Search in Unstructured Peer-to-Peer Networks based on Dynamic Topology Adaptation

T.Arumuga Maria Devi[1] ***Sabitha.S***[2]

Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Tirunelveli

Abstract

Peer-to-Peer networks (P2P) have gained great attention and popularity. One key challenge aspect in Unstructured P2P resource sharing environments is efficient searching algorithm. Flooding and Random Walk (RW) are two typical search algorithms. Flooding searches aggressively and covers most nodes but it generates large amount of query messages. RW searches conservatively and at each hop it generates a fixed amount of query messages but it takes longer search time. This paper proposes the Dynamic Search Algorithm (DS), a generalization of flooding and RW. It resembles flooding for short-term search and RW for long-term search. This system describes a multi-database system as a 4 tiered Client-Server DBMS architecture. The proposed system provides remarkable changes in database servers which process thousands of concurrent queries. The load balancing mechanism is implemented by using a four tier model. The first component of this system is a web based interface or Graphical User Interface. The second component is a client application program running in an application server. The third component is a Global Database Management System which manages multiple database servers. The GDMS receives all queries, divides and forwards to respective databases. The fourth component is a remote homogenous local component database system server. More than one database component depends on the requirements of the application. . Transaction submitted from the client interface to a multi-database system server through an application server will be decomposed into a set of sub queries and will be executed at various remote homogenous local component database servers and also, in case of information retrieval, all queries will be composed and will get back result to the end users.

Index Terms

MOU Replacement Policy -- P2P Networks -- Dynamic Topology Adaptation.

I.INTRODUCTION

The proposed system develops techniques to render the search process in unstructured network file sharing systems more efficient and scalable due to the similar interests of peer nodes. P2P computing or networking is a distributed application architecture that partition tasks or workloads between peers. Peers make a portion of their resources such as processing power, disk storage or network bandwidth directly available to other network participants, without the need for central coordination by servers or stable hosts. Peers are both suppliers and consumers of resources, in contrast to the traditional client–server model where only servers supply, and clients consume. The peer-to-peer application structure was popularized by file sharing systems like Napster [1] which is centralized. The peer-to-peer computing paradigm has inspired new structures and philosophies in other areas of human interaction. In such social contexts, peer-to-peer refers to the egalitarian social networking that is currently emerging throughout society, enabled by Internet technologies in general.

The P2P networks that are decentralized can be divided as structured or unstructured. Structured Peer-to-Peer networks use specialized placement algorithms and directed search protocols e.g., Pastry [2], Tapestry [3]. Unstructured peer-to-peer networks have no accurate control over the file placement and generally flooding search protocols. E.g., Gnutella [4], Freenet [5]. Directed search protocols are efficient as they route queries towards the peer responsible for a given file precisely. They require only fewer communication steps and generate little traffic. The search fails only if no match is found. Flooding protocols are less efficient and may yield false negatives.

The proposed system uses dynamic topology adaptation which maintains the semantic communities to improve search efficiency based on the query patterns and the results of preceding searches. Semantic communities are self organized communities of peer nodes that share similar interests. The semantic communities may lead to hot-spot problem when many users are requesting for the same file i.e., the load increases.. So the system uses a load balancing mechanism that directs whenever possible the responsibility to answer a query to less loaded peer nodes. The system proposes a dynamic TTL (time-to-live) update scheme to further limit the network congestion without decreasing the query success rate.

The major advantages of the proposed system are

- Provides precise results.
- Use of unstructured P2P reduces the network set-up cost.
- Never produce false negatives.
- Load of the database is well managed.
- Highly efficient.

Manuscript received November 5, 2012
Manuscript revised November 20, 2012

- Semantic community creation yields successful results.
- Works well in long-term search and short-term search.

II.EXISTING METHODS

Flooding and Random Walk are two search protocols. Flooding searches aggressively and covers most nodes. In flooding, the queries are broadcasted indiscriminately in a whole neighborhood. So a large amount of query messages are generated. They have little management overhead and adapt to transient activity of P2P clients. Flooding allows users to perform more elaborate queries. In flooding, increasing the TTL by 1 may increase the space coverage exponentially. Flooding is favored by dynamic topology. Network congestion is a major problem.

Random Walk searches conservatively and can re-issue the same request several times. At each hop it generates only a fixed amount of query messages. It takes longer search times. It is better than flooding for clustered overlay topologies. Random Walks can be used to add new peers with constant overhead in the construction of dynamic P2P topologies. It is weakly decentralized.

