Geographically Clustered P2P File Sharing

Sanoj Saju[1], Praahas Amin[2]

[1]M. tech student, Sahyadri College of Engineering and Management, Karnataka, India.
[2]Assistant professor, Sahyadri College of Engineering and Management, Karnataka, India.

---***---

Abstract - **File sharing over the internet should be more efficient and there should not be any delay in searching a file. The main problem generated by P2P file sharing system is traffic, to avoid traffic in P2P an efficient query is required. An approach is made with the file sharing system to improve the efficiency of file query. It can improve the file query performance by clustering the peer systems, the clustering can be done on the physical juxtaposition. In this paper a structured P2P file sharing system is proposed based on the physical proximity, relation based on type of files at the nodes and the geographical location. GCFS (Geographically Clustered File Sharing) uses a technique of clustering the node based on proximity interest of the node. GCFS uses a novel lookup function that includes DHT (chord technique) and Intelligent Replication File algorithm to provide an efficient query. It creates a replicated file to reduce hotspots and results in efficiency of file query. Moreover GCFS has several approaches. It classifies the cluster based on interest and location. Then, a distributed intra sub cluster file querying method is used to improve the efficiency of file querying. A chord protocol is used to support the operation, by mapping the key onto a node and providing a key. Values will be stored in their key. Bloom Filter with Cam gives a quick and proficient method to seek the application/set and results in improving the file sharing efficiency. The experimental results shows an improvement in file searching.**

Key Words: Bloom Filter, Intelligent Replication File, Chord Technique,

1. INTRODUCTION

In the recent year's P2P file sharing system has gained popularity in the internet, For file sharing system more traffic is generated i.e. bit torrent. Efficient and trustworthy file query results in the overall performance of peer to peer file sharing. In P2P sharing system classified into two classes: structured and unstructured P2P network . In an unstructured P2P network depends either on flooding where nodes will be assigned to random selected neighbor node before locating the file. The main drawback of unstructured P2P network is that they don't have a particular structure they overlay over the nodes. They don't

assign responsibility to a particular node. To overcome this disadvantage an advance technique is used i.e. structured P2P network

Even though the key node continuously changes DHT is responsible to keep update of the files in the nodes.

To improve the file location efficiency several methods have being proposed, time taken to locate the file from the node depends the efficiency of the system. Techniques have being proposed to increase the efficiency, scalability and deterministic data location. To achieve consistency maintenance and load balanced in structured P2P, it uses super peer topology, there are many issue in locating the files they include client peer super peer relation, load balancing, file location etc. It is solved using super peer technology, this technique result in making file querying more reliable. In super peer topology they are classified as regular node and super nodes, they are mutually linked with each other, regular node is surrounded to a super node, which result in faster connectivity. Therefore, these construct a more consistent and constant backbone. Clustering of the nodes and file replication in the network could result in efficiency in the file location. Clustering technique involves proximity awareness, where physically close nodes are grouped together, within this physical close node are again classified and grouped based on interest, node with same interest are clustered together. To maintain the consistency in P2P system due to their frequent data updates, An overlay network is established with two layer for each replica group. The two layer are classified as chord replica nodes and ordinary replica nodes, these helps in the consistency maintenance. Bloom Filter with CAM technique is used to recognize the broken words in a nano memory.

2. PROBLEM DEFINITION

The efficiency of the P2P file sharing depends on the ability to locate the file with the least query search. Unstructured P2P doesn't provide an efficient query because it is difficult to realize due to strictly defined topologies. To improve the file querying performance clustering is required, clustering is based on the common interest and closeness of the node. According

to the geographical layout and closeness of the node it is clustered. To define both proximity and interest clustering is not possible in super peer technologies.

3. RELATED WORK

Super peer topology is used by fast track, morphens, due to firmly defined topologies it is harder to understand in P2P. Hierarchical structure of structured P2P helps us to group the peer based on interest and proximity. Proximity aware interest clustered have being grouped they are: (a)Super peer topology, (b)Proximity awareness, (c)Interest based file sharing. It uses Intelligent File Replication algorithm that replicate the requested file near the physical node. In structured P2P system the super peer network is for efficient and scalable file consistency maintains, super peer network is for load balancing. To improve the file location efficiency, they are clustered on the basis of closeness.

Clustering technique in the P2P file sharing system involves proximity awareness. Hierarchical secure load balancing scheme in a P2P cloud system is proposed by liu et al. super nodes are balanced first, then it depends on super node to balance other nodes /regular nodes. Each super node balances the node /regular node surrounded by it. Each super node caches the file recently requested by the other nodes and send a request to the super node to solve their request, self-organizing super node architecture to facilitate file querying.

