ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A Review on Efficient File Sharing in Clustered P2P System

Anju S Kumar[1], Ratheesh S[2], Manoj M[3]

[1]PG scholar, Dept. of Computer Science, College of Engineering Perumon, Kerala, India

[2]Assisstant Professor, Dept. of Computer Science, College Of Engineering Perumon, Kerala, India

[3]Assistant Professor, Dept. of Computer Science, UKF College of Engineering and Technology, 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, clustering technique can be used. Clustering peers by their common interests and by their physical proximity can improve file query performance. In this paper we introduce a clustered P2P file sharing System based on a structured P2P.The structured P2P system provide higher efficiency in file querying. 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. The clustering by their proximity and interest information will be helpful in each file searching due to the presence of other nodes with the same interest within the proximity of that node. The objective of this paper is to analyse how these approaches works in the file sharing in Peer-to-Peer network and what are the impacts of these approaches in file sharing after applying it.

KEYWORDS*:* P2P Networks, File sharing system, Superpeer Network, Proximity awareness, Interest Clustering, File replication.

I. INTRODUCTION

A significant amount of work has been done in the field of performance optimizing and efficiency of content sharing peer-to-peer (P2P) networks[3].There are two classes of P2P systems: unstructured and structured. Unstructured peer-to-peer networks do not impose a particular structure on the overlay network by design, but rather are formed by nodes that randomly form connections to each other, where file query method is based on either flooding, where the query is propagated to all neighbours of node, or random- walkers where the query is forwarded to neighbours which is chosen randomly until file is found in the search. Structured P2P networks Distributed Hash Tables (DHTs) [5],by which a node responsible for a key can always be found even if the system is in a continuous state of change.

The overall performance of peer-to-peer (P2P) file sharing systems lies on efficient and trustworthy file querying. Several methods have been proposed to improve file location efficiency. One another method is a super-peer network topology system. This super-peer network topology consists of super nodes with high connectivity and regular nodes with low connectivity. A super node connects with other super nodes and some regular nodes and they are of high capacity. The regular node connects with a super node. Another method to improve file location efficiency is through clustering of nodes in the network. File replication technology is also widely used to reduce hot spots and improve file query efficiency.

Although numerous clustering methods have been proposed with different features, few methods can efficiently enhance the file location efficiency. This paper presents a technique of clustering nodes based on their interest and nodes proximity. In this, physically close nodes in the network are formed into a cluster and further groups physically-close and common-interest nodes into a sub-cluster [1]. Files with the same interests are placing together and these files are made accessible through the DHT Lookup () routing function. Thereby the search for a file in the network will be more efficient.

In this clustering method there is a proactive information collection of each and every node in the network and these information are distributed among all the nodes in the network. This proactive information collection will be helpful for the user in such a way that the user can know about the locations of almost all files they want to search. A more efficient method of file replication is also incorporated with this system of node clustering that, a replication of each and every file that are frequently requesting are made and is distributed among all the nodes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

The most relevant works related to this efficient file sharing by node clustering are:

Super-peer topology
Proximity-awareness
Interest-based clustering
Consistency maintenance

II. SUPER-PEER TOPOLOGY

In structured P2P systems, consistency maintenance and load balancing can be achieved through the superpeer topology, which exploit the heterogeneity of nodes in a peer-to-peer (P2P) network by assigning additional responsibilities to high capacity nodes called super-peers. Weak peers submit queries to their super-peers and receive results from them. There may be several issues are associated with locating a file in the super-peer network including client peer-super peer relation, load balancing, file location etc. Such issues can be solved by the self-organized super-peer technology, in which the peer relations are automatically discovered, maintained and exploited without user intervention.

