Information Service Interoperation among Heterogeneous Grids

Yongcai Tao
School of Information Engineering, Zhengzhou University,
Zhengzhou Henan 450001, China
ieyctao@zzu.edu.cn

Song Wu
Services Computing Technology and System Lab, Cluster and Grid Computing Lab, Huazhong University of Science and Technology, Wuhan, 430074, China

Abstract—Information service is one of the key components of the grid system, and information service interoperation is the first step towards grid interoperation. Existing grid interoperability projects mainly focus on immediate bridge mechanisms, which is intricate and of poor scalability. A uniform information service interoperation framework (UISIF) is proposed in the paper, which adopts virtual layer and plug-in technology. There are two well-established grid systems in China, ChinaGrid and China National Grid, and they are built with two different grid middleware, CGSP and VEGA. With UISIF, these two grid systems make information service interoperation a reality. The experiments also show that the information service interoperation has low information querying latency and high accuracy; Moreover, UISIF has good scalability.

Index Terms—Grid, Interoperation, Information Service, plug-in, CGSP, VEGA

I. INTRODUCTION

A wide variety of grid middlewares and grid systems have been developed in numerous research projects all over the world, such as Globus toolkits, CGSP, UNICORE, VEGA, etc [1, 2]. Existing grid systems are mostly for particular application and there is no practical standard to follow, which makes interoperation among these grids very complex. In order to achieve interoperability between heterogeneous grids, some issues must be addressed such as security, information service, job scheduling and data management. Information service is one of the key components of the whole grid system, which connects all system modules and is responsible for providing a global view of all resources in grid [3]. So, implementation of information service interoperation is the first step towards grids interoperation.

Though grid resources are usually described according to the XML-based schema or information model, information service of different grids are not easily interoperable, which mainly results from two aspects. In one hand, different grid systems adopt specific XML-based schema or information model. In the other hand, different grid systems exploit different mechanisms to discover, organize and store resources information.

Existing grid interoperation projects mainly focus on immediate bridge mechanism, which modifies the original grid systems. This approach would become complicated if we want to interoperate a number of heterogeneous grids. Moreover, the interoperation is not bidirectional. In the paper, a uniform information service interoperation framework (UISIF) is proposed, which adopts virtual layer and plug-in technology. There are two well-established grid systems in China, ChinaGrid and China National Grid, and they are built with two different grid middleware, CGSP and VEGA [3, 4]. With UISIF, these two grid systems make information service interoperation a reality without changing the codes of current grid systems. UISIF translates and caches the information retrieved from grid platforms periodically, which guarantees low information querying latency and high accuracy.

The rest of paper is as follows: Section II reviews some related work. Section III introduces the design and architecture of UISIF. The implementation of key components of UISIF is proposed in section . A prototype is presented in section and experiments and evaluation are conducted in section VI. Finally, conclusion and future work are described in section VII.

II. RELATED WORKS

Recently, a number of efforts have been put to grid interoperability. GRIP (Grid Resource Interoperability Project) implements resources interoperation between UNICORE and Globus by adopting resource broker to some extent [4]. But the interoperation is not bidirectional, that is to say, only through UNICORE client, can users access resources information of Globus system. GPE (Grid Programming Environment) exploits GridBean technology to integrate heterogeneous grid systems [5]. Users can access resources information of multiple grid systems via GPE portal or client access interface which have integrated UNICORE and Globus. It
has good scalability but breaks the independence of underlying grid systems.

Aforementioned interoperability mechanisms belong to immediate bridge mechanisms which is direct, simple and effective when the number of platforms remains very small, but when the number of grid platforms becomes large, this approach would become inflexible and intricate since the number of bridges increase exponentially in \( O(n^2) \), where \( n \) is the number of different platforms to be interoperated.

### III. UISIF

#### A Architecture

UISIF designs a virtual information service layer by adopting plug-in technology. As shown in Figure 1, UISIF consists of three layers: underlying grid information service layer, plug-in service layer and virtual information service layer.

