Study the Model of Information Resource Classified Register and Discovery based on Hierarchy in Grid

Mingyong Li
College of Computer and Information Science, Chongqing Normal University, Chongqing, China
Email: limingyong@cqnu.edu.cn

Yan Ma
College of Computer and Information Science, Chongqing Normal University, Chongqing, China
Email: mayan@cqnu.edu.cn

Yuanyuan Liang
College of Physics and Electronic Engineering, Chongqing Normal University, Chongqing, China
Email: iangyuanlucky@126.com

***Abstract*—With the development of information and digital technology, various information resources will be more, and growth exponentially. However, currently information resources construction generally has some problems, the main problem is that the effective sharing of resources is not shaped. Registration and discovery of information resources is the core and basis of resources effective sharing. Only on the basis of resources registration and discovery, other applications can be realized. This paper firstly proposes an improved metadata description of information resource. Based on the existing discovery mechanisms, present the model of information resource classified register and discovery based on hierarchy in grid, classified register center is a key of this model, we give a full account of classification method of classified registry center, and gives the algorithm. We describe the process of registration and discovery of information resources, and gives the algorithm, and tested this algorithm using the GridSim simulation , the test results show that: total simulation time of resource classification registration based on hierarchy the paper proposed is much lower than the GridSim built-in scheduling algorithm (Optimis Cost and Time) in the same case, and with Gridlets increases, the time cost of algorithm the paper proposed is growing gently, Compared to the GridSim built-in scheduling algorithm (Optimis Cost and Time),it has more advantages of time cost .**

***Index Terms*—grid, information resource, resource register, resource discovery**

I. INTRODUCTION

With the development of informational and digital resource, a variety of information resources will be more and more, and growing exponentially. Not only in the number of information resources increases, but resource types will be more abundant, so resources sharing has become increasingly difficult, the study of information resources discovery is not a moment to lose and significant. But the construction of current information resources has the following questions: little quality resource; storage dispersed and lack of extensive cooperation; there is a large number of heterogeneous network structures; a wide range of information resources can't be shared; dynamic adaptability is poor and so on.

With the emergence of grid technology [1, 2], resources can be fully shared, because the nature of the grid is the sharing of resources. The characteristics of grid have been widely used in the educational field: sharing, dynamic, heterogeneous, supercomputing performance, strong service support. Grid technology innovate the disadvantages of traditional resource sharing.

Grid resource register and discovery is one of the key of the grid system, significantly affect the use and efficiency of grid resources. Current mature resource discovery mechanisms including: Centralized resource discovery mechanism [3], such as MDS [4] of Globus projects; resource discovery mechanism based routing forwarding [5, 6]; resource discovery mechanism based on P2P [7]; resource discovery mechanism based on Small world [8]; grid resource discovery mechanism based on Mobile Agent [9], etc. The resource discovery mechanisms have their advantages and disadvantages, have their application environment and field, but there is not mature and improve resource discovery model in the field of information resources.

In grid environment, we need the resource discovery mechanism that distributed resource dynamics, easy expansion and good performance. Can be considered a good resource discovery mechanisms need to have the following characteristics [10]: less expensive to achieve satisfactory position and have scalability; able to adapt to

Manuscript received July 28, 2011; revised August 1, 2011; accepted August 1, 2011.

This work is sponsored by Chongqing Education Committee (NO: KJ090809).

Corresponding author: Mingyong Li

doi:10.4304/jsw.7.7.1554-1561

dynamic changes of grid resources; can solve the load balancing in resource discovery process; can support a variety of grid resources and applications.

The traditional bottom-up query methods [11] and top-down query methods have some drawbacks, in order to solve register and search of wide range grid resources, the paper proposed the model of information resource classified register and discovery based on hierarchy in grid.

II. MODEL FRAMEWORK

The definition of information resource register [12]: information resources register in grid is that the existing information resources in network are packaged as resource service, and registered to the resource service Classified Register Center in the grid system, so any node in the grid system can access to the network information resources. Resources register includes three aspects: the register of information resources, the update of information resources register information, the unregister of information resources register information.

Model structure of classified register and discovery of information resource based on hierarchy in grid was shown in Figure 1. In order to illustrate the problems conveniently, we first explain following terms:

Resource Node: resource node can be both resource providers and resource requester.

