Journal of Southeast University (English Edition) Vol. 23, No. 1, pp. 57 – 62 Mar. 2007 ISSN 1003—7985

Construction of overlay network based on P2P technology in Grid environment

Zhuang Yanyan Liu Ye Niu Lin

(Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: Based on the advantages of both Grid and peer-to-peer (P2P) networks, an overlay network in the Grid environment is constructed by P2P technologies by a modified version of the Chord protocol. In this mechanism, different nodes' accesses to different resources are determined by their contribution. Therefore, the heterogeneous resources of virtual organizations in large-scale Grid can be effectively integrated, and the key node failure as well as system bottleneck in the traditional Grid environment is eliminated. The experimental results indicate that this management mechanism can achieve better average performance in the Grid environment and maintain the P2P characteristics as well.

Key words: resource management; Grid computing; peer-to-peer environments; overlay network

Computational Grids[1] and peer-to-peer (P2P)[2] communities are both arousing great concern in the field of distributed resource sharing. While both technologies have the same final objective—the pooling and coordinated use of large sets of distributed resources. They followed different evolutionary paths; hence, different requirements and technologies exist.

The complementary nature of the strengths and weaknesses of the two approaches suggests that the design objective of the two environments will eventually converge[3]. On the basis of characteristics of Grid and P2P, and characteristics of distributed resources, this paper puts forward the idea of resource management of the Grid system by P2P technology, which combines the complexity of Grid with the scale and dynamism of P2P. Through the distributed management of P2P, Grid systems are free from key node failures and system bottlenecks resulting from central servers.

1 Grid and P2P Environments

Grid and P2P are two new approaches to distributed computing which have emerged during the past few years, both claiming to address the problem of organizing large-scale resource management. While both have undergone rapid evolution and widespread deployment, they bear certain limitations.

Grid has deployed relatively sophisticated services and applications of at least limited trust[4]. Resources owned by various administrative organizations are shared under locally defined policies. Such a set of individuals or institutions defined by these sharing rules is a virtual organization (VO). As a system increases in scale, Grid developers are facing and addressing problems relating to autonomic configuration and management. The Globus toolkit is one Grid software that is used worldwide. It provides the Grid infrastructure with command line utilities and APIs, but the task of discovering and deciding on what the optimal resource is, is left to users. So it is crucial to have a smart resource management system that accepts requests from multiple users and collects the results.

P2P communities have developed rapidly around unsophisticated but popular services, and are seeking to expand to more sophisticated applications as well as continuing the innovation of the large-scale autonomic system management[3]. Distributed lookup protocols, like Chord[5], are aimed at addressing the problem of locating decentralized resources efficiently. Given a key, it efficiently determines the node responsible for storing the key's value. It assigns each node and key an m-bit identifier using a base hash function such as SHA-1[6]. As shown in Fig. 1, the identifiers of each node and key make a ring of size 2^m. Key k is assigned to the first node whose identifier is equal to or follows k in identifier space, which is named successor (k).

Considering their interrelationship and future evolution, both Grid and P2P take the same general approach to solving resource sharing problems by the overlay structures that they coexist with, but need not correspond in structure to the underlying organizational structures.

Received 2006-05-16.

Foundation items: The National Natural Science Foundation of China (No. 60573133), the National Basic Research Program of China (973 Program) (No. 2003CB314801).

Biography: Zhuang Yanyan (1983—), female, graduate, zhuangyanyan@seu.edu.cn.

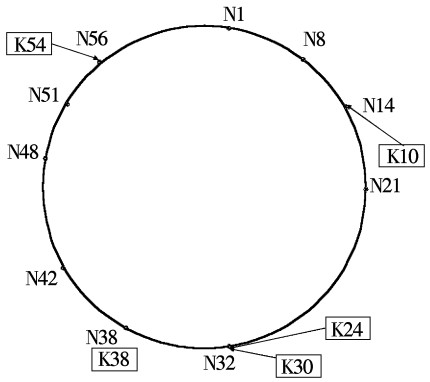

Fig. 1 Chord ring consisting of 10 nodes storing five keys

2 Related Work

As for the merger of the two environments, there is little previous literature for reference. In Ref. [3], the authors made some comparisons between Grid and P2P, including communities, incentives, applications, technologies, resources, and achieved scales. They also pointed out future directions. Over time, the scale of Grid systems is increasing as barriers to participation are lowered and as commercial deployments enable communities to be based on purely monetary transactions. Meanwhile, developers of P2P systems are becoming increasingly ambitious in their applications and services, as a result of both natural evolution and more powerful and connected resources.

