iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** XII **Month of publication:** December 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 3 Issue XII, December 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Survey on P2P File Sharing Systems Using Proximity-aware interest Clustering

Varalakshmi.T[#1], Arul Murugan.R[#2]
[#] *Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam*
[#] *Department of Information Technology, Bannari Amman Institute of Technology, Sathyamangalam*

Abstract*— Efficient file query is important to the overall performance of peer-to-peer (P2P) file sharing systems. Clustering peers by their common interests can significantly enhance the efficiency of file query. Clustering peers by their physical proximity can also improve file query performance. However, few current works are able to cluster peers based on both peer interest and physical proximity. Although structured P2Ps provide higher file query efficiency than unstructured P2Ps, it is difficult to realize it due to their strictly defined topologies. In this work, we introduce a Proximity-Aware and Interest-clustered P2P file sharing System (PAIS) based on a structured P2P, which forms physically-close nodes into a cluster and further groups physically-close and common-interest nodes into a sub-cluster based on a hierarchical topology. PAIS uses an intelligent file replication algorithm to further enhance file query 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. Further, the experimental results show the high effectiveness of the intra-sub-cluster file searching approaches in improving file searching efficiency.
Keywords*— P2P networks, file sharing system, proximity awareness, file replication, Bloom filter

I. INTRODUCTION

Over the past few years, the immense popularity of the Internet has produced a significant stimulus to P2P file sharing systems. There are two classes of P2P systems: unstructured and structured. Unstructured P2P networks such as Gnutella and Freenet does not assign responsibility for data to specific nodes. Nodes join and leave the network according to some loose rules. Currently, unstructured P2P networks' file query method is based on either flooding where the query is propagated to all the node's neighbors, or random-walkers where the query is forwarded to randomly chosen neighbors until the file is found. However, flooding and random walkers cannot guarantee data location. Structured P2P network si.e., Distributed Hash Tables (DHTs), can overcome the drawbacks with their features of higher efficiency, scalability, and deterministic data location. It strictly controlled topologies, and their lookup algorithms and data placement are precisely defined based on a DHT data structure and consistent hashing function. The node is responsible for a key can always be found eventhough if the system is in a continuous state of change. Most of the DHTs require Olog nÞ hops per lookup request with Olog nÞ neighbors per node, where n is the number of nodes in the system.

II. PROBLEM FORMULATION

Clustering Peers by their common interests can significantly enhance the efficiency of file query. Clustering peers by their physical proximity can also improve file query performance. However, few current works are able to cluster peers based on both peer interest and physical proximity. Although structured P2Ps provide higher file query efficiency than unstructured P2Ps, it is difficult to realize it due to their strictly defined topologies.Survey on Secure Query Processing. Proximity-awareness. Techniques to exploit topology information in P2P overlay routing include geographic layout, proximity routing, and proximity-neighbor selection. Geographic layout method maps the overlay's logical ID space to the physical network so that neighboring nodes in the ID space are also close in the physical network. It is employed in topologically-aware.

Volume 3 Issue XII, December 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Project Title	Algorithm/ Techniques	Remarks/ Problem Identification
Pastry: Scalable, decentralized object location and routing for	large-scale peer-to-peer systems	1. Pastry performs application-level routing and object location in a potentially very large overlay network of nodes connected via the Internet. 2. 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).
Semantic-aware metadata organization paradigm	Next-generation file systems	1. 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. 2. The key idea of SmartStore is to limit the search scope of a complex metadata query to a single or a minimal number of semantically correlated groups and avoid or alleviate brute-force search in the entire system..
An efficient and trustworthy P2P	social network integrated file sharing system	1. 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). 2. The 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.
A proximity-aware interest-clustered P2P file sharing system	PAIS algorithm	1. A proximity-aware and interest-clustered P2P file sharing system (PAIS) based on a structured P2P. It groups peers based on both interest and proximity. 2. PAIS supports sophisticated routing and clustering strategies based on a hierarchical topology. Theoretical analysis and simulation results demonstrate that PAIS dramatically reduces the overhead and enhances efficiency in file sharing.