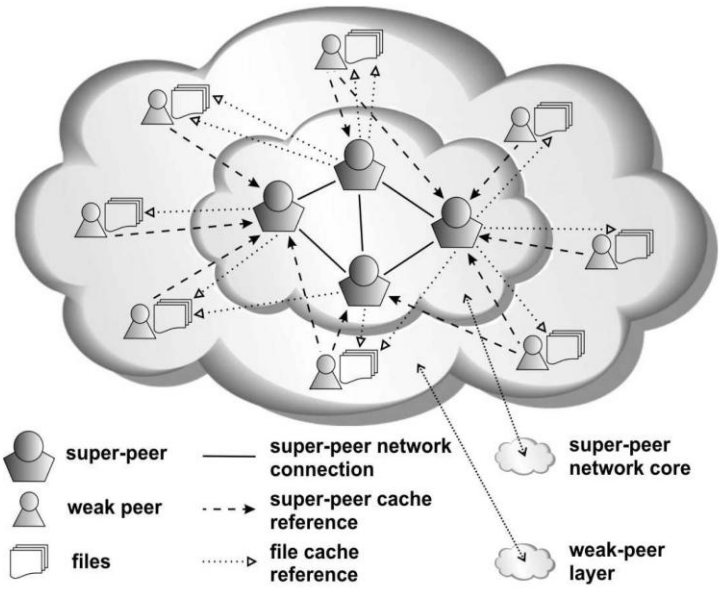

Fig.1. Structure of self-organizing superpeer network.

In this self-organizing super peer technology [2] there is an efficient caching technique that makes the file querying more reliable. I.e. a two level caching. The Fig.1 shows the structure of the self-organizing superpeer network. The super peer possess a file cache that containing pointers to files, which are requested by the peers and the client peer possess the super peer cache containing the super peers that offers the best performance [3].The super peer cache contains the identities of the super peer i.e., the IP address and port number. The cache and the file cache are assigned priorities and based on this priority only the searches are proceeding in the network. The priority of super peer in the super peer cache is increased by one after each positive response based on the LFU policy and the priority increment in the file cache is through the mixed LFU and LRU policy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

III. PROXIMITY-AWARENESS

Proximity Awareness is the first step in the clustering technique in the P2P file sharing system. The physically close nodes are grouped together and within this physically close nodes only further classifications of the nodes can be done as shown in fig.2. Node closeness can be represented by a method called Landmark method [1].The main intuition in this method is that the physically close nodes likely to have similar distance to the certain number of landmarks that are selected. With an assumption that m landmarks are randomly scattered in the internet and measures the distance of each nodes to these landmarks. Thus uses the vectors of distances $<d_1, d_2\ldots.., d_m>$ as Cartesian coordinates of the space. The nodes with the similar vector are considered as physically close nodes and are grouped together to form cluster.

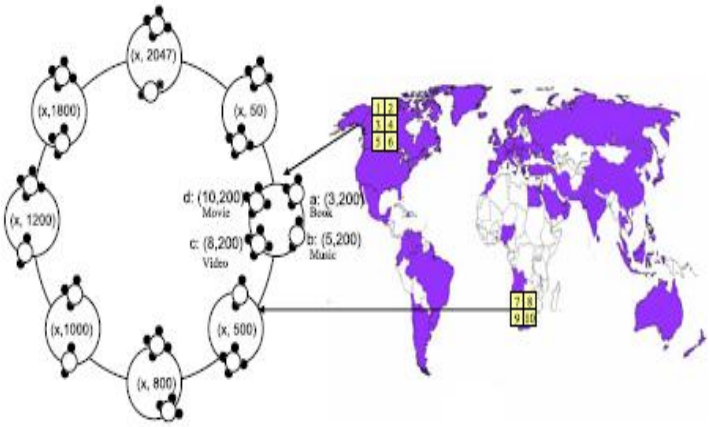

Fig.2. Structure of node clustering

In the Fig.2, the outer circle represents the cluster within a particular vicinity which is formed by proximity aware and within this cluster the nodes are formed to sub-cluster.

