Self Organized File Sharing In Proximity- Aware Interest- Clustered P2P System

Nibiya K J

PG Student, Dept. of Dual Degree Computer Applications, Sree Narayana Guru Institute of science and Technology N.Paravur, Kerala,India

---***---

Abstract – *The overall performance of peer-to-peer (P2P) file sharing lies on the efficiency of file query. To enhance the efficiency of file query in a structured peer-to-peer system, clustering technique can be used. Clustering peers by their common interests and by their physical proximity can improve file query performance. In the clustering technique the physically-close nodes are formed into a cluster and further physically-close and common-interest nodes are grouped into a sub-cluster based on a hierarchical topology. In the search mechanism, the file query will move to the nearest proximity cluster and to the corresponding interest cluster within that proximity cluster. In this paper, we propose a method, called the statistical matrix form (SMF), which improves the searching mechanism in the structured peer-to-peer system by selecting neighbors according to their capabilities. SMF measures the number of shared files, the content quality, the query service and the transmission distance between neighbor nodes. Based on these measurements, appropriate nodes can be selected by finding the rank of each nodes in the cluster, thereby reducing the traffic overhead significantly and enhance the file sharing efficiency in the structured P2P system.*

Key Words: **P2P Networks, File Sharing System, Sub Clustering, Clustering Technique, File Sharing.**

1. INTRODUCTION

A computer network or data network is a telecommunications network which allows computers to exchange data. In computer networks, networked computing devices exchange data with each other using a data link. The connections between nodes are established using either cable media or wireless media. The best-known computer network is the Internet.

Network computer devices that originate, route and terminate the data are called network nodes. [1] Nodes can include hosts such as personal computers, phones, servers as well as net-working hardware. Two such devices can be said to be networked together when one device is able to exchange information with the other device, whether or not they have a direct connection to each other.

Peer-to-peer (P2P) computing or networking is a distributed application architecture that partitions tasks or workloads between peers. Peers are equally privileged, equipotent participants in the application. They are said to form a peer-to-peer network of nodes.

2. PROBLEM STATEMENT

In a traditional file search mechanism, such as flooding, a peer broadcasts a query to its neighbors through an unstructured peer-to-peer (P2P) network until the time-to-live decreases to zero. The server maintains the index of all files in its sub-cluster. Efficient file query is important to the overall performance of peer-to-peer (P2P) file sharing systems. Clustering peers by their common interests can significantly enhance the efficiency of file query. Clustering peers by their physical proximity can also improve file query performance. However, few current works are able to cluster peers based on both peer interest and physical proximity.

Although structured P2Ps provide higher file query efficiency than unstructured P2Ps, it is difficult to realize it due to their strictly defined topologies. Every time a server receives a request, the server performs the search in two stages: the intra cluster searching and the inter-cluster searching. First the server perform the intra-cluster search. It consists of intra-sub-cluster searching and inter-sub-cluster searching. The server looks for the key of the requested file. If the key is found, the node sends the location of the file to the requester. If it is not found, the server performs the inter-cluster searching. Different files are classified into different sub-clusters based on their keys. For example music can be classified as pop, rock, jazz, classic etc., each with its own key. If there is a requested file in its sub cluster, the requester receives the location of the file from the server. Otherwise, the request is routed along its own cluster.

3. EXISTING AND PROPOSED SYSTEM

3.1 Existing System

A key criterion to judge a P2P file sharing system is its file location efficiency. To improve this efficiency, numerous methods have been proposed. One method uses a super peer topology which consists of super nodes with fast connections and regular nodes with slower connections. A super de connects with other super nodes and some regular nodes, and a regular node connects with a super node. In this super-peer topology, the nodes at the center of the network are faster and therefore produce a more reliable and stable backbone. This allows more messages to be routed than a slower backbone and, therefore, allows greater scalability. Super-peer networks occupy the middle-ground between centralized and entirely symmetric P2P networks, and have the potential to combine the benefits of both centralized and distributed searches. Another class of methods to improve file location efficiency is through a proximity-aware structure. The third class of methods to improve file location efficiency is to cluster nodes with similar interests which reduce the file location latency.

