Available online at http://www.mecs-press.net/ijem

Research on P2P Service Discovery Model Based on QoS

Gao Xiao Yan

Department of Computer Science North China Institute of Science and Technology Beijing, China

Abstract

The existing P2P network service discovery lack of service quality assurance consideration, and can not fully utilize the network of some node powerful service quality attributes , in order to improve the efficiency of P2P networks service discovery,this paper proposes a P2P service discovery based on QoS algorithm model by combining QoS attributes and P2P characteristics. It defines P2P service description of based on QoS and introduces semantic information into service description to improve accuracy of service matching by using these service description. On this basis, it builds up P2P service discovery model of QoS guarantee and provides genetic algorithm on P2P service discovery. At last, analysis the feasibility and effectiveness of the service discovery algorithm through the proof.

Index Terms: QoS; P2P; service discovery

1. Introduction

Currently, as a kind of abstract form of information processing ability on network P2P network services obtain wide attention. Service discovery technologies refer to the mechanism that network nodes automatically obtain services relevant information which provide by other nodes. There are two main ways of P2P service discovery[1]: one is that service provider publishes description of resources it provided and the demand select required service from the number of services when necessary. The other is that the demand of service publishes description of services it required,and then provider matches these requires.However,the dynamic,heterogeneous,geographical distribution,openness,scalability and uncertainty of the network environment making the design of service discovery model algorithm with QoS guarantee a challenge. How to obtain service resources with QoS guarantee has become a focus of research in field of current service discovery[2].

Existing service discovery model is mainly to solve two following problems: First, how to describe the service capabilities accurately and painstakingly to support more precise match operation between user requires

* Corresponding author.
E-mail address: yanxiaogao@163.com

and service description; second, how to storage, index and exchange service metadata, it can not only guarantee search span of service discovery but also limit search time within an acceptable range. Researchers of service discovery introduced many semantic model. By means of ontology ,description logic and other logic inference system to enhance machine comprehensible of service description information also to support logic reasoning match between user demand and service capabilities.

However, because of the increasing attention to demand of service resources with quality assurance, there are still several major problems as follows in the existing service discovery model: (1) Since it has not fully consider the dynamic、uncertain and other characteristics of the service and lack of quality attributes description of service resources,   there are some inaccuracy while selecting the service. (2) Currently, service discovery mechanism based on keywords is widely used, it does not provide support for semantic description which affects the efficiency and accuracy of P2P service discovery to a great degree. (3)Most existing model search the service that match goals by description of service function, there is no more in-depth of using relevant information of service qualities for further marking and selecting within the searched services. These discovery strategy has a certain blindness. Especially in the P2P network, as a service node it has a greatly casualness when it joins and exits. And it can not guarantee the reliability of service quality that provided by service node.

According to the appeared problems of existing service discovery, this paper binds QoS attributes and P2P network characters and proposes a P2P Service Discovery Model Base on QoS. First, it defines the description of QoS-based P2P service on research of Web service and by combining with semantic web technology[4]. It also introduces semantic information into service description to improve accuracy of service matching by using these services description[5,6].Second, it builds up P2P service discovery model of QoS guarantee and provides discovery algorithm on P2P service discovery. At last, it analyses performance of service discovery algorithm.

2.   QoS Description of P2P Service Network

According to QoS definition of P2P services, a service is regarded as a operation. All the inner process is shielded, only its input and output interfaces is exposures to the outside[3].Therefore, we can define P2P service and P2P service request as the following form:

Definition 2.1 (P2P services) P2P service is an abstract description of a operation which it supports, it can be expressed as a triple PS=(N,In,Out),which:

(1) N ,as the unique identification of the service, regards the

regards the name of service. (2) In={In1,In2,…,Ink} is the input set of services; (3) Out={Out1,Out2,…,Outl} is the output set of services;

Definition 2.2 (Service Request) service request is an abstract description of interface requirements, it can be expressed as a pair PSR = (I,O), which:

(1) I={I1,I2,…,Im} is input set that can be provided;

(2) O={O1,O2,…,On}is a output set that is needed. Definition 2.3 (QoS-based P2P service) QoS-based P2P service is an abstract description of an operation which it supports, it can be expressed as a four-tuple PSQ = (N, In, Out,Q),which:

(1) N ,as the unique identification of the service, regards the servicename

(2) In={In1,In2,…, Ink} is the input set of services;

(3) Out={Out1,Out2,…, Outl} is the output set of services;

(4) Q represents QoS factors of the service, Q={q1,q2,…,qk} is used as quality attributes of the service.

Definition 2.4 (QoS-based P2P service request) QoS-based P2P service request is an abstract description of interface requirements, it can be expressed as a triple PSR=(I, O,Q), which:

(1) I={I1,I2,…,Im} is input set that can be provided;

(2) O={O1,O2,…,On}is a output set that is needed.