However, there have not been any feasible mechanisms for the implementation of this merger so far, thus the motivation of this paper. In the following sections, we propose a resource management mechanism based on P2P technology that cooperates with Grid, of which the results are evaluated and validated by extendable experiments.

3 P2P Approach to Resource Management in Grids

3.1 XML in Grid and P2P environments

Resources are heterogeneous and distributed geographically, which makes their description extremely difficult. XML makes it possible to define data in such a way that both the sending and receiving party using it will understand the sort of data that has been sent.

In this paper, XML documents are both the Grid service descriptions and the input of the consistent hashing algorithm of Chord, making the Globus and Chord API compatible and effectively cooperative. Fig. 2 shows the functional framework of XML in Grid and P2P environments.

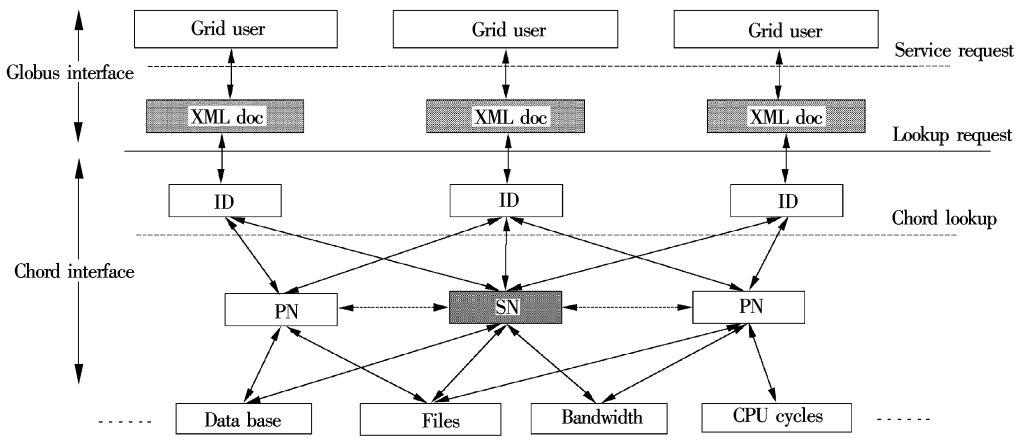

Fig. 2 Grid and P2P function framework

The following is an example of XML schema. It describes the structure of an XML document used for resource description.

```
〈xs: element name = "resource"〉
    〈xs: complexType〉
        〈xs: sequence〉
            〈xs: element name = "IP" type = "xs: string" minOccurs = "1"/〉
            〈xs: element name = "file" type = "FileType" minOccurs = "0" maxOccurs = "un-bounded"/〉
            〈xs: element name = "CPU" type = "xs: string" minOccurs = "0"/〉
            〈xs: element name = "memory" type = "xs: integer" minOccurs = "0"/〉
            〈xs: element name = "bandwidth" type = "xs: decimal" minOccurs = "0"/〉
        〈/xs: sequence〉
        〈xs: attribute name = "Differentiated" type = "xs: boolean" use = "required"/〉
    〈/xs: complexType〉
〈/xs: element〉

〈xs: complexType name = "FileType"〉
    〈xs: sequence〉
        〈xs: element name = "name" type = "xs: string" minOccurs = "1"/〉
        〈xs: element name = "size" type = "xs: decimal"/〉
    〈/xs: sequence〉
```

〈/xs: complexType〉

With this schema, resources can be described as follows:

```
〈resource〉
    〈IP〉172. 18. 13. 22〈/IP〉
    〈file〉
        〈name〉friends. rmvb〈/name〉
        〈size〉76〈/size〉
    〈/file〉
    〈CPU〉2. 01 GHz〈/CPU〉
    〈memory〉1024〈/memory〉
〈resource〉
```

3.2 Construction of overlay network

As Grids are composed of VOs with different local policies, the Chord protocol based overlay network can be organized within each VO. The construction of the substrate overlay network based on P2P technology is shown in Fig. 3, which demonstrates the intra-VO and inter-VO constructions in the Grid environment.