Virtual Organization (VO): A number of resource nodes that have adjacent location or similar content form a whole, and become a Virtual Organization (VO) [13]. Divide Virtual Organizations is to share and query resource in a small areas.

Super Node (SN): In order to facilitate management and inquiry resources information in VO, we select a high-performance resource node as Super Node in VO, Super Node using Globus MDS to organize and manage resources within VO, and provide local register and resources discovery services.

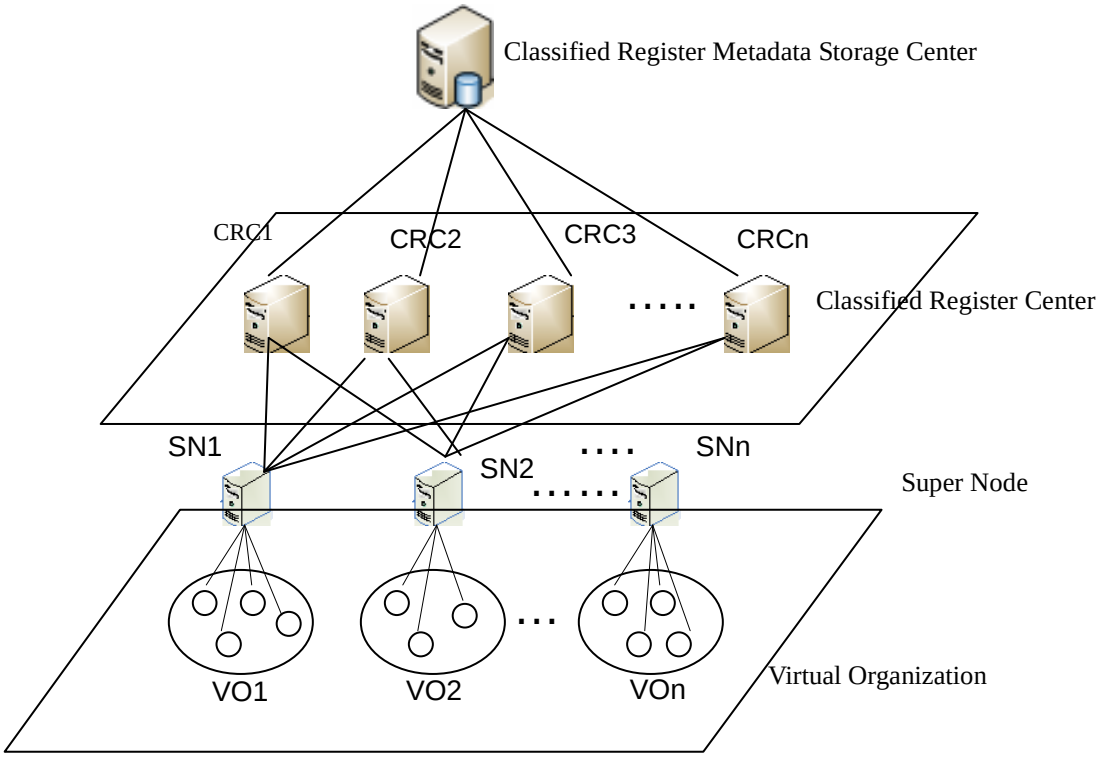

Figure 1 the model of information resource classified register and discovery based on hierarchy in grid

Classified Register Center: Each Classified Register Center in this system is used to register resources service metadata that have same nature, types and properties.

Classified Register Metadata [14] Storage Center: It Classified Register Center used to store Classified Register Center information; we can see it as a Classified Register Center of Classified Register Center object. The information stored by Classified Register Center Metadata Storage can be used to locate the different types of Classified Register Center.

III. THE DESCRIPTION AND ACHIEVEMENT OF CLASSIFIED REGISTER CENTER

It is a good way to classification when resource register, resource classification register is conducive to the rapid discovery of resources [15]. Classified Register Center's role is to realize resource classification register. can also Classification discovery when query resource.