The disadvantages of existing system include higher management overhead. They produce false negatives. Hot spot problem may arise if many users are requesting for the same file thus increasing the load. They are less efficient and are less precise. They support exact matched queries. The bigger TTL values may increase the success rate but may lead to network congestion.

To avoid these disadvantages, the system proposes semantic communities, load balancing mechanism, dynamic topology adaptation, and dynamic TTL update mechanisms.

III.SEARCHING P2P NETWORKS EFFICIENTLY

In this paper, a new system is proposed which uses dynamic topology adaptation to improve search efficiency. The main objective of this proposed system is to develop techniques to render the search process in unstructured network file sharing systems more efficient by taking advantage of similar interests of peer nodes (semantic communities) and implementing a direct flooding search protocol. Figure 1 shows the mechanisms proposed in this system.

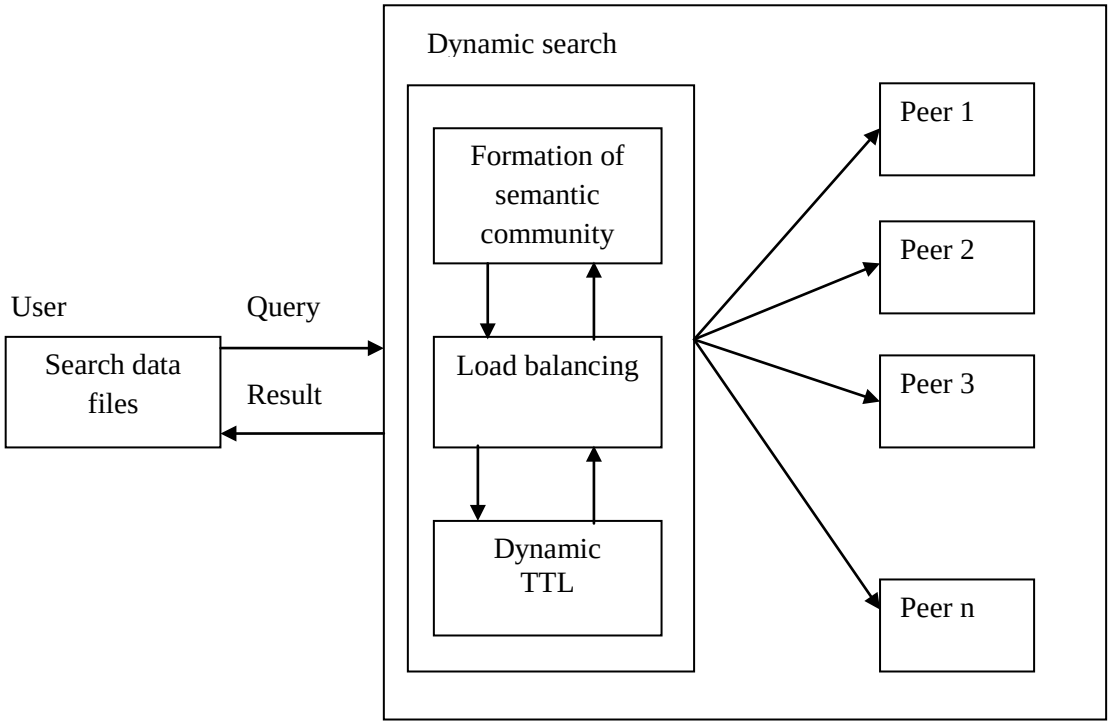

Figure 1. Flow diagram

Dynamic Topology Adaptation

Dynamic Topology Adaptation is implemented by directing Acquaintance links towards the peers. The basic principle consists of dynamically adapting the topology of P2P network so that peers sharing similar interests form well-defined semantic communities. Users are interested only in specific types of content [6]; therefore forming semantic communities will increase the search efficiency and the success rate. Replacing a peer from the list depends on the responses to previous requests issued by each peer. In this paper, a replacement policy is evaluated.

The *Most Often Used Policy (MOU)* maintains ranking of the peers and select the peer with highest ranking as the acquaintance. A peer has a highest ranking if it answers to many queries or if it is well-connected. After each successful query, each peer on the path followed by the query has its rank increased by an exponentially decreasing value. An aging factor is introduced that gives more weight to recent answers by decreasing the rankings over time. This gives preference to peers that stay longer in the system and are stable (algorithm 1).

