ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Proximity Knowledge on Passion Clustered P2p Distributed File System



R.Lakshmana Rao
PG Scholar,
Dept of CSE,
Pydah College of Engineering and Technology, Visakhapatnam, AP, India.

K.Saran Kumar
Assistant Professor,
Dept of CSE,
Pydah College of Engineering and Technology, Visakhapatnam, AP, India.

Dr.Ramesh Challagundla
Professor & Principal,
Dept of ECE,
Pydah College of Engineering and Technology, Visakhapatnam, AP, India.

Abstract:
Proficient document inquiry is critical to the general execution of distributed (P2P) record sharing frameworks. Grouping peers by their normal advantages can essentially upgrade the effectiveness of record inquiry. Bunching peers by their physical closeness can likewise enhance document question execution. Notwithstanding, couple of current works can bunch peers in view of both companion intrigue and physical vicinity. Albeit organized P2Ps give higher record inquiry proficiency than unstructured P2Ps, it is hard to acknowledge it because of their entirely characterized topologies.

In this work, we present a Proximity-Aware and Interest-bunched P2P document sharing System (PAIS) in light of an organized P2P, which shapes physically-close hubs into a group and further gatherings physically-close and normal intrigue hubs into a sub-bunch in view of a progressive topology. PAIS utilizes a canny record replication calculation to further upgrade document inquiry proficiency. It makes imitations of records that are often asked for by a gathering of physically close hubs in their area. In addition, PAIS upgrades the intra-sub-group record seeking through a few methodologies. To start with, it facilitate characterizes the enthusiasm of a sub-bunch to various sub-interests, and bunches regular sub-intrigue hubs into a gathering for record sharing. Second, PAIS assembles an overlay for every gathering that interfaces bring down limit hubs to higher limit hubs for conveyed record questioning while keeping away from hub over-burden. Third, to lessen record looking deferral, PAIS utilizes proactive document data accumulation so that a document requester can know whether its asked for document is in its adjacent hubs. Fourth, to diminish the overhead of the record data accumulation, PAIS utilizes sprout channel based document data gathering and relating dispersed document looking. Fifth, to enhance the record sharing proficiency, PAIS positions the sprout channel brings about request. 6th, considering that an as of late went by record has a tendency to be gone by once more, the blossom channel based approach is improved by just checking the recently added sprout channel data to diminish document looking deferral. Follow driven exploratory results from this present reality PlanetLab testbed show that PAIS significantly diminishes overhead and upgrades the proficiency of record imparting to and without beat. Advance, the trial comes about demonstrate the high adequacy of the intra-sub-group record looking methodologies in enhancing document seeking proficiency.

Keywords:
P2P System; Replication ; Topologies; Cluster.

ISSN No: 2348-4845
International Journal & Magazine of Engineering,
Technology, Management and Research
A Peer Reviewed Open Access International Journal

I. INTRODUCTION:

Distributed computing is a field of computer science that studies distributed systems. A distributed system is a software system in which components located on networked computers communicate and coordinate their actions by passing messages. The components interact with each other in order to achieve a common goal. There are many alternatives for the message passing mechanism, including RPC-like connectors and message queues. Three significant characteristics of distributed systems are: concurrency of components, lack of a global clock, and independent failure of components.

An important goal and challenge of distributed systems is location transparency. Examples of distributed systems vary from SOA-based systems to massively multiplayer online games to peer-to-peer applications. A computer program that runs in a distributed system is called a distributed program, and distributed programming is the process of writing such programs. Distributed computing also refers to the use of distributed systems to solve computational problems. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more computers, which communicate with each other by message passing.

II. LITERATURE SURVEY:

1) Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems

AUTHORS: A. Rowstron and P. Druschel

This paper presents the design and evaluation of Pastry, a scalable, distributed object location and routing substrate for wide-area peer-to-peer applications. Pastry performs application-level routing and object location in a potentially very large overlay network of nodes connected via the Internet. It can be used to support a variety of peer-to-peer applications, including global data storage, data sharing, group communication and naming.

