

Searching In Unstructured P2p Networks on Optimized Overlay Topology

*Vimal S, S. K. Srivatsa

Department of Information Technology, Jeppiaar Engineering College, Tamilnadu, Chennai, Sathyabama University, Chennai

Department Of Computer Science and Engineering, Prathyusha Institute of Technology and Management, Chennai

*Corresponding author: E-Mail: vimalshan@gmail.com

ABSTRACT

At present a popular unstructured peer-to-peer network depends on flooding query messages to find the object and produce striking network traffic because here the peers interconnect without particular order and have alike preferences. The network cannot function productively as it creates network without performance guarantees. In order to overcome this we have proposed a novel overlay formation algorithm that has accurate performance guarantees. From the theoretical performance we conclude the similarity of peers has been used by search and the object can be searched efficiently. From simulation result our proposal competes other algorithm by the no of message required for resolving query, hop count of routing a query message, message overhead for maintaining and formatting the overlay and the successful ratio of resolving the query.

KEY WORDS: flooding, searching, P2P network.

1. INTRODUCTION

Peer-To-Peer network has been used effectively in the internet and it provides various services. Previous studies reveal that P2P applications have strong influence up to 30% of traffic in internet. Gnutella a popular unstructured P2P search protocol has peers participating in networks that connect without particular order by message flooding. The broadcasted messages by the peers have a positive integer time-to-live (TTL) value. Gnutella have some orthogonal techniques to improve the search performances and the techniques include indexing replication super architectures and overlay topologies from present studies show that peers are probably to determine the queries produced by the peers sharing the common preferences. Our study plans to organize the participating peer to make use of their similarity.

The proposals in put forward combining a native random network in P2P network to decrease the overlay path length so that query response time is reduced. A semantic small-world random graph have been used in studies in contrast to

In this paper we view the existing P2P file-sharing network displays a quality power-law file sharing pattern so that we can produce a novel overlay construction algorithm to make the efficiency and effectiveness of search in unstructured P2P networks. Our proposal has the following special features:

- Approximately 100 percent in a constant probability the peers make use of their similarity on search path from querying peer to destination peer effectively.
- The search hop count between any nodes in constant probability is $O(\ln^{c1} N)$, where $1 < c1 < 2$ is a small constant and N number of active peers participating.
- Some previous studies require centralized servers whereas our proposal does not need it. Similar to decentralized algorithm our solution is provable mathematically that also provides performance guarantees for enhanced search.

Performance analysis shows that semantic P2P networks not only cluster the closet peers but also interconnect the selected peer without order so that it cannot work effectively by making use of similarity of peers. On comparing our simulation result with algorithms in we have the following

- the number of messages required for resolving the query
- the hop count of routing the query message
- message overhead for maintaining and formatting the overlay
- successful ratio of resolving the query

From similarity-aware protocol we say that the overlay networks reduce the query traffic by making use of similarity of participating peer than the search protocol that use bind flooding.

Related Work: Here we mainly discuss P2P networks that aim to make use of similarity of participating peers due to space limitation. GES, a semantic overlay in has no foreign indices and it depends on message flooding to find the requested objects. In this the peers are clustered since the peers with similar preference provide objects to one another and the peers connect to selected peer “uniformly at random” from the system. It is to develop a analytical model to depict that the participating peer partially select the neighbors uniformly at random from the network as in then the resultant network cannot function effectively and efficiently. The works in develop small-world based semantic overlays to intensify efficiency and effectiveness.

Semantic search based on distributed hash tables as provided by pSearch and SSW each published object described by indexed first in a semantic vector needs into the network where the participating peer are arranged in a ordered manner and host a disjoint key subspace. Participating peer has foreign indices and to find the object a requesting peer directs the message towards the peer responsible for key subspace where object is indexed. In our study we have presented a network where the peers need not arrange themselves in a rigid topology structure and it has no foreign indices, eliminates storage and bandwidth overhead for managing indices.

Chen et al heavily depend on centralized entity to format the P2P network. Centralized servers are required to calculate probabilistic model based on principle of maximum entropy that enables to calculate the probability of sharing an object o_1 when subjected on another shared object o_2 . This proposal suffers from certain disadvantages and it has no analytical properties. Our proposal does not depend on centralized server and provides rigorous performance results.

