Impact Factor (SJIF): 3.632

International Journal of Advance Research in Engineering, Science & Technology

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

Volume 3, Issue 10, October-2016

An Efficient P2P File Sharing System Using an Intelligent File Replication Algorithm

G .Swapna[1], Swapna Gundu[2]

[1, 2] **Department of computer science and engineering, Siddhartha Institute of Technology and Science – Narapally**

Abstract — efficient file query is very important performance of the peer to peer file sharing system. Significantly enhance the efficiency of file query by using a Clustering peers and physical proximity can improve the file query performance. Some of the working procedure has able to cluster peers based on both peer interest and physical proximity. Here the structured P2Ps will give the higher precedence file query efficiency than unstructured P2Ps, and can't define topologies. In this topic we introduced (PAIS) means a Proximity-Aware and Interest-clustered P2P file sharing System and this is completely based on the structure of P2P system, and it has form the physically closed nodes into a cluster and remaining groups are physically close and common interest nodes into a sub cluster based on a hierarchical topology. We have used intelligent file replication algorithm because enhance file query efficiency. It has created replicas of files which are requested by a group of physically close nodes in particular location. Searching the several approaches by using a PAIS enhance the intra sub cluster file. Initially it has to classify the interest of a sub-cluster to a number of sub-interests, and clusters common cub interest nodes into a group for file sharing. And then later it connects from lower capacity nodes to the higher precedence capacity nodes and distributing the file querying and avoids node overload. After that file searching delay is decreases. PAIS uses proactive file information collection so file requester can only kwon the where the file and whether it is nearby nodes or not. Overhead of the file information collection is reduced. Corresponding distributed file searching and file information collection are used by PAIS used bloom filter based. However, to improve the best performance of file sharing efficiency, and to get the blooms filter results in based on their rank order. Bloom filter approach has to use because of only check the newly collected bloom filter information to decrease file searching delay.

Keywords- P2P Networks, file sharing system, Bloom filter

I. INTRODUCTION

Distributed computing is to study the detail description about the field of computer science. Distributed system are the software system which basic components are located on general networked computers, those actions are coordinated and communicated by using message passing and these components has been interact with each other, because of achieving a common goal. Here some of the alternative mechanisms we have used for the message passing. Includes RPC like message queues and connectors. There are three significant characteristics of the distributed system,

- Concurrency of components
- Lack of a global clock, and
- Independent failure of components.

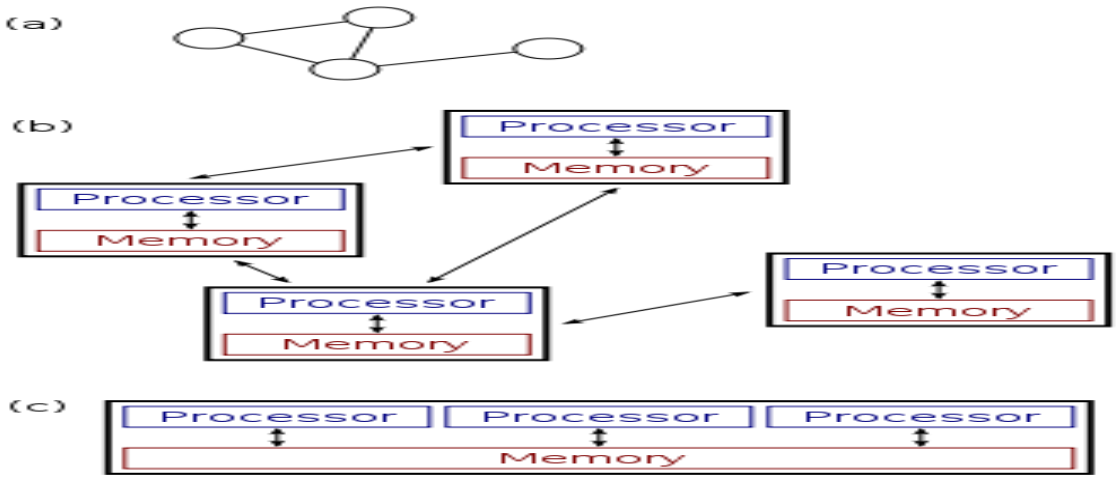
Location transparency is one of the most important challenges of distributed system. This distributed system is varying from SOA- based system to massively multiplayer online games to P2P application. A computer programs are running in the distributed system is called as distributed program, and it is write a programs. This distributed system solves the computational problems. Problem is too divided into many tasks, and these are solve the many problems one or more computers, these are communicated by passing message. Computer networks are referred to some terms such as a "distributed Algorithm", "Distributed programming", and "Distributed system". And each computer is distributed within some geographical location. Some of the following properties are normally used:

- There are several autonomous computational entities, each of which has its own local memory.
- The entities communicate with each other by message passing.