2) Semantic-aware metadata organization paradigm in next-generation file systems

AUTHORS: Y. Hua, H. Jiang, Y. Zhu, D. Feng, and L. Tian

Existing data storage systems based on the hierarchical directory-tree organization do not meet the scalability and functionality requirements for exponentially growing data sets and increasingly complex metadata queries in large-scale, Exabyte-level file systems with billions of files. This paper proposes a novel decentralized semantic-aware metadata organization, called SmartStore, which exploits semantics of files' metadata to judiciously aggregate correlated files into semantic-aware groups by using information retrieval tools.

3) An efficient and trustworthy P2P and social network integrated file sharing system

AUTHORS: G. Liu, H. Shen, and L. Ward

Efficient and trustworthy file querying is important to the overall performance of peer-to-peer (P2P) file sharing systems. Emerging methods are beginning to address this challenge by exploiting online social networks (OSNs). However, current OSN-based methods simply cluster common-interest nodes for high efficiency or limit the interaction between social friends for high trustworthiness, which provides limited enhancement or contradicts the open and free service goal of P2P systems. Little research has been undertaken to fully and cooperatively leverage OSNs with integrated consideration of proximity and interest. In this work, we analyze a BitTorrent file sharing trace, which proves the necessity of proximityand interest-aware clustering. Based on the trace study and OSN properties, we propose a SOcial Network integrated P2P file sharing system with enhanced Efficiency and Trustworthiness (SoNet) to fully and cooperatively leverage the common-interest, proximity-close and trust properties of OSN friends. SoNet uses a hierarchical distributed hash table (DHT) to cluster common-interest nodes, then further cluster proximity-close nodes into subcluster, and connects the nodes in a subcluster with social links.

III. SYSTEM ANALYSIS:

EXISTING SYSTEM:

- A key criterion to judge a P2P file sharing system is its file location efficiency. To improve this efficiency, numerous methods have been proposed. One method uses a super peer topology which consists of supernodes with fast connections and regular nodes with slower connections. A supernode connects with other supernodes and some regular nodes, and a regular node connects with a supernode. In this super-peer topology, the nodes at the center of the network are faster and therefore produce a more reliable and stable backbone. This allows more messages to be routed than a slower backbone and, therefore, allows greater scalability. Super-peer networks occupy the middle-ground between centralized and entirely symmetric P2P networks, and have the potential to combine the benefits of both centralized and distributed searches.
- Another class of methods to improve file location efficiency is through a proximity-aware structure.
- The third class of methods to improve file location efficiency is to cluster nodes with similar interests which reduce the file location latency.

DISADVANTAGES OF EXISTING SYSTEM:

- Although numerous proximity-based and interest-based super-peer topologies have been proposed with different features, few methods are able to cluster peers according to both proximity and interest.
- In addition, most of these methods are on unstructured P2P systems that have no strict policy for topology construction.
- They cannot be directly applied to general DHTs in spite of their higher file location efficiency.

PROPOSED SYSTEM:

- This paper presents 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 further groups physically-close and common-interest nodes into a sub-cluster. It also places files with the same interests together and make them accessible through the DHT Lookup() routing function. More importantly, it keeps all advantages of DHTs over unstructured P2Ps. Relying on DHT lookup policy rather than broadcasting, the PAIS construction consumes much less cost in mapping nodes to clusters and mapping clusters to interest sub-clusters. PAIS uses an intelligent file replication algorithm to further enhance file lookup efficiency.
- It creates replicas of files that are frequently requested by a group of physically close nodes in their location. Moreover, PAIS enhances the intra sub-cluster file searching through several approaches
- First, it further classifies the interest of a sub-cluster to a number of sub-interests, and clusters common-sub-interest nodes into a group for file sharing.
- Second, PAIS builds an overlay for each group that connects lower capacity nodes to higher capacity nodes for distributed file querying while avoiding node overload.
- Third, to reduce file searching delay, PAIS uses proactive file information collection so that a file requester can know if its requested file is in its nearby nodes.
- Fourth, to reduce the overhead of the file information collection, PAIS uses bloom filter based file information collection and corresponding distributed file searching.
- Fifth, to improve the file sharing efficiency, PAIS ranks the bloom filter results in order. Sixth, considering that a recently visited file tends to be visited again, the bloom filter based approach is enhanced by only checking the newly added bloom filter information to reduce file searching delay.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