Classified Register Center metadata attributes shown in Table 1. Classified Register Center usually has many, Classified Register Center number is determined according to the size of the system and its use classification. Specific areas of different classification methods. For educational resources, and better classification is based on the different subjects. For a subject, the type and form of its resources are limited, it is no problem that a Classified Register Center of resources

dealing with a discipline service information. Table 1 has seven fields, "Category", "Format" and "Address" is the most important, "Category" category defines Classified Register Center property, resources registration request and the user's query through match "Category" to identify Classified Register Center in Classified Register Metadata Storage Center. "Format" defines media formats, it is very useful media format that requires a specific user. "Address" is a Classified Register Center's address, only through the "Address" can find Classified Register Center. We use an example to illustrate how to use metadata to describe Classified Register Center. We describe a Classified Register Center that used to register high school sophomore physical resources, all resources are animated or multiple media formats, Chinese, all information is free and service resources open to all people. Metadata is described as the program list 1:

the program list 1
< Registers>
……
<Register>
<Category> Physics, Senior Two </Category>
< Format > Animation, multi-media </Format >
< Language > Chinese </ Language >
< Address > http://127.0.0.1< /Address >
< Cost > free </ Cost >
< Rights > For All </ Rights >
< Description > This is a classification registry that used to registe resources of Senior Two。
</ Description >
</Register>
……
< /Registers>

According to metadata description of Classified Register Center, we can accurately identify a particular resource corresponding Classified Register Center position, after positioning the corresponding Classified Register Center, you can register this resource. If we query resource, it can accurately locate the resources required by Classified Register Center. After Classified Register Center generates its own metadata description, send it to the Classified Register Metadata Storage Center, Classified Register Metadata Storage Center will be stored in its own table of contents, Resource registration and discovery to locate the appropriate category of Classified Register Center by Classified Register Metadata Storage Center, ultimately, to be successful registered and discovery of resources.

Table 1 Classified Register Center metadata attributions

No.	Name	Description	Whether mandatory need
1	Category(Category1< subject >, Category2< object of application >, Category3, Category4)	Category, up to four Classifications, different levels of Classification keywords separated by commas, described Classified Register Center category. The previous two words is fixed subject, object.	yes
2	Format(Format1, Format2, Format3)	Media formats, up to three Classifications, different levels of Classification keywords separated by commas, described media formats of Classified Register Center.	yes
3	Language	File language	yes
4	Address	address of Classified Register Center	yes
5	Cost	whether the Classified Register Center charge	no
6	Rights	permissions of Classified Register Center	no
7	Description	the basic information description of Classified Register Center	no

IV. INFORMATION RESOURCE REGISTER

A . Register process of information resources

a. Resource nodes first need to install the client program; users fill in the resource metadata description attributions in the client program interface, and generate the metadata description of resources.

b. Clients registered to the local SN by using generated resource metadata description.

c. The client sends resource metadata description that has generated to Classified Register Metadata Storage Center . Match a specific Classified Register Center according to submitted information resource metadata "Classified" field property.

d.Get "Address" field from the matched Classified Register Center metadata information, Classified Register Center's address is returned to the resource register node.

e.According to the received Classified Register Center address, client launch requests to this Classified Register Center, if Classified Register Center is running correctly, register their own resources. For a resource node, it may be many resources register at the same time and is registered to a different Classified Register Center.

f.The results of registered success (or failure) are returned to resource node

B . Algorithm description of information resources register

Before introducing the algorithm of information resources register and discovery, explain some basic classes and functions to facilitate description of the algorithm.

CenterEnrollCenterDB: Classified Register Metadata Storage Center

SourceInfoDB: Resource Information Database of Classified Register Center

SearchEnrollCenter():search Classified Register Center

SearchSource():search resources

RegisterSource():register resources

UpdateSource():update resources

Algorithm description of information resources register was shown in Table 2.