Algorithm 1 MOU replacement policy at requester pr
Variables:
AcqList: ordered list of N Acquaintances
CandList: {peer, ranking} pair list
_: aging factor with value [0,1]
Upon successful query answered by pa, (pr, p1. . . pn-1, pn=pa)
for all {p;r} 2 *CandList* **do**
{p; r} {p; _r}
end for
i 1.0
for j **from** n **downto** 1 **do**
if {pj;r} 2 *CandList* **then**
{pj;r} {pj;r+i}
Else
Insert {pj;i} in *CandList*
end if
i i/2
end for
AcqList first N peers of *CandList*

Dynamic Search

It uses TTL flooding scheme and exhibits improved efficiency. The semantic communities improve response time by increasing the chances that matching files are found inside the community within short distance of the requester. Therefore smaller TTL values are used for queries and thus reduce the network congestion without decreasing the success rate. The peers also maintain index of the files stored on their neighbouring peers. Thus the peer can explore other peers with similar interests at no communication cost. This also increases the success rate and reduces network congestion. The semantic communities has several disadvantages such as the network may get subdivided into sub networks and also peer searching for contents that are less frequently searched may give very low success rate.

Load Balancing

Flooding algorithms naturally direct much of the traffic towards highly connected peers. In this proposed system, a peer that has many neighbours can quickly become a hot-spot, not only because it receives more queries, but also because it typically sends more files to requesting peers. Although the issue of file transfers is not explicitly addressed in this project, it is a large source of overhead in P2P file-sharing networks and should not be overlooked. Therefore the following mechanism is used to better balance the file traffic. Before successfully answering a query, a peer “p” first checks if any of its neighbours also has the requested file. If so, it delegates the responsibility for answering the query to the peer among those serving the file that has the smallest in-degree (note that this peer may be ”p”). Otherwise, “p” sends the file itself. The rationale behind this approach is that well-acquainted peers are likely to be more loaded. i.e., receive more requests and serve more files than peers with fewer neighbours. Further, there is a good probability that some of the neighbours of a peer also have the same files. Therefore, force the less loaded peer to assume part of the load.

Dynamic TTL

Semantic communities reduce the query traffic. The query entering the semantic community to which the requested file belongs is answered by that peer with fewer hops. If the query does not match that community then it traverse more peers. Therefore the proposed system reduces the TTL value twice instead of 1 when the received query falls within the semantic community. This also reduces the network congestion without decreasing the success rate.

IV.EXPERIMENTAL SETUP

Simulator is used for this purpose which proceeds through a subset of peers. This consists of two phases. In the first phase the, serving peer issues queries at a rate proportional to the storage capacity and create local copies of files that they request. Acquaintance link is dynamically updated based on replacement policy. More popular files are found closer to the requester (figure 2).

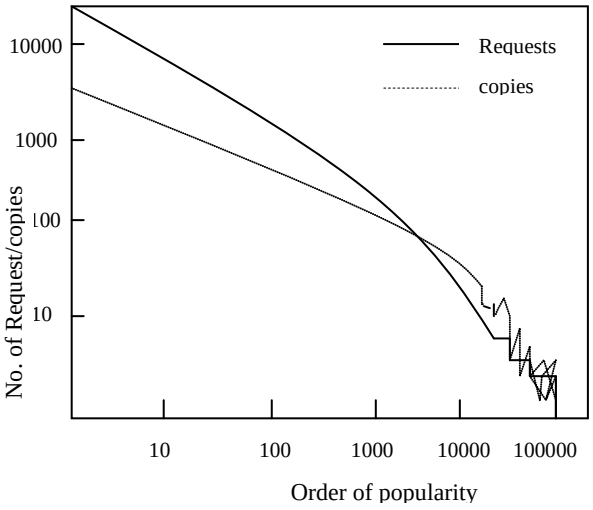

Figure 2. The number of requests for, and copies of each file depends on file popularity

In the second phase, the network behaviour under traffic load is measured. Table I shows the parameters used in the simulation.

Table I PARAMETERS USED IN THE SIMULATIONS.

