Journal of Southeast University (English Edition) Vol. 22, No. 3, pp. 330 – 333 Sept. 2006 ISSN 1003—7985

Searching scheme in P2P system based on semantic overlay network

Huo Ying[1,2] Chen Zhigang[1]

([1]College of Information Science and Engineering, Central South University, Changsha 410083, China)

([2]College of Information Engineering, Shaoguan University, Shaoguan 512005, China)

Abstract: In consideration of the limitation of super-peer overlay network, the semantic information was introduced into the super-peers' organization. A novel P2P (peer-to-peer) searching model, SSP2P, was put forward. The peers in the model were organized in a natural area autonomy system (AAS) based on the small-world theory. A super-peer was selected in each AAS based on power law; and all the super-peers formed different super-peer semantic networks. Thus, a hierarchical super-peer overlay network was formed. The results show that the model reduces the communication cost and enhances the search efficiency while ensuring the system expansibility. It proves that the introduction of semantic information in the construction of a super-peer overlay is favorable to P2P system capability.

Key words: peer-to-peer; searching; semantic; super-peer; small world

An admixture distributing peer-to-peer (P2P) network which adopts super-peer overlay, such as KaZaA, Morpheus and LimeWire, can balance well between centralized P2P and decentralized P2P. But the overlay constructing protocol is poor in efficiency and depresses searching performance. How to organize the super-peer efficiency and obtain robust and expansibility at the same time has become a hot question.

In this paper we studied the searching scheme with the guidance of a semantic network and the small-world theory and put forward an efficient, robust, scalable searching model—SSP2P. The model can obtain adequate efficiency with little cost. To the best of our knowledge, this work is the first one to organize super-peers with semantic information.

1 Related Work

Li[1] used a centralized server to maintain all the information of super-peers for determining each peer's identity and neighbor aggregates. It picks up the delicacy of overlay. Pyun et al.[2] utilized E-R model to establish degree-balanceable and low-diameter random super-peer overlay networks. But it has high costs owing to the use of broadcast to estimate the network scale. Montresor[3] achieved a minimal super-peer aggregate through ceaseless replacement of small capacity peers with large capacity peers. But it can conduce super-peers to change frequently and produce abundant traffic of resource indices and has a slow convergence and great cost. Yang et al.[4] used redundancy super-peers to solve the delicacy of overlay. But it is easy to bring assembly costs and the redundancy number is two at most.

Ling et al.[5-6] also introduced semantic information into their P2P systems. But they organized all peers with semantic information. Their costs in constructing semantic overlay networks are very great, and their overlay are instable because the peers frequently join and quit.

2 Searching Model—SSP2P

2.1 Architecture of SSP2P

We put forward a searching model in a P2P network—SSP2P based on a semantic network and the small-world theory. The model is shown in Fig. 1. It is composed of three parts:

1) The underlay is composed of a varied area autonomy system (AAS). It is found upon investigation that the topology of the Gnutella exhibits typical power law and small-world characteristics. Hereby, we organize peers into varied AASs according to their physical topology based on the small-world theory, and choose the peer with the highest degree as the super-peer by considering that the higher the degree, the more contact with other peers, and thus, the greater the probability of successful search. The peers with hypo-highest degree are selected as candidate super-peers. Each AAS is managed with a super-peer. Each peer in one AAS registers its own resources describing information on the

Received 2006-04-25.

Foundation items: The National Natural Science Foundation of China (No. 60573127), Specialized Research Fund for the Doctoral Program of Higher Education (No. 20040533036).

Biographies: Huo Ying (1975—), female, graduate; Chen Zhigang, (corresponding author), male, doctor, professor, czg@ csu. edu. cn.

super-peer in the same AAS. Each AAS only maintains and administrates resources in its own area. The relative stability of the super-peer guarantees the stability of physical connection and semantic connection and cuts down the influence of the peer's joining and quitting as well as the change of share resource. It can adapt to the variability of peers and resources.

2) The mesosphere is composed of many super-peer semantic networks (SSNs). Super-peers which have the same sort of resource semantically make up of an SSN. All theses are the core of SSP2P. It is only a logistic structure. One super-peer can belong to several SSNs at the same time. Owing to the fact that the overlay network is formed from super-peers owning the same sort of resources, the query would be spread in only one SSN. So the performance of organization and discovery is improved.

3) The stratosphere is the global name server (GNS), the entrance of query. The function of GNS is to parse the name of a resource type to the IP address of an SSN. Owing to only one entrance address of one sort of SSN, the query would be directed to the area most likely to obtain the answer. And the communications brought by the peer's first joining can be reduced substantially. To avoid the disable issue led by the frequent updating of a single host's address, the GNS neither saves the address of one super-peer nor saves the IP address of one AAS. It saves the addresses of one sort of SSN. This is different from other systems.

It can obtain the whole oriented ability of a structure system and the abundance query ability of an unstructured system by adopting this structure. It can adapt to precision query because of the broadcast in SSN after inquiring about GNS. To boost the robustness of the system, the GNS can be composed of several peers, backups for each other. For the convenience of expression, we present two GNS backups for each other (see Fig. 1).

