Volume 3, Issue 8, August 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Optimizing Search Performance in Unstructured Peer to Peer Networks Using Enhanced Search Protocol

E.Ramesh Babu [*]
Department of CSE
AITS, Tirupati
India

K.Sumalatha
Department of CSE
AITS, Tirupati
India

B.Rupa Devi
Department of IT
AITS, Tirupati
India

K.Chandra Babu
Department of CSE
AITS, Tirupati
India

Abstract— Unstructured Peer-to-peer (P2P) networks are popular in mass network. These networks can be constructed very efficiently without precise rules and therefore considered suitable to the Internet environment. But, the search techniques used in these networks usually perform poorly with a large network size. To optimize the search performance in unstructured P2P networks through exploiting the similarity among the participating peers in the network. A search protocol is further proposed in order to expedite the search process through self organizing the P2P network into a small world. Both theoretical and experimental analyses are conducted and demonstrated the effectiveness and efficiency of the approach.

Keywords— Peer to Peer Networks, Optimizing, Peer similarity, Search.

I. INTRODUCTION

Peer-to-peer (P2P) networks have been widely deployed in the internet[1]. These networks are largely classified into two categories, namely Structured Peer to Peer and Unstructured Peer to Peer. The structured P2P networks, constructed based on distributed hash tables(DHT)[2]. Unstructured peer to peer networks based on 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 attracted far more practical use in the internet. They provide many services like file sharing, information retrieval and media streaming. P2P applications are popular because they provide low entry barriers. Prior studies (e.g.,[3]) reveal that P2P services may dominate up to around 20 percent of internet traffic. Object search is the essential building block in several P2P applications (e.g., file sharing, media streaming). Gnutella is the popular P2P search protocol, because Gnutella networks are unstructured and the peers participating in the network are connected to one another randomly as shown in Fig.1, and the peers search the objects in the networks through message flooding. The basic 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 peer to peer networks.

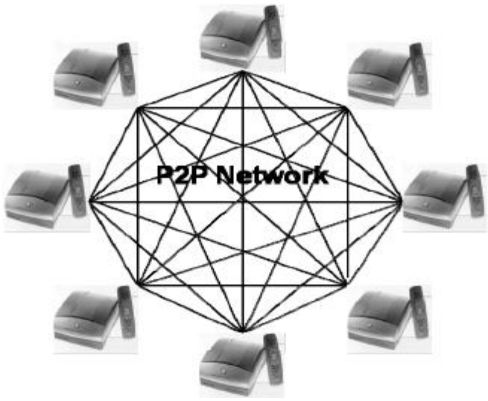

Fig.1 Unstructured Peer to Peer Network

II. RELATED WORK

In this section we mainly discuss about P2P networks that aim to exploit the similarity of participating peers. Li and Wu in [4] and Zhu and Hu in [5] conducted surveys on searching technique in P2P network. pSearch[6] and SSW[7] are content based P2P networks providing semantic search. Similar to most P2P networks based on distributed hash tables (e.g., Chord[8]) in pSearch and SSW, the participating peers need to maintain foreign indices, that is, the indices of the object stored in the remote peers. To locate an object, a requesting peer routes a message toward peer responsible for the key subspace where the object is indexed. In contrast, we can construct an unstructured peer to peer networks where the participating peers need not to organize themselves into a rigid, deterministic topology structure, considerably reducing the maintenance overhead of overlay topology. Unlike pSearch [6] and SSW[7], the peers in the network hosts the objects of interest and maintain no foreign indices, eliminating storage bandwidth overheads for publishing and managing such indices. However, due to space complexity, we concentrate on optimizing searching protocol. We intend to offer semantic searches exploiting the similarity of participating peers. Finally, prior studies on maximizing the coverage of a flooding-based search can be found in the literature (e.g., [9], [10]). While maximizing the coverage of a query message may not guarantee the efficiency and effectiveness of searches, ours aim at exploiting the similarity of participating peers.

III. EXPLOITING PEER SIMILARITY

A. Definition

The Let V be the set of peers participating in a P2P network. The peer similarity function measures the degree of the similarity between any two peers u€V and v€V in the system.

$$F: V \times V \rightarrow IR_0^+$$

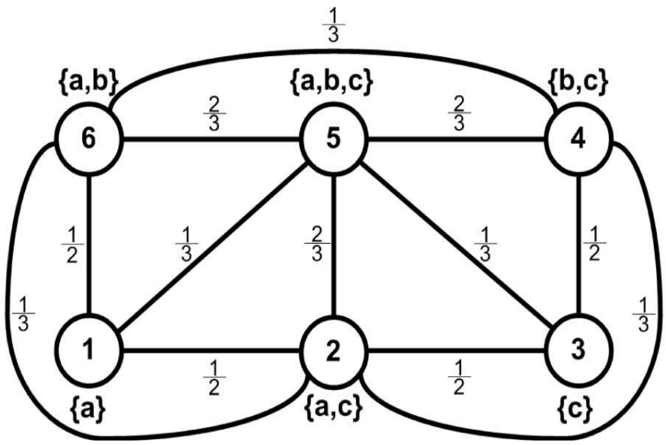

Fig. 2 An Example of a Peer Similarity Graph.

The Graph G = (V,E), here V={1,2,3,4,5,6}. Peers 1,2,3,4,5 and 6, respectively, host set of objects O_1={a}, O_2={a,c}, O_3={c},O_4={b,c},O_5={a,b,c},O_6={a,b}. Any two peers u and v have an edge in E if both peers share at least one common object. That is, $F(u,v) = \left| O_u \cap O_v / O_u \cup O_v \right|$, the value nearby an edge (u,v) indicates F(u,v).

For example, in [11], F(u,v) is defined as the inverse of the cosine angle of the two summarized latent semantic vectors representing any two vectors u and v in a p2p network where each element in a summarized vector for any peer i calculates the total frequency of the corresponding keyword appearing in the data items stored in i.

We note in this paper that designing peer similarity function is orthogonal to, but is out of the scope of, our study. Our proposal may refer to the peer similarity functions presented in the literature (e.g., [11], [12], [13], [14]). To guide the design and performance analysis of semantic unstructured P2P networks given any similarity function F(.). with the peer similarity graph, the theoretical performance analysis for the proposals in [11],[12] can be performed.

TABLE I
NOTATIONS FREQUENTLY USED

NOTATIONS	DESCRIPTION
V	Set of participating peers
G=(V,E)	Peer similarity graph
F(u,v)	Similarity function measuring the level similarity between vertices u and v

IV. ENHANCED SEARCH PROTOCOL

The second research issue states that how the search protocol is optimized.

Consider any query Q requesting an object O. Denote the peer issuing Q and the peer hosting O by s and d, respectively. Instead of blindly flooding query messages as in the typical Gnutella protocol, our search protocol conceptually operates as follows. Among the neighbouring peers in I_s and $Ø_s$, s selects the peers, each t Є I_s U $Ø_s$ has F(t,d) > F(s,d). s then broadcasts Q to the peers. Upon receiving Q, t performs similarly by selecting its neighbours who are more similar to d than t, hoping that there is atleast one overlay path P towards the destination peer d, such that any two neighbouring peers u and v on P (v receives Q sent from u) have F (v,d) > F (u,d).

Clearly the path P intends to explore progressively the peers that are likely to resolve the query Q. Here we tend to describe an Algorithm: peer t forwards a query message Q.

Input: I_t, $Ø_t$ and a query message Q
Step 1: if t receives Q for the first time and
Q.TTL ≤ MAXTTL then
Step 2: for each t'Є I_t U $Ø_t$ do
Step 3: if F (t',Q) > F (t,Q) then
Step 4: Q.TTL ← Q.TTL+1;
Step 5: t forwards Q to t'

The algorithm details the search protocol. Notobly as the peer,t,which forward the query, Q, is unware of the peers hosting the requested object, t simply broadcasts Q to its neighbours that are mostly to resolve the Q. Here by F(t,Q), similar to Definition, we refer to measure of the similarity level between a peer, t and a query message,Q.

Similar to navie Gnutella protocol, each query message can be relayed for a predefined maximum number of times. Possibly, a query message without being relayed for MAXTTL times can reach a dead end peer that cannot relay the message further because the neighbours of the dead end cannot improve F(.).

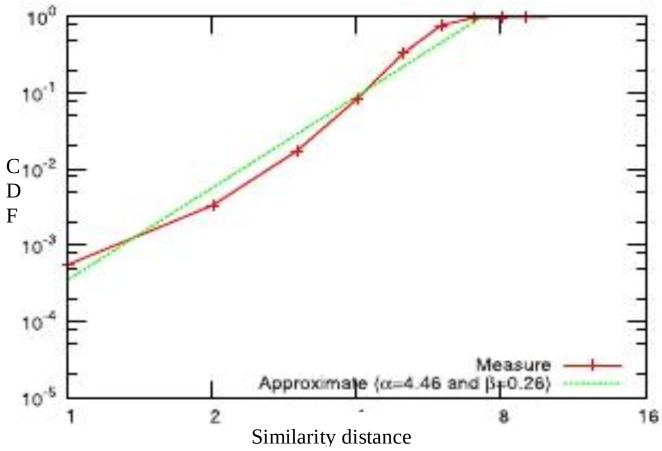

Fig.3 The log-log plots of the distribution of $s_u(d)$ for the eDonkey data set.

V. PERFORMANCE EVALUATION

B. Empirical Data Set

We have investigated the empirical data set, the files shared by edonkey users , and concluded that the data set exhibits the power law property. Before we study such property, we define the following,

The scope of a peer uЄV in a peer similarity graph G=(V,E) within a given peer similarity distance d is defined as,

$$Su(d)=\{v \mid D(u,v)\leq d, V\ v\text{Є}V\}.$$

For example, in Fig. 2, $S_1(1)$ = {1,2,5,6} and $S_2(2)$ = {1,2,3,4,5,6}. Thus , $|S_1(1)=4|$ and $|S_2(2)=6|$.

The edonkey data set mainly includes the set of files shared by each participating peer. we thus define the peer similarity function as $F(u,v) = |O_u \cap O_v / O_u \cup O_v|$, where O_u and O_v are the sets of files shared by peers v and u.

C. Analytical Results

We have observed that the empirical data set exhibits the power-law property in terms of distribution of $|S_u(K)|$. Thus our analytical model is based on such distribution.

Given a peer similarity graph G=(V,E) and two positive constants α and β, the set of nodes V, follows the α power-law Similarity distance expansion if for each node uЄV, the number of nodes with a similarity distance no more than d to u is $|S_u(d)|=\beta d^{\alpha}$. We summarize the analytical results as follows,

- Our proposal progressively and effectively exploits the peers to resolve a query. More specifically, let p be the overlay path originating from the requesting node s to the destination node d that can provide the requested object.
- The path length of P is no more than $O(\ln^{c1}N)$ in a constant probability , where $1 < c_1 < 2$.
- P can be successfully constructed in a constant probability approximating 100 percent.

VI. SIMULATION REULTS

The hop counts of routing the successful queries, despite the unsuccessful ones, are shown in Fig. 4 for GES, SocioNet.

As shown in Fig. 4, our proposal performs very well as most queries can be forwarded to their destination in no more than 10 hops (≈ ln N), validating our analytical results. In contrast , some queries in GES and SocioNet may take more than 25 and 40 hops , respectively. Notably, GES performs better than SocioNet in terms of hop count of routing a query message. SocioNet performs poorly, as the resultant overlay quality may trap in a local optimum, where as GES depends on random walks to discover similar peers that may help bypass a local optimum. However, GES does not guarantee exploiting the most similar neighbours for any participating peer. Overall both GES and SocioNet do not perform well, that is, peers cannot efficiently exploit the similarity of participating peers in structuring the overlay geometry and effectively take advantage of peers similarity in routing the queries.

We implement the search protocol based on blind flooding (i.e., Gnutella) over GES and SocioNet. Notably, in this study, the TTL values for a query message equal to 3 and 7 are investigated . With a TTL of 7, a query message reaches all peers in the system in this study.

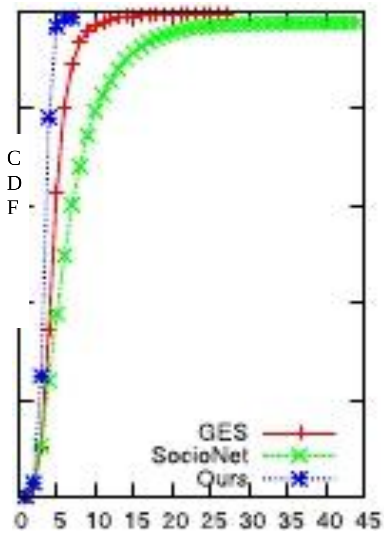

Fig. 4 The query hop count.