Parameter	Value
Avtive peers	20000
70% share	0 file ≡ free-riders
20%share	[1…..100]files (uniform)
7% share	[101…..1000]files (uniform)
3%share	[1001…..2000]files (uniform)
Distinct files	200000
Categories	50
Categories per peer	[1…..6] (uniform)
Link per peer	6
Acquantance links	{0,1,3,5}
Query TTL	6
Query rate	10%

The in-degree of a peer p is defined as the number of other peers that have chosen p as acquaintance. The number of queries received by peer is proportional to its in-degree. The query success rate depends on the chosen TTL value. This value should simultaneously provide high success rate and limit the number of messages sent over the network. Cumulative success rate is used here. Higher success rate for smaller TTL values indicate that the search is more efficient. Figure 3 shows the result.

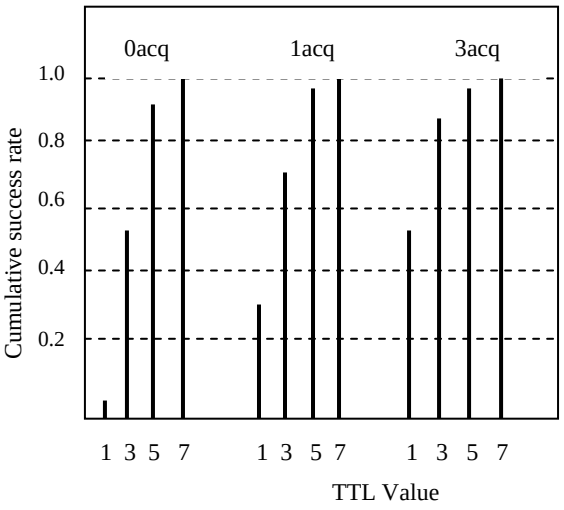

Figure 3. Cumulative success rate with MOU replacement policy

V.CONCLUSION

In the system, Search in Unstructured Peer-to-Peer Networks based on dynamic topology adaptation, the dynamic topology adaptation is used to improve search efficiency. The main merits of this proposed system is to develop techniques to render the search process in unstructured network file sharing systems more efficient and scalable by taking advantage of the common interests of the peer nodes i.e. semantic communities and effectively implementing a “directed flooding” search protocol. The system performs well in the unstructured peers to provide the desired results. The system needs more configurations at peer level if there is a need to increase the number of peers in the network. This needs entire network structure to be considered.

ACKNOWLEDGMENT

The authors would like to thank the members of the Dept. of CITE, M S University, Tirunelveli for various algorithms used in this research, and all the people who helped in preparing and carrying out the experiments and data collection.

References

[1] “Napster”, http://www.napster.com
[2] A. Rowstron and P. Druschel, “Pastry: scalable, distributed object location and routing for large scale peer-to-peer systems” in proceedings of the IFIP/ACM International conference on distributed system platforms, Nov 2001.

[3] B.Y. Zhao, J. Kubiatowicz and A.D. Joseph, "Tapestry: An infrastructure for fault tolerant wide area location and routing" Tech, Rep UCB/CSD-01-1141, Computer Science Division, University of California, Berkeley, April 2001.
[4] "Gnutella", http://gnutella.wego.com
[5] I. Clarke, O. Sandburg, B. Wiley and T.W. Hong, "Freenet: A Distributed anonymous information storage retrieval system", in proceedings of the Workshops on design issues in Anonymity and Unobservability. July 2000
[6] A. Crespo and H. Gracia Molina, "semantic overlay networks for P2P systems", Tech. Rep, Computer Science Department, Stanford University, 2003

T. Arumuga Maria Devi received B.E. Degree in Electronics and Communication Engineering from Manonmaniam Sundaranar University, Tirunelveli India in 2003, M.Tech degree in Computer and Information Technology from Manonmaniam Sundaranar University, Tirunelveli, India in 2005 and worked as Lecturer in the Department of Electronics and Communication in Sardar Raja College of Engineering. Currently, she is doing Ph.D in Computer and Information Technology and also the Assistant Professor of Centre for Information Technology and Engineering of Manonmaniam Sundaranar University. Her research interests include Signal and Image Processing, Multimedia and Remote Communication

S. Sabitha received B.Tech. Degree in Electronics and Communication Engineering from Mahatma Gandhi University, Kerala in 2010. Currently, she is doing M.Tech in Computer and Information Technology in Manonmaniam Sundaranar University, Tirunelveli. Her research interests include Image Processing, Software Engineering and Peer to Peer Networking.