Underlying grid information service layer is referred to the various information service centers belonging to different heterogeneous grid systems, which usually have their own uniqueness. Plug-in module periodically translates the information data retrieved from underlying grid information centers to a uniform information format. In virtual information service layer, consistent information repository (CIR) and service agent modules are responsible for caching and managing service information data. Service agent modules provide common information services such as service register and service query.

The remarkable characteristics of UISIF is that firstly system scales easily for the number of plug-ins only grows linearly in \( O(n) \), where \( n \) is the number of grid platforms to be interoperated. That is to say, adding a new platform only requires adding one plug-in module. Secondly, the CIR in virtual layer can cache the information of each heterogeneous grid platforms. Finally, UISIF keep the independence of original grid platforms without any change to them.

#### B Scalability Consideration

UISIF adopts hybrid topology as shown in Figure 2, which integrates the advantage of tree mode and flat mode. Intuitionallly, tree mode is easy to scale but has poor query response. Whereas flat mode has good query response but not easily scales.

In Figure 2, an autonomous UISIF system is called a domain which integrates some heterogeneous grid information services and can work independently. Domains can be interconnected in tree mode, namely, a domain can join another domain as its child. In addition, each domain at the same hierarchy keeps a same copy of the information of its sibling domains. Here, node represents underlying heterogeneous grid information centers, which are integrated in domain through corresponding plug-in.

In UISIF, information service maintains two kinds of information: One is domain information, which mainly includes the topology information of domains, indicating how the domains are connected, and the configuration information of each domain. The other is node information, which provides a global view of all resources in nodes.

### IV COMPONENTS

The specific design of UISIF is shown in Figure 3.

#### A Plug-in

Plug-in is one of the key components of UISIF, which serves as a bridge between CIR and underlying grid information service center. The main function of plug-in is to retrieve the information data from underlying grid information centers and translate the data to consistent data representation such as GLUE. In addition, plug-in is responsible for converting the service access command between different information service management mechanisms.
The introduction of plug-in to UISIF improves the scalability and extensibility of system. When underlying grid information centers register with UISIF, corresponding plug-in will be deployed and invoked. In order to provide real time information of various grid resources for end users and other grid systems, plug-in module periodically probes underlying grid information centers and updates the information stored in CIR.

B Consistent Information Repository

CIR is hides the heterogeneity of underlying grid resource information and provides consistent information description. CIR acts as a cache to address the long latency problems resulted from that almost every message is translated twice by plug-in modules. Currently the resource descriptions used by grid middlewares are different from each other and there is no mature standard available that all systems would adopt. In order to unify various types of resource description mechanisms and be compatible with other description models, GLUE, widely accepted as resource description model internationally, is adopted in CIR. GLUE can describe information about site, service, cluster and storage resources, etc.

CIR keeps two types of information: domain information service (DIS) and node register service (NRS). DIS is used to record the whole domain topology information including the access point of information center of each domain, the father-son relation of domains, the number of nodes in each domain, etc. NRS is to maintain the service information of inter-domain and sibling domains including nodes hardware information, deployed services information, relative information of sibling domain, etc.

C Service Agent

Service agent mainly takes charge of encapsulating the consistent information data and providing a uniform access point for various grid systems and end-users. Service agent is implemented using WSRF technology and collaborates with plug-ins to realize the translation of access commands of various grid information services.

V. Prototype

Our prototype is based on Globus toolkit [2]. CGSP and VEGA are two representatives of heterogeneous grid systems [6, 7]. They vary greatly in architecture and management mechanism, especially in information service management.

A CGSP

The current version, CGSP 2.0, is based on the core of Globus Toolkit 4.0, and is WSRF and OGSA compatible [6]. CGSP exploits hierarchical architecture to manage resource information. In CGSP, domain represents an autonomous grid system which can run independently and node provides resources. There is just one information center in each domain, and information centers of different domains can be integrated to provide a global view of a collection of domains, each of which is a collection of nodes. And each node is a collection of resources. In addition, CGSP utilizes XML-based document and XPath language to keep and query the resources and services information.