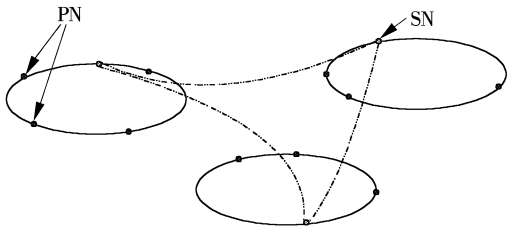

Fig. 3 Construction of P2P based overlay network

In Fig. 3, peer nodes (PN) make up the Chord ring for intra-VO resource sharing. Lookup messages are propagated by PNs within the local ring.

As VOs are made up of workstations with high performance, most queries can be met within the local domain. When different VOs want to share resources, the sharing node (SN), which joins a larger Chord ring, is elected.

3.2.1 Construction of overlay network intra-VO

1) Service deployment

Different from traditional Grid systems such as Globus, the model in this paper contains no service registration center which acts as a central server (such as UDDI in web services). Each node issuing its service is in charge of its service deployment.

Before issuing a certain resource, the corresponding resource description, i. e., the XML document provided by some user, is hashed into an m-bit identifier ID. And the service issuing and deploying algorithm is as follows:

① Assign XML document k an m-bit identifier using hash function: ID = hash(k);

② If (successor(ID) = = hash(IP of local host)), the relevant resource is deployed locally; otherwise go to step ③;

③ Invoke Chord routing process in search for successor(ID); deploy the resource on the ultimate node just as step ② does.

The algorithm above tends to balance load, since each peer in VO receives roughly the same number of keys by using consistent hashing. In the steady state of a system containing N peers, all lookups can be resolved via $O(\log N)$ messages to other nodes.

In practice, VO needs to deal with peers joining the system as well as peers that fail or leave voluntarily. The Chord protocol handles these situations skillfully in a decentralized manner such that the inconsistent state caused by concurrent joins is transient, and neither the success nor the performance of resource lookups is to be affected. It is scalable as well as robust with large numbers of peers.

2) Resource lookup

Resource lookup will be the same as Chord as long as the required resource can be located within local VO. When met with lookup failure, the query is forwarded outwards to another VO through SN:

Node i maintains a triple as $(\mathrm{ID}_i, \mathrm{ID}_j, t_q)$, where ID_i stands for the identifier of node i, ID_j is the corresponding node that responds to node i' s query, and t_q is the submission time of node i' s query.

① At the time of t_q, node i launches its query. It first checks whether its requirement can be met locally. If so, the lookup process is returned and the triple is logged as $(\mathrm{ID}_i, \mathrm{ID}_i, t_q)$. If not, go to step ②;

② Invoke Chord routing process to find the successor (for example, ID_j) of the correspondent key;

③ Forward the lookup message to ID_j. If the key is found then return, and log the triple as $(\mathrm{ID}_i, \mathrm{ID}_j, t_q)$. Otherwise, return FAILURE message and go to next step;

④ All nodes in the VO start the computing process:

$$B_k = \{(\mathrm{ID}_q, \mathrm{ID}_r, t) \mid 0 \leqslant \mathrm{ID}_q, \mathrm{ID}_r \leqslant n-1, \mathrm{ID}_r = \mathrm{ID}_k\} \tag{1}$$

$$N_k = |B_k|;\ t_k = \max\{t \mid (\mathrm{ID}_q, \mathrm{ID}_r, t) \in B_k\} \tag{2}$$

B_k takes the record of all triples when node k successfully returns the required resource in accordance with other node queries (including ID_k itself). Altogether node k satisfies N_k times, in which t_k stands for the latest time of successful responses. Meanwhile, all nodes calculate their weight W_k as follows:

$$W_k = \omega_1 N_k + \frac{\omega_2}{t_q - t_k} \qquad 0 \leqslant i \leqslant n-1;\ \omega_1 + \omega_2 = 1 \tag{3}$$

⑤ Elect nodes with the largest W_k as SNs, which

are responsible for communication with the outside.

