Grid Resource Discovery Algorithm Based on Distance

Zhongping Zhang[1,2], Long He[1], Chao Zhang [1]
[1]The School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei, 066004, China
[2]The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao, Hebei, 066004, China
Email: zpzhang@ysu.edu.cn, hl_helong1988@126.com, zcmysky310@163.com

***Abstract*—Resource discovery in a Grid environment is a critical problem and also a fundamental task which provides searching and locating necessary resources for given processes. Based on the domain and resource routing nodes, a grid resource discovery model of multilayer overlay network is given in this paper. And on this basis, using a linear combination of the block distance and chessboard distance instead of Euclidean distance, grid resource discovery algorithm based on distance is proposed. Experiment shows that the algorithm has low cost, fast response and can obtain better success rate of lookup, as well as the effectiveness of the resource discovery algorithm.**

***Index Terms*—Grid; model; distance; resource discovery; multi-layer overlay network**

I. Introduction

Now, the grid is the international frontier of research topics. With the development of Grid technology, the Grid begins to be used in various fields. The most important one is data grid which is widely used in data-intensive industries. Data Grid can access, storage, move and manage the data which is heterogeneous, distributed and mass. It has a very broad application prospects.

Grid computing connects a large amount of geographically distributed heterogeneous resources to form a virtual network environment, enabling the users to share resources dynamically, which can effectively improve the utilization rate of resources and system performance. In the dynamic grid environment, resources are huge and have strong heterogeneity. The problems of resource management and lookup under ensuring high efficiency and use as possible as low cost become very complicated. Grid resource discovery mechanism in grid system is the key to realize resources sharing. Grid resource discovery mechanism performance directly determines the performance of the grid system. So it is necessary to design grid resource discovery algorithm with a low cost, good scalability and high efficiency.

In order to develop an efficient and scalable solution for Grid resource discovery, a series of challenges must be faced, due to the following reasons [1]:

1. dynamic property of Grid resources;
2. absence of central authority;
3. heterogeneous and large scale character of Grid environments;
4. unpredictability of faults;
5. difficulty in handling complex multi-attribute range queries.

To deal with these challenges, a resource discovery mechanism must have the following important characteristics [1]:

1. support for intermittent resource availability;
2. independence from any centralized/global control and global knowledge;
3. support for attribute-based discovery;
4. support for multi-attribute range queries;
5. scalability in terms of number of users and resources, and type of resources;
6. provide excellent query performance.

The existing fully centralized or fully distributed resource discovery method is not very ideal. They have some flaws respectively. However, the grid resource discovery method which is based on hierarchical model focuses on the advantages of centralized and distributed discovery method. So, using the combination of the layered structure and tree structure, the model based on Multi-Layer Overlay Network to discover resources is given in this paper. And on this basis, using a linear combination of the block distance and chessboard distance instead of Euclidean distance, grid resource discovery algorithm based on distance is proposed.

The rest of the paper is organized as follows. Related work on grid resource discovery is given in section II. A Multi-Layer Overlay Network(MLON) Model is proposed for resource discovery in data grid in section III. Grid Resource Discovery Algorithm Based on Distance(BC-DIS) is proposed in section IV. The performance evaluation is presented in Section V. Finally, the conclusion and proposed future work are discussed.

II. Related Work

At present, many organizations and researchers both at home and abroad study grid resource discovery model

Manuscript received March 2014, revised April 1, 2014, accepted April 15, 2014.

doi:10.4304/jsw.9.11.2966-2973

from different perspectives and different levels. The unified strategy to manage resources in the grid is adopted in the Globus centralized model proposed in [2]. It has a good global control and high efficiency of resource discovery, but it lack adaptability and extensibility in grid. Resources are managed through the interaction of different resource management systems in the P2P distributed model proposed in [3-5]. And just the opposite of the centralized model, it has a good scalability and lacks in the overall control. Besides, the communication costs of whole model are far higher than these of the centralized model. The flooding technology to find the resources is adopted in [6]. In the face of the grid dynamic environment, it has good fault tolerance and availability, but poor extensibility. The distributed resource discovery based on the virtual structure is adopted in the resource discovery mechanism based on Overlay Network (ON) proposed in [7] and Tree-Structured Overlay Network (TSON) in dynamic grid environment proposed in [8]. The ON and TSON can shield dynamic characteristics of grid resources and heterogeneity effectively. But they also have some deficiencies: because they partition Organizational virtual structured domain based on the physical area and use attribute matching method to find the resources, they cause large network consumption and poor efficiency.

