Improved optimal Searching Techniques in Distributed P2P networks using Guided Protocol Approach

Mr. S.Bhaskaran ,
Assistant Professor II,

Dept. of Computer Applications, SASTRA University, Thanjavur , INDIA

E mail : boss.sngc@gmail.com

Summary

peer-to-peer (P2P) networks establish loosely coupled application-level overlays on top of the Internet to facilitate efficient sharing of resources. They can be roughly classified as either structured or unstructured networks. Without stringent constraints over the network topology, unstructured P2P networks can be constructed very efficiently and are therefore considered suitable to the Internet environment. However, the random search strategies adopted by these networks usually perform poorly with a large network Size. In this paper, we seek to enhance the search performance in unstructured P2P networks through exploiting users' common interest patterns captured within a probability-theoretic framework based on DHT . A search protocol and a routing table updating protocol are further proposed in order to expedite the search process through self organizing the P2P network into a small world.

Keywords: ***P2P , DHT***

1.Introduction

PEER-TO-PEER (P2P) networks establish loosely coupled application-level overlays on top of the Internet to facilitate efficient sharing of vast amount of resources. One fundamental challenge of P2P networks is to achieve efficient resources discovery. In the literature, many P2P networks have been proposed in an attempt to overcome this challenge. Those networks can be largely classified into two categories, namely, structured P2P networks based on a distributed hash table (DHT) and unstructured P2P networks based on diverse random search strategies (e.g., flooding). Without imposing any stringent constraints over the network topology, unstructured P2P networks can be constructed very efficiently and have therefore attracted far more practical use in the Internet than the structured networks. Peers in unstructured networks are often termed blind, since they are usually incapable of determining the possibility that their neighbor peers can satisfy any resource queries. An undesirable consequence of this is that the efficiency of distributed resource discovery techniques will have to be compromised. In practice, resources shared by a peer often exhibit some statistical patterns. The fundamental idea of this Paper is that the statistical patterns over locally shared resources of a peer can be explored to guide the distributed resource discovery process and therefore enhance the overall resource discovery performance in unstructured P2P networks. Three essential research issues have been identified and studied in this Paper in order to save peers from their blindness. For ease of discussion, only one important type of resources, namely, data files will be considered in this Paper.

2.Related Work

The first research issue questions the practicality of modeling users' diverse interests. To solve this problem, we have introduced the model based on a general probabilistic modeling tool termed Condition Random Fields (CRFs). With this, we are able to estimate the probability of any peer sharing a certain resource (file) fj upon given the fact that it shares another resource (file) fi. This estimation further gives rise to an interest distance between any two peers. The second research issue considers the actual exploration of users' interests as embodied by the model. To address this concern, a greedy file search protocol is presented in this project for fast resource discovery. Whenever a peer receives a query for a certain file that is not available locally, it will forward the query to one of its neighbors that posses the highest probability of actually sharing that file. The third research issue has been highlighted with the insight that the search protocol alone is not sufficient to achieve high resource discovery performance. This project proposes a routing table updating protocol to support our search protocol through self organizing the whole P2P network into a small world. Different from closely related research works that are also inspired by the small-world model, in order to reduce the overall communication and processing cost, in this project, the updating of routing tables are driven by the queries received by each peer. In a P2P network, queries handled by a peer may be satisfied by any peer in the network with uneven probability. This uneven distribution has a significant impact on our routing table updating protocol and is to be demonstrated in this project. To ensure the effectiveness of our protocol, a filtering mechanism is to be introduced to mitigate the impact of uneven updating.

3.Guided Search

This section presents and analyzes the guided search solution that we proposed for resource discovery in unstructured P2P networks.