From Eq. (3) we can see that the more times node k responds successfully, and the sooner it makes such a response, the higher W_k it will attain. Therefore, node k has a higher probability that it will be elected SN.

3.2.2 Construction of overlay network inter-VO

In the inter-VO scenario, XML makes sure that the heterogeneous resources in different domains be expressed in the same form, and the Chord protocol based overlay network makes sure the resources be distributed in the same fashion. Failed lookup is forwarded outwards to another VO by some local nodes with certain "authority", and the failed lookups of other VOs are also taken in by these authoritative nodes.

In this case, peer nodes can elect one or more leaders with the highest performance, the maximum trust or the best resource offers so as to efficaciously contribute local resources to outsiders with minimum transaction overhead. W_k (see section 3.2.1) amongst PNs may be a reasonable choice. PNs with the largest W_k are elected the SNs. Via applying Peterson's election algorithm[7], elected SNs join a larger, global Chord ring and carry out inter-VO communication once local service falls short of appropriate resources.

As presented in section 4, propagation of lookup messages will be first confined within VO, which is relatively moderate-scaled. Gradually, messages are distributed throughout the entire overlay network. In this way, most queries may be restricted within a local VO rather than immediately forwarded to the outside, which reduces network flows.

4 Evaluation

In this section, we evaluate the performance of our resource management mechanism. The SPIS[8] system is used for simulation since it has been implemented by the P2P research group of Southeast University and the code is written in Java with good portability.

4.1 Assumption

For simplicity, we assume that there are two VOs in the system. Nine different kinds of resources are deployed in VO_1, and eleven in VO_2. However, peers in both VOs randomly launch resource lookup requests for ten different kinds of resources. Therefore, 10% of users' requests cannot be met in VO_1. If VO_1 and VO_2 start to share resources for mutual use, then only 5% of all requests will be refused. Notice that the results obtained in this section may also be applied to more complicated scenarios which consist of more VOs.

4.2 Simulation setup

In order to test the performance of our mechanism, SPIS based simulation acts as follows. Peer nodes, which are free to join and leave a local VO, randomly launch resource lookup requests. If lookup failure occurs, SN is elected by Peterson's election algorithm among these PNs. SNs from different VOs together join a larger Chord ring.

We compare the simulation results in two different scenarios. In the former, VOs do not share any resource information (as shown in Fig. 4(a), VO_1 and VO_2 work separately so about 90% requests can be met in VO_1), while in the latter, resources are shared by elected SNs (see Fig. 4(b)) between the two VOs.

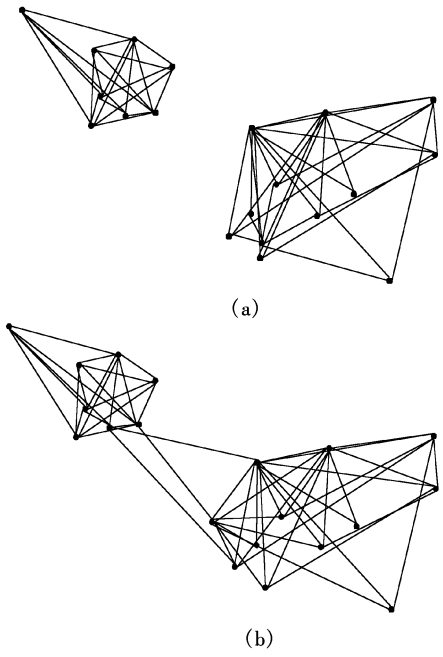

Fig. 4 Two simulation scenarios. (a) No SNs in VO; (b) SNs exist in VO

4.3 Performance analyses

There are two counters, namely SuccessCounter and FailedCounter, taking the record of successful and failed numbers of lookups. As a result, we get the success rate of random lookups in the two scenarios:

$$\text{success rate} = \frac{\text{SuccessCounter}}{\text{SuccessCounter} + \text{FailedCounter}} \quad (4)$$

First, we consider the case in which no node joining or leaving occurs. Tab. 1 and Tab. 2 show the counters' value and success rates respectively.