A peer-to-peer architecture for resource discovery in a large and dynamic collection of resources has been proposed in [9]， it is similar to Gnutella combined with more sophisticated query forwarding strategies taken from the Free-net overlay network. Requests are forwarded to one neighbor only based on experiences obtained from previous requests, thus trying to reduce network traffic and the number of requests per peer compared with simple query flooding as used by Gnutella. Because a suitable peer was not reached simply, the approach suffers from higher numbers of required hops to resolve a query compared to our approach and provides no lookup guarantees.

A model called the Hierarchical Resource Organizational Model has been proposed in [10]. The model consists of three layers to process the information in a grid. These three layers include: Physical Network, Resource Information and Index Information. The Physical Network Layer is at the lowest level containing the physical resources linked with each other on the Internet. For each resource, a resource node is placed in the Resource Information Layer. Therefore, the Resource Information layer contains virtual organizations (VO), which is a group of resource nodes in a star topology with a super node in the center. The super node stores all the information regarding the resources of a VO as adjacent lists. The Index Information Layer stores information pertaining to all super nodes of the middle layer and composition of the multiple layers form the basis for hierarchical resource discovery.

At present, the resource discovery mechanism is mainly divided into two types: one is centralized style; one is distributed style. Traditional centralized resource discovery mechanism has certain advantages [11].

1. the topology structure of the system is relatively simple and it is easy to build and maintain, besides the consumption is small;
2. services focus, easy to resources sharing and system security is better;
3. there is no problem of inconsistent of information resources and the efficiency of resource discovery is higher in a small area.

Meanwhile, there are same disadvantages [11] as follows:

1. Reliability is Relatively poor. When the central server fails, the system will not work properly and the system has not fault-tolerant;
2. Scalability of the system is Relatively poor.

However, the characteristics of fully distributed resource discovery mechanism and centralized resource discovery mechanism is opposite, the main drawback is that:

1. the disorder and structurelessness of resource information space make resource discovery have a certain blindness;
2. Nodes can join or leave at any time, which makes the security of system difficult to control.

III. Resource Discovery Model

Overlay network is built on a physical network virtual network connected by virtual or logical links. It can provide more reliable and better fault-tolerant application services without changing existing large-scale network architecture. Each layer of overlay network uses a structured P2P technology, which is advantageous to dynamically join, leave and forward service of the virtual node. Multi-Layer Overlay Network(MLON) resources organization mechanism proposed in this paper is make full use of the overlay network technology and resources which are large, widely distributed, heterogeneous and dynamically changed in the grid of P2P technology organizations to better meet the dynamic environment of the grid system.

A. Model Organization Structure

MLON uses the combination of the hierarchical structure and tree structure to organize overlay network. As the service nodes at the bottom, all the resources in the grid constitute the physical resource layer of the model framework. A large number of grid resources are classified according to the type of resources. The virtual organization with an internal structure is managed by the corresponding overlay network node. Multi-layer overlay network structure is shown in Figure 1.

Overall, MLON structure will be divided into two levels: the upper layer is the overlay network layer, or MLON layer; the lower layer is the physical resource layer, or the Internet layer. MLON consists of multiple virtual layers. All nodes in each layer are classified by the resource type and belong to different domains respectively. Each domain is independent, and each node in the domain uses the graph structure to connect each other in the form of structured peer-to-peer. Each layer adopts the method of the upper managing the lower. A virtual node of the upper manages a domain of the layer. Looking down from the top of a node, virtual nodes adopt tree structure to organize them.