3.2 Proposed System

In a traditional file search mechanism, such as flooding, a peer broadcasts a query to its neighbors through an unstructured peer-to-peer (P2P) network until the time-to-live decreases to zero. A major disadvantage of flooding is that, in a large-scale network, this blind-choice strategy usually incurs an enormous traffic overhead. This paper presents a self organized file sharing in proximity aware interest clustered P2P system (SOPAIS). SOPAIS propose a method, called the statistical matrix form (SMF), which improves the flooding mechanism by selecting neighbors according to their capabilities. The SMF measures the following peer characteristics: 1) the number of shared files; 2) the content quality; 3) the query service; and 4) the transmission distance between neighbors. Based on these measurements, appropriate peers can be selected, thereby reducing the traffic overhead significantly. Our experimental results demonstrate that the SMF is effective and efficient. For example, compared with the flooding search mechanism in dynamic unstructured P2P networks, the SMF reduces the traffic overhead by more than 80 percent. Moreover, it achieves a good success rate and shorter response times.

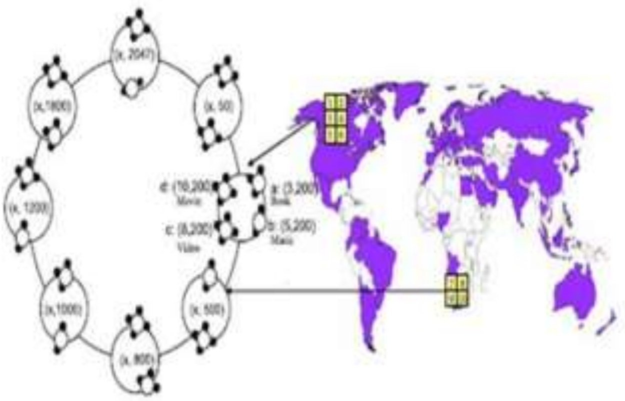

Fig -1: Peer Cluster

First classifies the interest of a sub-cluster to a number of sub-interests, and clusters common-sub-interest nodes into a group for file sharing. Second, SOPAIS builds an overlay for each group that connects lower capacity nodes to higher capacity nodes for distributed file querying while avoiding node overload.

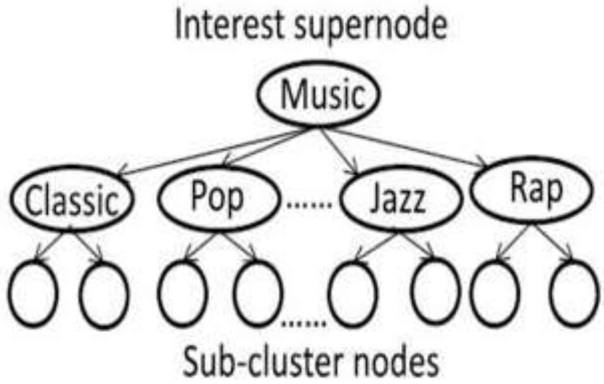

Fig -2: Sub Clustering

The SMF of query peer u is comprised of two matrixes: the left-hand matrix and the right-hand matrix. The left-hand matrix, called the feature matrix (FM), is an n x 4 matrix, where n is the number of neighbors of u. To derive the entries for the four columns of the FM, we compute the PA, ES, I, and TE scores for n neighbors of u. The PA of peers is analyzed to determine which peers leech the most resources without giving feed- back. The ES refers to the number of files that a peer shares, and can be used to classify a peer's sharing capability. It has been shown that, in a network, very few peers share a large number files, so that the quality of the files influences the sharing capability. The IP measures the number of files that a peer records in the index cache, and can also be used to analyze the number of responses in the cache content. Finally, the TE is utilized to measure the distance between peers in order to prevent inefficient routing. The right-hand matrix, called the weight matrix (WM), is a 4 x 1 matrix in which each peer can set the proper weights according to the derivation degree of each feature. Finally, each query peer u computes a scoring matrix (SM), which is an nx1 matrix obtained by the matrix multiplication FM x WM. To deliver queries for u, we obtain the score of each of u's neighbors in the SM and then select the neighbors with the top-k scores to send query messages. Since the query peer u has have neighbors, v, w , x, y, and z, its feature matrix is a 5 x 4 matrix; the weight matrix is a 4 x 1 matrix; and the score matrix is computed by the formula SM = FM x WM, which is a 5 x 1 matrix.