In this paper, we are calling entities as a computers or nodes. In distributed system main goal is to solve the large computational problem. Each computer may have its own clients with individual requirements, its main intension is to shared resources or to provide the communication services to the clients. And other some properties of the distributed system which includes the followings are:

- The system has to tolerate failures in individual computers.
- The structure of the system (network topology, network latency, number of computers) is not known in advance, the system may consist of different kinds of computers and network links, and the system may change during the execution of a distributed program.
- Each computer has only a limited, incomplete view of the system. Each computer may know only one part of the input.

Distributed systems are together with networked computers, they have some goals of their work. Followings have a lot of overlap those are "concurrent computing", "parallel computing", and "distributed computing" and doesn't have any clear distinction exists between them. System may be characterized both as distributed and parallel. In parallel system number of processor run concurrently. This parallel computing may be seen as a particular tightly coupled form of distributed computing and this distributed computing may be seen as a loosely coupled form of parallel computing. Concurrent systems are roughly classified into two types: "parallel" or "distributed". For using some follows criteria: In parallel computing, all processors may have access to a shared memory to exchange information between processors. In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.



This figure tells us the difference between the distributed and parallel systems. Figure (a) is a schematic view of a typical distributed system; this system is represented as a network topologies and each node is computer and each of the line is connected to the nodes are called communication link. Figure (b) this shows the more details about distributed system: each computer system has its own local memory; information may be exchange only by passing a message. Figure (c) its shows the parallel system and which each processor has a direct access to the shared memory.

II. RELATED WORK

In paper[1] 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. Each node in the Pastry network has a unique identifier (nodeId). When presented with a message and a key, a Pastry node efficiently routes the message to the node with a nodeId that is numerically closest to the key, among all currently live Pastry nodes. Each Pastry node keeps track of its immediate neighbors in the nodeId space, and notifies applications of new node arrivals, node failures and recoveries. Pastry takes into account network locality; it seeks to minimize the distance messages travel, according to a to scalar proximity metric like the number of IP routing hops Pastry is completely decentralized, scalable, and self-organizing; it automatically adapts to the arrival, departure and failure of nodes. Experimental results obtained with a prototype implementation on an emulated network of up to 100,000 nodes confirm Pastry's scalability and efficiency, its ability to self-organize and adapt to node failures, and its good network locality properties.

In paper [2] Existing information stockpiling frameworks in light of the various leveled registry tree association don't meet the versatility and usefulness necessities for exponentially developing information sets and progressively complex metadata inquiries in vast scale, Exabyte-level record frameworks with billions of documents. This paper proposes a novel decentralized semantic-mindful metadata association, called SmartStore, which abuses semantics of documents' metadata to prudently total connected records into semantic-mindful gatherings by utilizing data recovery instruments. The key thought of SmartStore is to restrain the pursuit extent of an intricate metadata question to a solitary or an insignificant number of semantically related gatherings and maintain a strategic distance from or reduce savage constrain look in the whole framework. The decentralized plan of SmartStore can enhance framework adaptability and lessen question inertness for complex inquiries (counting reach and top-k questions). Besides, it is additionally helpful for building semantic-mindful storing, and customary filename-based point inquiry. We have executed a model of SmartStore and broad trials in light of true follows demonstrate that SmartStore altogether enhances framework adaptability and diminishes inquiry inertness over database approaches. To the best of our insight, this is the primary study on the execution of complex inquiries in vast scale record frameworks.

In paper[3] Effective and dependable record questioning is imperative to the general execution of distributed (P2P) document sharing frameworks. Rising techniques are starting to address this test by abusing on the web informal

organizations (OSNs). Be that as it may, current OSN-based strategies essentially bunch normal intrigue hubs for high productivity or utmost the association between social companions for high dependability, which gives restricted improvement or repudiates the open and free administration objective of P2P frameworks. Little research has been attempted to completely and helpfully influence OSNs with coordinated thought of nearness and premium.