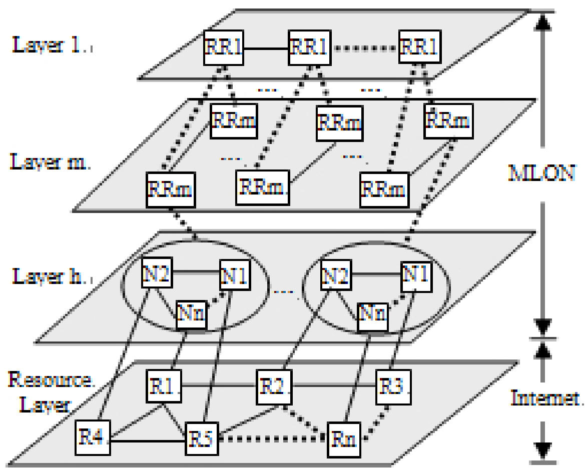

Figure 1. The frame of MLON

Domain and layer as the basic logic unit is used in the MLON. A vast amount of resources are logically divided into several domains through the classification. Each domain is managed by the Resource Router. Resource router holds a large number of service information of the domain nodes. Resources belong to different domains by type. When virtual nodes of a domain increases to a certain number, the domain was divided into two or more smaller domains, which leads to the granularity of resource type small; on the contrary, when the nodes of a domain is reduced to a certain number, the domain nodes is merged, which leads to the granularity of resource type large. Thus a multilayer overlay network is formed.

B. The Basic Definition

Definition 1: Grid Resource $R=<ID, A, op()>$, where *ID* represents the resource *R* identifier; *A* represents the attribute collection of the resource *R*; *op()* indicates the set of operations of the resource *R*. When $r\in R$: $a_i(r)=\{a_{i1}, a_{i2}, \cdots, a_{ij}\}$, $op_i(r)=\{op_{i1}(r), op_{i2}(r), \cdots, op_{ik}(r)\}$.

Definition 2: Resource Management Model $G=<N_G, E, R, T>$, where N_G represents the set of all the nodes in the grid, that is to say $N_G=\bigcup_{i=1}^{n} n_i$; *E* denotes the set of edges in the network; *R* denotes the set of all resources in the grid, that is to say $R=\bigcup_{i=1}^{n} r_i$; *T* indicates the collection of the type of resource, that is to say $T=\bigcup_{k=1}^{m} t_k$. As $r_i\in R$, for $\forall$ i , $j\in[1,n]: r_i\cap r_j=\varnothing$ .When the type of resource r_i is represented by $r_i(a)$,i.e. $r_i(a)=t_i$, $t_i\in T\Rightarrow R(A)\in T$,for $\forall$ i , $j\in[1,m]: t_i\cap t_j=\varnothing$.

Definition 3: Domain For each type of resource, the connected and structured P2P, called domain, is constructed by lots of information nodes which register with the type of resources.

Definition 4: Overlay Network Node $n_V=<VID, des, indexlist>$, where *VID* represents the virtual node n_V identifier in the overlay network; *des* indicates the description of the node n_V ; *indexlist* indicates the pointer list of the node n_V , it contains pointers of linking to the parent nodes, the child nodes and the brother nodes, namely $indexlist=\{index\_f, index\_s, index\_b\}$ .Each resource is assigned a globally unique *VID* address by overlay network, which is bundled with resource node *ID*. *VID* includes two parts of domain number (VID_r) and intra-domain node number (VID_s), then two parts are divided and each segment represents a certain type of resources.

Definition 5: MLON Model $G_V=<N_V, rf, IndexList>$, where N_V indicates the set of all the nodes in MLON, namely $N_V=\bigcup_{i=1}^{n} n_{vi}$; rf is the function of resources registration; *IndexList* means the set of the pointer list in MLON.

Definition 6: Resource Router Define the non-leaf nodes in the MLON model as *RR*(Resource Router)

Definition 7: the function of resources registration rf The domain of the MLON is divided by resource type which is determined by resource properties. For each resource, whether it is static attribute or dynamic attribute, it is only active in a particular area[12]. If $a_j(r)\in[c,d]$, m control points are inserted in the interval $[c,d]$, when $c=c_0<c_1<\cdots<c_{m-1}=d$ and $b=(b_0, b_1, \cdots, b_{m-1})=(0,0,\cdots,0)$, for any $a_j(r_i)$:

$$b_k=\begin{cases}1, & a_j(r_i)=[c_k, c_{k+1}) \\ 0, & else\end{cases} \tag{1}$$

On the basis of formula (1): $rf_j(a_j(r_i))=b=(0,\cdots,1,\cdots,0)$, then $rf(a(r_i))=\bigcup_{j=1}^{n} rf(a_j(r_i))$

According to the definition 1 and 2, the following assumptions are made:

Assumption 1: A resource belongs to only one type of resource, i.e. $|r_i(a)|=1$. A resource can be registered on multiple resource routers, one is the resource registration node, others are nodes of the copy.

