Survey on Cost Effective Replication in Locality Aware Interest Cluster Peer to Peer File Sharing System

V.Mani[1,] S.Gowsika[2]
[1] Assistant Professor, [2] PG Scholar,
Department of Computer Science and Engineering, M.Kumarasamy college of Engineering, Karur, India

_____

Abstract - **In peer to peer networking, in order to improve the overall performance of the file sharing system the competent file querying is essential, for this a method called Locality aware interest cluster p2p file sharing system (LAIC) is introduced. In this work, based on the locality the physically close nodes are clustered and based on the interest sub cluster is formed. Locality aware interest clustering enhances the process through various steps. First, the proximity nodes are clustered and the interest nodes are classified into sub-cluster as a group. Second, it constructs an overlay to diminish node overload. Third, it reduces the file sharing setback by using proactive file information collection. Fourth, the overhead of the file information collection is reduced by using bloom filter based information. Fifth, the Cluster leader is elected by using JSD method. In this work, file querying and sub-interest file querying mechanism is proposed to increase the efficiency in interest cluster. To further enhance the file query efficiency, a cost effective file replication algorithm is introduced In-order to reduce the cost as well as to improve the query efficiency.**

Index Terms - P2P, LAIC, Bloom filter, File replication, cluster.
_____

I. INTRODUCTION

P2P stands for peer to peer network. The term peer means that the computer system or nodes. The vast popularity of the internet has shaped a major motivation to P2P file sharing systems. A P2P network is a type of decentralized and distributed architecture. In this type of network, information is shared between multiple peers that access the resources directly with the help of other network with any servers. P2P networking utilize the Internet for client-based computing tasks and in modern world personal computers have speedy processor, huge memory, and a bulky hard disk so that they can perform the common computing tasks like e -mail and web browsing.

A server computer usually has large number of resources and must respond to the resources for the data transaction between many client computers. Client computers kick off requests for resources or data from server computer. Many major works has been done for optimizing the performance and for the efficient content sharing in peer to peer network. Peer to Peer network is classified into two types, Unstructured Peer to peer network and structured peer to peer network and the main principle involved in P2P network are sharing resource, self organization and decentralization.

Unstructured Peer to Peer network: This type of network does not inflict on a particular structure by their design on the overlay network. In unstructured P2P the file querying process is done either by flooding or by the random walker's. In this method the query is propagated to all the neighbor nodes which is chosen at random, the query forwarding will continue until the file found during this searching process. Although it process the query well, it does not guarantee the data location.

Structured Peer to Peer network: It uses Distributed Hash Table (DHT)in which the neighbor relationship between the node and the data locations are strictly defined. Distributed hash table is a kind if decentralized and distributed system that provides lookup service. As in the hash table the lookup services are similar. (value, key) pair is stored in DHT in which the values associated with a given key can be efficiently retrieved by the participating node. The node is in charge for a key can always be found even though if the system is in a permanent status of change.

Among the structured system, the DHT is implemented using the different data structure. The DHT structure is classified into abstract key space and random unique key(identifier). The abstract key space consists of large inter values (range from 0 to 2^{128}-1). From this random unique (identifier) key space is assigned to each participant. All node maintains a small routing table consisting of its neighboring peer. These routing information are linked together to form a overlay network. The routing procedure will traverse among the nodes based on the routing information to reach the destination. This process is sometimes called as key based routing.

II. LITERATURE REVIEW

Robert Morrisz [7] provides a solution to the problem that confronts Peer to Peer application ie. The problems like decentralization, availability, load balancing and supple naming. To overcome these kind of problems he proposed a structured p2p network protocol called chord. It is one of the most popular, scalable, lookup protocol in structured Peer to Peer network. Chord has some library functions. It provides lookup (key) which yields IP address of a peer with respect to its key. Chord uses finger table as its routing table and skip list as its routing data structure and SHA 1 hash function. Query will be passed to the successor node when a client queries a key to find its successor (k). In case key is not found the query time will be increased to O

(N). To avoid this situation by using a finger table a fastest searching method is implemented. The main disadvantage of this method is that it handles the queries which match exactly and maximum request is for replicated files.