III. IMPLIMENTATION

MODULES:

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

MODULES DESCSRIPTION:

a) PAIS Structure

"Cycloid structured P2P network is developed by using PAIS. 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".

b) Node proximity representation

"A land marking 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".

c) 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".

d) 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 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".

e) File Distribution

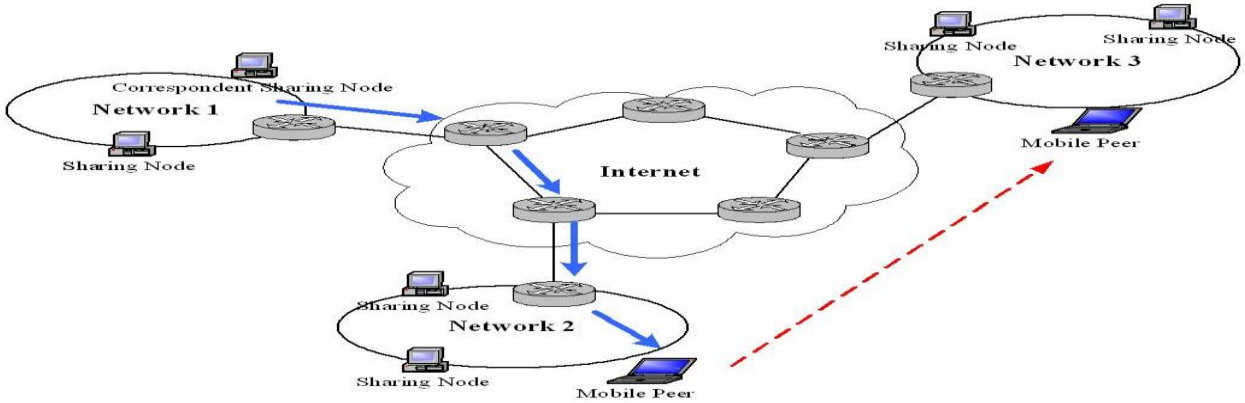
"As physically close and common-interest nodes form a sub-cluster, they can share files between each other so that a node can retrieve its requested file in its interest from a physically close node. For this purpose, the sub-cluster server maintains the index of all files in its sub-cluster for file 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".

IV. PROPOSED METHODOLOGY

This paper exhibits closeness mindful and intrigue grouped P2P document sharing Framework (PAIS) on an organized P2P framework. It shapes physically-close hubs into a bunch and further gathering's physically-close and normal intrigue hubs into a sub-group. It additionally puts documents with similar interests together and make them available through the DHT Query() directing capacity. All the more critically, it keeps all focal points of DHTs over unstructured P2Ps. Depending on DHT query strategy as opposed to broadcasting, the PAIS development devours considerably less cost in mapping hubs to groups and mapping bunches to intrigue sub-groups. PAIS utilizes a keen document replication calculation to further upgrade record query productivity. It makes copies of documents that are much of the time asked for by a gathering of physically close hubs in their area. Besides, PAIS upgrades the intra sub-group record seeking through a few methodologies.

