International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 24 Issue 9 – JUNE 2017.

An Intelligent algorithm to reduce the overhead and enhance the efficiency of Peer-to-Peer file sharing system

Nowduri Pujitha [#1] and B. Mouleswara Rao [*2]

[#] *STUDENT, DEPT OF C.S.E, NRI INSTITUTE OF TECHNOLOGY,AGIRIPAALI, A.P, INDIA*

[*2] *Asst. Prof., DEPT OF C.S.E, NRI INSTITUTE OF TECHNOLOGY,AGIRIPAALI, A.P, INDIA*

***Abstract*— In internet p2p file sharing system generates more traffic. To get better file query performance cluster the common interested peers based on physical proximity. In this project proposing a proximity aware interest –clustered p2p file sharing system implemented in structured p2p file system. It forms a cluster based on node proximity as well as groups the nodes which having common interest into sub cluster. A narrative lookup function named as DHT and file replication algorithm which supports resourceful file lookup and access. To diminish overhead and file searching delay it keeps up file information collection. Bloom filter technique is used to cut file sharing delay. Finally proposed approach shows competence in file search, sharing and overhead.**

***Index Terms*— P2P networks, file sharing system, proximity awareness, file replication, Bloom filter.**

I. Introduction

A key principle to judge a P2P file sharing system is its file location competence. To get better this efficiency, numerous methods have been proposed. One method uses a super peer topology which consists of super nodes with speedy connections and regular nodes with slower connections [1,2,3,4,5,6,7]. A super node connects with other super nodes and some regular nodes, and a regular node joins with a super node. In this super-peer topology, the nodes at the centre of the network are faster and therefore construct amore consistent and constant backbone. This allows more messages to be in retreat than a slower backbone and, therefore, allows better scalability. Super-peer networks reside in the middle-ground between centralized and utterly symmetricP2P networks, and have the probable to coalesce the benefits of both centralized and distributed searches [8,9].

II. RELATED WORK

Liu et al. planned a hierarchical secure load complementary scheme in a P2P cloud system. It first balances the load among super nodes, and then depends on every super node to equilibrium the load among nodes under its management. Garbacki et al. proposed self-organizing super node structural design to make possible file querying. Each super node caches the files newly requested by its children, and other peers send requests to the super nodes that can crack most of their requests[10,11].

III. PROBLEM DEFINITION

A key rule to judge a P2P record sharing framework is its file location fitness. To show signs of improvement this effectiveness, various strategies have been proposed. One technique utilizes a super companion topology which comprises of super hubs with quick associations and ordinary hubs with slower associations[12]. A super node connects with other super nodes and some regular nodes, and a regular node joins with a super node. In this super-peer topology, the nodes at the centre of the network are sooner and as a result produce a more reliable and stable backbone. This allows more messages to be routed than a slower backbone and, therefore, allows better scalability. Super-peer networks reside in the middle-ground between central and fully symmetric P2P networks, and have the budding to mingle the benefits of both centralized and distributed searches[13,14,15].

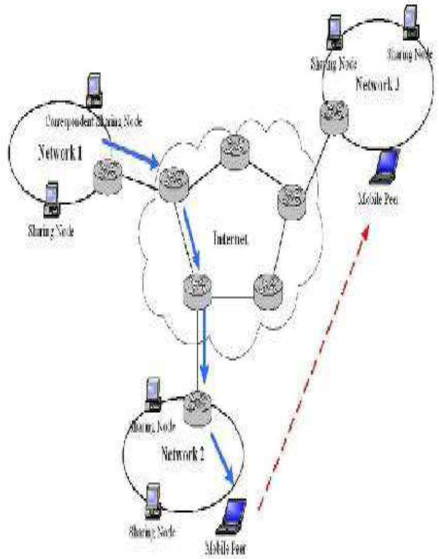
IV. PROPOSED APPROACH

It is a nearness mindful and premium bunched P2P record sharing System (PAIS) on an organized P2P framework. It shapes physically-close hubs into a bunch and extra gathering's physically-close and normal interest hubs into a sub-group. It likewise puts records with the comparative interests together and make them simple to get to through the DHT Lookup () directing capacity. All the more vitally, it keeps all points of interest of DHTs over unstructured P2Ps. Depending on DHT lookup administer as opposed to television, the

PAIS development expends substantially less cost in mapping hubs to bunches and mapping groups to intrigue sub-bunches. PAIS utilizes a canny record replication calculation to encourage enhance document lookup ability[16].