(3) Q represents QoS factors of the service, Q={q1,q2,…,qk} is used as quality attributes of the service.

Definition 2.5 (P2P atomic services) P2P atomic service is a service which finished through an interaction, it can be considered as an atomic, indecomposable process.

Definition 2.6 (P2P composite service) P2P composite service is a service which needs several atomic services to finish it.

3. P2P service discovery model base on QoS

A. Formal description of service discovery's problem model

When we research on the service discovery problem, we need to take path cost and the service's cost price of discovering a service into account according to the description of the QoS-based P2P service. In order to give out a formal description of the model we give the following definition:

It is assumed that PS is the set of the service which is provided by the nodes, and then $PS = \{ps_1, ps_2...\}, ps \in PS, ps$ is the specific service of PS.

Definition 2.7 service flag function (the function of service flag), it is used to represent whether the node has the required services.

$$x : V \to \{0,1\}, x_{ps}(v) = \begin{cases} 1 & \text{if node } v \text{ find ps service} \\ 0 & \text{else} \end{cases}$$
(1)

Definition 2.8 Service cost function (the function of service cost)

$$f : V \to [0, \infty], f_{ps}(v) = ps \text{ service cost in node } v$$
(2)

Definition 2.9 Service discovery flag function (the function of service discovery flag)

$$x_{ps}(v) = \begin{cases} >0, & \text{if find ps in node } v \\ 0, & \text{else} \end{cases}$$
(3)

Definition 2.10 Problems of service discovery (Cluster Service Discovery Problem, CSDP for short) Given P2P network structure as $G = (V, E)$, V set represents set of nodes of the P2P network, E set represents set of node's links of the P2P network. The problem of service discovery is that through searching the user's requires to find a strategy which can find out user demand services with least cost. $\mathrm{p}(s,d)$ represents paths which the discovery service has passed, these paths starting from any point within the P2P network to d. Problems of CSDP can formally repents as the following optimized model:

$$\underset{v \in V}{\text{Min}} \{ x_{ps}(v) * f_{ps}(v) \}$$

$$\text{s.t.} \begin{cases} x_{ps}(v) > 0 \\ delay(p(s,d)) \leq d^{req} \\ l_{p(s,d)} \leq N - 1 \end{cases}$$
(4)

In summary, the target of model solution is that the required cost $\{x_{ps}(v) * f_{ps}(v)\}$ of service discovery is minimum. Among them, the first constraint represents that it can always find the required service ps that start from point s to point d. $ps \in PS, ps$ is a particular service in PS; the second constraint represent delay of

paths that discovery service have passed, general delay can not exceed the user's required delay d^{req} ; the third constraint represents that it needs to go through a sufficient number of links while discovery services. $l_{p(s,d)}$ represents total path length of discovery services.

B. Genetic solving algorithm of CSDP

Genetic algorithms [7,8,9] is a general optimization algorithm that the optimization is not constrained to restrictive condition. And it has implicit parallelism . This paper will apply genetic algorithm to solve problem of discovery service The genetic algorithm of solving CSDP problem is as follows:

1) Initializationofgroup

Given a source node and a destination node,each chromosome of the group representing a path. The initial group generates randomly. First a ch destination node $d \in D$, find the all the path that can discovery service from source node s to d , that is $x_{ps}(v) > 0$. And for each destination node d∈D, use Dijkstra k-shortest path algorithm to find the path set that composed of all paths that can meet the maximum delay limitation. Then use the selected as an alternative path set of genetic algorithm's coding space. Let Q_v be a path set of destination node v then $Q_v = \{ P_v^1, ...P_v^j ...P_v^k \}$, which P_v^j is the j th path for destination of v. Select any path from path set and use it as chromosome of initial group. Obviously, such selected path can find all the required services.

2) fitnessfunction

For each path, the fitness function is defined as the reciprocal of the cost of discovery service, that is

$$f(P) = \frac{1}{\sum_{v \in V} x_{ps}(v) * f_{ps}(v)}$$

(5)

3) Select

Select Options adapt the optimal solution preservation method. First, it use fitness proportional selection method to select and then directly copy the individual of highest fitness into next group. The selection probability of each individual is proportional to its fitness. The greater the individual fitness is, the higher the probability of being selected. Set group size is n , the probability of being select of an individual v to be selected is that:

$$p_v = \frac{f(v)}{\sum_{v=1}^{n} f(v)}$$

(6)

4) the general of next group

Crossover operation uses the method of preserving same link and mutations operator uses bit mutation. Suppose given two parent paths, P_f and P_m , through the crossover operation generates offspring path P_c . If P_f and P_m is selected, it explains that their relatively high fitness value. Then the link they have in common is the main factor of resulting in higher fitness and represents the outstanding features of the parent generation. For the different links part of P_f and P_m ,we can reselect in the alternative. Thus we constitute a complete path.