IV.INTEREST BASED CLUSTERING

Within the physically close nodes, further classification of nodes are done based on the interest each node is sharing. Nodes with similar interest are grouped together to form clusters [1] [8].The clustering technique starts with the clustering of the physically close nodes and continues by clustering the nodes with their interest i.e., nodes with same interest. The Fig. 3 given below shows the clustering of peers in the peer-to-peer network.

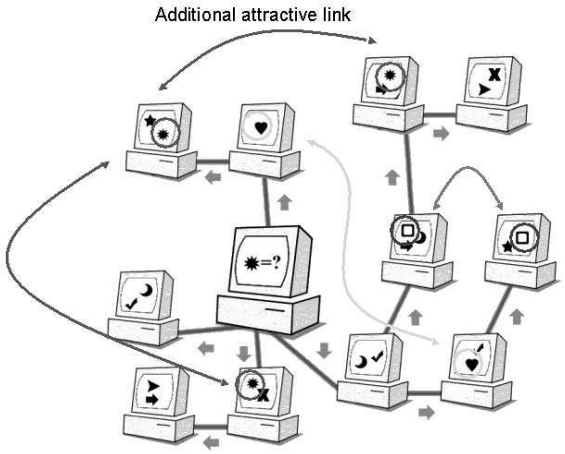

Fig.3. Peer Clustering

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

For the interest identification a method of signature calculation is used. Each peer is having a collection of data and these data collection is prepossessed and calculate a signature value SIG to characterize their data properties. Thus each signature value represents the interest associated with each node. Then in the clustering method the signatures values are analysed and the nodes with similar signature values are formed to a cluster.

V. CONSISTENCY MAINTENANCE

Consistency maintenance mechanism is necessary for the peer-to- peer (P2P) system due to their frequent data updates. For each replica group an overlay network is established with two layers [4] [9] as shown in Fig.4

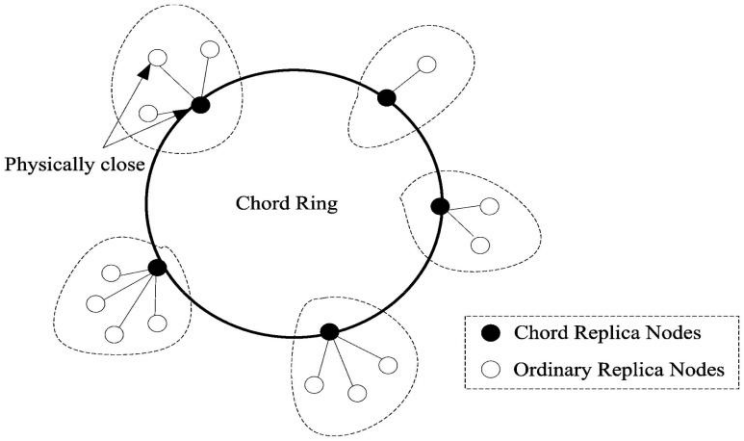

Fig.4.Hierarchical structure of nodes with CRNs and ORNs

The upper layer is Distributed Hash Table (DHT) based and consists of powerful and stable replica nodes called Chord Replica Nodes (CRN).The lower or the second layer consists of the Ordinary Replica Nodes(ORN).
Lower layer is attached to physically close upper layer. When an updated replica emerges on a node the upper layer initialize a tree called Update Message Propagate Tree (UMPT) by partitioning DHT identifier space. On specific update, the update massage passes through the tree and every replica nodes receive the update message. The CRN are the stable nodes and the ORN are attached to physically close CRN. These two constitute the cluster and the consistency is maintained in each updates in the cluster.