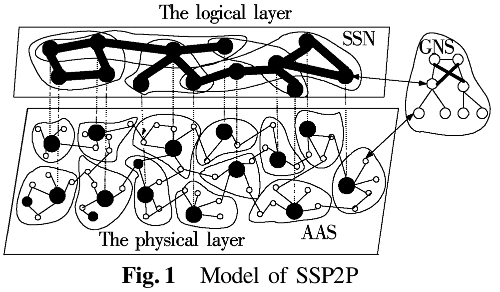

Fig. 1 Model of SSP2P

2.2 Construction of SSN

We present the notion of SSN. Suppose that the set of all super-peers in Internet is $N=\{n_i \mid 1\leqslant i\leqslant N\}$, the set of resources is $R=\{r_j \mid 1\leqslant j\leqslant M\}$. All the resources are mapped in K sorts by the sorter of GNS. The set of sorts is $C=\{c_k \mid 1\leqslant k\leqslant K\}$, namely $S: R\rightarrow C$. The set of resources of each super-peer $n_i (1\leqslant i\leqslant N)$ is $R(n_i)$. The set of sorts of each super-peer n_i is $C(n_i)$, $C(n_i)=\{S(r_j) \mid r_i\in R(n_i)\}$.

Definition 1 Super-peer semantic network. The overlay network formed by all the super-peers containing sort c is SSN(c), expressed in graph form as SSN(c) = (NC(c), E), where the set of super-peers is NC(c) = $\{n_i\in N \mid c\in C(n_i)\}$; the set of edges is $E=\{e \mid \forall p_i\in \text{NC}(c),\ \text{exit}\ n_j\in \text{NC}(c),\ \text{make}\ n_i\ \text{and}\ n_j\ \text{connected by}\ e\}$.

The process of constructing SSN is as follows:

1) To all super-peers that are connected by physical connect edges (CE), the resources of the same sort are conjunct and formed into a semantic network (SSN). The number of SSNs is set up as the number of resources.

2) If any given resource has no physical neighbor which shares the same sort of resource, then the following process is adopted:

① When a new isolated super-peer wants to join the SSN, it will send a request message with its resource information to GNS.

② GNS classifies this resource and finds the entrance address of this resource in SSN and returns a response message to the source.

③ Then the super-peer joins the SSN. It sends a request message to SSN. When a super-peer v in the SSN receives this message, it returns an affirmation message.

④ This super-peer joins the SSN through super-peer v and appends neighbors according to this SSN information.

Each super-peer for each sort of interest maintains several most-similar neighbors according to the sort of resource being saved, the history of search, the similitude of interest and the comparability of behavior. And the more similar two super-peers are, the closer they are in logical distance.

3 Scheme of Searching

In SSP2P, all the searching is carried out in a certain SSN after being parsed by GNS. The resource, if it exits in SSP2P, can only be found in the SSN that has the same sort of the resource because of the same classification method we adopt in the join and the search of the resource. A resource search algorithm after entering an SSN is presented as follows.

Algorithm 1 Resource search arithmetic

```
Search _ for _ Result(P, query, TTL)//P is the search peer, query is a data type of Call _ For _ Probe
{
    Send(query, GNS);//send the description of searching resource to GNS
    R _ ID = parse(query);//obtain the number of resource type by parse of GNS
    IP = resolve(R _ID);//obtain the IP address of entrance in SSN through GNS
    Result = Search _ in _ SSN(query, IP, TTL);//enter into the SSN and obtain the result
    Return result;
}
Search _ in _ SSN(query, IP, TTL)//the search arithmetic after enter into SSN
{
    tempresult = localquery(query);//search in AAS. If the super-peer in this AAS has the resource record searched then return the IP address of the peer which has this resource.
    if(TTL! =0)
    {
        totalresult = MergeResult(totalresult, tempresult);
        for(each of my neighbors P_i)//to each super-peer neighbor of P_i. P_i is a super-peer
    {if(P_i is not visited)
  Search _ in _ SSN(query, P_i _ IP, TTL - 1);}
    }
        else return totalresult;
}
```

4 Results and Analyses

The detail of KaZaA is not exoteric, so we compare our model with LimeWire, which also adopts super-peer structure.

We simulate 1 000 peers in a computer and produced a large quantity of files of about 10 KB each. These files belong to four sorts of semantics. Each peer administers 100 files belonging to two sorts of semantics, the proportion is 80%∶20%. So the whole SSP2P has four SSNs and each peer belongs to two SSNs.

Each time, 100 files were selected randomly. We sent the search requests for these 100 files in sequence. Each request was sent from a random peer. Each search request was performed 10 times. The average value of these 10 times was taken as the result.

