
A Cluster Based Locality Aware P2P File Sharing System Using Super Peers

Divya.V¹, Prof. Rabindranath.S², Poornima.Pujar³ & Rekha. Patil⁴
4th Semester M.Tech, Software Engineering,
AMC Engineering College Bangalore

Abstract- *The efficient and effective querying of file is important in increasing the overall performance of p2p sharing of a file in a system. Grouping of each node through their common interest can effectively enhance the filing of query and it can also be done through the physical proximity. The proposed cluster based locality is based on p2p, which combines physically closest nodes into a cluster and henceforth, combine the common interest nodes into another sub-cluster based on their hierarchical topology. Cluster based locality system involves the method of intelligent file replication algorithm to enhance the file query efficiency. It commonly creates the replicas of each file that are frequently accessed and create a common clustering in particular nearest location. Initially, it classifies all the interest nodes into sub-cluster as a group. Second, it builds an overlay to reduce node overload. Third, it reduces the file sharing delay by using proactive file information collection. Forth, using bloom filter based information it reduces the overhead of file information collection. Fifth, it also improves efficiency in file sharing. Finally, when the visited node tends to visit again then the bloom filter will make updating for information and as to reduce search delay.*

Keywords: *file replication, bloom filter, file sharing system, cluster based locality, sub-interest, p2p network.*

1 INTRODUCTION

From the past few years, the huge popularity of the internet is producing a significant stimulus in p2p file sharing system. The example of Bit Torrent [1] constitutes only about 35% of overall traffic on the internet. We can find 2 classes of p2p system. (i) Structured network (ii) unstructured p2p network. As we know the nodes join and leave each of networks according to some basic rules. At present,

unstructured p2p network file query method is based on either random walker where each query packet is forwarded to random picked neighbour and either by flooding where each query is propagated to all of neighbour node until the file is found. But these random walkers and flooding don't guarantee data location. But in structured p2p network (DHT's) can overcome the drawbacks with each of features of high scalability and efficiency along with deterministic file location. Each of them have strict and control topology and data placement, lookup hashing function. The node that is responsible for key can always be found even when the system is in continuous state of change.

2 RELATED WORKS

We discuss about the related work that are mostly related to each of the cluster based locality and that also uses various topology and approaches to enhance the file location efficiently.

Proximity Awareness: Various techniques to exploit each topology information in p2p overlay routing include geographic layout, proximity neighbour selection and also proximity routing. The geographic layout maps each of the logical ID space to their physical network. It is already employed in topologically aware CAN[11]. In this proximity routing method, the logical overlay is constructed without considering the underlying physical topology.

Eg. Pastry[5], each of constraint will be the node ID prefix). This method can also been adapted to Chord[4] and CAN[27]. They select the routing table entries pointing to the topologically nearest among all the nodes with node ID in the desired portion of each ID space. Bit Torrent downloading scheme is also been proposed by Gross et.al[35] which has the scheme of locality aware file searching and the replication in order to supply a fast downloading and robust. Each of Bit Torrent contains various components such as data content, the tracker, original content provider, Meta info file and the end host/clients.

3 EXISTING SYSTEM

The key criterion to judge a file in p2p sharing system can be its file location.

Super peer topology: It consists of super nodes with wide range of fast connections and each of regular nodes with slower connection. Each of super nodes can connect with other of super node and some of the regular nodes and in turn to a super node. The nodes that are at a centre of network are faster and always produce a more stable and reliable backbone. This also allows more messages to be routed than a slower backbone and allows good stability.

It occupies middle-place between entirely symmetric and centralized p2p network and they also have the potentiality of combining the benefits of both the distributed and centralized searches. Cluster Based Locality: it is used to group a closed peer physically to effectively improve the efficiency. The next methods so as to improve file location efficiency by cluster node with similar interests, which reduce the file searching latency.

Disadvantages:

- The proximity abstraction derived from p2p system does not necessarily match the physical proximity information in reality.
- According to routing protocol the shortest path chose is not necessarily the shortest physical path. Hence, this mismatch becomes a very big obstacle for the process of deployment and performance optimization of p2p file sharing system.
- Only certain methods are able to cluster the peers according to both interest and proximity.
- They don't have any strict policy for the construction of topology.
- These methods can't be directly applied to general DHT.

4 PROPOSED SYSTEM

A cluster based locality and interest clustered p2p file sharing system is on structured p2p system. Each of the cluster based locality form physically close nodes as a cluster based on their interest. They also create a replica of each file that is accessed frequently requested by a group of physically closed nodes in their location. It also places each of files with same interest together and make them accessible easily through DHT lookup () routing function. The following fig (4) represents the cluster based location of the network.

Initially, it classifies all the interest nodes into sub-cluster as a group.

Second, it builds an overlay to reduce node overload.

Third, it reduces the file sharing delay by using proactive file information collection.