VII. CONCLUSIONS

In this study, we have presented an unstructured P2P network with rigorous performance guarantees to enhance search efficiency and effectiveness. In constant probability, a querying peer takes O($\ln^c$ N) hops (where c is a small constant) to reach the destination node capable of resolving the query, where as the query message can progressively and effectively exploit the similarity of the peers. The query can be successfully resolved in an approximate probability of 100 percent. Notably, the theoretical analysis further reveals that the competitive decentralized solutions (e.g., those in [11], [12], [13], [14]) do not perform well as the hop count of routing a query message in such networks, considering the exploitation of the similarity of participating peers, is in the polynomial of system size N. The Search protocol we have suggested in this paper, which takes the advantage of similarity of peers exploited by overlay network can considerably reduce the search traffic. Peers participating in a Peer-to-Peer network are often heterogeneous in terms of their network bandwidth, storage space, and computational capability.

REFERENCES

[1] S. Androutsellis-Theotokis and D. Spinellis, “A Survey of Peer-to-Peer Content Distribution Technologies,” ACM Computing Surveys, vol. 36, no. 4, pp. 335-371, 2 004.

[2] A. Rowstron and P.Druschel, “Pastry: Scalable, Distributed Object Location and Routing Routing for Large Scale Peer-to-Peer Systems,” Proc IFIP/ACM Int’1 Conf.Distributed Systems Platforms (middleware), 2001.

[3] S. Sen and J. Wang, “Analyzing Peer-to-Peer Traffic Across Large Networks,” IEEE/ACM Trans. Networking, vol. 12, no. 2, pp. 219-232, Apr. 2004.

[4] X. Li and J. Wu, “Searching Techniques in Peer-to-Peer Networks,” Handbook of Theoretical and Algorithmic Aspects of Ad Hoc, Sensor, and Peer-to-Peer Networks, pp. 613-642, Auerbach, 2006.

[5] Y. Zhu and Y. Hu, “Semantic Search in Peer-to-Peer Systems,” Handbook of Theoretical and Algorithmic Aspects of Ad Hoc, Sensor, and Peer-to-Peer Networks, pp. 643-664, Auerbach, 2006.

[6] C. Tang, Z. Xu, and S. Dwarkadas, “Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks,” Proc. ACM SIGCOMM ’03, pp. 175-186, Aug. 2003.

[7] M. Li, W.-C. Lee, A. Sivasubramaniam, and J. Zhao, “SSW: A Small World-Based Overlay for Peer-to-Peer Search,” IEEE Trans. Parallel and Distributed Systems, vol. 19, no. 6, pp. 735-749, June 2008.

[8] I. Stoica, R. Morris, D. Liben-Nowell, D.R. Karger, M.F. Kaashoek, F.Dabek, and H.Balakrishnan,“Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications,” IEEE/ACM Trans. Networking, vol.11,no.1,pp.17-21,Feb.2003.

[9] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, “Making Gnutella-like P2P Systems Scalable,” Proc. AC SIGCOMM ’03,pp.407-418,Aug.2003.

[10] S. Schmid and R. Wattenhofer, “Structuring Unstructured Peer-to- Peer Networks,” Proc. Int’l Conf. High Performance Computing (HiPC’07), pp.432-442,Dec.2007.

[11] Y. Zhu and Y.Hu, “Enhancing Search Performance on Gnutella- Like P2P Systems,” IEEE Trans. Parallel and Distributed Systems, vol. 17, no.12, pp. 1482-1495, Dec. 2006

[12] K.C.-J. Lin, C.-P. Wang, C.-F.Chou, and L. Golubchik, “SocioNet: A Social-Based Multimedia Access System for Unstructured P2P Networks,” IEEE Trans. Parallel and Distributed Systems, vol. 21, no.7, pp.1027-1041, July 2010.

[13] H. Jin, X. Ning, and H. Chen, "Efficient Search for Peer-to- Peer Information Retrieval Using Semantic Small World," Proc. 15th ACM Int'l Conf. World Wide Web (WWW '06), pp. 1003- 1004, May 2006.

[14] W. Ke and J. Mostafa, "Strong Ties versus Weak Ties: Studying the Clustering Paradox for Decentralized Search," Proc. Seventh Int'l Workshop Large-Scale Distributed Systems for Information Retrieval (LSDS-IR'09), pp. 49-56, July 2009.