In order to better describe the structure of the MLON model, according to the definition, the following assumptions can be made:

Assumption 2: The neighbor of the node n_v is only its brother. i.e. $index\_f(n_v) = index\_f(index\_b(n_v))$.

According to the definition and description, node n_{vi} in MLON has the following characteristics:

(1) $\forall n_{vi} \in N_V$,if $indexlist_i(n_{vi}) \neq \varnothing$, then $\exists indexlist_i \in IndexList$.

(2) $\forall n_{vi} \in N_V$, if $index\_s(n_{vi}) \neq \varnothing$, then $n_{vi} \in RR$.

(3) $\forall n_{vi} \in N_V$, if $index\_f(n_{vi}) = \varnothing$, then n_{vi} is the top-level resource router node in MLON.

(4) $\forall n_{vi} \in N_V$, if $index\_s(n_{vi}) = \varnothing$, then n_{vi} is the leaf node in MLON.

Non-leaf node (i.e. resource router) in MLON is only responsible for managing resource sub-tree which uses it as the root and provides the sub-tree information for the upper node. The nodes on the bottom are called leaf nodes and the leaf nodes correspond to resource nodes of the physical network, which contain all the information of resources. In MLON, leaf nodes and non-leaf nodes only distinguish logically.

IV. Resource Discovery Algorithm

Resource discovery algorithm is focused on resource searching. If the grid dynamic changes like the joining or exiting of the nodes are mastered, good resource discovery algorithm is able to resource search. Considered with the dynamic change of grid environment from the resource routing nodes service deployment and the resource nodes register, the resource discovery algorithm of the Multi-Layer Overlay Network model based on distance is obtained.

Because of the dynamic of the resources in the grid environment, MLON must be able to self organized and maintained. It mainly includes the registration of resources, the cancellation of resources, the update of resources information and the failure and exception handling of resource nodes. These issues have been discussed in other articles, so they are not described in detail here

A. The Problem Description

Definition 8: Grid resource discovery model for request processing $K = \{G, request(R(A)), l\}$, where $request(R(A))$ signifies the search request, it defines the various attribute constraints which the resources of finding should be met; l denotes the request-forward strategy in the process of resource searching; K indicates the collection of resources found in resource discovery process, which matches the search request.

It has a lot of similarity between resource nodes in each domain of virtual network. When looking for resources, users usually not concerned about a resource or a certain type of resources, but only pay attention to the most relevant k results with the needs. The proposed resource discovery algorithm in this paper combines Source Searching Technique Based on Style Matching Routing(SMRT) with Source Searching Technique Based on Topk Algorithm(TopkT) to finally find the collection of resources K which match the search request of user.

Definition 9: Euclidean distance The $P_1 = (x_1, x_2 \cdots, x_n)$ and $P_2 = (y_1, y_2 \cdots, y_n)$ are regarded as the point in V_n , so we can define the formula of Euclidean distance $d(P_1, P_2)$ as follows:

$$d(P_1, P_2) = \left\| P_1 - P_2 \right\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3}$$

Definition 10: Blocks distance For any $P_1 = (x_1, x_2 \cdots, x_n)$ and $P_2 = (y_1, y_2 \cdots, y_n)$, $P_1, P_2 \in R^n$, so we can define the formula of Blocks distance $d_s(P_1, P_2)$ as follows:

$$d_s(P_1, P_2) = \sum_{i=1}^{n} \left| x_i - y_i \right| \tag{4}$$

Definition 11: Chessboard distance For any $P_1 = (x_1, x_2 \cdots, x_n)$ and $P_2 = (y_1, y_2 \cdots, y_n)$, $P_1, P_2 \in R^n$, so we can define the formula of Chessboard distance $d_c(P_1, P_2)$ as follows:

$$d_c(P_1, P_2) = \max_{1 \le i \le n} (\left| x_i - y_i \right|) \tag{5}$$

Theorem 1: For any $P_1 = (x_1, x_2 \cdots, x_n)$ and $P_2 = (y_1, y_2 \cdots, y_n)$, $P_1, P_2 \in R^n$,then $d_c(P_1, P_2) \le d(P_1, P_2) \le d_s(P_1, P_2)$.