Paul Francis [9] proposes one of the other DHT system called Scalable Content Addressable Network. The main motive behind this paper is to provide a scalable indexing system for large scale decentralized storage application on the internet. CAN uses a d dimensional cartesian space coordination system, where the keys are hashed to map on to the values. The main operations performed in this DHT overlay protocol are insert(key,Value) and retrieve(key). In this paper francis deals with CAN construction, routing, and maintenance. The performance are simulated and it is compared.

There are some techniques to develop each topology information in P2P overlay steering includes proximity awareness *i.e.* neighbor selection, routing and geographic layout. Antony Rowstron [1] proposed design and development of pastry, which is alike to that of chord. PASTRY assigns 128 bits node id to each node in the system and each node is responsible for handling and routing requests for numeric keys to node with closest node id($B=2^b$). All node in the pastry maintains structures namely, neighborhood set, a routing table, and a leaf set. Routing table consists of ($\log_B N$) rows and B-1 columns (B is the configuration parameter with typical value 4). The row in each cell contains routing information eg: the IP address of all nodes whose identifier has same n first digit as recent node. The locality n set contains the node that are close in its proximity to its current node (2*B). The leaf set contains |L|/2 nodes which is numerically closest and greater than its current node. Using the node Id which is in the desired portion of the id space the entries in the routing table are selected. The selection process in pastry is recursive and not iterative and it uses tree list as its data structure.

The work of LAIS is mostly related to the cluster based on the locality and some approaches that enhance the location efficiently. The super peer network is mainly referred for their scalability, efficiency and file consistency maintenance in structured peer to peer network. D.H. Epema [4] implements an architecture called self organizing super peer network architecture (SOSPNET). It is usually constructed on the top of unstructured topology with semantic association between the nodes and the files. There are two different peers involved in this architecture one is super peer and the weak peer. Here the information related to the content is stored in the super peer. Weak peer will sort the super peer. The sorting is done according to the number of positive response to the quires and connection is done to the super peers which suit the majority of their request. It solves some of the issues in a fully decentralized approach: how super peer situate the files, how client peers are linked to each other, how the load balancing is done between the super peers and how the system deal with the peer failure.

Fanbin Meng uses a hierarchical clustering Peer to Peer network model. This model is based on user interest to improve search efficiency by using the topological algorithm. This algorithm processes the received query and will send back the reply with decreased copious arbitrary penetrating process and gains the appropriate supply quicker than usual searching algorithm.[3]

Haiying Shen *et all.* [6] proposed a hash based proximity clustering, which is based upon the consistent hashing function. This method balances the work load among all the nodes in the network. Here the clustering happens in the physical network and in virtual network. The cluster in the physical network is termed as pcluster and virtual network is termed as vcluster. In pcluster the regular node are associated to their physically nearest super node and occasionally reports the load information to the super node whereas in v cluster regular nodes are connected to the rationally close super node as in novel DHT network. Although these clusters are self organized, it still need for least amount preservation as in DHT network.

Sylvia Ratnasamy[8] comes up with a scheme called distributed binning scheme. The main motive is to optimize routing path in network. This scheme requires a well known landmark machine stretch across the internet. For extracting the proximity information network latency is used. Here, the nodes are partitioned into bins. In this partition the peers that drop within a given bin will be relatively close. From a bin of nodes, it measures the round trip time in the increasing order of RTT to each landmark. In this paper Ratnasamy et all deals with three applications of distributed binning scheme, they are on structured overlay, unstructured overlay and server selection and compared these three application with their experimental result and concluded that the nature of underlying topology affects the behavior of the scheme.

Yu Hua [10] proposed a semantic grouping method to identify the files effectively that is linked by their physical or behavioral attributes. Hau uses a novel technique to store all the files called as smart store. By using the retrieval tool the file information that are correlated on the group can be retrieved.