In the proposals the peers select their neighbors “heuristically” and does not offer rigorous performance guarantees. in order to overcome this our proposal has algorithm that performs with accurate mathematical guarantees and successfully discovers the requested object. The algorithm in Gossip-based peer membership protocol requires sampling peers from the system and it allows the peers to gossip one another and sample participating peer uniformly at random Our algorithm requiring a no uniformly distribution of sampled peers is designed and Implemented based on Metropolis-Hastings method Theoretical analysis in our paper is inspired by Kleinberg model that helps us to inquire the performance of our network (Zhao, 2004) in detail and in contrast to that our model works in fully distributed manner.

Previous studies maximize the coverage of a query message that (Stoica, 2003) does not guarantee efficiency and effectiveness of search. Our proposal aims at making full use of similarity of peers participating in the network. We format P2P topologies that offer efficient and effective search compared to network topologies.

Proposed System: For instance let us take any given unstructured P2P network say $G = (V, E)$. Here,

V = set of participating peers

E = set of overlay connections linking peers in V .

The peers are interconnected without any particular order and our aim is to satisfy the following properties by restructuring G .

C1. (High Clustering). Each peer u connects \max_u peers in V and those neighbors, selected the peers among in V , the top - \max_u nodes similar to v at most.

C2. (Low Diameter). Let us consider two distinct peers u and v in V and it should have at least one overlay path P connecting u and v . In order to enable a query message to be quickly propagated from u to v the hop count of P should be small and here hop count of P refers to number of overlay links in it.

C3. (Progressive). Let us consider the peers issuing query as s and peers resolving query as d with an overlay path P such that on receiving a query message u forwards it to v that is more similar to d than u .

In the following section we will produce the search with properties C1, C2, and C3 that are effective and efficient.

Peer Similarity Graphs: Here we take V to be set of peers participating in a P2P network.

Definition 1: The peer similarity function measures the degree of similarity between any two peers $u \in V$ and $v \in V$ in the system.

$$F: V \times V \rightarrow \mathbb{R}_0^+$$

Definition 2: Given $G=(V,E)$, the peer similarity distance between two distinct peers $u \in V$ and $v \in V$, denoted by $D(u,v)$, is defined as the length of shortest path in E that connects u and v .

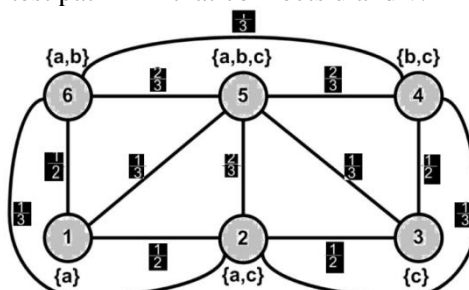


Figure.1. An example of peer similarity graph $G=(V,E)$. here $V=\{1,2,3,4,5,6\}$. No of Peers 1,2,3,4,5 and 6 respectively host sets of objects $O_1=\{a\}$, $O_2=\{a,c\}$, $O_3=\{c\}$, $O_4=\{b,c\}$, $O_5=\{a,b,c\}$ and $O_6=\{a,b\}$. any two peers u and v have an edge in ε if both peers share at least one common object.

Definition 3: A peer similarity graph $G=(V, E)$ is a graph where V denotes 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.

Overlay Formation

Exploiting Similar Peers: In this each peer u will connect to the peers selected among all peers in $V - \{u\}$ that are most similar to u to satisfy property C1. A_{current} is defined as representing the averaged peer similarity value u and v 's neighbors in I_u .

$$A_{\text{current}} = \frac{\sum_{v \in I_u} F(u, v)}{|I_u|}$$

On exploiting the peers most similar to u that seeks a peer $w \in V - I_u - \{u\}$ and invites w as its neighbor such that

$$\frac{A_{\text{update}} - A_{\text{current}}}{e} \geq \frac{1}{T}$$

Algorithm 1: details our proposal

Input : I_u and I_u^2

Output : I_u

$q \leftarrow \arg \min_{v \in I_u^2} F(u, v)$

if there is a $w \in I_u^2$ and w is willing to link to u then

if $|I_u| < \max_u$ then

$I_u \leftarrow I_u \cup \{w\};$

else

$I_u \leftarrow I_u \cup \{w\};$

$I_u \leftarrow I_u - \{q\};$

else

u randomly picks a $w \in I_u^2$;

if w is willing to link to u then

if $|I_u| < \max_u$ then

u performs $I_u \leftarrow I_u \cup \{w\}$ with a probability of eq;

else

u performs

$I_u \leftarrow I_u \cup \{w\};$

$I_u \leftarrow I_u - \{q\};$

With a probability of eq;

return I_u ;