After the user requests resource, grid system can quickly find the resource domain which matches the request of user, because the *RR* in the resource domain is responsible for assigning the most appropriate resources. Resources in the resource domain can be uniformly represented with a point of the m-dimensional coordinates. Searching a resource can be seen seeking a point that satisfies the query conditions in the m-dimensional coordinates. So, $request(R(A))$ can be converted into seeking the k resources which satisfy the query conditions and whose distance is nearest from enquiry point in the m-dimensional coordinates.

By the Euclidean distance formula, m-dimensional resource attribute query can be converted into a one-dimensional range query to meet the grid resource discovery mechanism principles of multiple attribute queries. According to the formula (3), (4), (5), it can be seen that the calculation of d_S and d_C is simpler than the calculation of d. According to theorem 1, if we use d_S instead of d, the calculated value will be larger; if we use d_C instead of d, the calculated value will be smaller; deviation value of the two methods are often large. So we can consider using an appropriate linear combination of the d_S and d_C instead of d. This will not only improve the efficiency of calculation, but also make the deviation value smaller. According to the proposed method in [13], we can calculate $\alpha(d_S + d_C)$ instead of d.

When searching for resource, users often have a preference for a certain attribute, so the resource attributes are weighted according to users' need when they query. Let $W(w_{m-1}, \cdots, w_1, w_0)$ be the attribute weight of resource r_i, where $\sum_{i=0}^{m=1} w_i = 1$. So the distance between resource r_i and enquiry point r is given below.

$$d_i = \alpha(\sum_{j=0}^{m-1}(\left|a_j^i - a_j\right| \times w_j) + \max_{0 \le j \le m-1}\{\left|a_j^i - a_j\right| \times w_j\}) \quad (6)$$

The method that uses an appropriate linear combination of Blocks distance and Chessboard distance instead of Euclidean distance reduces multiplication operation and has a high operation speed, thereby the efficiency of resource discovery has been improved.

B. Algorithm Description

Resource discovery algorithm based on MLON can be divided into two parts: the user's request is passed between nodes *RR*; the request is transferred and searching for resource in matching types of resources domain.

After $request(R_i(A))$ is submitted, the Resource Router in the network generates VID_i by the function rf, then it makes and operation with bitwise between VID_i and VID of node *RR*. If the result is equal to VID, it shows that the resources of query are in resource domain which is managed by node *RR*, i.e. if $\exists(VID_i \wedge VID) = VID$,then $r_i \in R(VID)$. Otherwise, the request is passed, until the resource domain that accords with the request is found or Time To Live (TTL) of request message is zero.

When the bottom resource router RR_j receives a request, $request(R_i(A))$ will be transponded towards each node of its management domain by the router. When receiving the request, the leaf node first tests itself whether meet the requirements of the query. If it meets, the distance d_i between local resource and query point is calculated. Then d_i is returned to the resource router. Based on the optimal matching principle, the nearest *k* resources with $request(R_i(A))$ will be returned to the user by the resource router according to d_i. The resource discovery algorithm of the Multi-layer Overlay Network model based on distance(BC-DIS) is given below.

Algorithm: The resource discovery algorithm of the Multi-layer Overlay Network model based on distance(BC-DIS)

Inputs: *W*, $request(R_i(A))$, *k*, *ttl* ;

Outputs: The optimal *k* resources.

BC-DIS (*W*, $request(R_i(A))$, *k*, *ttl*)

Begin

(1) user sends $request(R_i(A))$ to MLON;

(2) $r_i(a) = request(R_i(A))$;

(3) $VID_i = rf(r_i(a))$;

(4) While ($(VID_i \wedge VID) \neq VID$) { /*search the domain of matching resources*/

(5) *ttl*-1, if *ttl* is less than 0, then the identifier which represents the end of the life cycle of registration message is returned.

(6) If ($index\_f(VID) \neq Null$)

(7) Node *RR* delivers $request(R_i(A))$ and VID_i towards father node;

(8) Else

(9) Node RR delivers $request(R_i(A))$ and VID_i towards brother node by the best neighbor strategy; }

(10)While($index\_s(index\_s(VID)) \neq Null$) { /*search father node RR_j of matching resources */

(11) *ttl*-1, if *ttl* is less than 0, then the identifier which represents the end of the life cycle of registration message is returned.