3.2.1 Feature Matrix Construction

We define the d-collected scope of a query peer u as the set of peers that are at most d hop(s) away from u. In the construction of the FM presented in the following subsections, we collect relative information about the d-collected scope of a query peer. We could improve the search performance by increasing the value of d; however, it may increase the computational overhead because of the extra cost of collecting information.

Therefore, the problem is how to choose appropriate values of d to improve the performance by determining the acceptable extra-overhead incurred by the collection and exchange of information.

PROCESSING ABILITY In P2P networks, there are usually free loaders who down-load files without sharing any of their resources, which impacts the search performance of coadjutant[5] communities. To prevent free loaders, we utilize the PA to differentiate between leeching and enthusiastic peers.

Query Frequency(QF) In a P2P network, a query peer that generates a lot of queries may be a free loader. Let N(u) be the neighbors of a query peer u; that is, N(u) are peers that are one hop away from u. In addition, let NQ(v) be the number of queries sent by v. Each query peer u computes SQ1(u), which is the total number of queries (SQ) sent p.

from the peers that are one hop away from u. Formally, SQ1(u) = v2N(u) NQ(v)

The Query-Minus-Score (QMS) of a neighbor v of u is defined as QMS(u; v) = SQ1(u) NQ(v)

When NQ(v) increases, the possibility of v being regarded as a free loader also increases and peer v will be assigned a lower score. Next, each query peer u computes SQMS1(u) (resp. SQMS2(u)), which is the sum of the query-minus- scores (SQMS) of all peers that are one (resp. two) hop(s) away from u:

QF (u; v) = w1 QMS(v) + w2 SQMS1 (v)
SQMS1(u) SQMS2 (u)

Where w1 and w2 are two parameters used to adjust the influence of peers that are one hop away and two hops away from u respectively. A peer can determine the amount of resources that their neighbors leech from the network.

4. SYSTEM ARCHITECTURE

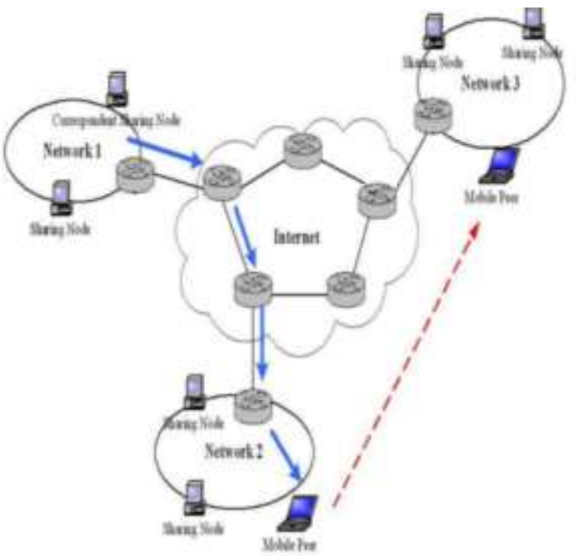

Fig -3: System Architecture.

5. MATHEMATICAL MODEL

Input: File query in DTN. Output: File query result in less time in DTN. Process: Let W be the whole system, W = {U, S, C, B, R, r, F}. Where,
1. U is the set of number users.

U = {U1,U2 . . . . Un}.

2. S is the system which contains the unstructured data to provide the service to user based on user request.

3. C is set of number of cluster based on user request.

C = {C1, C2, ... Cn}.

4. B be set of bloom filter which is required to filter the user requests based on user interest.

5. F be the set of files user is requesting.

F = {f1, f2 . . . . . . fn}.