It makes copies of records that are regularly asked for by a gathering of physically close hubs in their area. In addition, PAIS upgrades the intra subgroup record penetrating from end to end a few methodologies [17].

V. SYSTEM ARCHITECTURE:



VI. PROPOSED METHODOLOGY

A. *PAIS STRUCTURE*

PAIS is developed based on the Cycloid structured P2P network. A node's interests are explaining by a set of attributes with a globally known string description such as "image" and "music". The strategies that consent to the description of the content in a peer with metadata can be used to obtain the interests of each peer. Taking advantage of the hierarchical structure of Cycloid, PAIS gathers physically close nodes in one cluster and further group's nodes in each cluster into sub clusters based on their comfort.

B. *NODE PROXIMITY REPRESENTATION*

Landmark clustering has been widely accepted to produce proximity information. It is based on the intuition that nodes close to each other are possible to have similar distances to a few selected landmark nodes. We presume there are m landmark nodes that are aimlessly scattered in the Internet. Each node measures its physical distances to the m landmarks and uses the vector of distances as its coordinate in Cartesian space. Two physically close nodes will have comparable vectors. We use space filling curves, such as the Hilbert curve, to map the m-dimensional landmark vectors to real numbers, so the closeness relationship surrounded by the nodes is preserved. We call this number the Hilbert number of the node denoted by H. The closeness of two nodes' Hs indicates their physical closeness on the Internet[18].

C. *NODE INTEREST REPRESENTATION*

Dependable hash functions such as SHA-1 is generally used in DHT networks for node or file ID due to its collision-resistant nature. When using such a hash function, it is computationally infeasible to find two different messages that fabricate the same message digest. The consistent hash function is effective to cluster messages based on message dissimilarity.

D. *CLUSTERING PHYSICALLY CLOSE AND COMMON-INTEREST NODES*

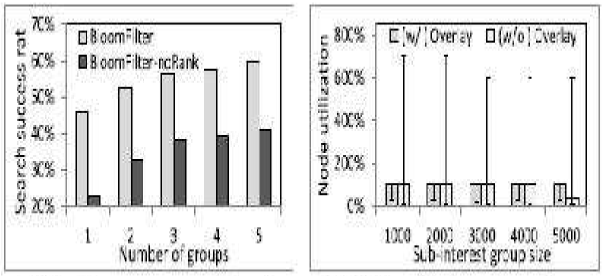
Based on the Cycloid topology and ID determination, PAIS brightly uses cubical indices to differentiate nodes in different physical locations and uses cyclic indices to further order physically close nodes based on their interests. Particularly, PAIS uses node i's Hilbert number, Hi, as its cubical index, and the dependable hash value of node i's interest as its cyclic index to generate node i's ID denoted. If a node has a number of interests,

it generates a set of IDs with different cyclic indices. Using this ID purpose method, the physically close nodes with the same H will be in a cluster, and nodes with similar H will be in close clusters in PAIS. Physically close nodes with the equivalent interest have the same ID, and they added compose a sub-cluster in a cluster[19].

E. *FILE DISTRIBUTION*

As actually close and common-interest nodes form a sub cluster, they can split files between each other so that a node can get back its requested file in its interest from a actually close node. For this purpose, the sub-cluster server upholds the index of all files in its sub-cluster for file sharing amongst nodes in its sub-cluster. A node's requested file may not exist in its sub-cluster. To help nodes find files not active in their sub-clusters, as in traditional DHT networks, PAIS re-distributes all files among nodes in the network for competent global search.

VII. RESULTS:



(a) Searching success rate with sub-interest clustering (b) Node utilization with node capacity classification

It shows the intra-group search success rate of Bloom Filter and Bloom Filter-no Rank versus the number of groups. From the figure, we can see that the search success rate of Bloom Filter is much larger than Bloom Filter-no Rank[20].

