implemented and aggregate requests are processed. Global processing refers to the requests are transferred among VOs. In LRDM, when a user or an application on a node needs to search the required resource information, the procedure of resources discovery is as follows.

- Local Processing
Firstly, the required resources or services are searched inside node. If the results are found and the requirements for application are met, the appropriate resources or services information will be returned and provided immediately; else the request will be transferred to MN in VO to search. The algorithm description is shown as Fig. 3.

```
When a RR receives a request from itself
if(no matched resources)
{
    send the request to its MN;
}
else
{
    return the results to RR;

    finish;
}
```

Figure 3 Local processing

- Global Processing

Firstly, the VO is formed according to the registered resources type and the required resources should be matched with processing experience, if successfully the results are returned. Otherwise, the required resource information will be searched inter-VOs through MN in every VO. Suitable algorithm for message diffusion is adopted until the matched information is found. Distributed query approach is used on VO layer, Resource request message spread over the constructed overlay network and the reliable discovery comes turn. Algorithm description is shown as in Fig. 4.

```
When an MN receives a request from a RR or other MN
if(no matched resources&&not reach the max hop)
{
    send the request to other MN;
}
else
{
    return the results to RR;

    finish;
}
```

Figure 4 Global processing

5) Resource Selection

Resource selection indicates that the actually used resources are selected from a set of resources which meet the user requirements as proper standard, such as the highest system performance, or minimum user fees and so on. During the resource discovery process, the system may make a choice according to their own configuration, and it should be some simple, common standards. However, during the resource selection process, it should be more complex, professional and specific. Resource can be selected during resource discovery, that is eligible resource is returned selectively; Resource can be selected closing to client, that is all eligible resources are returned and then someone is selected. All the selection criteria and strategies are independently implemented and deployed, co-existence among the grid system, and are used in combination.
3.3 Model Features

LRDM in grid has the following characteristics:

- The combination of centralized and distribution. The model has advantages of centralization and distribution. Compared with fully central mode, it has better scalability, scale and reliability. Compared with fully distributed mode, it has a relatively small system size and efficient resource discovery service.

- The grid which is divided into VOs can improve security, performance and scalability of the system. The MN in each VO would not be the bottleneck of the grid system compared to central searching. Other VOs would not be affected because of a single VO which is breakdown or under attack. The resources management is convenient and existing resources and services can be made rational use.

- High resource discovery efficiency. The VOs are formed according to different types of resources. During the resources searching, the respective VO space can be targeted directly from the resources types. The benefits are the search scope is reduced, a more accurate result set can be gotten, and the query time is reduced.

4. Model examples

Having configured the MDS4 services of GT4, resource information is described, stored, organized, the MDS searching is realized by tools of searching GRIS and GIIS service which is provided by Java Commodity Grid (COG) kit. The classes are in the org.globus.mds package, give supports to connect to the MDS server, search, and print results and disconnect the server, and provide an intermediate application layer which is adapt to different LDAP client libraries, such as JNDI, Netscape SDK and the Microsoft SDK.

- The main function core codes of MDS searching are shown in Fig. 5. The local and global resource discovery is achieved by use of GRIS and GIIS respectively. The code fragments are shown in Fig. 6. The value of ‘host’ and ‘baseDN’ determines whether the GRIS is running or GIIS.

```java
MDSService gridInfoSearch = new MDSService();
String filter = "(&(objectclass=MdsOs)(Mds-Os-name=Linux))";
System.out.println("Your search string is: " + filter);
gridInfoSearch.HostInfoSearch(filter);
```

Figure 5 Main function of MDS searching

```java
String baseDN = "Mds-Vo-name=globus, o=grid";
String host = "linux.whut.zz";

// searching for GIIS
// &OBJECTCLASS=* represents that filter wants to get all the data
MDSService mds = new MDSService(host, "8866", baseDN);
String xml = mds.HostInfoSearch("(objectclass=*)");
System.out.println("GIIS output
" + xml);

// searching for GRIS
MDSService mds1 = new MDSService();
mds1.HostInfoSearch("(objectclass=*")");
System.out.println("GRIS output
" + xml);
```

Figure 6 GRIS and GIIS searching fragments
Part of the output of MDS searching results which are transformed from multi-valued hash table into XML string is shown in Fig. 7.

![XML string](image)

Figure 7 MDS searching results

Meanwhile, resources information can be searched by use of ‘grid-info-search’ command in Globus. Fig. 8 shows part of results searching a local node, ‘-h’ represents the searching MDS host, ‘-p’ represents the port number, and ‘Mds-Vo-name = local, o = grid’ represents searching on the GRIS. It is the corresponding code that the Globus graphical searching results in Fig. 8 and the output of MDS searching results which are transformed into XML string in Fig. 7. ‘Fs’ represents the file system, ‘Memory-Ram’ is the memory, and ‘Memory-Vm’ represents virtual memory. The contents which are not shown in Fig. 8 include more detailed information about all attributes of the node, such as the resource operating system, CPU information, file system information and so on.

![Graphical searching results fragments](image)

Figure 8 Graphical searching results fragments
5. Summary

Based on the centralized and distributed resource discovery mechanism, a layered resource discovery model in grid was proposed. The model which is composed of physical resources layer and virtual organization layer overcome the shortcomings of the bottleneck in the centralized mechanism using a central node; eliminated the irregularity of resource information space, unstructuredness and blindness of resource discovery in distributed mechanism; improved the efficiency and accuracy of resource discovery.

Finally, the key components of resource discovery were designed and analyzed which include resource description, resource information storage, resource information organization, and resource request processing and resource selection. A LDRM example was shown by using GT4. The LRDM describes a simple and special situation. Actually, the gird is astronomical; resources are diverse and heterogeneous; resource organization and the network of VOs are complicated and various, therefore, the research will go on.

References