V. INFORMATION RESOURCE DISCOVERY

A . The process of information resource discovery

After information resources in the resource grid classified register based on hierarchy, resource discovery becomes quite simple and convenient. Resource discovery process in grid is as follows:

Table 2. ALGORITHM DESCRIPTION OF INFORMATION RESOURCES REGISTER

```
setp1: UserClient generate all resources of the resource Node to Metadata-info,
setp2: UserClient register Metadata    to the local SN
setp3: UserClient request to the CenterEnrollCenterDB with the metadata-info of a resource, if
(valid)
setp4: SearchEnrollCenter ()
//Locate similar ClassifiedEnrollCenter,return the address of ClassifiedEnrollCenter;
setp5: if( ClassifiedEnrollCenter=valid)    goto setp6;
else goto setp9
setp6: RegisterSource()
//Register the metadata-info of a resource to Locate ClassifiedEnrollCenter
setp7: if (Unregistered resources=true) goto setp3
else
setp8: if (Resource nodes have updated) UpdateSource()
setp9:exit
setp10: end;
```

a. After users install the client, they enter their query in the client (in normal cases, the user node is also a resource node, belong to a VO)

b. The client generate metadata description according to the need the users provide, first, submit query to SN of VO, if find the required resources, then extract the location information of resources service node, communicate directly with the resource node. If the VO has not the required resources, then go to step c.

c. Query location information of Classified Register Center suitable for customer needs in Classified Register Metadata Storage Center.

d. location information Classified Register Center that meet the will be returned to the client

e. The client transmit the resources metadata description that need to Classified Register Center, according to the returned location information of Classified Register Center,

f. Classified Register Center inquire the appropriate resources metadata information according to needs, returne the corresponding resources metadata information to the client

g. The client extract location information of resources node according to the resource metadata information, communicate directly with the resource nodes.

h. Resource nodes provide resources services, and returne service results to the client. The client informs the results of query, and shows searchers the retrieved resource.

As method eliminate the invalid queries based on hierarchy information resource discovery, so efficiency is very high. Compare with the previous resource discovery methods, this mechanism has obvious advantages [16].

B . Algorithm description of information resources discovery

Discovery algorithm of information resources shown in Table 3:

Information resource discovery mechanism based on hierarchy retained the advantages of centralized query, a number of Classified Register Centers share a large number of grid resource service register pressure. also reduce the performance impact of a large database search. Although the centralized query has many faults, but it still has an obvious advantage that query rapidly. The

Classified Register Metadata Storage Center use centralized query, but all of the Classified Register Centers also are distributed storage, so not only retained the advantages of centralized query , but also overcome the shortcomings of its difficult to expansion.

C . Resource similarity search method

To be able to choose the users most need resources, we propose a resource similarity search method based on Probabilistic models.

Probabilistic model [17]: If the document is sorted according to relevance of the size of the query probability, Top surface of the document is most likely to be acquired. The main task of obtain information is to calculate the correlation between document and query .A query composed by keywords from a fixed space, A document composed by collection of words from the same keyword space, namely Doc = <term1,term2,…, termn>. If the document meets the formula (1), the document will be accessed.

$$P(Rela|Doc) >= P(Dissimi|Doc) \quad (1)$$

P (Rel | Doc) represents the conditional probability that the document Doc is related to the query, P (Dissimi | Doc) represents the conditional probability that the document Doc is not related to the query.

Reference Probabilistic model, we propose the following two definitions:

Suppose there are two variables: X {A (ax), B (bx), C (cx), D (dx)}, Y {A (ay), B (by), C (cy), D (dy) }, X represents the user input keywords quantized value, Y is the resource metadata keywords quantized value of Classified Register Center, they have the same four key A, B, C, D, four keyword weights for the X-ax, bx, cx, dx, in the Y for ay, by, cy, dy.

Definition 1. The dissimilarity of X, Y：

$$\text{Dissimi}=\tfrac{1}{2}(\sum|X-Y|)=\tfrac{1}{2}(|ax-ay|+|bx-by|+|cx-cy|+|dx-dy|) \quad (2)$$

