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

Location Efficient Proximity and Interest Clustered P2p File Sharing System

B.Ajay Kumar
M.Tech,
Dept of Computer Science & Engineering,
Usharama College of Engineering & Technology,
A.P, India.

B.Sri Vara Prasad
Assistant Professor,
Dept of Computer Science & Engineering,
Usharama College of Engineering & Technology,
A.P, India.

Abstract:
The overall performance of peer-to-peer (P2P) file sharing systems lies on efficient and trustworthy file querying. Several methods have been proposed to improve file location efficiency. One another method is a super-peer network topology system. This super-peer network topology consists of super nodes with high connectivity and regular nodes with low connectivity. A super node connects with other super nodes and some regular nodes and they are of high capacity. The regular node connects with a super node. Another method to improve file location efficiency is through clustering of nodes in the network. File replication technology is also widely used to reduce hot spots and improve file query efficiency.

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:
File sharing system, proximity awareness, file replication, P2P networks.

INTRODUCTION:

There are two classes of P2P systems: unstructured and structured. Unstructured peer-to-peer networks do not impose a particular structure on the overlay network by design, but rather are formed by nodes that randomly form connections to each other,

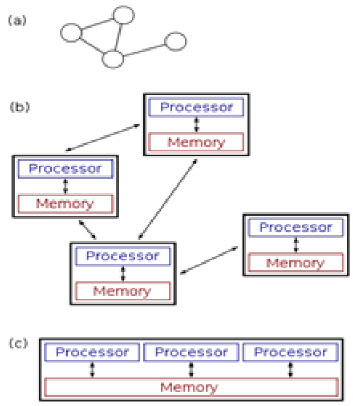
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

where file query method is based on either flooding, where the query is propagated to all neighbours of node, or random- walkers where the query is forwarded to neighbours which is chosen randomly until file is found in the search. Structured P2P networks Distributed Hash Tables (DHTs) [5], by which a node responsible for a key can always be found even if the system is in a continuous state of change. The overall performance of peer-to-peer (P2P) file sharing systems lies on efficient and trustworthy file querying. Several methods have been proposed to improve file location efficiency. One another method is a super-peer network topology system. This super-peer network topology consists of super nodes with high connectivity and regular nodes with low connectivity. A super node connects with other super nodes and some regular nodes and they are of high capacity. The regular node connects with a super node. Another method to improve file location efficiency is through clustering of nodes in the network.

File replication technology is also widely used to reduce hot spots and improve file query efficiency. Although numerous clustering methods have been proposed with different features, few methods can efficiently enhance the file location efficiency. This paper presents a technique of clustering nodes based on their interest and nodes proximity. In this, physically close nodes in the network are formed into a cluster and further groups physically close and common-interest nodes into a sub-cluster [1]. Files with the same interests are placing together and these files are made accessible through the DHT Lookup () routing function. Thereby the search for a file in the network will be more efficient. Distributed systems are groups of networked computers, which have the same goal for their work. The terms "concurrent computing", "parallel computing", and "distributed computing" have a lot of overlap, and no clear distinction exists between them. The same system may be characterised both as "parallel" and "distributed"; the processors in a typical distributed system run concurrently in parallel.

Parallel computing may be seen as a particular tightly coupled form of distributed computing, and distributed computing may be seen as a loosely coupled form of parallel computing. Nevertheless, it is possible to roughly classify concurrent systems as "parallel" or "distributed" using the following 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.



The figure on the right illustrates the difference between distributed and parallel systems. Figure (a) is a schematic view of a typical distributed system; as usual, the system is represented as a network topology in which each node is a computer and each line connecting the nodes is a communication link. Figure (b) shows the same distributed system in more detail: each computer has its own local memory, and information can be exchanged only by passing messages from one node to another by using the available communication links. Figure (c) shows a parallel system in which each processor has a direct access to a shared memory. The situation is further complicated by the traditional uses of the terms parallel and distributed algorithm that do not quite match the above definitions of parallel and distributed systems; see the section Theoretical foundations below for more detailed discussion.

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

Nevertheless, as a rule of thumb, high-performance parallel computation in a shared memory multiprocessor uses parallel algorithms while the coordination of a large-scale distributed system uses distributed algorithms.

PROPOSED SYSTEM DESCRIPTION:

In this paper, A cluster based locality and interest clustered p2p file sharing system is on structured p2p system. Each of the cluster based locality form physically close nodes as a cluster based on their interest. They also create a replica of each file that is accessed frequently requested by a group of physically closed nodes in their location. It also places each of files with same interest together and make them accessible easily through DHT lookup () routing function. The following fig (4) represents the cluster based location of the network. Initially, it classifies all the interest nodes into sub-cluster as a group. Second, it builds an overlay to reduce node overload. Third, it reduces the file sharing delay by using proactive file information collection.