6. R be the user request for file to S.

7. r be the rank assigned to file based user request.

Step 1: User U login to the system and request for particular f1 to the system.

Step 2: The system S will process the user request R from the unstructured data. In this the bloom filter will filter the user request to check whether the same file request has come before or not if not it will rank that file.

Step 3: The system will process the user request R based on ranks assigned to files by using bloom filter.

Step 4: If same file request R is come at system more than 2 times (assigning threshold) then system will create an interest-cluster for that requested file to minimize the searching time as system will search the requested file from unstructured data.

6. THE RELATED WORK

The super-peer network i s for efficient and scalable file consistency maintenance in structured P2P systems. Our previous work built a super-peer network for load balancing. We proposed a self-organizing super-peer network architecture that solves four issues in a fully decentralized manner: how client peers are related to super-peers, how super-peers locate files, how the load is balanced among the super-peers, and how the system deals with node failures. Mitra et al. developed an analytical framework, which explains the emergence of super-peer networks on execution of the commercial P2P bootstrapping protocols by incoming nodes. Chordella is a P2P system that is particularly designed for heterogeneous environments such as wireless networks. Sachez-Artigaz et al. investigated the feasibility of super-peer ratio maintenance, in which each peer can decide to be a super-peer independently of each other. Liu et al. proposed a hierarchical secure load balancing scheme in a P2P cloud system. It first balances the load among super nodes, and then depends on each super node to balance the load among nodes under its management. Garbacki et al. A proposed a self-organizing super node architecture to facilitate file querying. Each super node caches the files recently requested by its children, and other peers send requests to the super nodes that can solve most of their requests.

7. RESULT AND ANALYSIS

7.1 Results

We implemented a prototype of SOPAIS on a real-world distributed test bed, to measure the performance of PAIS in comparison with other P2P file sharing systems. We randomly selected 350 Planet Lab nodes all over the world. Among these nodes, we randomly selected 30 nodes as landmark nodes to calculate the Hilbert numbers of nodes. We clustered all nodes into 169 different locations according to the closeness of their Hilbert numbers. The files are randomly assigned to a sub-cluster with the files' interest over the total 160 locations, and then randomly assigned to nodes in the sub-cluster. Eighty percent of all queries of a requester target on files with owners within the same location, among which 70 percent of its queries are in the interests of the requester. According to, 80 percent of all requests from a peer focus on its interests, and each of other requests is in a randomly selected interest outside of its interests. A request in an interest means a request for a randomly selected file in this interest. We also let each file have a copy in another peer in a different location in order to test the proximity-aware file searching performance.

