An Effective Service Discovery Approach Based on Field Theory and Contribution Degree in Unstructured P2P Networks

Wanchun Yang
School of Electronics and Information Engineering
Tongji University
Shanghai, China
yangwch1982@126.com

Chenxi Zhang* and Jiangfeng Li
School of Software Engineering
Tongji University
Shanghai, China
chenxizhang10@aliyun.com

***Abstract*—With the growing number of services in the Cloud environment, service discovery is a critical issue. A number of works study structured P2P overlay network for services discovery, which can not describe the real world. In this paper, we propose an effective service discovery approach based on field theory and contribution degree in unstructured P2P networks. The services are discovered from three dimensions which are functional dimension, QoS dimension and transactional dimension. The proposed field theory consists of capability field and interest field. The flow of request and service can be guided by the capability field and interest field respectively. During the discovery process the request is submitted to the nodes whose service is more semantic similar to the request. An active service replication based on the interest field is applied to improve the service discovery. To address the free-riding problem, a contribution degree model is presented. The experimental results show that our proposed approach is feasible and effective.**

***Keywords*—service discovery; field theory; flow mechanism; contribution degree; unstructured P2P**

I. Introduction

Cloud computing is emerging as a new paradigm for developing distributed applications. Service discovery enables consumers to find the desired ones from the growing number of services in the Cloud environment. Current service selection approaches still have several limitations. Firstly, traditional service discovery approaches are based on UDDI which may easily suffer from problems such as performance bottleneck[1]. The structured P2P technology is adopted to improve the reliability, scalability of distributed applications[2], but need to set constraints on the network topology, so they can not describe the real world. Secondly, lots of service discovery approaches are based on the QoS dimension, neglecting other dimensions[3][4]. Furthermore, the free-riding problem exists in the service discovery problem[5].

In this paper, the semantic similarity, QoS parameters and transactional properties are all considered in the service discovery problem. The field includes capability field and interest field, while the flow is divided into service flow and request flow. A contribution degree model is presented to solve the free-riding problem.

II. Service Discovery in Unstructured P2P Networks

A service is uniformly expressed as a tuple < *Category, Name, Input, Output, QoS , Transaction*> to semantically describe the characteristics from multiple aspects. The QoS parameters include cost, execution time, availability and reliability. The transactional properties of services in this paper are *pivot(p)*, *compensatable(c)*, *retriable(r)* and their possible combination.

A. Semantic Matching

The semantic similarity $Sim(q, S_i)$ between request q and candidate service S_i is defined as

$$Sim(q,S_i) = (I(q,S_i) + O(q,S_i)) / 2 \qquad (1)$$

Where $I(q, S_i)$ indicates the semantic similarity degree of the inputs between q and S_i ; $O(q, S_i)$ indicates the semantic similarity degree of the outputs.

B. Field Theory and Flow Mechanism

A node n which has a group of semantic services $\{S_1, S_2, \ldots, S_m\}$ arouses a capability field. If the distance between request q and node n is less than TTL, then the request q is influenced by the capability field, under which there is a force f. The field force f is computed as follows:

$$f = \sum_i (QoS(S_i) * Sim(q,S_i)) \qquad (2)$$

Where $QoS(S_i)$ indicates the quality of service S_i. The transactional property of S_i chooses from $\{p, c, r, cr\}$.

As shown in Fig.1(a), n_1 and n_2 are two nodes, q is the request, L_1 and L_2 are two field lines. q is influenced by the capability field of n_1 and n_2. In Fig.1(b), $n_1,\ldots,n_7$ are several nodes which are influenced by the interest field of n_0. If n_1 has the appropriate service, then the service will flow to n_0 after negotiation. A cluster which has a group of nodes $\{n_1, n_2, \ldots, n_m\}$ arouses a complex field cf. A request q can be influenced by the complex field cf.

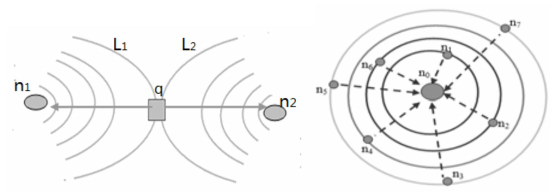
$$cf = \sum_{i=1}^{m} f(n_i) - \sum_{1 \le i < j \le m} f(n_i \cap n_j) + \sum_{1 \le i < j < k \le m} f(n_i \cap n_j \cap n_k) + \ldots + (-1)^{m-1} f(n_1 \cap n_2 \cap n_3 \cap \ldots \cap n_m) \qquad (3)$$

Where $f(n_i)$ indicates the field force f between n_i and q,

* The Corresponding Author.

978-1-4673-8590-9/15/$31.00 ©2015 IEEE

$f(n_i \cap n_j)$ represents the force f aroused by the common services in node n_i and n_j.