Tab. 1 Counter values and success rate when no SNs in VO

Times	SuccessCounter	FailCounter	Success rate
1	54	8	0. 871 0
2	156	30	0. 838 7
3	209	40	0. 839 4
4	202	39	0. 838 2
5	306	57	0. 843 0
6	360	64	0. 891 1
7	454	62	0. 879 8
8	422	70	0. 857 7
9	408	69	0. 855 3
10	480	86	0. 828 0
11	413	86	0. 888 0
12	486	25	0. 951 1
13	484	20	0. 959 9
14	515	21	0. 960 8
15	532	22	0. 960 3
16	505	21	0. 960 1
17	551	23	0. 960 3

Tab. 2 Counter values and success rate when SNs exist in VO

Times	SuccessCounter	FailCounter	Success rate
1	86	13	0. 868 7
2	127	9	0. 933 8
3	155	12	0. 928 1
4	182	27	0. 870 8
5	178	20	0. 899 0
6	214	22	0. 906 8
7	266	20	0. 930 1
8	241	13	0. 933 0
9	298	18	0. 944 0
10	296	22	0. 930 0
11	327	18	0. 947 0
12	303	17	0. 947 0
13	305	7	0. 977 1
14	321	1	0. 997 2
15	310	0	1
16	313	0	1
17	334	0	1

In the no SNs scenario (see Fig. 5(a)), the success rate is increasing as time goes on. However, it never reaches 1 if no such required resource exists in a local VO (average success rate is 0. 895 6 in Fig. 5 (a), which approximates the 0. 9 we assumed in section 4. 1). In this circumstance, not all requirements can be met, and it barely guarantees the quality of service in the Grid environment.

In the second scenario, the success rate gradually increases and finally reaches 1. Although it takes a bit longer time to stabilize, Fig. 5(b) indicates that the average success rate is higher (0. 942 0 compared to 0. 895 6). Thus, the system achieves better performance and users get better service, which is not affected by the slightly longer response time.

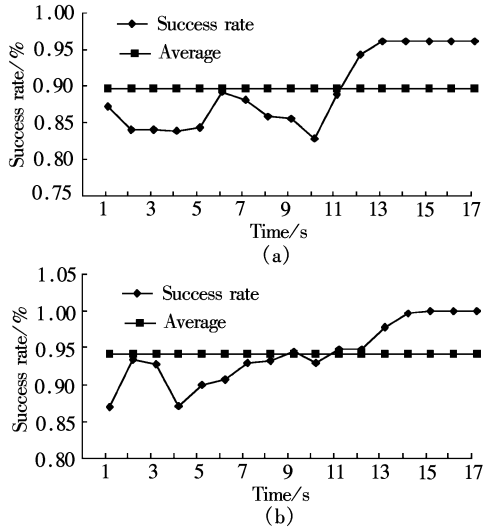

Fig. 5 Success rate. (a) No SNs in VO; (b) SNs exist in VO

When considering node joins, transitory low performance inevitably occurs (see Fig. 6).

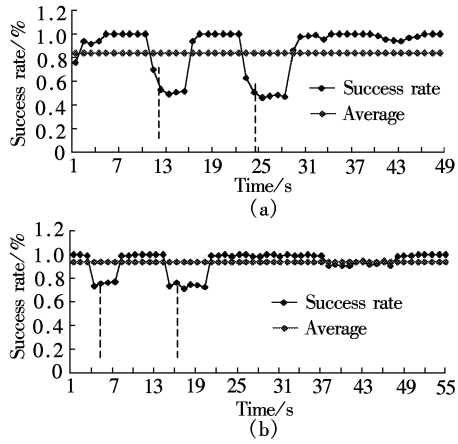

Fig. 6 Success rate with node joins. (a) No SNs in VO; (b) SNs exist in VO

In the first scenario, at the 12th and 24th second, new node joins the Chord ring. Similarly, node joining happens at the 4th and 16th second in the latter scenario. Indeed, system performance suffers a little, as can be seen in Fig. 6. However, the inconsistent state is transient, only occurring at the moment the node enters. As for the node departure or failure, results are much alike. Both lead to transient performance fluctuations but little influence in the long run.