C. Performance analysis of solution algorithm

Theorem 2.1 In the topology of P2P network in which there are a number of constrains, for the service discovery with given QoS request, genetic algorithm can search all the required service in sufficiently large genetic populations and evolutionary generations if the required service exists.

Proof2.1. Crossover probability p_c and mutation probability $p_m \in (0,1)$. In sufficiently large genetic populations and sufficiently long iteration times, a lot of crossover can discovery excellent solution and the service that searched by mutation mechanism can include every node of the network topological graphs. All these ensure the range of search space. At the same time, it copy individuals according to proportion the individual fitness accounts for in group fitness. According to the research on convergence standard genetic algorithm, it can search all the required service in sufficiently large genetic populations and evolutionary generations if the required service exists.

Theorem 2.2 The time complexity of algorithm is $\mathrm{O}\left(kSS'G + N^2 S\left(1 + P_m G\right)\right)$.

Proof2.2. The calculations of solution algorithm's time complexity include two parts, they are the iteration time of genetic algorithm and the complexity of using Dijkstra algorithm to calculate shortest path. Genetic algorithm compares the ralatively optimal relation and defines all the constrains as a unified   function   which can be equal to an added subtotal items. Suppose the number of sub goals is m, then for the group which scale is s the times of comparison is $\mathrm{O}\left((k+1)S\right)$. Advantages and disadvantages comparison carry out between it and individuals (S')in external set, it totally occur $\mathrm{O}\left((k+1)SS'\right)$ times. As the evolutional generation is G then the complexity of it is $\mathrm{O}(kSS'G)$. For the network topological graphs which P2P network size is N ,since the time complexity of Dijkstra algorithm is $\mathrm{O}(N^2)$ and the times that is needed to execute to initial groups is S, then the initial individual would require a total time $\mathrm{O}(N^2 S)$; Every implementation of the mutation in evolutionary would need to execute once Dijkstra algorithm. So when mutation probability is p_m, the required time is $\mathrm{O}(N^2 S p_m G)$. The totally demand complexity of executing Dijkstra algorithm is $\left(N^2 S\left(1 + p_m\right)G\right)$.

Above all , it can proof that the time complexity of solution algorithm is $\mathrm{O}\left(kSS'G + N^2 S\left(1 + P_m G\right)\right)$.

D. Conclusion

In this paper, combining QoS attributes and P2P characteristics,we proposed a P2P service discovery model based on QoS ,and applied genetic algorithm to solve problem of discovery service on P2P services,then analyzed the feasibility and effectiveness of the service discovery algorithm through the proof. the next research is that how to perfect model by experiments.

Acknowledgment

References

[1] Yan F, Zhan SY. A peer-to-peer approach with semantic locality to service discovery. In: Proc. of the 3rd Int'l Workshop on Grid and Cooperative Computing. LNCS 3251, Berlin: Springer-Verlag, 2004,pp: 831~834

[2] Hu Jianqiang. Research on Some Key Technologies of Web Service Discovery [D]. Changsha: National University of Defense Technology, 2005,pp: 42~52, in Chinese.

[3] Kuang shuo,Deng shuiguan,Li yin. Optimization of inverted index used for the combination of semantic service discovery. [J] Journal Of Software，2007，18（6）pp: 1911~1921,in Chinese.

[4] M Liu Jie, Zhuge Hai. A semantic link based infrastructure for web service discovery in P2P networks[C] roceedings of International World Wide Web Conference. Japan: Chiba, 2005,5 ,pp:940~941.

[5] Paolucci M, Kawamura T, Payne TR, Sycara K. Semantic matching of Web services apabilities. In: Goos G, Hartmanis J, van Leeuwen J, eds. Proc. of the Int'l Semantic Web Conf. (ISWC). LNCS 2342, Sardinia: Springer-Verlag, 2002,6: pp:333~347

[6] Burstein M, Bussler C, Zaremba M, Finin T, Huhns M, Paolucci M, Sheth A, Williams S. A semantic Web services architecture. IEEE Internet Computing, 2005,9,pp: 52~61

[7] Li YH, Bandar ZA, McLean D. An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. on Knowledge and Data Engineering, 2003,15,pp: 871~882

[8] Lo C-C,Chang W-H. :A Multiobjective Hybrid Genetic Algorithm for the Capacitated Multipoint Network Design Problem. IEEE Transactions on systems, Man, and Cybernetics-Part B:Cybeernetics, 2000,30 (4),pp:461~470.

[9] S. Caminiti, I. Finocchi, and R. Petreschi, A unified approach to coding labeled trees, in Proceedings of the 6th Latin American Symposium on Theoretical Informatics (LATIN'04), LNCS 2976,2004, pp: 339~348