B VEGA

VEGA is mainly based on standard Web Service technology, which views a grid as a distributed computer system. VEGA adopts three-layer architecture to manage grid resources. Based on the routing concept of router on Internet, VEGA adopts grid router and agora to manage resources which can be categorized into physical resource, virtual resource and efficient resource. User can access them through physical address, virtual address and effective address [7, 8]. In addition, VEGA adopts Mysql database and SQL to store and query the information.

![Figure 4. Information integrating between CGSP and VEGA](image)

![Figure 5. Information mapping between CGSP and VEGA](image)

C Implementation

The information interoperation between CGSP and VEGA comprises of two phases: information integrating and information mapping. Figure 4 shows the process of information integrating, the query module retrieves the information from CGSP and VEGA periodically; then the translator module translates the information to a uniform format described by GLUE schema; finally through subscription and publish mechanism, the resource information of CGSP and VEGA are stored in CIR. Figure 5 shows the process of information mapping, through the plug-in and virtual information service layer, resources are mapped between CGSP and VEGA.

VI. Experiments and Evaluation
The experimental environment comprises eight grid sites which are deployed VEGA or CGSP grid systems, and five UISIF domain sites which are based on clusters. The configurations are shown in Table I.

<table>
<thead>
<tr>
<th>TABLE I. CONFIGURATIONS OF UISIF DOMAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Root Domain</td>
</tr>
<tr>
<td>Domain A</td>
</tr>
<tr>
<td>Domain B</td>
</tr>
<tr>
<td>Sub Domain 1</td>
</tr>
<tr>
<td>Sub Domain 2</td>
</tr>
</tbody>
</table>

A. Accuracy

As shown in Table II, we deploy 10 GRS applications (Laplace and Image services) in each CGSP platform. Because VEGA system doesn’t support GRS, there are 0 values in Table II. In addition, 10 types of Web Service applications are deployed in CGSP and VEGA platforms respectively. Physical resources in Table II denote the hardware resources which the grid platforms provide such as CPU frequency, memory capacity, storage, etc. Then we query service information in terms of some key words (e.g., laplace, iterative, image, CPUFrequency, MemoryCapacity) in three ways: through UISIF, through CGSP and VEGA. The experimental result is shown in Table II, from which we can conclude that UISIF can achieve desirable information query accuracy, sometime even higher than CGSP and VEGA.

<table>
<thead>
<tr>
<th>TABLE II. INFORMATION QUERYING ACCURACY COMPARISON OF CGSP, VEGA AND UISIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanism Services</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Laplace</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Web Service</td>
</tr>
<tr>
<td>Physical Resources</td>
</tr>
</tbody>
</table>

B. Latency

We evaluate the information querying latency of UISIF in terms of the metric of information query response time (QRT). QRT refers to the average amount of time that elapsed from when a client sends a request for query until it receives the query results. We adopt the same experimental setting as used in above section. Similarly, we query service information using some key words through UISIF, CGSP and VEGA respectively for 30 times with a waiting period of one minute between receiving a request response and issuing the next query.

The results are shown in Table III. Note that there is not big difference in information query response among CGSP, VEGA and UISIF. This is because that by adopting caching technology, UISIF can eliminate the latency incurred by translating the information description between heterogeneous grids and UISIF. UISIF translates and caches the service information of various grid platforms and periodically updates them, which make user finish the query in UISIF instead of in underlying grid platforms.

<table>
<thead>
<tr>
<th>TABLE III. INFORMATION QUERYING RESPONSE COMPARISON OF CGSP, VEGA AND UISIF (SECOND)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanism Services</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Laplace</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Web Service</td>
</tr>
<tr>
<td>Physical Resources</td>
</tr>
</tbody>
</table>

VII. Conclusion and Future Work

In this paper, the challenges confronted with the interoperability of heterogeneous grid information services are discussed. UISIF is proposed and analyzed in the paper. It adopts virtual layer and plug-in technology. With UISIF, two well-established grid systems (CGSP and VEGA) make information service interoperation a reality. The experiment also shows that the information service interoperation has low information querying latency and high accuracy. In future work, ontology based representation will be considered and applied into UISIF to improve the interoperability.

REFERENCES