Reducing Overlay Semantic Diameter: In order to minimize the overlay diameter each peer $u \in V$ will develop a number of extra overlay links. To find t in a probability of $\Pr(u, t)$, u will issue a biased random walker based on metropolis-hasting method to visit peers in the system. Each peer u maintains its I_u and ϕ_u independently. Our proposal creates an overlay network.

Search Protocol: For instance consider any query Q requesting an object O and denote peer issuing by s and hosting by d . conceptually our protocol works as follows. Neighbouring peers selects the peers s and then broadcasts Q to the peers. t performs likely by selecting its neighbor similar to d than t , hoping at least one overlay path P .

$$F(v, d) > F(u, d)$$

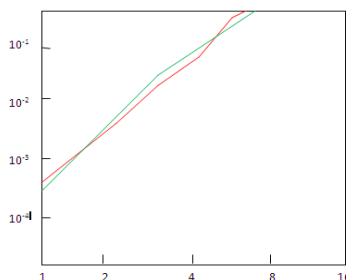


Figure.2. The log-log plots the distribution e-Donkey Data set

Theoretical Performance Analysis

Empirical Data Set: On investigating the files shared by eDonkey users (Crespo, 2002) we have concluded that the data set follows power-law property.

Definition 4 : the scope of a peer $u \in V$ in a peer similarity graph $G=(V,E)$ within a given peer similarity distance d is defined as

$S_u(d) = \{v \mid D(u, v) \leq d, v \in V\}$ Fig.2 represents the distribution of eDonkey data set and in that each peer is

limited to connect upto eight neighbors because in typical P2P network peer can connect to only small number of neighbors. The peer u likely to follow power-law distribution called Pareto distribution $f(x) = \beta x^{-\alpha}$ with $\alpha \approx 4.46$ and $\beta \approx 0.26$.

Analytical Results: My analytical model is based on power-law distribution.

Definition 5: 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 $u \in V$, 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,

- The peers in our proposal is exploited to resolve the query effectively.
- The path length of P is no more than $O(\ln^c N)$.
- In a constant probability approximately 100% peers can be successfully constructed.
- Compared to other proposal our proposal performs efficiently and effectively by exploring the path P .

Simulations

2. EXPERIMENTAL SET UP

In order to evaluate the performance we have developed event-driven simulator in our proposal. eDonkey data set is given as input and it has the files shared by the participating peers. The files shared by are recorded as data set. In the simulations participating peer connects to the arbitrary peers where the peers connect to at least four neighbors (Shen, 2006).

The performance metrics are,

- The query traffic overhead
- Query message hop count
- The traffic overhead for maintaining and rewriting the overlay topology
- Successful query ratio

Compared to GES and Socionet our proposal maintains same number of neighbors where n_s are similar and n_i are dissimilar and we measure the performance of investigated algorithm after it becomes stable.

The default value of TTL is 0.03 and we will calculate the performance by varying T

3. SIMULATION RESULTS

Comparative studies: My proposal performs well as most queries can be forwarded to the destination in no more than 10 hops whereas queries in GES and Socionet takes 25 and 40 hops respectively. The successful ratio of routing queries are improved in GES and our proposal so that it works effectively and the success ratio is approximately 100 percent. At around 3500 algorithmic rounds my proposal becomes stable whereas GES requires 13000 rounds. Compared to GES and my proposal Socionet does not perform well because it has lengthy hop count and low success ratio for resolving query.

My proposal outperforms GES and Socionet where the biased random walker may remain at a peer without advancing to a newly dissimilar peer that introduce less network traffic. The TTL values in our experiment are 3 and 7. Although the TTL value 7 produce successful ratio of 100 percent it will produce query traffic and when the TTL values will be reduced to 3 and the traffic is reduced in the successful ratio $\approx 20\%$.