Definition 2. the similarity of X, Y：

$$\text{Rela=100-Dissimi} = 100-\frac{1}{2}(\sum|X-Y|) \quad (3)$$

D.Time complexity of information resources discovery

Suppose there are n network resource services, to be equally distributed to the m Resource Register Center. Query is divided into two parts, the first query Resource Register Center, and then query resources. In the query you first need to locate Resource Register Center, In m Register Center Metadata Storage Center, the query time complexity is O (m). In n / m resource services Resource Register Center, the query to a specific resource needs O (n / m) time complexity.. Therefore, time complexity of information resources discovery is:

$$O(N) = O(n/m)+O(m) \quad (4)$$

Table 3 ALGORITHM DESCRIPTION OF INFORMATION RESOURCES DISCOVERY

setp1: UserClient generate query to Metadata-info,
setp2: UserClient send query to the local SN, UserClient query the local VO in a small area
setp3: if (valid) return the query results, goto setp 10
elsesetp4: request to CenterEnrollCenterDB, SearchEnrollCenter(),
//Locate similar ClassifiedEnrollCenter,return the address of ClassifiedEnrollCenter;
setp5: if(ClassifiedEnrollCenter=valid) goto setp6;
else goto setp10
setp6: Send query to ClassifiedEnrollCenter, SearchSource(),look for the appropriate resource information
setp7: return results
setp8: UserClient extract resources Node's address, communicate with the resource Node directly;
setp9: Resource Node return the results to UserClient
setp10: end;

VI. SIMULATION TEST

At present the main grid scheduling simulation tools: Bricks [18], MicroGrid [19], SimGrid [20], GridSim [21], etc. Grid simulator used here is GridSim[22], GridSim been developed by Rajkulnar Buyya[23] of Melbourne University, Australia. GridSim is a grid simulator that can run in windows and Linux platforms. In this study, we were using the scheduling algorithm built-in GridSim (Optimis Cost and Time) and the proposed task scheduling algorithm, compare the two algorithms.

Number of experiment task scheduling began in 100, the number increased by 100 per scheduled task once, stop at 2000. We compared time-cost of two methods in dealing with the grid task, whether the proposed algorithm has advantages.

A. Installation and setup GridSim environment

It also need to set the environment variable of GridSim after installation is complete, the method set In WindowsXP as follows:

Right Click on My Computer→ [Properties] → [Advanced] → [environment variable].

Variables in the system, the new variable [GRIDSIM], the value of installation path is gridsimtoolkit, D:\gridsimtoolkit-3.3

Edit [classPath], the value of the addItem is: %GRIDSIM%\application\gridbroker\gridbroker.jar;%G

RIDSIM%\jars\gridsim.jar;%GRIDSIM%\jars\simjava2.jar;%GRIDSIM%\application\visualmodeler\visualmodeler.jar;

GridSim package tasks as Gridlet, Gridlet has parameter MI (Million Instructions), MI is Gridlet's task length, task length means Running time in 1000MIPS processor.

B. The impact of Total Simulation Time in the case of Group and Ungroup

We set: Average MI value for 20, Deviate% value for 10, Granularity time (s) for 10, Overhead time for grouping (s) is 10. The results shown in Figure 2.

Figure 2 shows: When using Gridlets Ungroup method, as Gridlets increases, the total simulation time increases rapidly, the curve slope is steep; when using Gridlets Group method, as Gridlets increases, the total simulation time increases slowly. When using Gridlets Group method, since a large number of Gridlets are grouped into a small number of Gridlets group, the total transmission time decreased, because only a small amount of Gridlets group is processed, Gridlets processing time decreased.

C. Different Granularity time impact on the simulation time when using the Group method

We set: Average MI value for 20, Deviate% value for 10, Granularity time (s) for 5,10,15,20,25. The results shown in Figure 3.

Figure 3 shows: When Gridlets grouping, Granularity time the higher the total simulation time shorter. Before the start of the simulation, given Granularity time multiplied with the MIPS, and the results will determine MI in a given Granularity time, therefore, if the higher granularity of time, each resource can support more of the MI.