3.1 P2P Network Architecture

This paper considers a loosely connected P2P network. We use p to denote a single peer in the network. P is further utilized to denote the set of all peers in the network. The main type of resource, namely, a data file, is represented by f. For every peer p, Fp is used to represent the group of files shared by p. In order to conduct distributed search over the P2P network, every peer p maintains locally a list of neighbor peers. This list serves as the routing table for peer p, denoted by Rp. There is an upper bound Br on the size of any routing table Rp, while the size is measured in terms of the number of entries in Rp. An entry Ep of Rp is a tuple of two elements: < p0; f0 > . It represents a link from peer p to another peer p0 that shares file f0. In order to locate (or discover) any file under request, the user of a peer p, denoted by up, sends out a query to the network. A query that originated in peer p is represented by qp and is a tuple of six elements: < p;f;hq;TTL; ts; te > . Here, p stands for the peer that issued the query qp. f is the file requested by the query. hq records the search history , which is a list of peers that have processed the query previously, including peer p itself. In order to prevent a query from incurring too much traffic in the network, timeto-live (TTL) in a query defines an upper bound on the allowable size of hq. ts refers to the time when the query is issued, while te is the time when the query is completed. A query is completed successfully if the requested file f has been identified. On the contrary, the query is failed if the size of hq exceeds the TTL. Upon receiving a query q, a peer p needs to perform several basic operations: 1) append itself to the search history hq, 2) search the requested file f among its locally shared files (i.e., local repository), and 3) forward the query to one of its neighbor peers. Each forwarding operation is termed a hop. At the time when query q is finished, the number of hops NOP becomes an important measure of the search performance. In practice, we hope that NOP for average search tasks could be as low as possible, which essentially implies that only a small group of peers will be involved in processing any query. To summarize, there are two widely used performance metrics for resource discovery in P2P networks: NOP and search success rate. Search success rate refers to the proportion of queries that have been successful among all the queries issued by network users.

3.2The Search Protocol

In this section, a file search protocol is presented to regulate the activities of every peer p in a P2P network upon receiving a query q ¼< p;f;hq;TTL; ts; te > . The local decision involved in the search protocol demands three main types of information: 1) the search history hq stored in the query q, 2) the routing table Rp of the peer p that handles the query q, and 3) the UIM. hq and Rp, which are readily available in many P2P networks. UIM, however, is a new source of information introduced in this paper. The workflow of our search protocol is shown in Fig. 1. It follows essentially a greedy search strategy. Upon receiving a query q, a peer p will try to find, among its neighbor peers, the one that is most probable of having the requested file fq according to UIM. In case that the peer chosen to forward the query already exists in the search history hq, another peer with a relatively high probability of satisfying the query will be chosen instead. If all neighbor peers have been visited before, then the peer that enjoys the highest probability will be selected again for query processing.

1. Add peer p to the search history hq
2. If filw fq is locally stored in peer p, inform peer pq that search is successful
3. Else if the size of hq is greater than TTL, inform peer pq that search is failed
4. Else
a. Order all the routing entries Ep=<pv,fv> of Rp decreasingly based on pr(fq|fe)
b. Iterate over every routing entry Ep=-<pe,fe> starting from the one with highest pr(fq|fe)
c. If no entry is chosen at step 4.b forward query q on to pe with highest pr(fq|fe)

Fig 1. The guided search protocol.

4.Protocol for Updating Routing Tables

Whenever the search process driven by any query q ¼< pq; fq; hq;TTL; ts; te > is completed successfully, a new routing entry Ep ¼< pi; fq > , indicating that peer pi shares the queried file fq, will be temporarily added into the routing table Rp of every peer p recorded in the search history hq. If Rp is not full, no entries of Rp will be removed. Otherwise, the size of Rp will be reduced to below Br by deleting one or more selected entries.

1. Repeat for every peer p in the search history hq
2. Add a new entry Ep=<pi,fq> to the routing table Rp of peer p
3. While the size of Rp is greater than Br
   a. With respect to each entry Ep=<pe,fe> , calculate the interest distance d(pe,p)
   b. Select an entry Ep=<pe,fe> with probability proportional to d(pe,p)
   c. Remove Ep from the routing table Rp

Fig. 2. The Protocol for updating routing tables

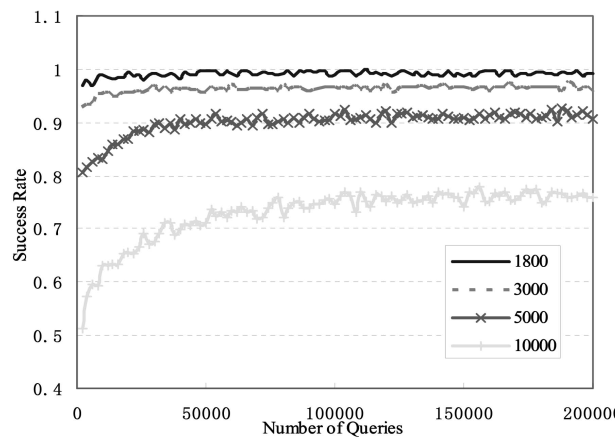
updating routing tables are summarized as follows:

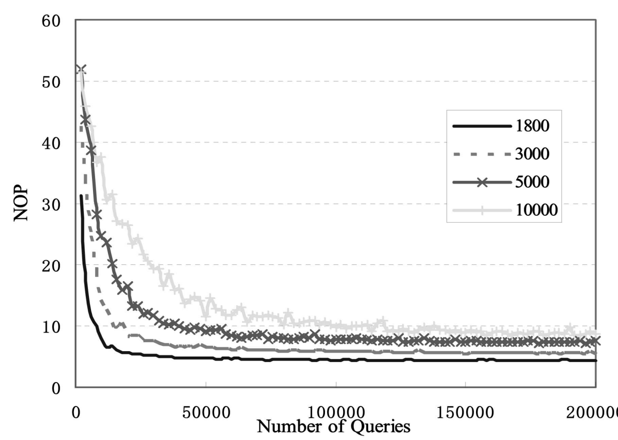
. **The LRU strategy.** The routing entry that is least recently used to forward queries will be dropped.

. **The ECCR scheme** . With a certain probability Pre, the least recently used routing entry will be dropped. Otherwise, the neighbor peer p0, which has the longest interest distance from peer p, will be removed from Rp.

. **The distance-centric (DC) strategy.** Either the peer p0, which has the longest interest distance from peer p, or another peer p00, which has the second longest distance, will be removed from Rp of peer p, depending on a probability Prd.

5.Performance Measure





In this section, the effectiveness of our protocols under varied routing table sizes Br will be evaluated first. All experiments are performed in a network of 300 peers, each sharing 10 files. Fig. 8 depicts the search performance obtained when Br is equal to 10, 15, 20, and 25, respectively. As shown in Fig. 8, with increasing Br, the search performance will be consistently improved, both for NOP and success rate. When Br _ 20, nearly all queries will be completed successfully. Moreover, when Br hits 25, each search process driven by a query, on the average, comprises only 4.75 hops.

Real P2P networks usually contain far more than 300 peers. In order to examine the capability of our protocols in handling large-scale networks, a group of experiments have been performed, with the number of peers in the network ranging from 1,800 to 10,000. Each peer in these experiments maintains locally two files. Files shared by separate peers do not overlap. To achieve a good search performance in networks of such a scale, the routing table size Br is set to 40. For all queries, TTL ¼ 200. The simulation continues until 200,000 queries have been completed.

6. Conclusion

In this paper, it is shown that the search performance in unstructured P2P networks can be effectively improved through exploiting the statistical patterns over users' common interests. Our solution toward enhancing search performance was presented in three steps:

1. A Guided Search has been introduced in order to capture users' diverse interests within a probability-theoretic framework. It leads us to introduce a concept of interest distance between any two peers.
2. Guided by a greedy protocol it has been proposed to drive the distributed search of queried files through peers' local interactions.
3. Finally, a routing table updating protocol has been proposed to manage peers' neighbor lists. With the help of a newly introduced filtering mechanism, the whole P2P network will gradually self organize into a small world.

References

[1] S. Tewari and L. Kleinrock, "Optimal Search Performance in Unstructured Peer-to-Peer Networks with Clustered Demands," IEEE J. Selected Areas in Comm., vol. 25, no. 1, 2007.

[2] M. Khambatti, K. Ryu, and P. Dasgupta, "Structuring Peer-to-Peer Networks Using Interest-Based Communities," Proc. Int'l Workshop Databases, Information Systems and Peer-to-Peer Computing (P2PDBIS), 2007.

[3] E. Cohen, A. Fiat, and H. Kaplan, "Associative Search in Peer-to-Peer Networks: Harnessing Latent Semantics," Proc. IEEE INFOCOM, 2008.

[4] H. Jin, X. Ning, and H. Chen, "Efficient Search for Peer-to-Peer Information Retrieval Using Semantic Small World," Proc. Int'l Conf. World Wide Web (WWW '06), pp. 1003-1004, 2006.

[5] H. Zhang, A. Goel, and R. Govindan, "Using the Small-World Model to Improve Freenet Performance," Computer Networks, vol. 46, pp. 555-574, 2004.

Bhaskaran S obtained his Bachelors degree in Electronics from Bharathiar University during 1996 and Masters Degree in Computer Applications during 1999. He has completed his M.Tech degree from Anna University Coimbatore during 2007. He is currently working in SASTRA University Thanjavur.