(12) Node *RR* delivers $request(R_i(A))$ and VID_i towards child node, the request is passed by the best neighbor strategy in the domain; }

(13) RR_j transmits the request towards child node;

(14) The distance d_j between leaf node r_j and $r_i(a)$ is calculated by leaf node, then returns d_j ;

(15)Return TopK = $SelectMinD(d_0, d_1, \cdots, d_n)$;

End

C. The Algorithm Time Complexity Analysis

Assuming the number of grid resource type is n_t, namely $|T|=n_t$, the number of nodes with the lowest layer resource domain that requests to be matched is n_v, the height of MLON is *h*, the number of resource return is *k*.

From the above algorithm, we can get to know that the resource discovery process can be divided into two stages: (1) the first stage is that user requests to find resource domain. The worst case is that the requests of user are transferred from the leaf nodes to the top of MLON. Then it traverses the top of the resource routers. Finally, it sent a request to (*h*-1) layer resource routers of the MLON. The number of comparison between requests and nodes is ($2h+n_t$). As the number of resources managed by the top-level resource routers has a threshold, the depth *h* is a constant. (2) the second stage is that the request is propagated in the matching resource domain. When all the resources in the domain satisfy the request, the number of comparisons is ($n_v+k(n_v-k)$). Therefore, the time complexity of the algorithm is $O(n_t+n_v)$.

V. The Experimental Evaluation

Experiment employs GridSim[14] simulation toolkit to simulate experiment about the grid resource discovery algorithm BC-DIS and grid resource discovery algorithm based on flooding[15]. Use the JDK and JCreator as a programming environment. The program is running on Windows XP platform.

A. Experimental Environment Settings

The experiment only simulates computing resources in the grid, not considering other grid resources. In the process of simulation, we need to ignore change of network topology in resource discovery at a time, and all the resource information should be kept stable and effective in the network. According to the macroscopic statistical properties showed by the system, we will simulate. When we ignore the dynamic changes of the various details of the system, we can simplify the problem under the condition of grasping the essence of the system.

Supposing the number of resource nodes of the model is 50000, namely $|N_G|=50000$, $|R|=50000$; the number of resource type is 10, $|T|=10$, Various types of resources are evenly distributed and $Max(N_{vi}(R(A)))=5000, Min(N_{vj}(R(A)))=50$;

According to the experimental environment, the node n_{vi} of the MLON must meet the following two conditions:

(1) $\exists f(n_{vi}) \in \varnothing, then 50 \le |N_{vi}| \le 5000, de(n_{vi}) = 4$

(2) $\forall S(n_{vi}) \in \varnothing, then 50 \le |N(f(n_{vi}))| \le 500, de(n_{vi}) = 4$

where $de(n_i)$ represents the degree between nodes n_i and its sibling nodes in a certain layer of the MLON. $f(n_i)$ represents the parent node of node n_i, $S(n_i)$ represents the child node of node n_i.

In flooding resource discovery mechanism, an average degree between nodes is 4 with life cycle in the way of TTL controlling request information in the network.

In order to make the comparability between the two kinds of resource discovery mechanism, the concept of Resource Density [16] is introduced:

Definition 12: Resource Density is the proportion between the resources that meet the search conditions and the total number of all the resources, that is to say $d = \frac{|N_R|}{|N|}$, where *d* indicates the density of the resource *r* which is corresponded with *request (R (A))*, its unit is 1/1000. N_R indicates the resource node set that matches with *request(R(A))*. *N* represents the set of all resource nodes.

B. Simulation Analysis

In the process of simulation experiment, experiment runs many times with different user requests, and simulation results are the average of the experiment results. There are three important performance indexes of the resource discovery algorithm: the number of nodes involved in the process of resource discovery, response time and the success rate of lookup. This section mainly compares BC-DIS resource discovery algorithm with flooding resource discovery algorithm through the three indexes.

B1. The Comparison of the Number of Nodes Involved in the Process of Resource Discovery

The comparison of the number of nodes in the resource searching for the two kinds of resource discovery algorithm is shown in Figure 2. Seen from the Figure 2, with the increase of density of resources, the number of nodes involved in BC-DIS resource discovery algorithm is far less than flooding algorithm. Especially when the density is small, BC-DIS algorithm has more obvious advantages.