4. PROPOSED SYSTEM

The time taken by the system /software to search /locate a file in an network determines the efficiency of the system. For an efficient and scalable P2P network sharing an efficient file retrieval technique is needed. It propose a technique based on the geographical area and the closeness of the node based on the interest and clustering. It is based on cycloid structure of P2P network. It is an hierarchical structure, where they are clustered based on physical closeness of the node, then they are sub clustered based on their interest. The clustered node are again grouped on the basic of geographical location, to maintain the consistency for the peer to peer system due to their frequent data update, replica of the file is created.

An overlay network is established with two layer, the upper layer is based on the Distributed Hash Table and it is a powerful and consist of a stable replica nodes

called Chord Replica Node(CRN) and the lower layer consist of an Ordinary Replica Nodes. Ordinary Replica Node is connected to the Chord Replica Node. When a file is to be replicated a message is initialized on the upper layer and the upper layer initialize a tree called Update Message Propagate Tree(UMPT).

By partitioning the DHT identifier space, after the update message is passed to all the node regarding the new entry/deletion. Thus CRN nodes are stable nodes thus helps in maintaining the stable cluster after each update. The update entry as well as entry of each nodes is maintained in the server, when an request is made to search, the server check its index file and perform its search operation. The search is made in two stages they are inter cluster and intra cluster. If the search found success, the node sends the location of the file to the requested client, if not found it will make an inter cluster search. The efficiency of file searching depends on the DHT lookup among the nodes, cluster are sub clustered based on the key. The File Replication helps to improve the file location efficiency, Depending on the hash function such as SHA1 is used in the DHT lookup because it has a collision resistant nature. Computation is infeasible to find different message that facilitate the same message digest. Hash function is efficient to cluster message based on message dissimilarity. File is searched in a distributed manner, It forwards to the interest super node. It check for time to live to find out how long the message can survive, every message is associated with TTL, its value decreased by one at each time when a message passes through peer, if TTL is zero message will be dropped and no longer it will be forwarded, thus we can conclude that file querying is efficient in interest clustered and DHT lookup method can enhance file searching, for the lookup efficiency files are replicated to neighbor physical node if it is frequently requested. Bloom Filter allows to check an element set and the reduant representation of the set. Redundancy is used for error detection and correction. Bloom Filter is used to correct and detect the error in the element set. CBF can be used to correct the error, it enable an cost efficient solution. To improve the file sharing efficiency the Bloom Filter result content are ranked to reduce the delay in file searching it check only the new added Bloom Filter information.

5. CONCLUSION

Due to firmly defined topologies it is harder to understand in structured P2P. In recent year several techniques used to improve the file

sharing efficiency. The interest clustering and proximity aware clustering results in perk up performance of P2P system. In this paper we introduce a clustering based on structured P2P. Hierarchical structure helps in grouping the structured P2P based on interest and proximity of the client. A group of client based on their interest are grouped when they want to download a same file each of the client downloading will help the other client by sharing the folder /data, which results in faster downloads, thus result in decrease conjunction in the central server. Bloom filter using CAM based structure helps in resuming the broken downloads thus clustering based on geographic location results in lookup efficiency in node dynamism, it also results in reduces overhead

REFERENCES

[1] Haiying Shen Senior Memeber,IEEE, Guoxin Liu , Student Member,IEEE, Lee Ward"A Proximity Aware Interest Clustered P2P File Sharing System" ,Vol 26, No 6 ,Jun 2015.

[2] G.Rajashekaram, Nimmakayala Madhavi, Chittala Kamala, Kotha Sumanth, "A New Approach to Intra Sub Cluster File Searching Implemented in P2P File System".IJSEAT,Vol 4,Issue 4,April 2016.

[3] Anju S Kumar, Ratheesh S, Manoj M, " A Review on Efficient File Sharing in Clustered P2P System",IJIRCE,Vol 3,Issue 11,Nov2015 .

[4] Suravarapu Anusha,M.Srihari, M.Nagendra Kumar, "Implementation of Bloom Filter Using Cam Based Structure",IJSEAT,Vol.4,Issue 11,Nov 2016.

[5] K.Bhargavi, B.Soumya"An Innovative Peer to Peer Information Querying Structure to Decrease Overheads",IJITR Vol,4,Issue 4,June 2016.

[6] Talari Meena,V, Sundar Ratnam, "A Profitless Data Utilization Theory for Hierarchical Structures ",IJITR Vol 4,Issue 5,Sept 2016.

[7] Karthi Naveen Abyudav, V Krishna Chaitanya "A closeness Alert Interest Gathered P2P Folder Sharing System",IJITR Vol 4, Issue 6,Nov 2016.

[8] M Satya Priyanka, V.Ravi Kishore "Improved File Search and Sharing Mechanismin Structured P2P System",IJRCCT,Vol 5,Issue 8,Aug 2016.

[9] Varalakshmi .T, Arul Murugan .R"A Survey on P2P File Sharing Systems using Proximity Aware Interest Clustering ", IJRASET ,Vol 3 ,Issue 12, Dec 2015.