Locality-preserving clustering	Discover of wide-area grid resources	1. The HCO framework, new performance metrics, and simulation experimental results. This HCO scheme compares favorably with other resource management methods in static and dynamic grid applications. In particular, it supports efficient resource clustering, reduces communications cost, and enhances resource discovery success rate in promoting large-scale distributed supercomputing applications

A. Existing System Model

1) 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.
2) In addition, most of these methods are on unstructured P2P systems that have no strict policy for topology construction.
3) Another class of methods to improve file location efficiency is through a proximity-aware structure.
4) The third class of methods to improve file location efficiency is to cluster nodes with similar interests which reduce the file location latency

III. THE RELATED WORK

The super-peer network is for efficient and scalable file consistency maintenance in struc- tured P2P systems. Our previous work built a super-peer network for load balancing. We proposed a self-organizing super-peer network architecture that sol- ves four issues in a fully decentralized manner: how client peers are related to super-peers, how super-peers locate files, how the load is balanced among the super-peers, and how the system deals with node failures. Mitra et al. Developed an analytical framework, which explains the emergence of super-peer networks on execution of the com- mercial P2P bootstrapping protocols by incoming nodes. Chordella [23] is a P2P system that is particularly designed for heterogeneous environments such as wireless networks. Sachez-Artigaz et al. investigated the feasibility of super-peer ratio maintenance, in which each peer can decide to be a super-peer independently of each other. Liu et al. proposed a hierarchical secure load balancing scheme in a P2P cloud system. It first balances the load among supernodes, and then depends on each supernode to balance the load among nodes under its management. Garbacki et al. A proposed a self-organizing supernode architecture to facilitate file querying. Each supernode caches the files recently requested by its children, and other peers send requests to the supernodes that can solve most of their requests.

IV. PROPOSED SYSTEM

In this paper represents 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.

Advantages of proposed system

A. 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.
B. 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.
C. It groups peers based on both interest and proximity by taking advantage of a hierarchical structure of a structured P2P.
D. 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.
E. PAIS enhances the file searching efficiency among the proximity-close and common interest nodes through a number of approaches.

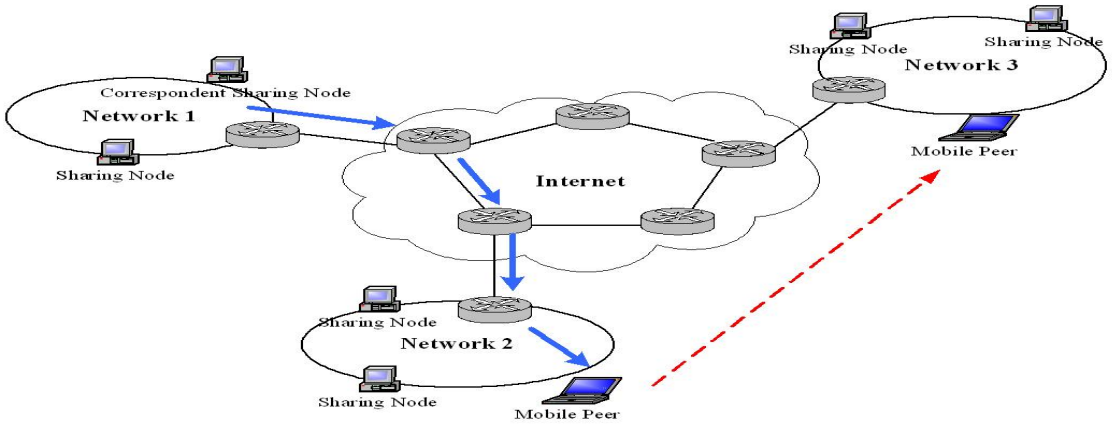


Fig2. System Architecture

A. 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. In PAIS, file ID is determined using the same way in Cycloid. That is, a file's cyclic index is its key's hash value modulated by d and its cubical index is set to the hash value divided by d, represented as ðH%d; H=dÞ, where H is consistent hash value of its key. A file's key can be its name or a string describing its contents. The file key must be consistent with the node interest. A node stores its files to the system via Cycloid interface Insert(fileID, file). According to Cycloid key assignment policy, each sub-cluster is responsible for the files whose cyclic indices fall into the key space sector it supervises. Thus, files with similar keys will be in the same sub-cluster in a cluster. The supernode in a sub-cluster further distributes files among its clients in balance. For example, in Fig. 1, a file with key "book" has ID ð3; 200Þ, then it will be stored in a node in sub-cluster a. In node joins and departures, the files are transferred between nodes based on the key assignment policy.