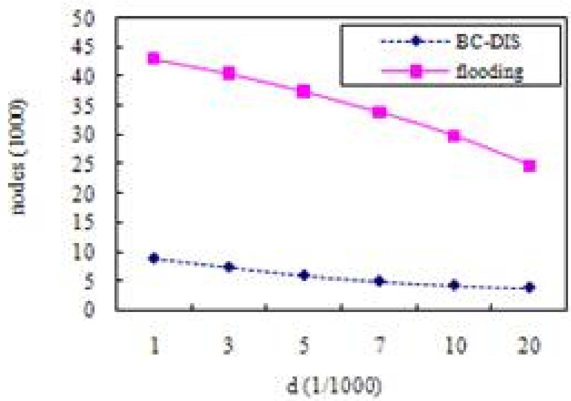

Figure 2 Referenced nodes

B2. The Comparison of Response Time

Response time is the average time from a resource request to receiving a successful response by user. Judge the response time according to hops in the path in which service request is forwarded. The comparison of the number of hops in the resource searching for the two kinds of resource discovery algorithm is shown in Figure 3. Seen from the Figure 3, the hops of the BC-DIS resource discovery algorithm are less than the hops of flooding algorithm. But with the increase of density of the network resources, resources satisfying the requests in the system increase, and the hops of two kinds of resource discovery algorithm become smooth and consistent.

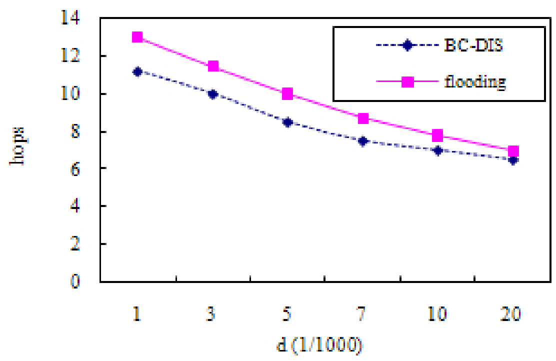

Figure 3 Hops

B3. The Comparison of the Success Rate of Lookup

The comparison of the success rate of lookup for the two kinds of resource discovery algorithm is shown in Figure 4. Seen from the Figure 4, two algorithms can always find the resources with the increasing of lifecycle of information. The BC-DIS algorithm discovers resources more effectively when TTL value is small.

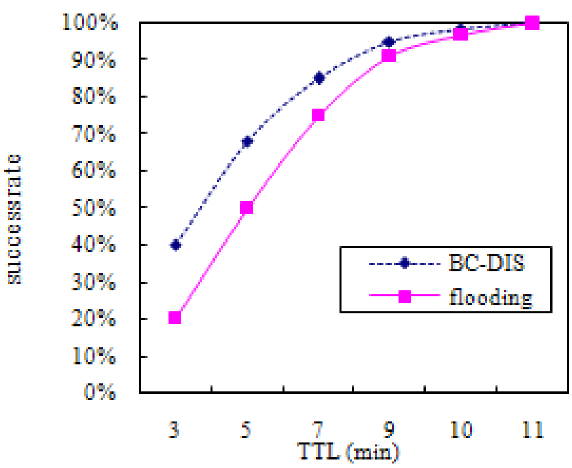

Figure 4 Research success rate

It can be seen from the above simulation results that the use of resource discovery mechanism proposed in this paper can achieve better query efficiency with less traffic and node hops in large quantities of certain resources.

In addition, resource search results of two algorithms are found: grid resource discovery algorithm based on flooding that adopts the resource search method of comparison of the attributes often returns to results with deviation of attribute too much; grid resource discovery algorithm based on BC-DIS supports multiple attributes and range search, and it can return to the k most appropriate resources according to customer request, so it has the highest customer satisfaction.

VI. The Conclusion And Future Work

Because of disadvantage that layered resource discovery mechanism is strongly dependent on resource routing nodes in the grid, a grid resource discovery model based on MLON is given. And on the basis of MLON, the grid resource discovery algorithm based on distance is proposed by using a linear combination of the block distance and chessboard distance instead of Euclidean distance. Through performance analysis and simulation results, the model has fault tolerance and scalability. And it also meets the requirements of the grid dynamics, distribution and scalability. Then it can shield heterogeneity among the resources. The next step is to deploy the model to experiment in real grid environment.

Acknowledgements

This work was financially supported by Hebei Provincial Natural Science Foundation of China (F2012203087), National Natural Science Foundation of China (61272124),and National Natural Science Foundation of China (61073060).