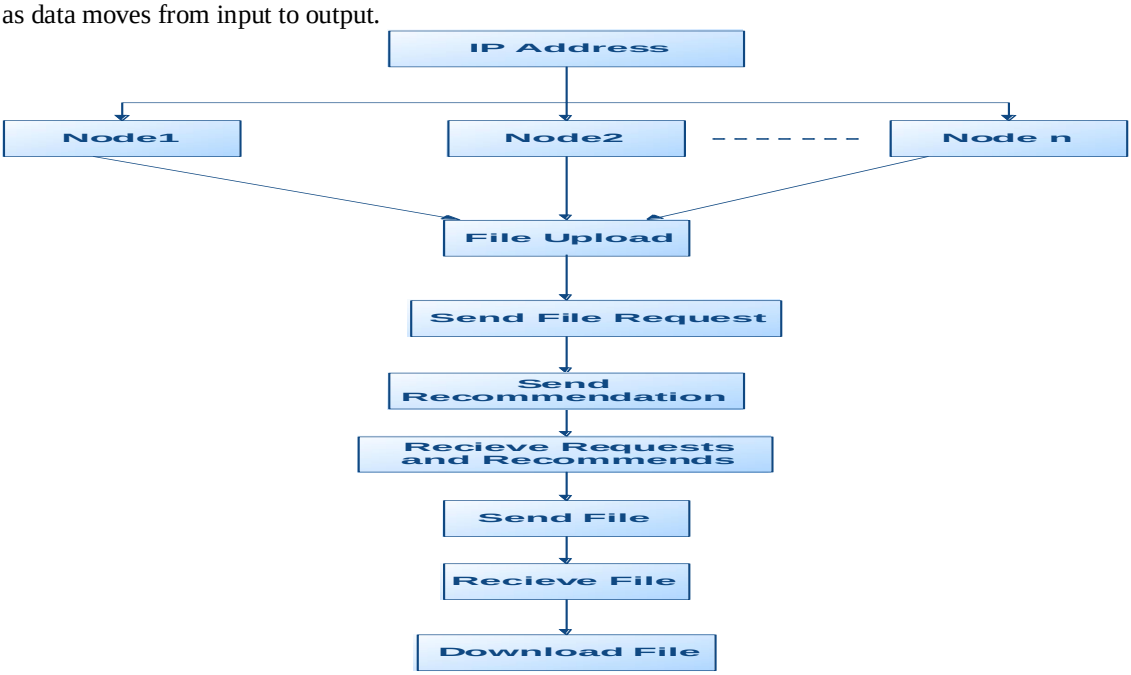
- It further classifies the interest of a sub-cluster to a number of sub-interests, and clusters common-sub-interest nodes into a group for file sharing.

- PAIS builds an overlay for each group that connects lower capacity nodes to higher capacity nodes for distributed file querying while avoiding node overload.
- 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.
- To reduce the overhead of the file information collection, PAIS uses bloom filter based file information collection and corresponding distributed file searching.
- 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.



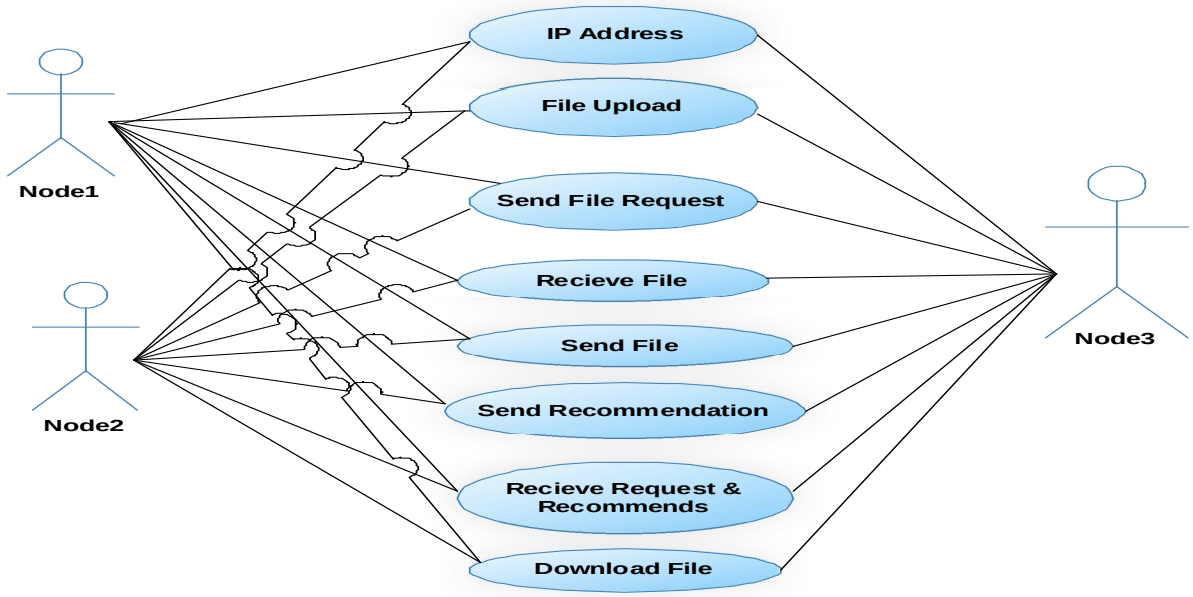
DATA FLOW DIAGRAM:

- This data flow is also called as "Bubble chart". This is graphically represented as a term which gives the input to the system, it takes various stages, and it gives the output to system.
- The DFD is very important modeling tool. This model has to develop the system components. This component is called as a system process. External entity interacts with a system and information flow in the system.
- DFD shows how this data moves through the system and what it required to modify the 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.