- ❖ 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.

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.

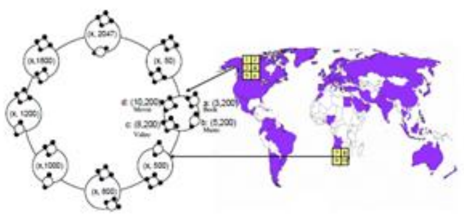
ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

- PAIS enhances the file searching efficiency among the proximity-close and common interest nodes through a number of approaches.

RELATED WORK:

Proximity Awareness is the first step in the clustering technique in the P2P file sharing system. The physically close nodes are grouped together and within this physically close nodes only further classifications of the nodes can be done. The main intuition in this method is that the physically close nodes likely to have similar distance to the certain number of landmarks that are selected. With an assumption that m landmarks are randomly scattered in the internet and measures the distance of each nodes to these landmarks. Thus uses the vectors of distances as Cartesian coordinates of the space. The nodes with the similar vector are considered as physically close nodes and are grouped together to form cluster. Proximity Awareness: Various techniques to exploit each topology information in p2p overlay routing include geographic layout, proximity neighbour selection and also proximity routing. The geographic layout maps each of the logical ID space to their physical network. It is already employed in topologically aware CAN[11].

In this proximity routing method, the logical overlay is constructed without considering the underlying physical topology. 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 neighbour selection. Geographic layout method maps the overlay's logical ID space to the physical network so that neighbouring nodes in the ID space are also close in the physical network. It is employed in topologically-aware.



PAIS-ARCHITECTURE

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. 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 neighbours 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.

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. 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. The decentralized design of SmartStore can improve system scalability and reduce query latency for complex queries (including range and top-k queries). Moreover, it is also conducive to constructing semantic-aware caching, and conventional filename-based point query. We have implemented a prototype of SmartStore and extensive experiments based on real-world traces show that SmartStore significantly improves system scalability and reduces query latency over database approaches. To the best of our knowledge, this is the first study on the implementation of complex queries in large-scale file systems.

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 analyse a BitTorrent file sharing trace, which proves the necessity of proximity and 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 sub-cluster, and connects the nodes in a sub-cluster with social links. Thus, when queries travel along trustable social links, they also gain higher probability of being successfully resolved by proximity-close nodes, simultaneously enhancing efficiency and trustworthiness. The results of trace-driven experiments on the real world PlanetLab test bed demonstrate the higher efficiency and trustworthiness of SoNet compared with other systems.

4) PAIS: A proximity-aware interest-clustered P2P file sharing system

AUTHORS: H. Shen

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. On the other hand, clustering peers by their physical proximity can also improve file query performance. Few current works are able to cluster peers based on both peer interest and physical proximity. It is even harder to realize it in structured P2Ps due to their strictly defined topologies, although they provide higher file query efficiency than unstructured P2Ps. In this paper, we introduce 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. 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.

5) Locality-preserving clustering and discover of wide-area grid resources

AUTHORS: H. Shen and K. Hwang

In large-scale computational or P2P grids, discovery of heterogeneous resources as a working group is crucial to achieving scalable performance. This paper presents a hierarchical cycloid overlay (HCO) architecture with resource clustering and discovery algorithms for efficient and robust resource discovery in wide-area distributed grid systems. We establish program/data locality by clustering resources based on their physical proximity and functional matching with user applications. We further develop randomized probing and cluster-token forwarding algorithms. The novelty of the HCO scheme lies in low overhead, fast speed and dynamism resilience in multi-resource discovery. The paper presents the HCO framework, new performance metrics, and simulation experimental results. This HCO scheme compares favourably 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.

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 PlanetLab 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 .In future we can provide more the user specific search such as file name and their sub-interest. Hence, the requester will get specific file of his interest. It can be enhanced by creating a new super peer node which contains only the information about the sub-interest division of each file and also it reduces the searching time delay.

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] R. Cuevas, N. Laoutais, X. Yang, G. Siganos, and P. Rodriguez, "BitTorrent locality and transit traffic reduction: When, why and at what cost?" IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 5, pp. 1177–1189, May 2014.

Author's Details:

B. Ajay Kumar is a student of Usha Rama College of Engineering & Technology, Telaprolu, Near Gannavaram, Vijayawada. He is presently pursuing his M.Tech degree from JNTU, Kakinada. He has obtained B.Tech, degree from JNTU, Kakinada.

B. Sri Vara Prasad is presently working as Assistant professor in CSE department, Usha Rama College of Engineering & Technology, Telaprolu, Near Gannavaram, Vijayawada.