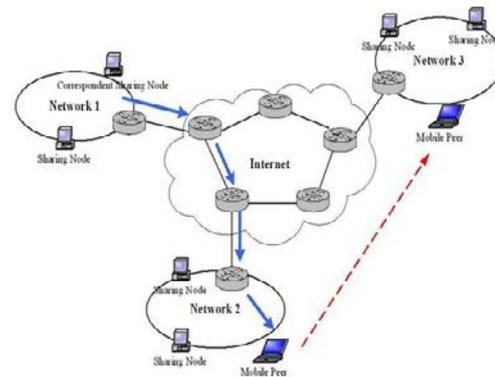
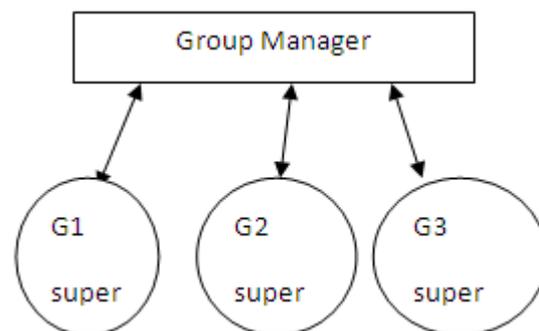


Fig (4) System Architecture of Cluster Based Locality Network

Forth, using bloom filter based information it reduces the overhead of file information collection.

Fifth, it also improves efficiency in file sharing. Finally, when the visited node tends to visit again then the bloom filter will make updating for information and as to reduce search delay.

fig (4)a Group Based Negotiation:



In this paper we propose two different approaches to increase the performance of a file searching mechanism. They can be as follows:

- File Querying Mechanism
- Sub-Interest File Querying Mechanism

File Querying Mechanism: It is used to attain the high efficiency of file searching policy. It includes several stages such as:

- ❖ Intra-sub cluster searching

- ❖ Inter sub-cluster searching
- ❖ Inter cluster searching (DHT routing)

If the intra sub clustering fails then, cluster based node can depend on inter sub cluster searching. If this inter sub cluster searching also fails in turn, it will rely on DHT routing for file searching.

Pseudo code for searching file in cluster based node:

- Consider the file 'f'
- Identify the key for the file 'f'
- Get the 'f' ID
- If the key of file 'f' belongs to interests then,
- Continue sending request to the particular server of sub-interest
- Then if positive response then
- Exit;
- Else
- Look for the file 'f' ID and continue the search in other clusters.
- Endif;
- Endif;

Sub Interest File Querying Mechanism: The cyclic index that will be performed to identify the required file using various mechanism to reduce the time latency. To achieve high efficiency in both of the inter cluster and intra cluster searching. They uses the relative small path with gross grained interest classification and also uses additional method to improve the searching in intra cluster method. The chosen small path to identify the file/information would improve the file searching mechanism.

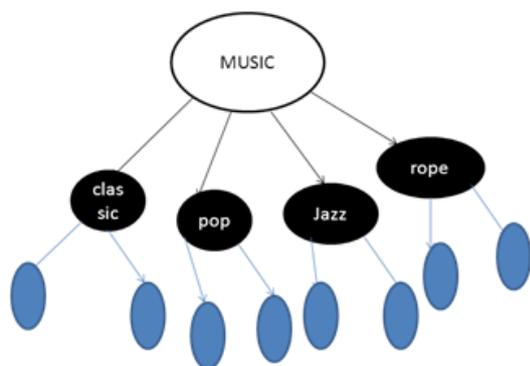


Fig 4(b)Sub Interest file querying

Advantages:

- Cluster based locality system uses an intelligent file replication algorithm to further enhance file lookup efficiency.

- They rely on DHT lookup policy instead and broadcasting, its construction consumes much less cost in mapping nodes to cluster and mapping back clusters to interest sub-clusters.

5 CONCLUSION

The enhancement of file location efficiency in p2p system, clustered locality super p2p and interest clustered network has been proposed. Although these strategies improve the performance, but with the introduction of physical nearest cluster and peer interest the efficiency can be still faster. In this cluster based locality p2p system and interest cluster p2p file sharing system is structured network would rapidly increase the performance rate of identifying file location. It groups the peers based on their interest and physical locality with the help of hierarchical structure. Cluster based locality uses an intelligent file replication to enhance physical locality of frequently accessed nodes for improving efficiency. Finally, clustered based locality enhances the file sharing efficiency among the nearest node and common interest node through several approaches as mentioned.

6 FUTURE ENHANCEMENTS

In the proposed paper, file distribution is done between the peers by FILE AND FILEID. It can be enhanced to provide more the user specific search such as file name and their sub-interest. Hence, the requester will get specific file of his interest. It can be enhanced by creating a new super peer node which contains only the information about the sub-interest division of each file and also it reduces the searching time delay.

REFERENCES

1. Morpheus. (2008) [Online]. Available: <http://www.musiccity.com>
2. P. Garbacki, D. H. J. Epema, and M. van Steen, "Optimizing peerrelationships in a super-peer network," in Proc. Int. Conf. Distrib.Comput. Syst., 2007, p. 31.
3. Y. Hua, H. Jiang, Y. Zhu, D. Feng, and L. Tian, "Semantic-awaremetadata organization paradigm in next-generation file systems,"IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 2, pp. 337-344, Feb.2012.
4. Bit Torrent user activity (2008). Available:<http://www.cs.brown.edu/palvo>