VI. FILE QUERYING

The server maintains the index of all files in its sub-cluster. 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 performs the intra-cluster search. It consists of intra-sub-cluster searching and inter-sub-cluster searching [1].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. In this search the DHT Lookup is done among the nodes which enhance the efficiency of file searching.
Different files are classified into different sub-clusters based on their keys. For example music can be classified as pop, rock, jazz, classic, melody 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. This clustered P2P system relies on file replication to further improve its file location efficiency. In this, nodes inside a large sub-interest group first search files among their neighbours in a distributed manner. The file requester forwards its request to the sub-interest super node as a complementary method if the search fails. When a new query appears, the peer checked it against a local cache for duplication. If it is found to be duplication, that is, the same message has passed through before, the message will not be propagated. Another one mechanism is the Time- To-Live value to indicate how long a message can survive. Similar to IP packets, every message are associated with a TTL. The TTL value is decreased by

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

one at each time the message passes through a peer. The message will be dropped and no longer forwarded if the TTL is zero.

VII. COMPARISON

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. Currently query routing is either flooding or random walkers, in which there is no guarantee for data location, because query is simply flooding or randomly propagate to any of the neighbours. 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 DHT Lookup method can efficiently enhance the file searching efficiency. Proximity and interest based clustering may help in the efficient location of the data in the network. The intelligent file replication method replicates frequently requested file by physically close nodes to enhance the file lookup efficiency.

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.

VIII. CONCLUSIONS

To enhance the efficiency of file query, clustering technique can be used. Clustering peers by their common5 interests and by their physical proximity can improve file query performance. This paper, we introduce a clustered file sharing System based on a structured P2P network, 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. This clustered system uses an intelligent file replication method to replicate a files that are frequently requesting by physically close nodes near their physical location to enhance the file lookup efficiency and thereby enhance the overall performance of file sharing in the P2P system.

REFERENCES

1. Haiying Shen, 'A Proximity-Aware Interest- Clustered P2P File Sharing System', IEEE transactions on parallel and distributed systems, vol. 26, no. 6, June 2015.
2. P. Garbacki, D. H. J. Epema, and M. V. Steen 'The design and evaluation of a self-organizing super-peer network', IEEE Trans. Comput., vol. 59, no. 3, pp. 317331, Mar. 2010.
3. P. Garbacki, D. H. J. Epema, and M. van Steen 'Optimizing peer relationships in a super-peer network', in Proc. Int. Conf. Distrib. Comput. Syst., 2007, p. 31.
4. Z. Li, G. Xie, and Z. Li 'Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems', IEEE Trans. Parallel Distrib. Syst., vol. 19, no. 12, pp. 16951708, Dec. 2008.
5. H. Shen and C.-Z. Xu, 'Hash-based proximity clustering for efficient load balancing in heterogeneous DHT networks', J. Parallel Distrib. Comput., vol. 68, pp. 686702, 2008.
6. C. Hang and K. C. Sia, 'Peer clustering and firework query mode', in Proc. Int. World Wide Web Conf., 2002.
7. G. Liu, H. Shen, and L. Ward, 'An efficient and trustworthy P2P and social network integrated file sharing system', Proc. P2P, 2012, pp. 203–213.
8. K. Elkhiyaoui, D. Kato, K. Kunieda, K. Yamada, and P. Michiardi,'A scalable interest-oriented peer-to-peer pub/sub network', in Proc. 9th Int. Conf. Peer-to-Peer Comput., 2009, pp. 204–211.
9. M. Yang and Y. Yang, 'An efficient hybrid peer-to-peer system for distributed data sharing', IEEE Trans. Comput., vol. 59, no. 9, pp. 1158–1171, Sep. 2010.

BIOGRAPHY

Anju S Kumar is a Master of technology (M-tech) student in Computer Science Department, College of Engineering, Perumon. She received her Bachelor of Technology (B-tech) degree from Cochin University of Science and Technology. Her research interests are Networking, Network security etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Ratheesh S is an Assistant professor in College of Engineering, Perumon. He received his Master of technology (M-tech) degree and Bachelor of Technology (B-tech) degree from MS University. His research interests are Image processing, Image restoration, Image editing, Multimedia processing etc.

Manoj M is an Assistant professor in UKF College of Engineering and Technology, Kollam. He received his Bachelor of Technology (B-tech) degree from Kerala University. His research interests are Networking, Network Security etc.