VIII. CONCLUSION:

In recent years, to augment file location efficiency in P2P systems, interest-clustered super-peer networks and proximity- clustered super-peer networks have been planned. Even though both strategies perk up the performance of P2P systems, few works cluster peers based on both peer interest and physical proximity at the same time. Moreover, it is harder to understand it in structured P2P systems due to their firmly defined topologies, although they have high competence of file location than unstructured P2Ps. In this paper, we introduce a proximity-aware and interest-clustered

P2P file sharing system based on a structured P2P. It groups peers based on both interest and proximity by taking advantage of a hierarchical structure of a structured P2P. PAIS uses an intellectual file replication algorithm that replicates a file normally requested by physically close nodes near their physical location to add to the file lookup efficiency. Finally, PAIS enhances the file searching effectiveness among the proximity-close and common interest nodes through a number of approaches. The trace-driven experimental results on PlanetLab exhibit the efficiency of PAIS in comparison with other P2P file sharing systems. It radically decrease the overhead and yields important improvements in file location.

REFERENCES

[1] BitTorrent. (2013) [Online]. Available: http://www.bittorrent. com/
[2] Gnutella home page. (2003) [Online]. Available: http://www.gnutella.com
[3] I. Clarke, O. Sandberg, B. Wiley, and T. W.Hong, “Freenet: A distributed anonymous information storage and retrieval system,” in Proc. Int. Workshop Des. Issues Anonymity Unobservability, 2001, pp. 46–66.
[4] I. Stoica, R. Morris, D. Liben-Nowell, D. R.Karger, M. F. Kaashoek, F. Dabek, and H.Balakrishnan, “Chord: A scalable peer-to peer lookup protocol for internet applications,” IEEE/ACM Trans. Netw., vol. 11, no. 1, pp. 17–32, Feb. 2003.
[5] A. Rowstron and P. Druschel, “Pastry: Scalable, decentralized object location and routing for large scale peer-to-peer systems,” in Proc. IFIP/ACM Int. Conf. Distrib. Syst. Platforms Heidelberg, 2001, pp. 329–350.
[6] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. Kubiatowicz, “Tapestry: A resilient global-scale overlay for service deployment,” IEEE J. Sel. Areas Commun., vol. 22, no. 1, pp. 41–53, 2004.
[7] H. Shen, C. Xu, and G. Chen, “Cycloid: A scalable constant-degree P2P overlay network,” Perform. Eval., vol. 63, pp. 195–216, 2006.
[8] 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. 1695–1708, Dec. 2008.
[9] 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. 686–702, 2008.
[10] FastTrack. (2003) [Online]. Available: http://www.fasttrack.nu/ index_int.html
[11] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, “Topologically-aware overlay construction and server selection,” in Proc. IEEE INFOCOM, 2002, pp. 1190–1199.
[12] M. Waldvogel and R. Rinaldi, “Efficient topology-aware overlay network,” in Proc. ACM Workshop Hot Topics Netw., 2002, pp. 101–106.
[13] Y. Zhu and H. Shen, “An efficient and scalable framework for content-based publish/subscribe systems,” Peer-to-Per Netw. Appl., vol. 1, pp. 3–17, 2008.
[14] C. Hang and K. C. Sia, “Peer clustering and firework query model,” in Proc. Int. World Wide Web Conf., 2002.
[15] A. Crespo and H. Garcia-Molina, “Routing indices for peer-topeer systems,” in Proc. 22nd Int. Conf. Distrib. Comput. Syst., 2002, pp. 23–32.
[16] W. Nejdl, M. Wolpers, W. Siberski, C.Schmitz, M. Schlosser, I. Brunkhorst, and A.Loeser, “Super-peer-based routing and clustering strategies for RDF-based peer-to-peer networks,” J.Web Semantics, vol. 1, no. 2, pp. 137–240, 2004.
[17] P. A. Bernstein, F. Giunchiglia, A.Kementsietsidis, J. Mylopoulos, L. Serafini, and I.Zaihrayeu, “Data management for peer-to-peer computing: A vision,” in Proc. 5th Int. Workshop World Wide Web Databases, 2002, pp. 89–94.
[18] A. Y. Halevy, Z. G. Ives, P. Mork, and I.Tatarinov, “Piazza: Data management infrastructure for semantic web applications,” in Proc. 12th Int. Conf. World Wide Web, 2003, pp. 556–567.
[19] K. Aberer, P. Cudre-Mauroux, and M.Hauswirth, “The chatty web: Emergent semantics through gossiping,” in Proc. 12th Int. Conf. World Wide Web, 2003, pp. 197–206
[20] Morpheus. (2008) [Online]. Available: http://www.musiccity. Com.