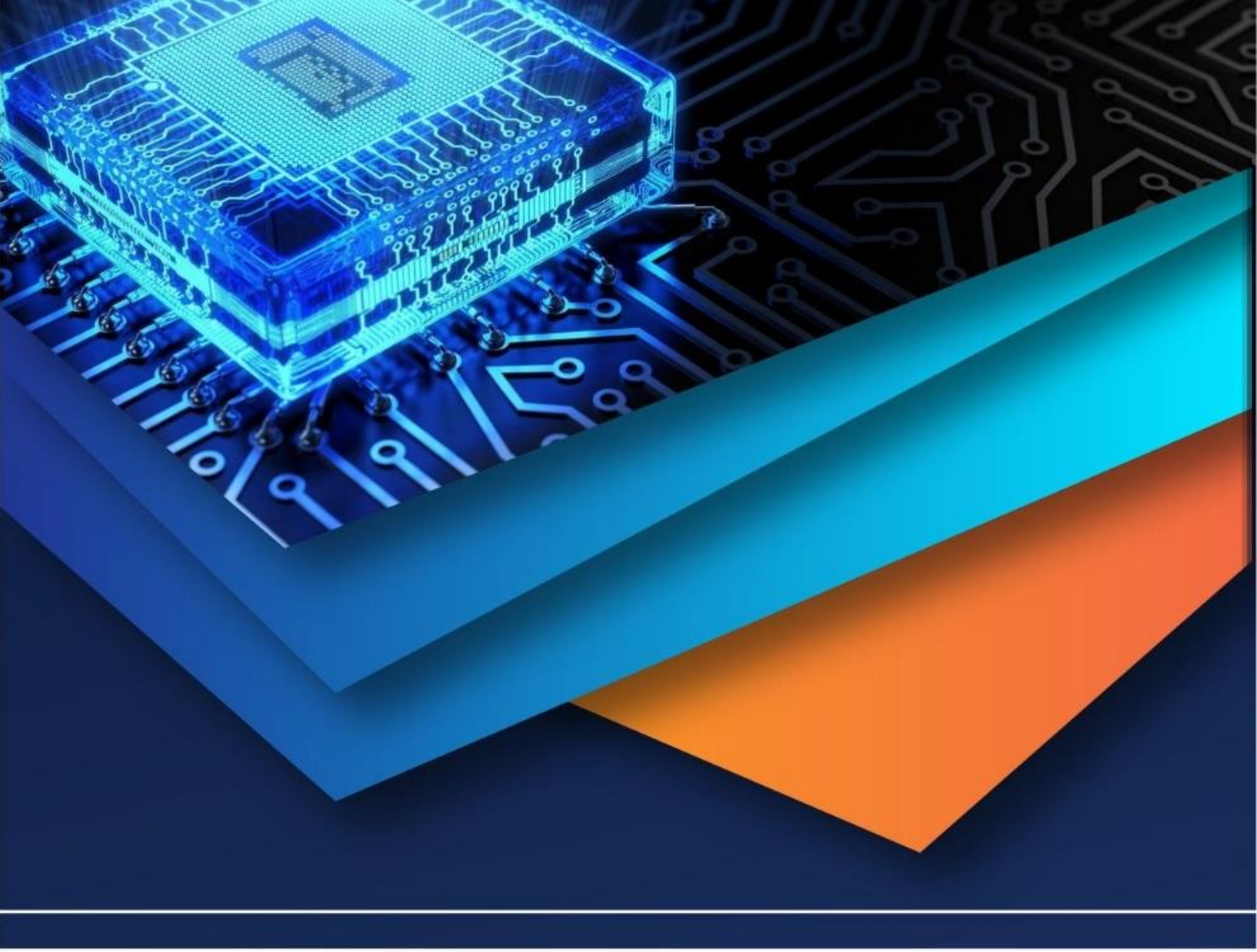
II. CONCLUSIONS

In recent years, to enhance file location efficiency in P2P systems, interest-clustered super-peer networks and prox-imity-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 effciency 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.

REFERENCES

[1] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system," Proc. Int. Workshop Des. Issues Anonymity Unobservability, 2001,pp. 46–66.
[2] 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.
[3] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," Proc. IFIP/ACM Int. Conf. Distrib. Syst. Platforms Heidelberg, 2001, pp. 329–350.
[4] 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.
[5] H. Shen, C. Xu, and G. Chen, "Cycloid: A scalable constant-degree P2P overlay network," Perform. Eval., vol. 63, pp. 195–216, 2006.
[6] 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.
[7] 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY