International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.3, Issue.2, March-April. 2013 pp-903-907 ISSN: 2249-6645

Topologies in Unstructured Peer To Peer Networks

B. Supraja, [1] N. Krishna Kumar, [2] U. Penchala Prasad, [3] Sk. Mohiddin[4]
[1]*M. Tech (CSE), PBRVITS, Nellore.*
[234] *Associate Professor, CSE, PBRVITS, Nellore*

Abstract: *The Peer-to-Peer (P2P) architectures that are most prevalent in today's Internet are decentralized and unstructured. 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, we 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 accurate performance guarantees. Based on the simulation results, our 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.*

Index Terms: *Peer-to-peer systems file sharing, unstructured overlay networks, search.*

I. Introduction

Peer-To-Peer (P2P) networks (overlay networks) have been widely deployed in the Internet, and they provide various services such as file sharing, information retrieval, media streaming, and telephony. P2P applications are popular because they primarily provide low entry barriers and self-scaling. Prior studies reveal that P2P applications may dominate up to around 20 percent of Internet traffic.

Gnutella is a popular P2P search protocol in the mass market. Specifically, because Gnutella networks are unstructured, and the peers participating in networks connect to one another randomly, peers search objects in the networks through message flooding. To flood a message, an inquiry peer broadcasts the message to its neighbors. (by the neighbors of peer i, we mean those peers that have end-to-end connections with i). The broadcast message is associated with a positive integer time-to-live (TTL) value. Upon receiving a message, the peer (say, j) decreases the TTL value associated with the message by 1 and then relays the message with the updated TTL value to its neighbors, except the one sending the message to j, if the TTL value remains positive. Aside from forwarding the message to the neighbors, j searches its local store to see if it can provide the objects requested by peer i. Conceptually, if j has the requested objects and is willing to supply them, then j either directly sends i the objects or returns the objects to the overlay path where the query message traverses from i to j.

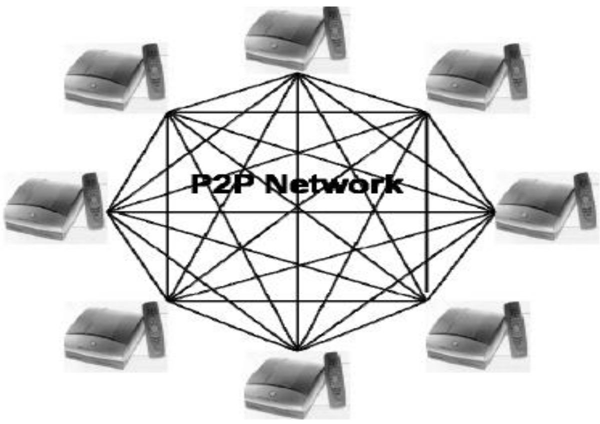

Fig: 1. Peer to Peer Overlay Network

In this paper, we first observe that existing P2P file sharing networks 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. Compared with previous proposals our proposal has the following unique features:

- In a constant probability, the search hop count between any two nodes is $O(\ln^{c_1} N)$, where 1 < c1 < 2 is a small constant, and N is the number of active peers participating in the network.
- In a constant probability of approximately 100 percent, the peers on the search path from the querying peer to the destination peer progressively and effectively exploit their similarity.

- Whereas some prior solutions require centralized servers to help organize the system, our proposal needs no centralized servers to participate in. Unlike most decentralized overlay construction algorithms for enhancing searches in unstructured P2P networks, our solution is mathematically provable and provides performance guarantees.

Moreover, we suggest a search protocol to take advantage of the peer similarity exhibited by our proposed overlay network.

II. Related Work

PSearch and SSW are content-based P2P networks providing semantic search. Similar to most P2P networks based on distributed hash tables in pSearch and SSW, each published object, which is represented by a latent semantic vector, needs to be indexed first into the network where the participating peers are formatted in a well-structured manner and host a disjoint key subspace. Therefore, the participating peers need to maintain foreign indices, that is, the indices of objects stored in remote peers. To locate an object, a requesting peer routes a message toward the peer responsible for the key subspace where the object is indexed.

Flooding and RW are two typical examples of blind search algorithms by which query messages are sent to neighbors without any knowledge about the possible locations of the queried resources or any preference for the directions to send. Some of search algorithms include modified BFS (MBFS), directed BFS, expanding ring and random periodical flooding (RPF). These algorithms try to modify the operation of flooding to improve the efficiency. However, they still generate a large amount of query messages. We propose a Light Flood algorithm, which is a combination of the initial pure flooding and subsequent tree-based flooding. DS and Light Flood operate analogously, but DS avoids the extra cost to construct and maintain the treelike sub overlay. Knowledge-based search algorithms take advantage of the knowledge learned from previous search results and route query messages with different weights based on the knowledge. Thus, each node could relay query messages more intelligently. Some examples are adaptive probabilistic search (APS) , biased RW, routing index (RI) , local indices , and intelligent search. APS builds the knowledge with respect to each file based on the past experiences. RI classifies each document into some thematic categories and forwards query messages more intelligently based on the categories. The operation of local indices is similar to that of super peer networks. Each node collects the file indices of peers within its predefined radius. If a search request is out of a node's knowledge, this node would perform a flooding search. The intelligent search uses a function to compute the similarity between a search query and recently answered requests. Nodes relay query messages based on the similarity. There are some other research works that focus on replicating a reference pointer to queried resources in order to improve the search time.

III. Our Proposal

Consider any given unstructured P2P network G = (V, E), 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_u peers in V, and these neighbors, selected among the peers in V, are the top- max_u nodes most similar to v.
C2. (Low diameter) - Consider any two distinct peers u and v in V. There should exists 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.

3.1. Peer Similarity Graphs

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

Definition 1:

The peer similarity function measures the degree of similarity between any two peer's u $\in$ V and v $\in$ V in the system.

$$\mathbf{F: V \times V \rightarrow R_0^+}$$

Definition 2:

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

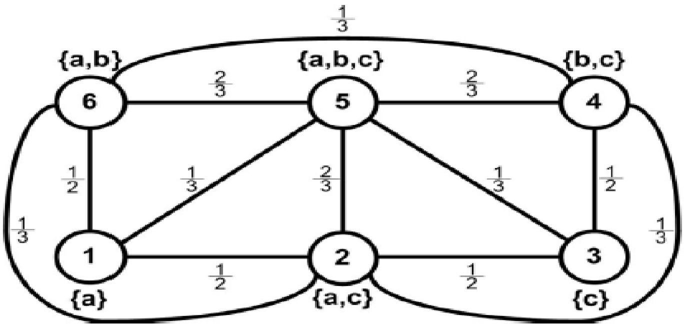

Fig. 1. An example of a peer similarity graph $\mathcal{G} = (V, \mathcal{E})$. Here, $V = \{1, 2, 3, 4, 5, 6\}$. Peers 1, 2, 3, 4, 5, and 6, respectively, host sets of objects $\mathcal{O}_1 = \{a\}$, $\mathcal{O}_2 = \{a, c\}$, $\mathcal{O}_3 = \{c\}$, $\mathcal{O}_4 = \{b, c\}$, $\mathcal{O}_5 = \{a, b, c\}$, 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 $\mathcal{F}(u, v)$.

3.2. Overlay Formation

3.2.1 Exploiting Similar Peers

As previously mentioned, each peer u will connect to the peers selected among all peers in V _ {u} that are most similar to u; that is, u intends to satisfy Property C1. Let I_u be the set of neighbors that u currently maintains in the network G = (V, E). Define A_{current} as representing the averaged peer similarity value of u and u's neighbors in I_u.

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

By exploiting the peers most similar to u, u seeks a peer w ∈ V – I_u – {u} and invites w as its neighbor such that

$$\mathcal{A}_{\text{update}} \geq \mathcal{A}_{\text{current}},$$

Where

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

Algorithm 1 details our proposal

input : $\mathcal{I}_u$ and $\mathcal{I}_u^2$
output: $\mathcal{I}_u$

$q \leftarrow \arg\min_{v \in \mathcal{I}_u} \mathcal{F}(u, v)$;
if there is a $w \in \mathcal{I}_u^2$ satisfying Eq. (3) and w is willing to link to u **then**
  if $|\mathcal{I}_u| < \max_u$ **then**
    $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{w\}$;
  else
    $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{w\}$;
    $\mathcal{I}_u \leftarrow \mathcal{I}_u - \{q\}$;
else
  u randomly picks a $w \in \mathcal{I}_u^2$;
  if w is willing to link to u **then**
    if $|\mathcal{I}_u| < \max_u$ **then**
      u performs $\mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{w\}$ with a probability of Eq. (6);
    else
      u performs $\begin{cases} \mathcal{I}_u \leftarrow \mathcal{I}_u \cup \{w\} \\ \mathcal{I}_u \leftarrow \mathcal{I}_u - \{q\} \end{cases}$ with a probability of Eq. (6);
return $\mathcal{I}_u$;

3.2.2 Minimizing Semantic Overlay Diameter

To minimize the overlay diameter in our proposal, each peer u ∈ V will create a number of extra overlay links. Denote such extra connections for u by Φ_u. Each t ∈ Φ_u is selected in a probability of Pr(u, t), where Pr(u, t) depends on the peer similarity distance between u and t, that is, D(u, t).

Algorithm 2: Peer t forwards a query message Q.

input : $\mathcal{I}_t$, Φ_t, and a query message Q

if t receives Q for the first time and $Q.TTL \leq$ MAXTTL **then**
  foreach $t' \in \mathcal{I}_t \cup \Phi_t$ **do**
    if $\mathcal{F}(t', Q) > \mathcal{F}(t, Q)$ **then**
      $Q.TTL \leftarrow Q.TTL + 1$;
      t forwards Q to t';

IV. Performance Evalution

Comparing with Guided Search, Routing protocol, filtering with routing updating table provides optimum results for the search performance. Initially when the queries are minimum guided search performance was good. When the queries are getting increased, filtering mechanism with routing updating table is the suitable one which gives the best results up to 90%.Hence it improve the searching performance of the peer. Routing updating table protocol contains the past successful search results and it is used for future references. Updating process can be taken place in each and every second.

V. Simulations

We have developed an event-driven simulator to evaluate the performance of our proposal. The input trace to our simulator is the eDonkey data set. The data set maintains the files shared by peers participating in the eDonkey file sharing network. Specifically, the files shared by each peer are recorded in the data set.

As the eDonkey data set lacks the details for describing each shared file (e.g., the keyword metadata), we measure the similarity level between any two peers u and v in the trace as the similarity function

$\mathcal{F}(u, v) = \frac{|\mathcal{O}_u \cap \mathcal{O}_v|}{|\mathcal{O}_u \cup \mathcal{O}_v|}$ Where O_u and O_v represent the files shared by peers u and v, respectively.

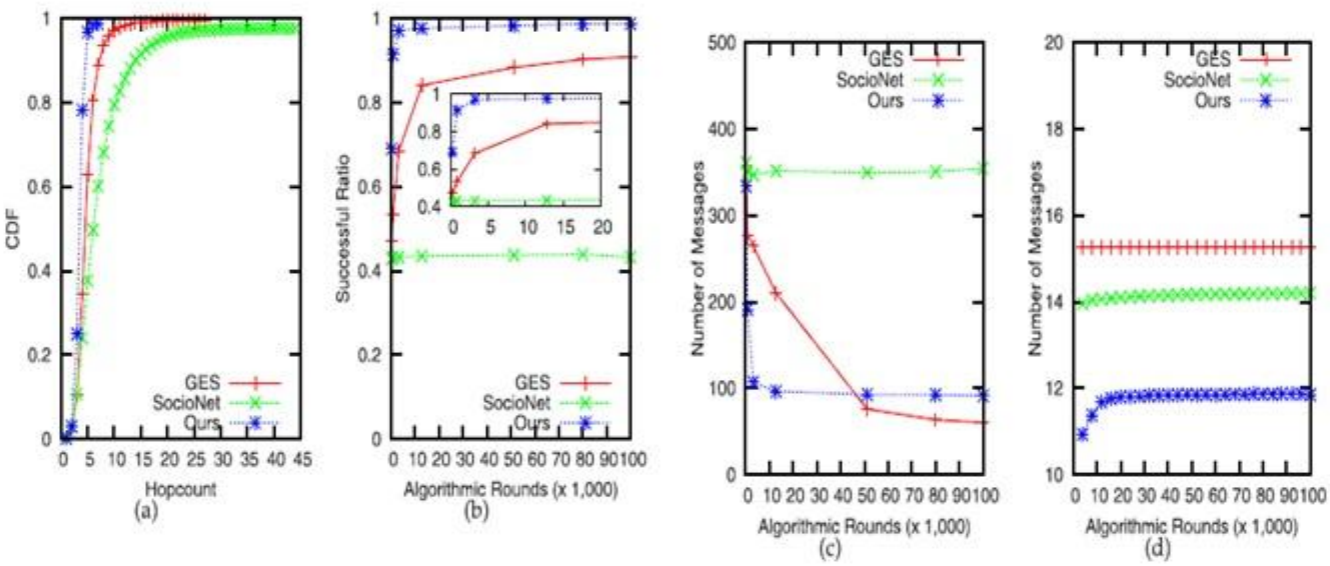

Fig: (a) The query hop count, (b) the successful query ratio. (c) The overhead for resolving a query, (d) The overhead of rewiring and maintaining the network.

VI. Summary and Future Work

We have presented an unstructured P2P network with rigorous performance guarantees to enhance search efficiency and effectiveness. In a constant probability, a querying peer takes $O(\ln^c \mathcal{N})$ 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.

We validate our proposal with simulations. The simulation results reveal that whereas GES and SocioNet, that is, the two representative distributed algorithms among, introduce fair traffic overhead to maintain and rewire their overlay topologies, ours clearly outperforms GES and SocioNet in terms of

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.

Moreover, we 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.

References

[1] 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.

[2] Gnutella, http://rfc-gnutella.sourceforge.net/,

[3] E. Cohen and S. Shenker, "Replication Strategies in Unstructured Peer-to-Peer Networks," Proc. ACM SIGCOMM '02, pp 177-190, Aug. 2002.

[4] 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.

[5] 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.

[6] 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.