References

[1] Anand Padmanabhan, Sukumar Ghosh, Shaowen Wang. A Self-Organized Grouping (SOG) Framework for Efficient Grid Resource Discovery[J]. Journal of Grid Computing, 2010, vol.8, no.3, pp.365–389.
[2] Karl Czajkowski, Steven Fitzgerald, Ian Foster. Grid information services for distributed resource sharing[C]. //Proceedings of the 10th IEEE HPDC. Washington, DC: IEEE Computer Society, 2001, vol.9, no.14, pp.181-194.
[3] Yingjie Xia, Mingzhe Zhu, Yang Li. Towards Topology-and-Trust-Aware P2P Grid[J]. Journal of Computers, 2010, vol.5, no.9, pp.1315-1321.
[4] Zhengzhen Zhou, Yonglong Luo, Liangmin Guo, Meijing Ji. A Trust Evaluation Model based on Fuzzy Theory in P2P Networks[J]. Journal of Computers, 2011, vol. 6, no.8, pp.1634-1638.
[5] Javad Akbari Torkestani. A distributed resource discovery algorithm for P2P grids[J]. Journal of Network and Computer Applications, 2012, vol.35, no.6, pp.2028-2036.
[6] Shikha Goyal. Flooding Algorithm Based Resource Discovery on Virtual Organization[C]. Proceedings of National Conference on Trends in Signal Processing & Communication(TSPC'13), 2013, pp.12-14.
[7] Francesco Palmieri. Introducing Virtual Private Overlay Network services in large scale Grid infrastructures[J]. Journal of Computers, 2007,vol.2,no2, pp.61-72.
[8] M Marzolla, M Mordacchini, S Orlando. Resource Discovery in a Dynamic Grid Environment. Proc. DEXA Workshop 2005, pp.356-360.
[9] A. Iamnitchi and I. T. Foster. On fully decentralized resource discovery in grid environments. In GRID '01: Proceedings of the Second International Workshop on Grid Computing, pages 51–62, London, UK, 2001. Springer-Verlag.
[10] Mingyong Li, Yan Ma, Yuanyuan Liang. Study the Model of Information Resource Classified Register and Discovery based on Hierarchy in Grid[J]. Journal of Software, vol.7, no.7,July 2012.
[11] Xing Liu, Weidong Xiao, Lei Xu; Grid Resource Discovery Mechanism Based on Complex Topology [J]; Computer Engineering and Applications; vol.9, pp. 132-136, 2005.
[12] Moreno Marzolla, Matteo Mordacchini, Salvatore Orlando. Resource discovery in a dynamic grid environment[C].

//Database and Expert Systems Applications(DEXA), 2005, pp.356-360.
[13] Wang Yuxuan, Li Haijun, Zhou Chunguang. Using a linear combination of the Blocks and Chessboard distance instead of computing Euclidean distance. The small microcomputer system, 2004, vol.25, no.12, pp.2121-2125.
[14] Weifeng Sun, Qiufen Xia, Zichuan Xu, Mingchu Li. A Game Theoretic Resource Allocation Model Based on Extended Second Price Sealed Auction in Grid Computing[J]. Journal of Computers, 2012, vol.7, no.1, pp.65-75.
[15] Noghabi, Hossein Boroumand; Ismail, Abdul Samad; Ahmed, Aboamama Atahar; Khodaei, Masoumeh. An Optimized Search Algorithm for Resource Discovery in Peer to Peer grid[C]. Proceedings - 1st International Conference on Informatics and Computational Intelligence, ICI 2011, pp.21-24, 2011.
[16] Cheng Zhu, A Decentralized Grid Resource Discovery Scheme based on Resource Classification, National University of Defense Technology, 2004.

Zhongping Zhang, Male, Born in 1972, professor, Ph.D., post-doctoral, CCF Senior Member (E20-0006458S). His main research interests are the grid computing, data mining and semi-structured data etc. He has undertaken 1 project of provincial level and participated in 2 projects funded by national natural science foundation of China. He rewarded the provincial Scientific and Technological Progress second-class Award. On the domestic and international academic conferences and journals, He published more than 80 papers, 15 of them are cited by EI.

Long He, Male, Born in 1988, Current Master Student, the main research interest is the grid computing.

Chao Zhang, Male, Born in 1988, Current Master Student, the main research interest is the grid computing.