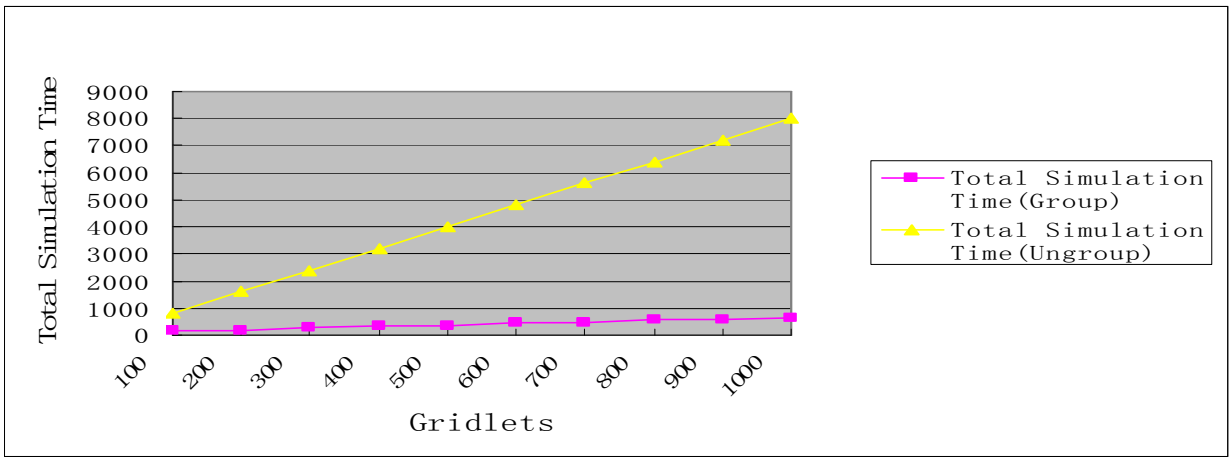

Figure 2 The impact of Total Simulation Time in the case of Group and Ungroup

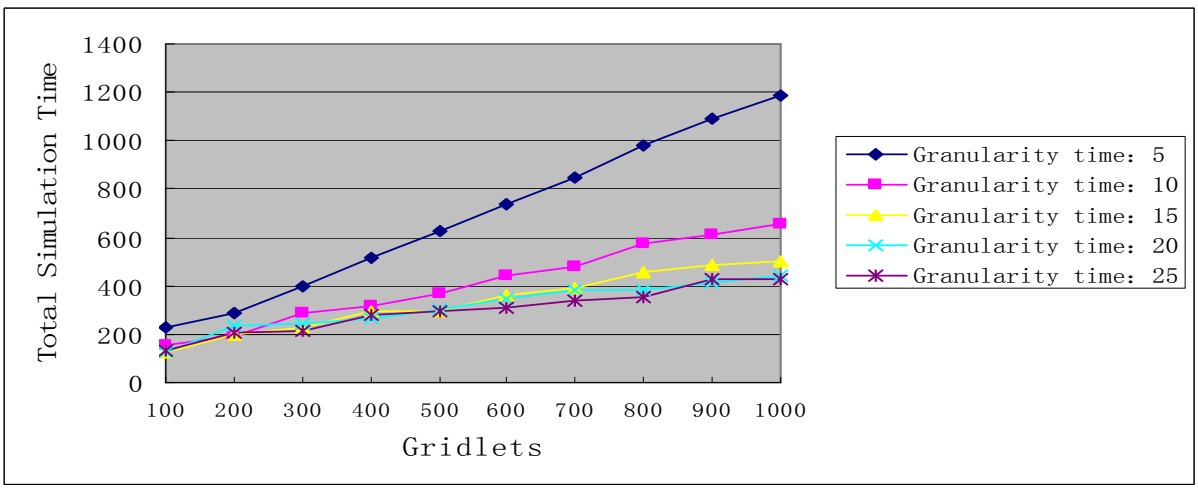

Figure 1. Figure 3 Different Granularity time impact on the simulation time when using the Group method

Table 4 Comparison of two scheduling algorithms

Number of Gridlets	Total Simulation Time(Optimis Cost and Time algorithm)	Total Simulation Time(hierarchy information resource classified register and discovery algorithm)
100	230.03	224.79
200	263.33	248.16
300	344.64	300.25

400	385.30	339.17
500	435.57	346.06
600	521.33	385.68
700	553.18	410.35
800	633.02	458.06
900	673.99	490.06
1000	721.17	528.25
1100	811.62	538.30
1200	844.43	538.55
1300	923.09	553.49
1400	963.29	575.34
1500	1008.19	585.61
1600	1094.65	634.00
1700	1129.27	672.61
1800	1211.49	672.61
1900	1234.96	698.00
2000	1298.99	730.00