Effects of Varying α : Based on approximated α and β the participating peer u connects to the dissimilar peer t with a probability $\Pr(u,t)$ and as suggested each peer operates with a positive and small α value so that it improves the approximation for α over time. The underestimated α value shows slight degradation in performance however our proposal still performs well.

Effects of System Dynamics: In my proposal each peer connects up to 8 neighbors and our system is stabilized for 100 minutes. The lifetime of the peer is 2.5 hours in expectation. The peers in dynamic system have a lifetime of 1.25 and 0.5 hours where the system operates for 18 hours. My proposal performs well when the expected lifetime of peer is 2.5 and 1.25 hours and it converges quickly under a dynamic environment. Compared to blind flooding our proposal performs well.

Effects of Varying T : My proposal depends on the parameter T and the default value is 0.03. My proposal is sensitive to the value of T . This peer must be optimized to connect to similar neighbors without finding the neighbors in the best effort method in GES and the Socionet.

Effects of Varying n_s and n_i : For the simulation let us consider the default similar neighbors (n_s) and dissimilar neighbors (n_i) are taken as 8. By varying the values of n_s and n_i we investigate the experiment. Fig 5 shows the simulation result for the query hop count averaged over all successful queries with respect to different number of algorithmic rounds. My proposal outperforms GES and Socionet and it cannot work well if $n_s = 4$. As a result our proposal may not improve the overlay interconnect geometry.

Summary And Future Work: The unstructured P2P network presented in our paper produce performance guarantees with search efficiency and effectiveness. Here the peer takes $O(\ln^c N)$ hops to reach its destination and the query message exploit the similarity of the peers. The query messages can be resolved successfully in approximate 100 percent. The simulation results reveals that GES and Socionet do not perform well since it introduce traffic in the network. My proposal outperforms others by

- Query message hop count
- Successful ratio of resolving query
- Query traffic overhead
- Overlay maintenance overhead.

4. CONCLUSION

My proposal reduce the network traffic by taking the advantage of similarity of peers exploited by our overlay network. The heterogeneity effects in our proposal are interesting for our future work and the overlay formation algorithm in the network introduce wide area network traffic . It is very challenging to design an algorithm aware of both similarity of participating peer and physical network topology.

REFERENCES

- Bernstein P.A, Giunchiglia F, Kementsietsidis A, Mylopoulos J, Serafini L and Zaihrayeu I. Data management for peer-to-peer computing, A vision. In Proc. of WebDB, 2002.
- Clarke I, Sandberg O, Wiley B and Hong T.W, Freenet, A Distributed Anonymous Information Storage and Retrieval System. In Proc. of the International Workshop on Design Issues in Anonymity and Unobservability, 2001, 46–66.
- Crespo A and H. Garcia-Molina H, Routing indices for peer-to-peer systems. In Proc. of ICDCS, 2002.
- Halevy A.Y, Ives Z.G, Mork P and Tatarinov I, Piazza, Data management infrastructure for semantic web applications. In Proc, 2003.
- Li Z, Xie G and Li Z, Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems, TPDS, 2008.
- Mitra B, Dubey A.K, Ghose S and Ganguly N, How do superpeer networks emerge? In Proc, 2010.
- Ratnasamy S, Handley M, Karp R, and Shenker S, Topologicallyaware overlay construction and server selection. In Proc, 2002.
- Rowstron A and Druschel P, Pastry, scalable, decentralized object location and routing for large-scale peer-to-peer systems. In Proc. of Middleware, 2001.
- Shen H and Xu C.Z, Hash-based proximity clustering for efficient load balancing in heterogeneous dht networks. JPDC, 2008.
- Shen H, Xu C, and Chen G. Cycloid, a scalable constant-degree P2P overlay network, Performance Evaluation, 2006.
- Stoica I, Morris R, Liben-Nowell D, Karger DR, Kaashoek MF, Dabek F and Balakrishnan H, Chord, a scalable peer-to-peer lookup protocol for internet applications, TON, 2003.
- Waldvogel M and Rinaldi R, Efficient topology-aware overlay network. In Proc. of HotNets-I, 2002.
- Zhao B.Y, Huang L, and Tapestry, an infrastructure for fault-tolerant wide-area location and routing, J-SAC, 2004.
- Zhu Y and Shen H, An efficient and scalable framework for content-based publish/subscribe systems, PPNA, 2008.