Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 4, Issue 6, June-2017

A Proximity-Aware Interest-Clustered P2P File Sharing System

[1]Ashwini Borole, [2]Prof. Pankaj Agarkar

[1,2]Computer Department, Ajeenkya D Y Patil, School of Engineering, Lohegaon, Pune

Abstract **---** Clustering peers by their property of proximity can also increase file query overall efficiency. Even so, many proximity-based and interest-based mostly super-peer topologies have been proposed with various characteristics, number of techniques are capable to cluster peers in accordance to each proximity and curiosity. In addition, most of these techniques are on unstructured P2P techniques that have no stringent policy for topology development. In this function, we introduce a Super-peer-primarily based routing and clustering scheme for file sharing in Delay tolerant Network based mostly on bloom filter which provides structured P2P, which are types of physically-near nodes into a cluster and additional groups physically-near and frequent-interest nodes into a sub-cluster primarily based on a hierarchical topology. PAIS makes use of an intelligent file replication algorithm to more additional improve file query efficiency. It generates replicas of files that are often requested by a group of end users. Super-peer topology consists of super-nodes with quick connections and typical nodes with slower connections. A super node connects with other super nodes and some typical nodes, and a typical node connects with a super node. In this super-peer topology, the nodes at the center of the network are more quickly and consequently generate a much more trustworthy and secure backbone. Additional, the experimental outcomes display the higher effectiveness of the file browsing approaches in enhancing the file browsing efficiency. Efficient file question is vital to the general performance of peer-to-peer (P2P) file sharing systems. Clustering peers by their common interests will considerably enhance the efficiency of file question. Clustering peers by their physical proximity also can improve file question performance.

Keywords --- Bloom Filter, P2P, file search query, interest cluster, super-peer.

I. INTRODUCTION

Nodes be a part of and leave the network in keeping with some loose rules. Currently, unstructured P2P networks' file question methodology is predicated on either flooding wherever the question is propagated to any or the entire node's neighbors, or random-walkers wherever the question is forwarded to willy-nilly chosen neighbors till the file is found. However, flooding and random walkers cannot guarantee information location. Organized P2P systems i.e., Distributed Hash Tables (DHTs), resolve stunned the problems by their decisions of upper potency, quantifiability, and settled information location. It strictly controlled topologies and their operation algorithms and information placement square measure exactly outlined supported a DHT organization and consistent hashing operates. The node is answerable for a key will perpetually be found despite the fact that if the system is during a continuous state of amendment. Most of the DHTs need O (log n) hops per operation request with (O log n) neighbors per node, wherever n is that the range of nodes within the system.

More than the previous couple of many years, the high quality of the World Wide Web has produced a significant input to P2P file sharing techniques. There are two classes of P2P techniques: unstructured and structured. Unstructured P2P networks like Gnutella and free net doesn't assign accountability for understanding to certain nodes. Nodes be component of and depart the network per some loose principles. At present, unstructured P2P networks' file query methodology relies on both flooding wherever the query is propagated to all or any the node's neighbors, or random-walkers wherever the query is forwarded to haphazardly selected neighbors until the file is identified. Even so, flooding and random walkers are not able to assure information area. Structured P2P networks i.e., Distributed Hash Tables (DHTs), will conquer the disadvantages with their possibilities of upper potency, measurability, and settled information area. It strictly managed topologies, and their search algorithms and information placement are specifically outlined supported a DHT arrangement and constant hashing complete. The node is liable for an important will invariably be located even if the technique is in an exceedingly constant state of amendment. Most of the DHTs require O (log n) hops per search request with (O log n) neighbors per node, wherever n is that the selection of nodes inside the method.

