An Efficient Topology Search Mechanism for Unstructured Peer-To-Peer Networks

B. Vijayanandh,[1] P.K.Yogeshwaran,[2] R. Sivanesan,[3] M.New Begin[4]

Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College

I.   Introduction

Unstructured peer-to-peer (P2P) file-sharing networks are popular in the mass market. As the peers participating in unstructured networks interconnect randomly, they rely on flooding query messages to discover objects of interest and thus introduce remarkable network traffic. Empirical measurement studies indicate that the peers in P2P networks have similar preferences, and have recently proposed unstructured P2P networks that organize participating peers by exploiting their similarity. The resultant networks may not perform searches efficiently and effectively because existing overlay topology construction algorithms often create unstructured P2P networks without performance guarantees.

Thus, I propose a novel overlay formation algorithm for unstructured P2P networks. Based on the file sharing pattern exhibiting the power-law property, our proposal is unique in that it poses rigorous performance guarantees. Theoretical performance results conclude that in a constant probability, 1) searching an object in our proposed network efficiently takes N hops (where c is a small constant), and 2) the search progressively effectively exploits the similarity of peers. In addition, the success ratio of discovering an object approximates 100 percent. We validate our theoretical analysis and compare our proposal to competing algorithms in simulations.

Based on the simulation results, my proposal clearly outperforms the competing algorithms in terms of 1) the hop count of routing a query message, 2) the successful ratio of resolving a query, 3) the number of messages required for resolving a query, and 4) the message overhead for maintaining and formatting the overlay.5.) Resultant set of finite N hops for data sharing in unstructured peer-to-peer (P2P) networks.

Keywords: *Topology structures, Search, Load Sharing, unstructured P2P networks*

II.   Related Work

In this study, I am interested in optimizing the search performance in Gnutella-like unstructured P2P networks. Existing orthogonal techniques in the literature for improving search performance in unstructured P2P networks include indexing (e.g., [5], [6], [7]), replications (e.g., [8], [9]), super peer architectures (e.g., [10], [11], [12], [13]), and overlay topologies (e.g., [14], [15]), among others. In this Paper, I primarily study the overlay topology formation technique for unstructured P2P networks, aiming to enhance search efficiency and effectiveness. In particular, as recent measurement studies [16], [17], [18] show that peers are likely to resolve the queries issued by the peers sharing the common preferences my study intends to organize the participating peers to exploit their similarity [19], [20]. Overlay construction algorithms intending to exploit the similarity of peers for enhancing search performance can be found in the literature (e.g., [21], [22], [23], [24], [25], [26], [27], [28]). In [21], [22], [23], [24], distributed algorithms are used for constructing unstructured P2P networks (or semantic overlays), such that peers having common preferences cluster together. In addition, the proposals in [21], [22], [23], [24] suggest mixing a naive random network [29] in the P2P network to minimize the overlay path length between any two peers to reduce the query response time. (In this paper, we refer to a random network with N nodes as the random graph formally defined in [29] that is, in the graph, a node links to another node in a probability of 1 N.)

In contrast to [21], [22], [23], [24], the studies in [25], [26], [27], [28] suggest organizing the P2P network as a semantic small-world random graph. Here, semantic small-world networks refer to the probability of peer j being the neighbor of peer i increasing if j shares more common interests with i. In this paper, I first observe that existing P2P file sharing networks (e.g., eDonkey) exhibit the power-law file sharing pattern. Based on such sharing pattern, we present a novel overlay construction algorithm to enhance the efficiency and effectiveness of searches in unstructured P2P networks.

Unlike [21], [22], [23], [24], the works in, for example, [25], [26], [27], [28], create small-world-based semantic overlays to enhance search efficiency and effectiveness. For example, in the semantic small-world P2P network [27], peer i is likely to connect to peer j, instead of to another peer k, if i and j have more common interests. However, Chen et al. [27] heavily rely on a centralized entity to help format the P2P networks. More specifically, in [27], a centralized server is employed to compute a probabilistic model based on the principle of

maximum entropy [16], which facilitates to estimate the probability of sharing an object o1 when conditioned on another shared object o2. As the number of shared objects in a P2P network is numerous, learning such a probabilistic structure (i.e., the feature functions) is computationally intensive, and the proposal in [27] thus demands substantial computation resources. Additionally, the centralized server requires disseminating the estimated probabilistic model to all participating peers. As a result, although the idea presented in [27] is interesting, the proposed solution may suffer from the performance bottleneck and introduce a single point of failure. Moreover, the probabilistic model in [27] possesses no analytical properties to enable tractable analysis. In contrast to [27], our proposal does not depend on any centralized server for computation-intensive tasks and guarantees rigorous performance results.

Prior efforts in developing distributed overlay formation algorithms for semantic small-world P2P networks can be found in the literature (e.g., [25], [26], [28]). As any peer in these proposals often selects its neighbors "heuristically," the proposals in [25], [26], [28] mainly depend on simulations for performance investigation. Although these proposals are pragmatic for dynamic, large-scaled distributed environments, they offer no rigorous performance guarantees. In contrast, our network formation algorithm performs very well with rigorously mathematical guarantees and is unique in that in a constant probability, a querying peer takes N hop count (where c is a small constant) to reach the destination peer by progressively and effectively exploiting the similarity of peers on the search path. Additionally, the probability of successfully discovering a requested object in our proposal is approximately 100 percent.

III. My Proposal

Consider any given unstructured P2P network P, where V is the set of participating peers, and E is the set of overlay connections linking the peers in V. The peers in G may be interconnected randomly. Our goal is to restructure G to satisfy the following properties: C1. (High clustering). Each peer u connects max peers in V, and these neighbors, selected among the peers in V, are the top-max nodes most similar to v. C2. (Low diameter). Consider any two distinct peers u and v in V. There should exist at least one overlay path P connecting u and v, and the hop count of P should be as small as possible, enabling a query message to be rapidly propagated from u to v. Here, the hop count of an overlay path P means the number of overlay links in P. C3. (Progressive). Let s be the peer that issues a query, and d be the peer that can resolve the query. There should exist an overlay path P connecting s and d such that for any two neighboring peers u and v on P, upon receiving a query message, u forward the message to v that is more similar to d than u. In Section 4, we will present that searches in an overlay with Properties C1, C2, and C3 are efficient and effective. More precisely, in a probability of approximately 100 percent, a query message in such an overlay takes a few hops by taking advantage of similar peers to reach a peer that can successfully resolve the query. Although the overlay formation algorithms in the literature (e.g., [21], [22], [23], [24]) construct networks with high clustering and low diameter, they possess no progressive, and thus cannot efficiently and effectively take advantage of short paths comprising similar peers toward query resolvers.

In the following, we first discuss peer similarity graphs used to quantify the similarity level of any two distinct peers in the system. Section 3.2 describes our proposed overlay formation algorithm. We then detail in Section 3.3 the search protocol implemented on top of our constructed overlay network. We conclude in Section 3.4 the analytical results of my proposal.

3.1 Peer Finding
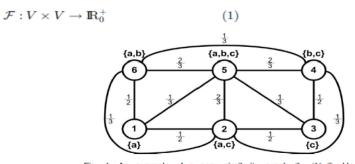Let V be the set of peers participating in a P2P network.

$$\mathcal{F} : V \times V \to \mathbb{R}_0^+ \qquad (1)$$

Fig. 1. An example of a peer similarity graph $\mathcal{G} = (V, \mathcal{E})$. Here, $V = \{1, 2, 3, 4, 5, 6\}$. Peers 1, 2, 3, 4, 5, and 6, respectively, host sets of objects $\mathcal{O}_1 = \{a\}$, $\mathcal{O}_2 = \{a, c\}$, $\mathcal{O}_3 = \{c\}$, $\mathcal{O}_4 = \{b, c\}$, $\mathcal{O}_5 = \{a, b, c\}$, and $\mathcal{O}_6 = \{a, b\}$. Any two peers u and v have an edge in $\mathcal{E}$ if both peers share at least one common object. That is, $\mathcal{F}(u, v) = \frac{|\mathcal{O}_u \cap \mathcal{O}_v|}{|\mathcal{O}_u \cup \mathcal{O}_v|}$, $\forall u \neq v \in V$, and if $\mathcal{F}(u, v) > 0$, then $(u, v) \in \mathcal{E}$. The value nearby an edge (u, v) indicates $\mathcal{F}(u, v)$.

For example, in [21], P is defined as the inverse of the cosine angle of two summarized latent semantic vectors representing any two peers 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. I note in this paper that designing peer similarity functions 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., [21], [24], [26], [28], and [27]).

3.2 Peer Buffer Graph Search

G = (V, E) is a graph where V denotes the set of participating peers, and E is the set of edges. Each edge (V, U) indicates that peers u and v are similar to some extent.

As previously mentioned, each peer u will connect to the peers selected among all peers in figure that are most similar to u; that is, u intends to satisfy Property C1. Let E be the set of neighbors that u currently maintains in the network G= (V, U) as,

$$\mathcal{A}_{current} = \frac{\sum_{v \in \mathcal{I}_u} \mathcal{F}(u,v)}{|\mathcal{I}_u|}, \qquad (2)$$

Representing the averaged peer similarity value of u and v neighbors in E. By exploiting the peers most similar to u, u seeks a peer E 2 V in the figure and invites w as its neighbor such that

$$\mathcal{A}_{update} \geq \mathcal{A}_{current}, \qquad (3)$$

where

$$\begin{cases} \mathcal{A}_{update} = \dfrac{\sum_{v \in \mathcal{I}_u \cup \{w\}} \mathcal{F}(u,v)}{|\mathcal{I}_u \cup \{w\}|}, & \text{if } |\mathcal{I}_u| < \max_u; \\[2mm] \mathcal{A}_{update} = \dfrac{\sum_{v \in \mathcal{I}_u \cup \{w\} - \{q\}} \mathcal{F}(u,v)}{|\mathcal{I}_u \cup \{w\} - \{q\}|}, & \text{otherwise}; \end{cases} \qquad (4)$$

And q (arg min v2u) in P hops is the peer among the neighbors of u that is most dissimilar to u. possibly, u cannot accommodate any new neighbor if u has reached its maximum number of connections (max). If so, u removes the most dissimilar neighbor and then invites a newly discovered similar one.

IV.   Experiments And Results

I have presented an unstructured P2P network with rigorous performance guarantees to enhance search efficiency and effectiveness. In a constant probability, a querying peer takes N hops (where c is a small constant) to reach the destination node capable of resolving the query, whereas the query messages 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 [21], [22], [23], [24]) 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. We validate our proposal with simulations. The simulation results reveal that whereas GES and SocioNet, that is, the two representative distributed algorithms among [21], [22], [23], [24], introduce fair traffic overhead to maintain and rewire their overlay topologies, ours clearly outperforms GES and Socio Net in terms of the following results are

1. the query message hop count,
2. the successful ratio of resolving a query,
3. the query traffic overhead, and
4. the overlay maintenance overhead.
5. the Resultant set of finite N hops for data sharing in unstructured peer-to-peer (P2P) networks.

V. Future Works

Moreover, In future the work may continue to find that together with a similarity-aware overlay topology, the search protocol we have suggested in this paper, which takes advantage of the similarity of peers exploited by our overlay network, can considerably reduce the search traffic. Peers participating in a P2P network are often heterogeneous in terms of their network bandwidth, storage space, and/or computational capability. It would be interesting for our future work to investigate how the heterogeneity affects our proposal. Moreover, the overlay formation algorithm presented in this paper is oblivious to the physical network topology, and this may introduce considerable wide-area network traffic [14], [29]. It would be challenging to design an overlay formation algorithm aware of both the similarity of participating peers and the physical network topology.

References

[1].   IPOQUE, "Ipoque Internet Study 2007: P2P File Sharing Still Dominates the Worldwide Internet," http://www.ipoque.com/ news-and-events/news/archive/2007, 2011.
[2].   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.
[3].   Gnutella, http://rfc-gnutella.sourceforge.net/, 2011.
[4].   Y. Liu, J. Han, and J. Wang, "Rumor Riding: Anonymizing Unstructured Peer-to-Peer Systems," IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 3, pp. 464-475, Mar. 2011.
[5].   H. Chen, H. Jin, Y. Liu, and L.M. Ni, "Difficulty-Aware Hybrid Search in Peer-to-Peer Networks," IEEE Trans. Parallel and Distributed Systems, vol. 20, no. 1, pp. 71-82, Jan. 2009.
[6].   A. Crespo and H. Garcia-Molina, "Routing Indices for Peer-to- Peer Systems," Proc. 22th IEEE Int'l Conf. Distributed Computing Systems (ICDCS '02), pp. 23-32, July 2002.
[7].   X. Shi, J. Han, Y. Liu, and L.M. Ni, "Popularity Adaptive Search in Hybrid P2P Systems," J. Parallel and Distributed Computing, vol. 69, no. 2, pp. 125-134, Feb. 2009.
[8].   E. Cohen and S. Shenker, "Replication Strategies in Unstructured Peer-to-Peer Networks," Proc. ACM SIGCOMM '02, pp. 177-190, Aug. 2002.
[9].   Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," Proc. ACM Int'l Conf. Supercomputing (ICS '02), pp. 84-95, June 2002.
[10].   Gnutella2, http://g2.trillinux.org/, 2011.
[11].   KaZaA, http://www.kazaa.com/, 2011.
[12].   Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, "Making Gnutella-like P2P Systems Scalable," Proc. ACM SIGCOMM '03, pp. 407-418, Aug. 2003.
[13].   L. Xiao, Z. Zhuang, and Y. Liu, "Dynamic Layer Management in Superpeer Architectures," IEEE Trans. Parallel and Distributed Systems, vol. 16, no. 11, pp. 1078-1091, Nov. 2005.
[14].   Y. Liu, L. Xiao, X. Liu, L.M. Ni, and X. Zhang, "Location Awareness in Unstructured Peer-to-Peer Systems," IEEE Trans. Parallel and Distributed Systems, vol. 12, no. 2, pp. 163-174, Feb.2005.
[15].   L. Xiao, Y. Liu, and L.M. Ni, "Improving Unstructured Peer-to- Peer Systems by Adaptive Connection Establishment," IEEE Trans. Computers, vol. 54, no. 9, pp. 1091-1103, Sept. 2005.
[16].   F.L. Fessant, S.B. Handurukande, A.-M. Kermarrec, and L. Massoulie´, "Clustering in Peer-to-Peer File Sharing Workloads," Proc. Third Int'l Workshop Peer-to-Peer Systems (IPTPS '04), pp. 217-226, Feb. 2004.
[17].   A. Iamnitchi, M. Ripeanu, E. Santos-Neto, and I. Foster, "The Small World of File Sharing," IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 7, pp. 1120-1134, http://doi.ieeecomputersociety.org/ 10.1109/TPDS.2010.170, July 2011.
[18].   K. Sripanidkulchai, B. Maggs, and H. Zhang, "Efficient Content Location Using Interest-Based Locality in Peer-to-Peer Systems," Proc. IEEE INFOCOM '03, pp. 2166-2176, Mar. 2003.
[19].   L.C. Freeman, "A Set of Measures of Centrality Based on Betweenness," Sociometry, vol. 40, no. 1, pp. 35-41, Mar. 1977.
[20].   M. Girvan and M.E.J. Newman, "Community Structure in Social and Biological Networks," Proc. Nat'l Academy of Sciences of USA, vol. 99, no. 12, pp. 7821-7826, June 2002.
[21].   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.
[22].   S. Voulgaris, M. Steen, and K. Iwanicki, "Proactive Gossip-Based Management of Semantic Overlay Networks," Concurrency and Computation: Practice and Experience, vol. 19, no. 17, pp. 2299-2311,Dec. 2007.
[23].   [23] P. Raftopoulou, E.G.M. Petrakis, and C. Tryfonopoulos, "Rewiring Strategies for Semantic Overlay Networks," Distributed and Parallel Databases, vol. 26, nos. 2/3, pp. 181-205, Dec. 2009.
[24].   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.
[25].   S. Merugu, S. Srinivasan, and E. Zegura, "Adding Structure to Unstructured Peer-to-Peer Networks: The Use of Small-World Graphs," J. Parallel and Distributed Computing, vol. 65, no. 2, pp. 142-153, Feb. 2005.
[26].   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.
[27].   G. Chen, C.P. Low, and Z. Yang, "Enhancing Search Performance in Unstructured P2P Networks Based on Users' Common Interest," IEEE Trans. Parallel and Distributed Systems, vol. 19,no. 6, pp. 821-836, June 2008.
[28].   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.
[29].   B. Bolloba´s, Random Graphs, second ed. Cambridge Univ. Press, 2001.
[30].   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.
[31].   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.