The average success rate in the second scenario is also higher, as there is more resource available when compared to individual VO.

Therefore, the Grid system is in possession of the attractive features of P2P. It is scalable as well as robust.

5 Conclusion and Future Work

Traditionally, centralized resource management in

Grid severely affects the robustness and flexibility of the scheduling system, which leads to such problems as key node failures and system bottlenecks. A natural tendency of Grid is to expand in size for larger communities and a natural tendency of P2P systems is the increase in service complexity. This paper introduces a P2P resource management mechanism, to cooperate with the Grid system so as to achieve better average performance and maintain the P2P characteristics. Finally, the experiments show that the results are quite satisfactory.

Extension of this work to further discussion on security issues is under investigation now and the refinement of the SPIS system in the ongoing procedure is left for future research.

References

[1] Foster I, Kesselman C. *The Grid*: *blueprint for a new computing infrastructure*[M]. San Francisco, CA, USA: Morgan Kaufmann Publishers, Inc, 1999.

[2] Iamnitchi A, Ripeanu M, Foster I. Locating data in (small-world) peer-to-peer scientific collaborations [C]//*The First International Workshop on Peer-to-Peer Systems*, *IPTPS* 2002. Cambridge: Springer-Verlag, 2002: 232 - 241.

[3] Foster I, Iamnitchi A. On death, taxes, and the convergence of peer-to-peer and grid computing[C]//*The Second International Workshop on Peer-to-Peer Systems*, *IPTPS* 2003. Berkeley, CA, USA: Springer-Verlag, 2003: 118 - 128.

[4] Czajkowski K, Fitzgerald S, Foster I, et al. Grid information services for distributed resource sharing [C]//*Proceedings of the* 10*th IEEE International Symposium on High Performance Distributed Computing* (*HPDC*-10). San Francisco, CA, USA: IEEE Computer Society Press, 2001: 181 - 184.

[5] Stoica I, Morris R, Liben-Nowell D. Chord: a scalable peer-to-peer lookup service for Internet applications [C]//*Proceedings of the ACM SIGCOMM'*01 *Conference*. San Diego, California, 2002: 149 - 160.

[6] Brown R H, Prabhakar A. Secure hash standard [EB/OL]. (1995-04-17) [2006-05-10]. http://www. itl. nist. gov/fipspubs/fip180-1. htm.

[7] Peterson G, Byungho Y. Average case behavior of election algorithms for unidirectional rings [C]//*Proceedings of the* 13*th International Conference on Distributed Computing Systems*. Pittsburgh, PA, USA, 1993 : 366 - 373.

[8] Research Group of P2P Technology of Key Laboratory of Computer Network and Information Integration of Ministry of Education. Detailed designing report of the SPIS system [R]. Nanjing: Southeast University, 2006. (in Chinese)

网格环境下基于 P2P 技术的覆盖网络构建

庄艳艳　刘　业　钮　麟

(东南大学计算机网络和信息集成教育部重点实验室,南京 210096)

摘要:在结合网格与 P2P 网络技术优势的基础上,提出一种在网格环境下基于 P2P 技术的覆盖网络构建机制. 该机制对传统的 P2P 网络资源管理协议 Chord 进行改进,并根据节点对系统贡献的大小决定其资源管理的权限,实现大规模网格中不同虚拟组织间异构资源的整合,旨在进行有效的资源管理,消除集中式的网格环境下单点失效和系统性能瓶颈的问题. 实验结果表明,该机制在使得网格系统获得良好性能的同时,也保持了 P2P 网络的动态性与网络规模的可缩放性.

关键词:资源管理;网格计算;P2P 网络;覆盖网络

中图分类号:TP393