II. LITERATURE REVIEW

Sr. No.	Paper Name	Author	Description
1	**An efficient and scalable framework for content-based publish/subscribe systems**	**Yingwu Zhu Haiying Shen**	This paper presents Chord, a distributed lookup protocol that addresses this issue. Chord offers assistance for just a single operation: provided an importance, it maps the essential onto a node. Data location can be effortlessly implemented on leading of Chord by associating a important with each and every information item, and storing the essential/data pair at the node to which the essential maps.
2	**Hash-based proximity clustering for efficient load balancing in heterogeneous DHT networks**	**Haiying Shena, Cheng-Zhong Xub**	Antony Rowstron and Peter Druschel in presents design and layout and evaluation of Pastry, a scalable, distributed object area and routing substrate for broad region peer-to-peer applications. Pastry performs application-degree routing and object area in a potentially very huge overlay network of nodes linked by means of the World Wide Web. It can be utilized to assistance a range of peer-to-peer applications, such as global data storage, information sharing, group communication and naming.
3	**Cycloid: A constant-degree and lookup-efficient P2P overlay network**	**Haiying Shena, Cheng-Zhong Xua, Guihai Chenb**	This paper a novel hash table information structure and algorithm which outperforms the typical hash table algorithms by delivering much better bounds on hash collisions and the memory access per lookup. Our hash table algorithm further extends the technique of multi-hashing, Bloom filter, to assist precise match.
4	**Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems**	**Antony Rowstron1 and Peter Druschel**	In this paper a binning scheme whereby nodes partition themselves into hins this kind of that nodes that fall inside a provided hin are relatively close to one particular yet another in terms of network latency. Our binning strategy is straight forward (requiring minimum assistance from any measurement infrastructure), scalable (requiring no type of glnhal understanding, each and every node only demands expertise of a tiny number of effectively-recognized landmark nodes) and fully distributed (requiring no communication or cooperation between the nodes being hinned)

III. PROPOSED SYSTEM

We're designing a proximity-aware and interest-clustered P2P file sharing System (PAIS) on a structured P2P system. It forms physically-close nodes into a cluster and any team's physically-close and common-interest nodes into a sub-cluster. It additionally places files with a similar interests along and create them accessible through the DHT Lookup routing perform. a lot of significantly, it keeps all benefits of DHTs over unstructured P2Ps. wishing on DHT search policy instead of broadcasting, the PAIS construction consumes abundant less price in mapping nodes to clusters and mapping clusters to interest sub-clusters. PAIS uses Associate in nursing intelligent file replication formula to any enhance file search potency. It creates replicas of files that are often requested by a gaggle of physically shut nodes in their location. Moreover, PAIS enhances the intra sub-cluster file exploring through many approaches initial, it any classifies the interest of a sub-cluster to variety of sub-interests, and clusters common-sub-interest nodes into a gaggle for file sharing. If upcoming time exact same query request comes then technique will searches that file curiosity cluster primarily based on bloom filter record.

3.1 Advantages of Proposed System:

1. The techniques planned during this paper will profit several current applications like content delivery networks, P2P video-on-demand systems, and information sharing in on-line social networks.

2. We tend to introduce the careful style of PAIS. It's appropriate for a file sharing system wherever files are often classified to variety of interests and every interest are often classified to variety of sub-interests.

3. It teams peers supported each interest and proximity by taking advantage of a hierarchical data structure of a structured P2P.

4. PAIS uses associate intelligent file replication rule that replicates a file often requested by physically shut nodes close to their physical location to reinforce the file search potency.

5. PAIS enhances the file looking out potency among the proximity-close and customary interest nodes through variety of approaches.