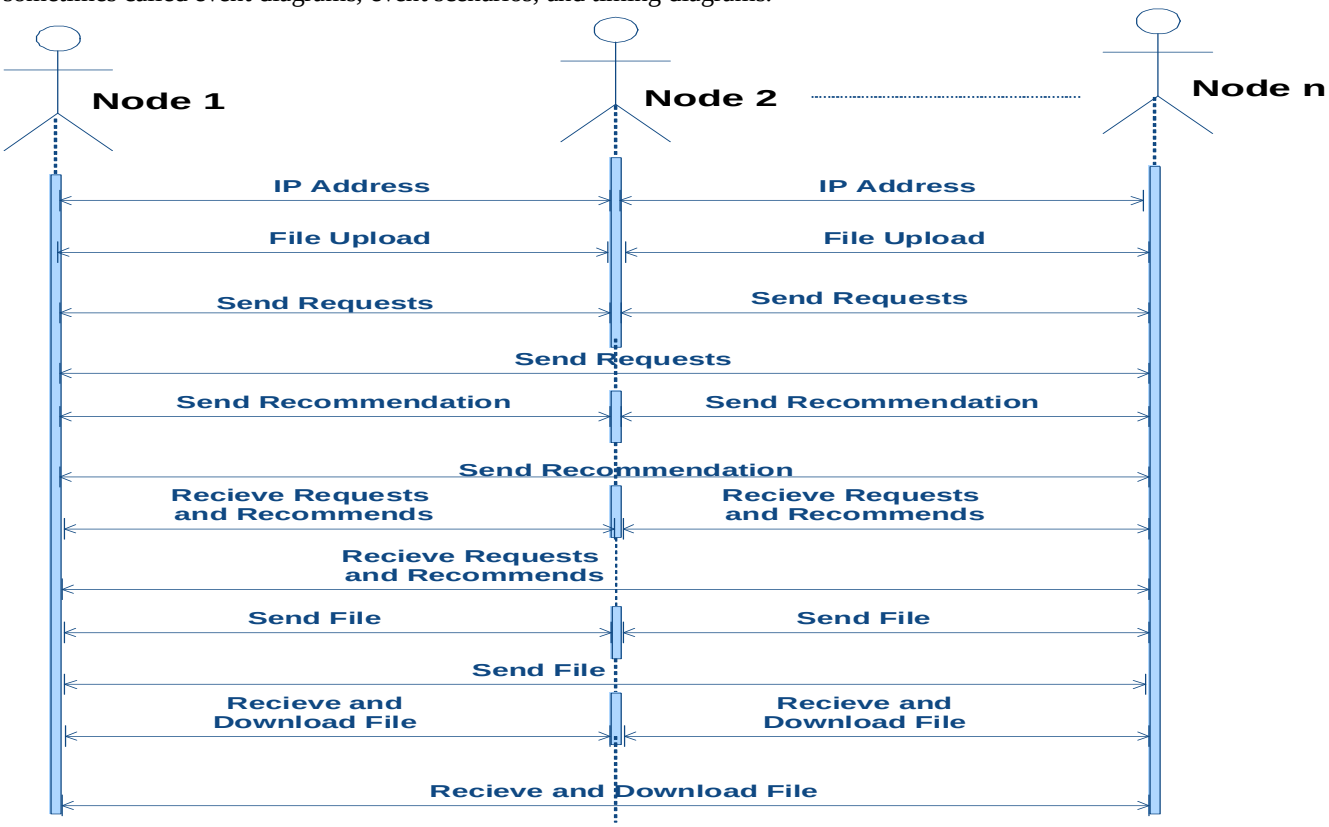
USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