(a) Capability Field (b) Interest Field

Fig.1. Capability Field and Interest Field

We define two parameters α and β which reflect the influence area of capability field and interest field. The field information is updated immediately when the capacity and interest of the nodes are changed. The capability field and interest field guide the flow of request and service respectively.

C. Contribution Degree

Each node has its contribution degree. $CD_i(t)$ indicates the contribution degree of node i during t unit time. $CD_i(t)$ is computed as follows:

$$CD_i(t) = \sum_{m=0}^{t-1} a*(S_{ip}(m) - S_{ic}(m)) + \sum_{m=0}^{t-1} b(I_{ip}(m) - I_{ic}(m)) + c*\sum_{m=0}^{t-1}(a*S_{ip}(m) + b*I_{ip}(m)) \quad (4)$$

Where a, b and c are constant parameters. $S_{ip}(m)$ indicates the number of services provided by node i in the m^{th} unit time, $S_{ic}(m)$ is the number of services download by node i. $I_{ip}(m)$ denotes the number of service invocations provided by node i, $I_{ic}(m)$ indicates the number of service invocations consumed by node i.

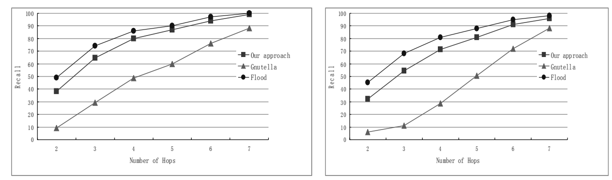
III. Experimental Analysis

In this section, we conduct a set of experiments to compare our approach with flood and Gnutella algorithms. The number of service categories is 9. There are 100 services in each category. In practice, every researcher only involved himself with a limited number of fields. So here, we supposed there are four categories on each node. One or two categories had much higher percentage than the others. The QoS attributes of the services are randomly generated expect for the cost, which is partially anti-correlated to the other QoS attributes. The value of recall is computed as follows:

$$\text{Recall} = S_{receive} / S_{total}$$

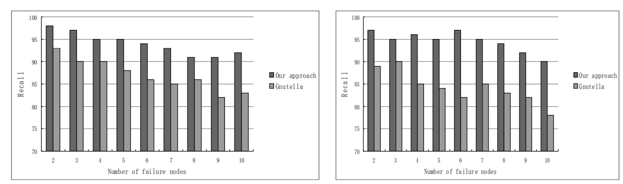
Every query was repeatedly sent five times from different nodes, and then we averaged the five results. Experiments were performed in overlay networks consisting of 1000 and 2000 nodes, in order to evaluate the performance on different scales. The TTL ranged from 2 to 7. We analyze the recall performance under different TTL and network size. The values of α and β are all fixed 3. To set up the volatile environment, we randomly select a fraction of nodes participating in the network to fail in the experiment, increasing from 2% to 10% by steps of 1%.

In Fig.2, we can see in contrast to flooding, queries were precisely sent to relevant services for response, reducing the TTL that it would take to answer the queries. In Fig.3, the recommended TTL is 7. The recall of our approach is much higher than that of Gnutella on different scales. That's because the active service replication method based on the interest field can solve the volatile network problem efficiency. Based on the contribution degree, the free-riding problem was prohibited.



(a) N=1000 (b) N=2000

Fig.2. Performance Comparison of Query Processing



(a) N=1000 (b) N=2000

Fig.3. Performance Comparison on Volatile Network

IV. Conclusion

This paper solves the service discovery approach in unstructured P2P networks. In this paper, the service description includes multiple aspects. To improve the service discovery efficiency, the field theory and contribution degree model are proposed. In our future work, we will consider the problem of efficient service composition in unstructured P2P networks.

Acknowledgment

The research is supported by the Fundamental Research Funds for the Central Universities, and National Key Basic Research Program of China under Grant(No.2010CB328106)

References

[1] L. Clement, A.Hately, C. von Riegen and T. Rogers, "UDDI Version 3.0.2," 2004, http://www.uddi.org/pubs/uddi_v3.htm.

[2] Q. He, J. Yan, Y. Yang, R. Kowalczyk and H. Jin, "A decentralized service discovery approach on peer-to-peer network," *IEEE Transactions on Services Computing*, vol.6, no.1, pp.64-75, 2013.

[3] K. Kritikos and D. Plexousakis, "Requirements for QoS-based web service description and discovery," *IEEE Transactions on Services Computing*, vol.2, no.4, pp. 320 - 337, 2009.

[4] Z.Zheng, Y. Zhang and M.R. Lyu, "Investigating QoS of real-world web services," *IEEE Transactions on Services Computing*, vol.7, no.1, pp. 1939-1374, 2014.

[5] M. Karakaya, I. Korpeoglu and O. Ulusoy, "Free riding in peer-to-peer networks," *Internet Computing*, vol.13, no.2, pp. 92–98, 2009.