IV. SYSTEM ARCHITETURE

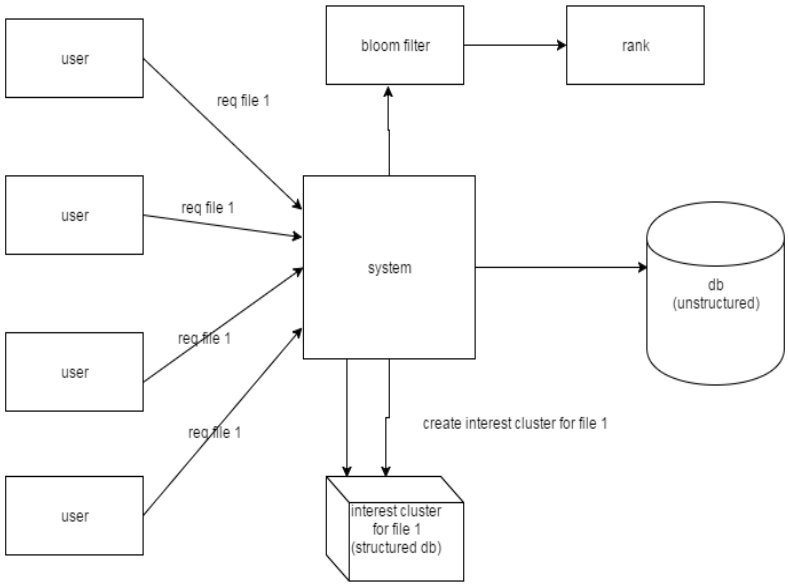


Figure 4.1 System Architecture of Proposed System

V. MATHEMATICAL MODEL

Input: File query in DTN.
Output: File query result in less time in DTN.
Process:
Let W be the whole system, W = {U, S, C, B, R, r, F}.
Where,

1. U is the set of number users.

   U = {U1, U2 . . . . Un}
2. S is the system which contains the unstructured data to provide the service to user based on user request.

3. C is set of number of cluster based on user request.

   C = {C1, C2, ... Cn}.
4. B be set of bloom filter which is required to filter the user requests based on user interest.

5. F be the set of files user is requesting .

   F = {f1, f2 . . . . . fn}.
6. R be the user request for file to S.

7. r be the rank assigned to file based user request.

Step 1: user U login to the system and request for particular f1 to the system.
Step 2: The system S will process the user request R from the unstructured data. In this the bloom filter will filter the user request to check whether the same file request has come before or not if not it will rank that file.
Step 3: The system will process the user request R based on ranks assigned to files by using bloom filter.
Step 4: if same file request R is come at system more than 2 times (assigning threshold) then system will create an interest-cluster for that requested file to minimize the searching time as system will search the requested file from unstructured data.

VI. RESULT ANALYSIS

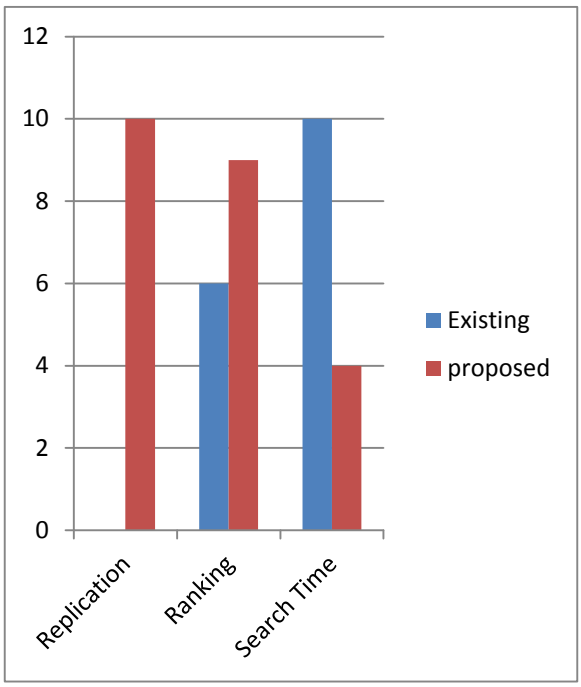
Input: Here, Entire Technique taken numerous much more attribute for the input function but right here writer mostly focuses on the Time and efficiency of technique primarily based on this attributes we obtaining following outcome for our proposed technique.

Expected Result:

Tabular Form :

	Existing	proposed
Replication	0	10
Ranking	6	9
Search Time	10	4

Graphical Form :



VII. CONCLUSION AND FUTURE SCOPE

In recent years, to support heading position force popular P2P schemes, interest-clustered super-peer systems and proximity-clustered super-peer webs are scheduled. However every customs recover the presentation of P2P schemes, insufficient mechanism collection nobles maintained both noble attention and corporeal immediacy on the like historical. Furthermore, it's rougher toward gain he popular organized P2P schemes because of their severely outlined topologies, though they need high potency of file location than unstructured P2Ps. This paper, we have a tendency to introduce a proximity-aware and interest-clustered P2P file sharing system supported a structured P2P. It teams peers supported each interest and proximity by taking advantage of a hierarchical data structure of a structured P2P.

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.