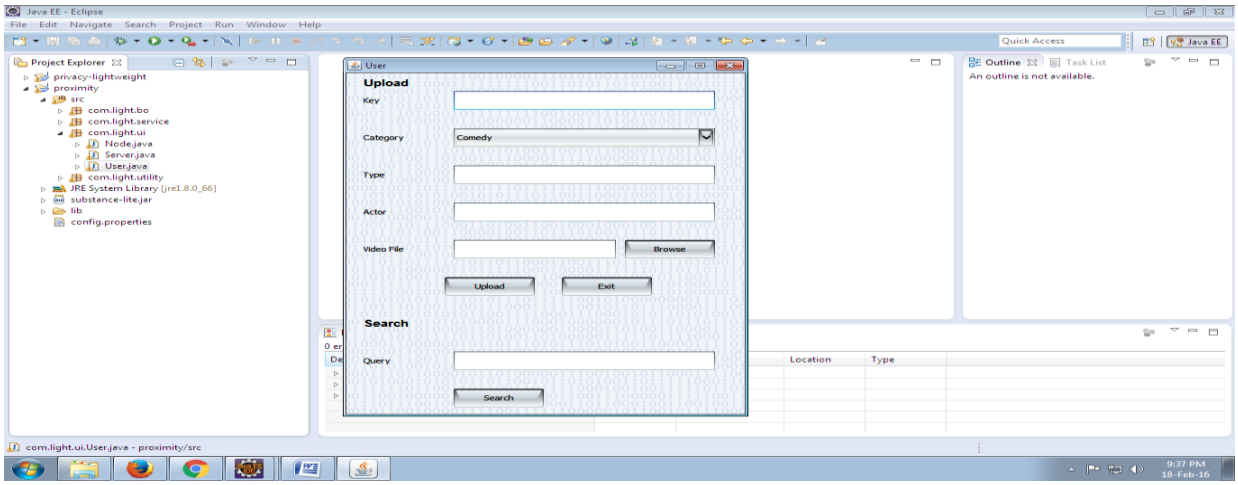
SEQUENCE DIAGRAM:

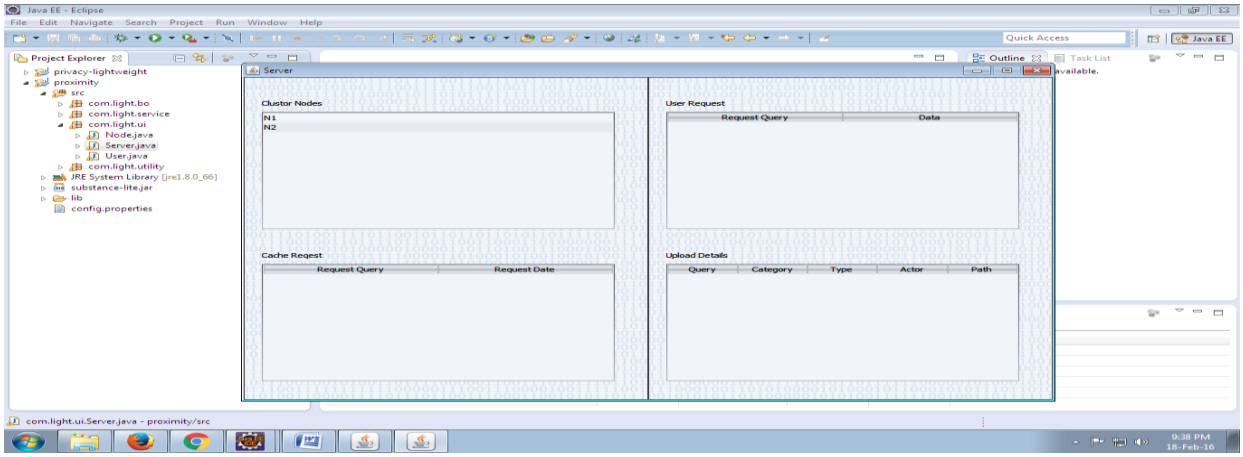
A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