ADVANTAGES OF PROPOSED SYSTEM:

- The techniques proposed in this paper can benefit many current applications such as content delivery networks, P2P video-on-demand systems, and data sharing in online social networks.
- We introduce the detailed design of PAIS. It is suitable for a file sharing system where files can be classified to a number of interests and each interest can be classified to a number of sub-interests.
- It groups peers based on both interest and proximity by taking advantage of a hierarchical structure of a structured P2P.
- PAIS uses an intelligent file replication algorithm that replicates a file frequently requested by physically close nodes near their physical location to enhance the file lookup efficiency.
- PAIS enhances the file searching efficiency among the proximity-close and common interest nodes through a number of approaches.

IV. IMPLEMENTATION:

MODULES:

- PAIS Structure
- Node proximity representation
- Node interest representation
- Clustering physically close and common-interest nodes
- File Distribution

MODULES DESCSRIPTION:

PAIS Structure:

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

Node Proximity Representation:

A landmarking method can be used to represent node closeness on the network by indices used. Landmark clustering has been widely adopted to generate proximity information. It is based on the intuition that nodes close to each other are likely to have similar distances to a few selected landmark nodes. We assume there are m landmark nodes that are randomly 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 similar 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 among 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.

Node Interest Representation:

Consistent hash functions such as SHA-1 is widely 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 produce the same message digest. The consistent hash function is effective to cluster messages based on message difference.

Clustering Physically Close and Common-Interest Nodes:

Based on the Cycloid topology and ID determination, PAIS intelligently uses cubical indices to distinguish nodes in different physical locations and uses cyclic indices to further classify physically close nodes based on their interests. Specifically, PAIS uses node i's Hilbert number, Hi, as its cubical index, and the consistent 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 determination method, the physically close nodes with the same H will be in a cluster, and nodes with similar H will be in

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

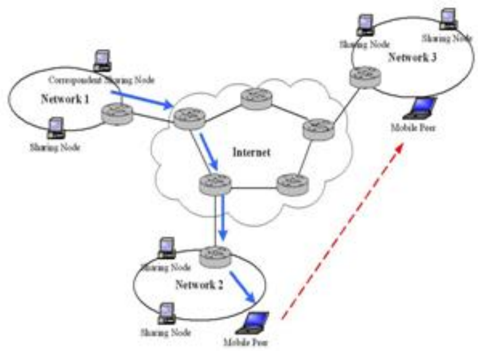
close clusters in PAIS. Physically close nodes with the same interest have the same ID, and they further constitute a sub-cluster in a cluster.

File Distribution:

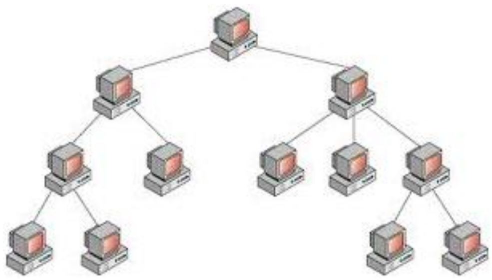
As physically close and common-interest nodes form a subcluster, they can share files between each other so that a node can retrieve its requested file in its interest from a physically close node. For this purpose, the sub-cluster server maintains the index of all files in its sub-cluster for file sharing among nodes in its sub-cluster. A node's requested file may not exist in its sub-cluster. To help nodes find files not existing in their sub-clusters, as in traditional DHT networks, PAIS re-distributes all files among nodes in the network for efficient global search.