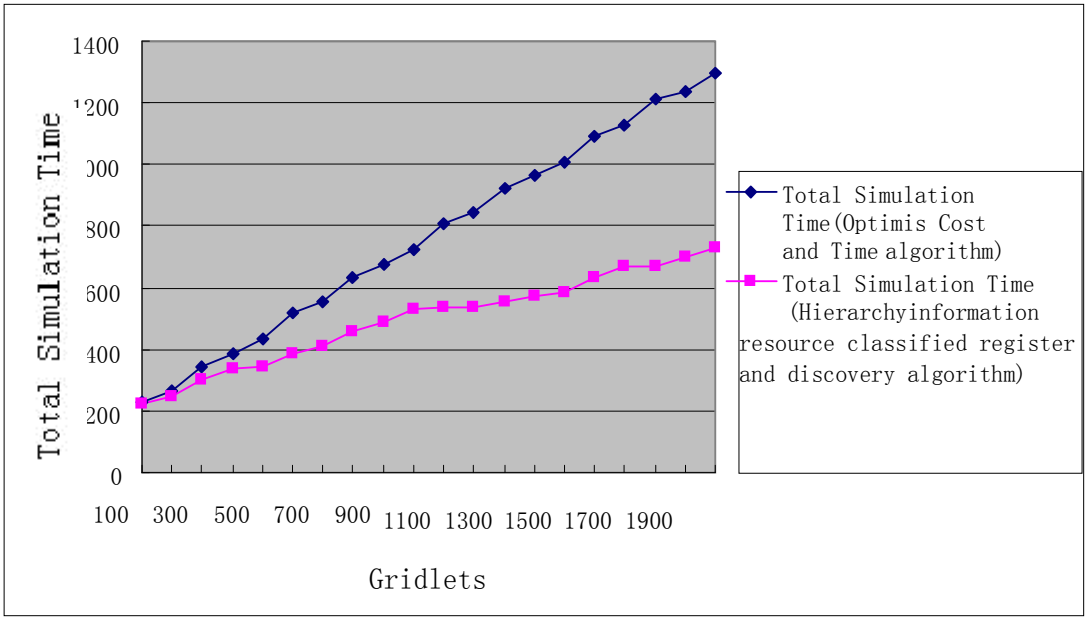

Figure 4 Comparison of two scheduling algorithms

To make results accurate, reasonable, we set: Average MI value for 40, Deviate% value for 10, Granularity time (s) for 20, Gridlet grouping method selected Group, Overhead time for grouping (s) is 10.

All scheduling data is shown in Table 4, the data in Table 4 plotted in Figure 4, Figure 2 shows: in same situation ,the proposed method that information resource classified register and discovery based on hierarchy in grid, it's Total Simulation Time is more lower than scheduling algorithm built-in GridSim (Optimism Cost and Time), and with the Gridlets increasing, the proposed algorithm time-cost growth smooth , compared with scheduling algorithm built-in GridSim (Optimis Cost and Time), the proposed algorithm is more time-cost advantages.

VII. CONCLUSIONS

This paper proposed the model of information resource classified register and discovery based on hierarchy in grid， this model combines the advantages of the existing resource discovery mechanisms, and avoid their disadvantages. This model has great advantages in the field of information resources discovery, has a smaller time-cost when query resources, this model can find the necessary resources quickly and accurately in a large number of information resources. This paper also gives algorithm description of register and discovery. Finally, we did performance testing using GridSim , prove that the proposed method is successful.

REFERENCES

[1] Foster, and C. Kesselman. The Grid: Blueprint for a Future Computing Infrastructure. Morgan Kaufmann Publishers, 1999, 36～40 .

[2] I. Foster. What is the Grid? A Three Point Checklist[J]. GRID Today, July 20,2002.