Daniel Baue[2] proposes a replica placement algorithm that works on the Distributed hash table. The novel method used here is replica enumeration in a fully decentralized manner. This algorithm does not depend upon the location. The random probing is used and the simulation result shows the performance of the system in a logarithmic sequence.

Haiying Shen *et all.* [5] proposed proximity aware interest clustering concept in structured peer to peer network. This method is urbanized based on the cycloid peer to peer network. Cycloid is a lookup competent stable degree superimposed network with $n=d.2^d$ (d is the dimension)of the peer. PAIS forms the cluster based upon the node proximity and interest by using the consistent hashing function.

III. PROBLEM STATEMENT

In Peer to Peer network, especially on the structured overlay there are some problems like decentralization, load balancing and availability exists. These issues reduce the performance of the peers while sharing the files among multiple peers. Although many methods are proposed for clustering the peers based on node closeness and interest of the user only some of the methods are able to cluster peers by both proximity and interest of a files. In this paper Locality aware interest clustering method is proposed to reduce the file querying delay by using the DHT routing process. To further reduce the delay cost effective file replication algorithm is used, where the underutilized replicas are removed.

IV. CONCLUSION

In this paper enrichment of file location efficiency, clustered locality p2p and interest clustering has been proposed. Although these approaches improve the recital, but the physical nearest cluster and peer interest the competence can be still faster. In this cluster based on location peer to peer system and interest cluster peer to peer file sharing system in structured network would swiftly increase the performance rate of identifying file location. LAIC method uses a cost effective file replication to enhance physical locality of frequently accessed nodes for improving competence. The cost effective replication improves the efficiency of the file querying process.

V. ACKNOWLEDGMENT

REFERENCES

[1] Antony Rowstron1 and Peter Druschel "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems"

[2] Daniel Bauer, Paul Hurley, Marcel Waldvogel "Replica Placement and Location Using Distributed Hash Tables, 32nd IEEE conference on local computer networks,2007 pp.315-324

[3] Fanbin Meng, Lei Ding, Sheng Peng and Guangxue Yue "A P2P Network Model Based on Hierarchical Interest Clustering Algorithm" in journal of software, vol. 8, no. 5, may 2013

[4] Garbacki, P., D.H. Epema and M. Van Steen,Self-Organizing Super-Peer Networks.

[5] Haiying Shen, Senior Member, IEEE, Guoxin Liu, Student Member, IEEE and Lee Ward "A Proximity-Aware Interest-Clustered P2P File Sharing System" in ieee transactions on parallel and distributed systems, vol. 26, no. 6, june 2015

[6] Haiying Shen and Cheng-Zhong Xu, "Hash-based Proximity Clustering for Load Balancing in Heterogeneous DHT Networks" in journal of parallel and distributed computing, vol.68 Issue 5, may 2008

[7] I.Stoica, R.Morris, D.Liben-Nowell, Chord:A scalable peer to peer look up protocol for Internet Application" in IEEE/ACm Transactions on Networking, vol.11 Issues: 1, Feb 2003

[8] Sylvia Ratnasamy, Mark Handley, Richard Karp, Scott Shenker , "Topologically-Aware Overlay Construction and Server Selection"

[9] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, "A scalable- Content Addressable Network"

[10] 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.

AUTHOR'S BIOGRAPHY

V. Mani received the B.Tech degree in Information Technology from Anna University, Chennai, Tamil Nadu, India in 2010 and the M.E. degree in Computer Science and Engineering from Anna university, Chennai, Tamil Nadu, India in 2015. He has teaching experience of about 4years 5 months. Presently working as Assistant Professor in the Department Of CSE. His area of interest is database management system, Object Oriented Programming Language and Computer Networks.

S.Gowsika received the B.Tech degree in Information Technology from Anna University, Chennai, Tamil Nadu, India in 2015. Currently, she is pursuing her Master degree in the area of Computer Science and Engineering in M.Kumarasamy College of Engineering, Karur. Her area of interest is database management system and Computer Networks