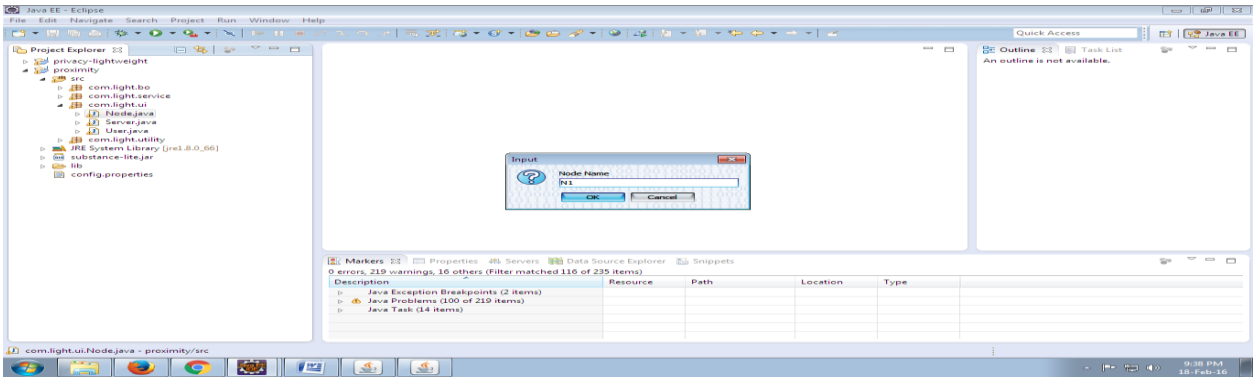


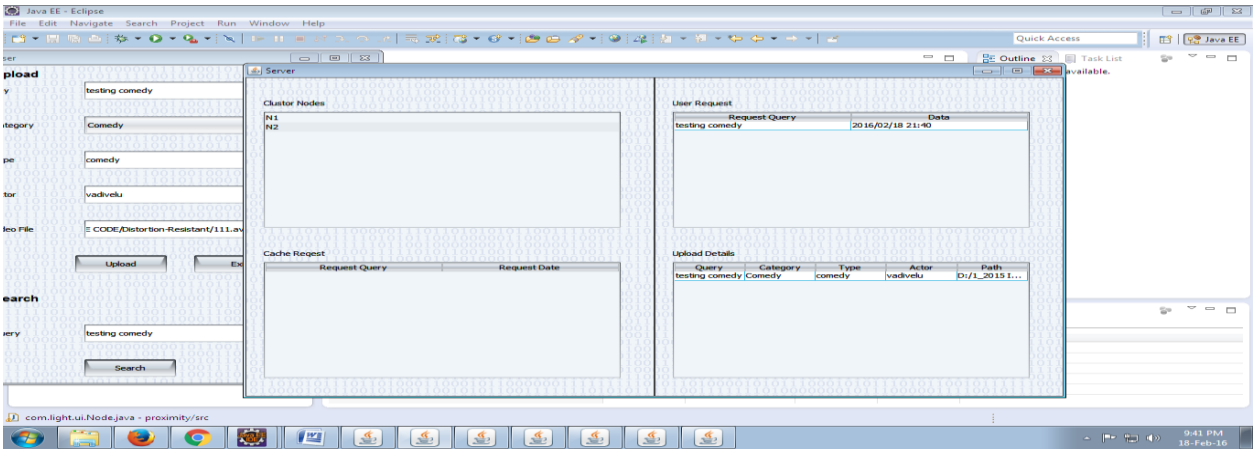
V. EXPERIMENTAL RESULTS

In this paper we discussed about the lots of benefits such as a current delivery networks, P2P Video on demand systems and data sharing in online social networks. Here also we have explained the design of PAIS structure. This is one of the suitable for file sharing system and file can classified into 1 or more interests and everyone is classified into number of sub-interests. It groups peers based on both interest and proximity by taking advantage of a hierarchical structure of a structured P2P. It uses the intelligent file replication algorithm that replicates a file recently requested by physically close nodes near their physical location. PAIS enhances the file searching efficiency among the proximity-close and common interest nodes through a number of approaches.









VI. CONCLUSION

Past decades, to enhance or to develop file location efficiently in P2P systems, also had proposed the proximity clustered super peer networks and interest clustered. These two have improved the performance of the P2P systems, few of them are works simultaneously. It's difficult to realize in structured P2P systems due to their based on topologies. We have introduced the new concept of proximity aware and interest clustered P2P file sharing system. It is mixture of interest and proximity by getting advantage of a hierarchical structure of a structured P2P. We used the intelligent file replication algorithm because that file is frequently requested by physically close nodes nearby its location and to enhance file lookup efficiency. At long last, PAIS upgrades the record looking effectiveness among the vicinity close and normal intrigue hubs through various methodologies. The follow driven test comes about on PlanetLab exhibit the effectiveness of PAIS in examination with other P2P record sharing frameworks. It drastically decreases the overhead and yields huge changes in record area.

REFERENCES

[1] 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.

[2] 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.

[3] H. Shen and K. Hwang, "Locality-preserving clustering and discover of wide-area grid resources," in Proc. IEEE Int. Conf. Distrib. Comput. Syst., 2009, pp. 518–525.

[4] H. Shen, C. Xu, and G. Chen, "Cycloid: A scalable constant-degree P2P overlay network," Perform. Eval., vol. 63, pp. 195–216, 2006.

[5] 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.

[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.