V. SYSTEM DESIGN:

SYSTEM ARCHITECTURE:

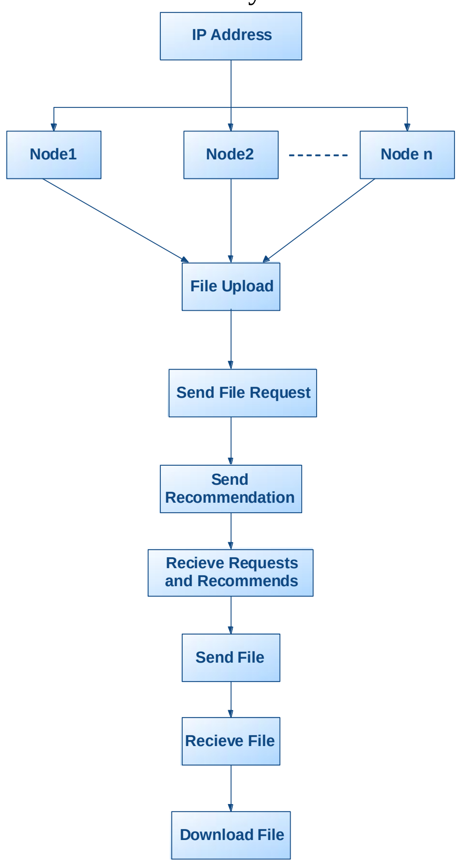


Clustered By Interest:



DATA FLOW DIAGRAM:

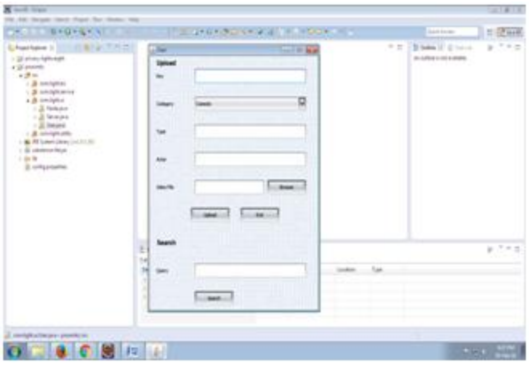
1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.
5. In the following DFD our system contains nodes, nodes are clustered by interest and proximity and upload file and store as by interest

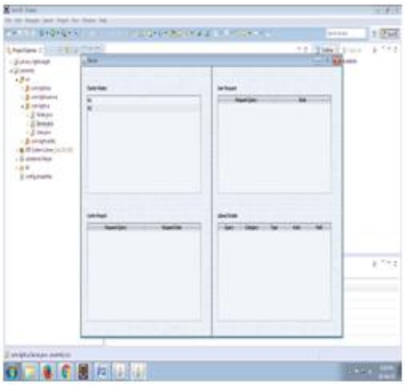


ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

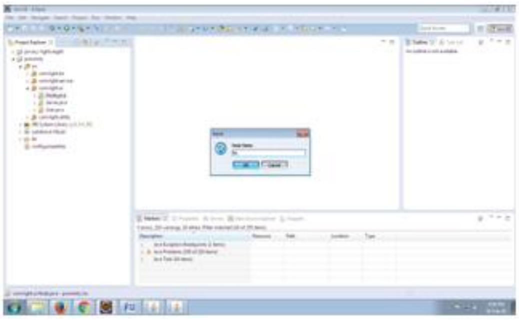
VI. SCREEN SHOTS:

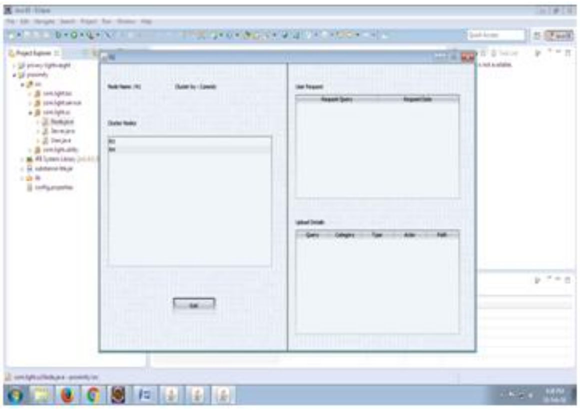
File Upload:

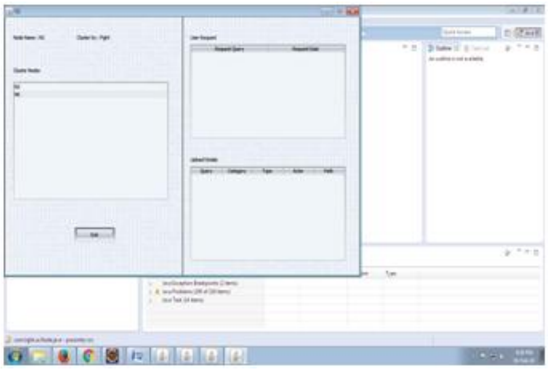


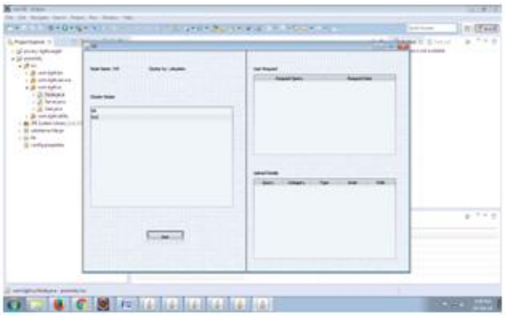


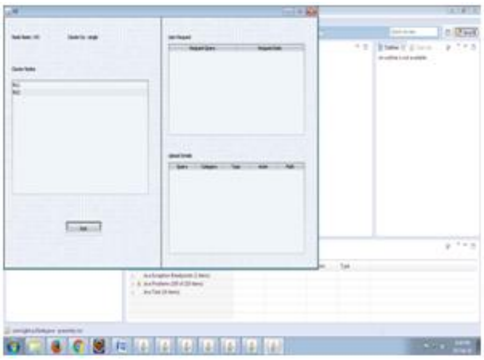
Node Creation:



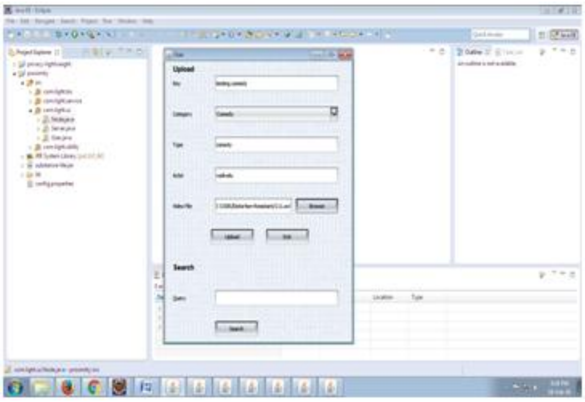


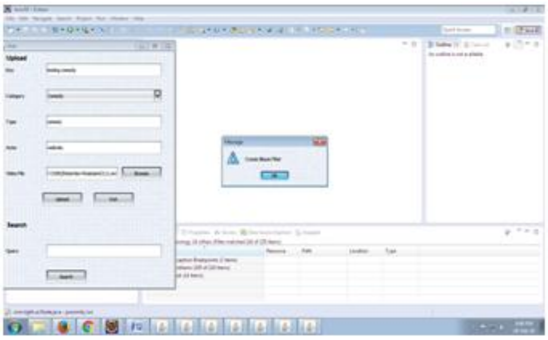




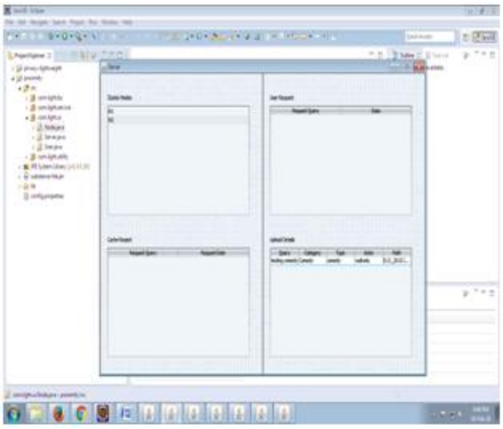


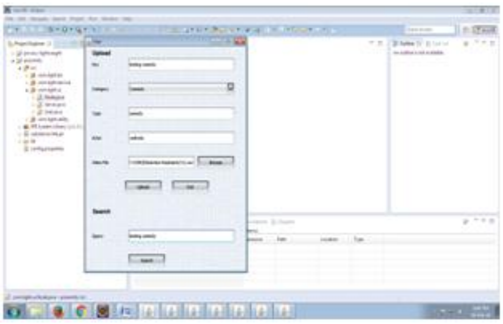
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal



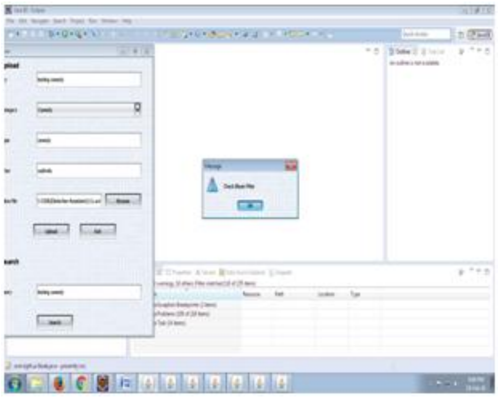


Clustered:

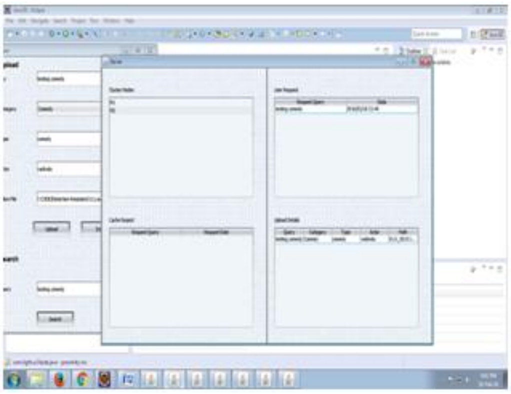




File Search:



Search Result:



VII. CONCLUSION:

In recent years, to enhance file location efficiency in P2P systems, interest-clustered super-peer networks and proximity- clustered super-peer networks have been proposed. Although both strategies improve the performance of P2P systems, few works cluster peers based on both peer interest and physical proximity simultaneously. Moreover, it is harder to realize it in structured P2P systems due to their strictly defined topologies, although they have high efficiency 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 intelligent file replication algorithm that replicates a file frequently requested by physically close nodes near their physical location to enhance the file lookup efficiency.

Finally, PAIS enhances the file searching efficiency among the proximity-close and common interest nodes through a number of approaches. The trace-driven experimental results on Planet Lab demonstrate the efficiency of PAIS in comparison with other P2P file sharing systems. It dramatically reduces the overhead and yields significant improvements in file location .

VII. 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-topeer 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.

Author's Profile's:

R.Lakshmana Rao,
M.Tech CSE, Pursuing M.Tech in Pydah College of Engineering & Technology , Visakhapatnam.

K.Saran Kumar
M.Tech, Assistant Professor, Pydah College of Engineering & Technology , Visakhapatnam.

Dr.Ramesh Challagundla
M.E, PH.D, MIE EE (USA), FIETE(IND), MIE(IND), MISTE is a former faculty of Birla Institute of Technology, Gitam University, ANITS and currently working as Professor and Principal of Pydah College of Engineering and Technology, Visakhapatnam.