Fig -4: Admin Login Form

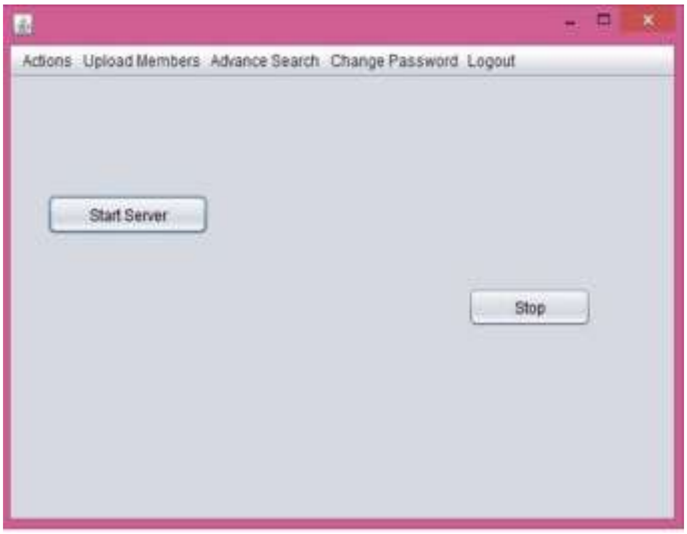

Fig -5: Admin Home Page

Fig -6: Admin Upload Form

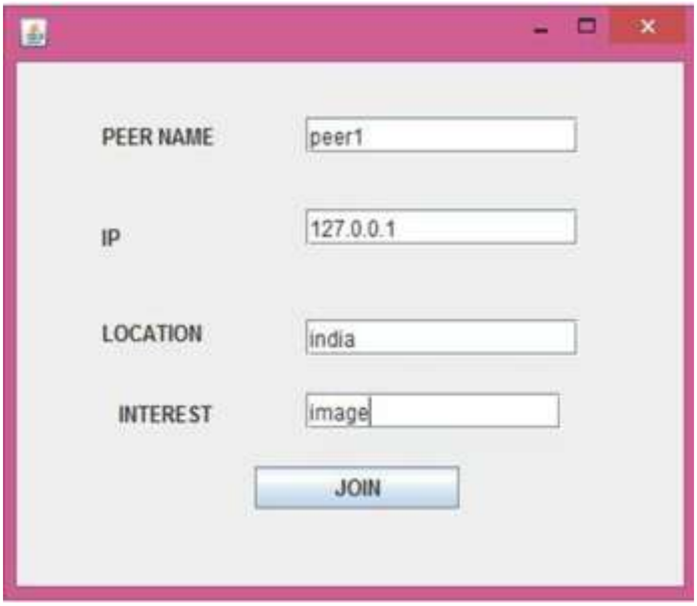

Fig -7: Peer Form

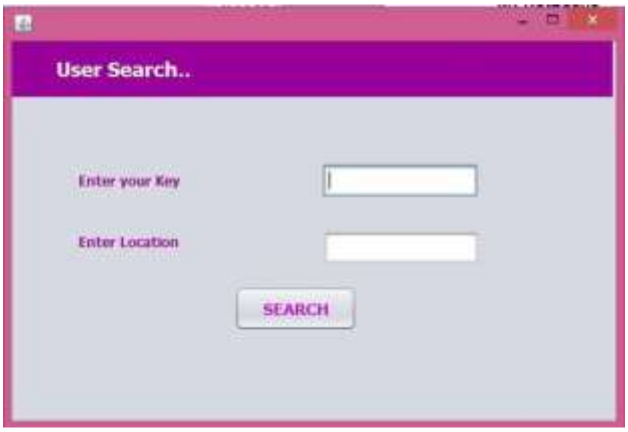

Fig -8: User Search Form

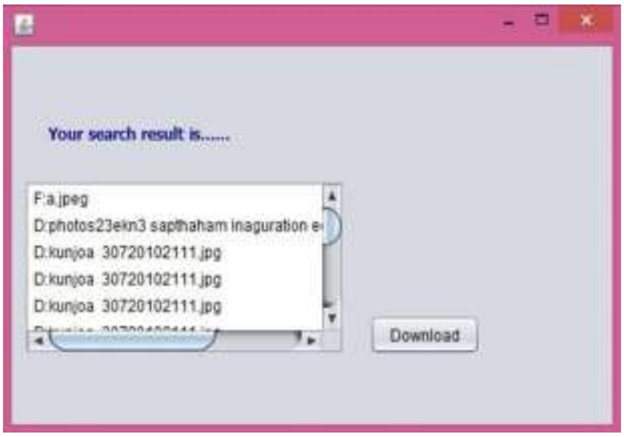

Fig -9: Search Result

7.2 Experimental Analysis

In this section the comparison of the search time for normal peer, proximity peer and SMF peer are done. From the below graph, we can analyze that the search time for SMF peer is much less than the other two peer type, thus we can conclude that this SMF based file sharing is more efficient than the existing methods.