[3] Wei Li, Zhiwei Xu, Fangpeng Dong. Jun Zhang. A Grid Resource Discovery Mode Based on The

Routing-Transferring Method, Grid Computing--GRID 2002
[4] http://www.globus.org/toolkit/docs/4.0/info/key/
[5] GONG YiLi, DONG FangPeng, LI Wei, et al .VEGA Infrastructure for Resource Discovery in Grids [J]. Journal of computer science & technology, 2003, 18 (4): 413-422
[6] W Li, Z Xu, F Dong, et al .Grid resource discovery based on a routing-transferring model [C]. Proc of the 3rd Int'l Workshop on Grid Computing. NewYork: Springer -Verlag, 2002.145-156
[7] Qing Wang, THE RESEARCH OF GIRD RESOURCE DISCOVERY BASED ON P2P, Proceedings of the 8~(th) International Symposium on Distributed Computing and Applications to Business, Engineering and Science, 2009.
[8] S Milgram. The Small World problem. PSYCHOLOGY Today, 1967, 67 (1):60 — 67.
[9] Lei Zhang, The research of mobile agent-based mobile grid resource discovery, 2010 International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2010), 2010.
[10] Judith Pearce, Janifer Gatenby, New Frameworks for Resource Discovery and Delivery, National Library of Australia Staff Papers, 2005.
[11] Yue Chen, Ying Li, Zhengxian Gong, Qiaoming Zhu. A Framework of a Tree-based Grid Information Service [C]. Proceedings of the 2005 IEEE International Conference on Services Computing (SCC'05). 255～256
[12] Zhi-yun Zheng; Jun-xia Zhao; Xing-jin Zhang; and Lun LI , Resource Description and Discovery of Campus Grid, 2010 The 3rd International Conference on Computational Intelligence and Industrial Application (PACIIA2010), 2010.
[13] Dion Hoe-Lian Goh , Alton Chua, Chei Sian Lee, Khasfariyati Razikin, Resource discovery through social tagging: a classification and content analytic approach, Online Information Review, Vol. 33 Iss: 3, pp. 568 - 583, 2009.
[14] Luis E. Anido, etc. Education metadata and brokerage for learning resources. Computers & Education 38(2002)
[15] Simone Mora and Babak A. Farshchian, A Unified Architecture for Supporting Direct Tag-Based and Indirect Network-Based Resource Discovery, Lecture Notes in Computer Science, Volume 6439/2010, pp. 197-206, 2010.
[16] Rafael Moreno-Vozmediano, A hybrid mechanism for resource/service discovery in ad-hoc grids, Future Generation Computer Systems, Volume 25, Issue 7, pp. 717-727, July 2009.
[17] Michael Cardosa, Abhishek Chandra, Resource Bundles: Using Aggregation for Statistical Large-Scale Resource Discovery and Management, IEEE computer society, vol. 21 no. 8, pp. 1089-1102, August 2010.
[18] Aida K, Takefusa A, Nakada Hetal. Performance Evaluation Model for Scheduling in a Global Computing System [J]. The International Journal of High Performance Computing Applications, 2000, 14(3):
[19] Song H J, Liu X, Jakobsen Detal. The MicroGrid: A Scientific Tool for Modeling Computational Grids [C]. Proceedings of the 2000 Conference on Supercomputing, 2000:4-10
[20] Legrand A, Marchal L, Casanova H. Scheduling Distributed Applications: the SimGrid Simulation Framework [C]. Proceedings of the third IEEE International Symposium on Cluster Computing and the Grid, 2003:138-145
[21] Buyya R, Murshed M. GridSim: a Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing [J]. Concurrency and Computation: Practice andExperience, 2002, 14(13-15):1175-1220
[22] http://www.gridbus.org/gridsim/gridsimtoolkit-3.3
[23] Buyya R, Murshed M. GridSim: a Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing [J]. Concurrency and Computation: Practice and Experience, 2002, 14(13-15):1175-1220

Mingyong Li was born on October 27, 1979 in Hubei Province, China. He received the Bachelor's degree in educational technology in 2003 at Central China Normal University, and he received the Master of Educational Technology in 2008 at Chongqing Normal University. His main research interests include computer networks, education grid, educational resources, the Semantic Web.

Yan Ma was born on October 13, 1960 in Yunnan Province, China. He received the Master of Educational Technology in 1993 at Central China Normal University, and he received the doctor's degree of Education in 2008 at Southwestern University. His main research interests include artificial intelligence, semantic meshes, education grid.

Yuanyuan Liang Henan Province, China. Birthdate: May, 1985. Study in Chongqing normal University. Major in Curriculum and Teaching Methodology. And research interests on Modern Educational Technology.