1) Recall

Recall is defined as a ratio between the number of eligible answers searched for and the number of eligible answers in all on-line peers in a scheduled time. From Fig. 2, we can see that the recall of SSP2P is higher than that of LimeWire in the same TTL (time to live). When the TTL is four, the SSP2P can search a certain SSN almost wholly and its recall is 93%. While the recall of LimeWire is 96% when the TTL is six owing to the ineffective organization of super-peers. So the effective organization of super-peers can improve the recall of searching without the increase in traffic.

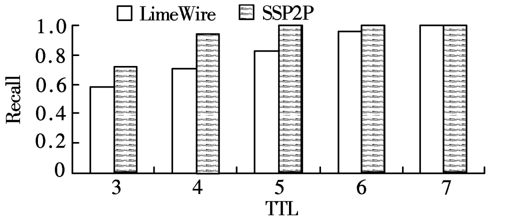

Fig. 2 Recall

2) The efficiency of bandwidth

The quantity of messages produced in finding the same answer is used to weigh the efficiency of bandwidth. The result is shown in Fig. 3. Where average message denotes the average quantity of messages produced by obtaining each eligible answer. Maximal message denotes the maximal quantity of messages produced by obtaining one eligible answer. From Fig. 3, we can see that the quantity of messages produced in SSP2P is smaller than that in LimeWire. It is mostly because of the adoption of AAS structure. When we search information in a super-peer, it corresponds to what we have searched in the whole AAS. So it greatly reduces the number of searching steps. Moreover, the search is restricted to a certain scope because of SSN. So the efficiency of bandwidth in SSP2P is better than that in LimeWire. The larger the scale is, the more distinct the superiority of SSP2P is.

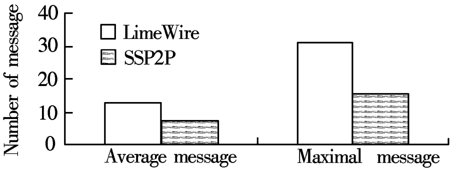

Fig. 3 Efficiency of bandwidth

3) Robust and dynamic adaptability of the model

Because of the stabilization of the super-peer, it eliminates the instability and turbulence caused by the frequent joining and quitting of peers in a self-configuration system or semantic network formed purely by peers. And the logical SSNs constituted by super-peers can adapt to the frequently joining and quitting of peers. When a certain peer wants to quit, SSP2P will only send a prune message to delete this peer. When a peer wants to join SSP2P, it can join an SSN in a constant time through GNS. Regardless of the increase or decrease in the number of peers, SSNs, as complete entities, can provide service perfectly. Isolated islands will not come into being. Moreover, the form of SSNs does not need to wait for some time after the information spreads. It can adapt to the global, dynamic, complicated next generation network.

5 Conclusion

An efficient, robust and expandable searching model in a P2P network—SSP2P is put forward in this paper. The rationality and validity of the model is proved by analysis and validation. In future work we will optimize the organization of SSN, combine the application of web service and P2P and experiment on a large scale.

References

[1] Li J. ECSP: an efficient clustered super peer architecture for P2P networks [D]. Columbia, British: University of British Columbia, 2003.

[2] Pyun Y J, Reeves D S. Constructing a balanced, (log(*n*)/loglog(*n*))-diameter super-peer topology for scalable P2P system[A]. In: *Proc of the 4th International Conference on P2P Computing*[C]. Zurich, Switzerland, 2004. 210 - 218.

[3] Montresor A. A robust protocol for building super-peer overlay topologies[A]. In: *Proc of the 4th International Conference on P2P Computing* [C]. Zurich, Switzerland, 2004. 202 - 209.

[4] Yang B, Garcia-Molina H. Designing a super-peer network [A]. In: *Proc of the 19th International Conference on Data Engineering*[C]. Bangalore, India, 2003. 49 - 74.

[5] Ling Bo, Lu Zhiguo, Ng Wee-Siong, et al. PeerIS: a peer-to-peer based information retrieval system[J]. *Journal of Software*, 2004, **15**(9): 1375 - 1384.

[6] Crespo A, Garcia-Molina H. Semantic overlay networks for P2P systems [R]. America: Stanford University, 2003.

基于语义叠加网的对等网搜索机制

霍　英[1,2]　　陈志刚[1]

([1] 中南大学信息科学与工程学院,长沙 410083)
([2] 韶关学院信息工程学院,韶关 512005)

摘要:针对目前超结点叠加网络研究中的不足,将语义信息引入到超节点的组织过程中,提出了一种新的对等网搜索模型——SSP2P. 模型中结点依据小世界理论在物理上形成自然的区域自治系统(AAS),各 AAS 依据幂规律选取各域内的超结点,超结点再根据语义关系形成多个超结点语义网(SSN),从而形成一个层次化的超结点叠加网络模型. 实验结果表明:该模型在保证系统扩展性的基础上有效地提高了搜索效率,减少了通信开销. 证明在超节点叠加网络的构建过程中语义信息的引入有利于对等网搜索性能的提高.

关键词:对等网;搜索;语义;超节点;小世界

中图分类号:TP393