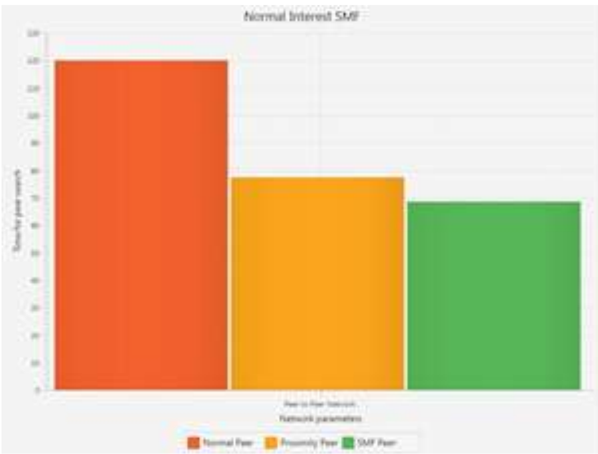

Fig -10: Search Time Comparison

8. CONCLUSION AND FUTURE WORK

To enhance the efficiency of file query, clustering technique can be used. Clustering peers by their common interests and by their physical proximity can improve file query performance. This paper we introduce a clustered P2P file sharing System based on a structured P2P, in which physically-close nodes are formed into a cluster and further physically-close and common-interest nodes are grouped into a sub-cluster based on a hierarchical topology. In the existing unstructured P2P system, there is no responsibility assignment for each nodes. The proposed structured P2P system has definite rules for node join and leave. In the proposed structured P2P system, file querying is efficient due to the proximity and interest clustered file sharing. Due to the strictly controlled topology, the data placement and lookup algorithms are precisely defined based on a DHT data structure and consistent hashing function. The statistical matrix form (SMF), which improves the flooding mechanism by selecting neighbors according to their capabilities. The SMF measures the following peer characteristics: 1. The number of shared files, 2. The content quality; 3. The query service, and 4. The transmission distance between neighbors. Based on these measurements, appropriate peers can be selected, thereby reducing the traffic overhead significantly. The self-organizing property of the super peer can solve the issues related to the client relation with the super peer, issues associated with the file location by the super peer, load balancing in the super peer topology, thereby enhancing the file sharing efficiency.

REFERENCES

[1] L. W. M Haiying Shen, Guoxin Liu, “A proximity-aware interest- clustered p2p file shar-ing system,” IEEE transactions on parallel and distributed systems.

[2] M. V. S. P Garbacki, “The design and evaluation of a self-organizing super-peer network" ieee trans. comput., vol. 59, no. 3,” pp. 317–331, 2010.

[3] M. v. S. P Garbacki, “Optimizing peer relationships in a super-peer network,” in Proc. Int. Conf. Distrib. Comput. Syst., 2007.

[4] G. X. Z. Li and Z. Li, “Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems,” in IEEE Trans. Parallel Distrib. Syst., vol. 19, no. 12, 2008.

[5] H. Shen and C.-Z. Xu, “Hash-based proximity clustering for efficient load balancing in heterogeneous dht networks,” in Parallel Distrib. Comput., vol. 68, pp. 206–224, IEEE, 2008.

[6] C. Hang and K. C. Sia, “Peer clustering and firework query model,” in Proc. Int. World Wide Web Conf., pp. 273–283, ACM, 2008.

[7] H. S. G. Liu and L. Ward, "An efficient and trustworthy p2p and social network integrated file sharing system," in Proc. P2P, 2012.

[8] K. K. K. Y. K. Elkhiyaoui, D. Kato and P. Michiardi, "A scalable interest-oriented peer-to-peer pub/sub network," in Proc. 9th Int. Conf. Peer-to- Peer Comput., ACMr, 2009.

[9] M. Yang and Y. Yang, "An efficient hybrid peer-to-peer system for distributed data shar-ing," in IEEE Trans. Comput., vol. 59, no. 9, IEEE, 2010.

[10] C. X. H. Shen and G. Chen, "Cycloid: A scalable constant-degree p2p overlay network," in Perform. Eval., vol. 63, 2006.

[11] J.-J. Z. CHIA-HUNG LIN and S.-Y. HSIEH, "Improving the search mechanism for un-structured peer-to-peer networks using the statistical matrix form," in Digital Object Iden-tifier 10.1109/ACCESS, 2015.

BIOGRAPHIES

Nibiya K J
D/O Johny K P
DDMCA Student at
SNGIST, N. Paravur,
Ernakulam
Kerala,India