Middle-East Journal of Scientific Research 24 (Techniques and Algorithms in Emerging Technologies): 444-449, 2016
ISSN 1990-9233;
© IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.TAET23611

Efficient Trace Aware Locality Clustered P2P File Sharing System

[1]V. Priya, [2]Dr. C. Nelson Kennedy Babu and [3]V. Avibala

[1]Assistant Professor, Department of Information Technology, K.S.Rangasamy College of Technology, India
[2]Professor, Department of Computer Science and Engineering, C.M.S College of Engineering, India
[3]PG Scholar, Department of IT, K.S.Rangasamy College of Technology, India

Abstract: ETAFS uses an intelligent file replication algorithm to further improve the query file efficiency. She creates replicas of files that are often requested by a group of physically close nodes in their place. In addition, the ETAFS improves intra-sub-cluster file searching through multiple approaches. First, it is classified further the interests of sub-cluster on a number of partial interests and cluster nodes in a direct part interest group for file sharing. Secondly ETAFS builds an overlay for each group, which connects lower capacity nodes to higher capacity nodes for distributed file query avoiding node load. Third, in order to reduce the file search delay, ETAFS used proactive file information collection, so that a requester may file know if its requested file in its nearby nodes. Fourth, to reduce the overhead of file information collection, use ETAFS bloom filter-based file information collection and corresponding distributed file search. Fifth, improve the file-sharing efficiency, ETAFS arranges the bloom filter results in order. Sixth, if you consider that a recently visited file be visited again tends to bloom filter approach will reduce only the verification of newly added bloom filter information file search amplified delay. Trace-driven experimental results from the real world show Planet Lab test that ETAFS dramatically reduces overhead and increases the efficiency of file-sharing with and without churn. Furthermore, the experimental results will show the high efficacy of intra-sub-cluster file search looking approaches in improving the efficiency file. An intuitive way to alleviate this problem is to create file replicas in the network. However, despite the efforts of the File Replication, no research has focused on the global optimum replica with minimum average query delay. In particular, current file replication protocols in networks have two drawbacks. First, they lack a rule to limited resources, to assign different files to the average query to minimize delay. Secondly, they just look at memory than available funds for the replicas, but neglect the fact that the file holder frequency of meetings with other nodes also plays an important role in determining file availability. Actually, a node that a higher frequency to meet other has a higher availability of their files. This is even more evident in sparsely distributed networks in which nodes make disturbing.

Key words: Cluster nodes · ETAFS · File sharing · Intra-sub-cluster

INTRODUCTION

In recent years, is the immense popularity of the Internet has produced a significant incentive for P2P file-sharing systems [1]. For example, sets the Bit Torrent approximately 35 percent of all met fix on the Internet. There are two classes of P2P systems: unstructured and assign structured. Unstructured P2P networks such as Gnutella and Free Net not assume any responsibility for the data on specific node. To connect nodes and let the network for some loose rules. Currently, unstructured P2P networks' File Scan method either to flooding, where the query to all neighbors of the node or random walker, where the query forwarded to random neighbors until the file propagated found based. However, flood in bulk random walker cannot guarantee data availability. StructuredP2P networks ie Distributed Hash Tables (DHT), can the disadvantages to overcome with their properties higher efficiency, scalability and deterministic data location. They were strictly controlled topologies and precisely defined on a DHT-data structure and function exists enchasing their data placement and lookup algorithms. The charge of a key node can always be found, even if the system to change in a steady state.

Correspondnig Author: V. Priya, Assistant Professor, Department of Information Technology, K.S. Rangasamy College of Technology, India.

Most of DHTs require Oolong NTH hops per lookup request with Oolong nth neighbors per node, where the number of nodes in the system.A to evaluate important criteria, a P2P file-sharing system is its file location efficiency [3]. In order to improve this efficiency, numerous methods have been proposed. One method uses a super-peer topology, consisting of super nodes with fast connections and regular nodes with slower connections. A super node connects with other super-node and several regular nodes and connects a regular node to a super node. In this super-peer topology, the nodes in the middle of the network are faster and therefore produce amore eliable and stable backbone. This can be passed as a slower backbone more messages and therefore allows for greater scalability. Upper-peer networks occupy the middle-ground between central and very symmetricP2P networks and have the potential to combine the benefits of centralized or decentralized searches. Another class of methods for file location efficiency's by a proximity-aware structure to improve. A logical near abstraction derived from a P2P system does not necessarily correspond to the physical proximity Information in reality. The shortest route to the Routing Protocol (i.e. the least hop routing) is not the shortest physical path. This mismatch is to big obstacle for the use and performance optimization of P2P file-sharing systems [2]. A P2P system should provide information to file utilize proximity query overhead and improve reduce their efficiency. In other words, the allocation or replicating file to a node that is physically closer to a requester can significantly help the applicant to retrieve the efficiently.Proximity-aware clustering file used to colleague group physically near to effectively improve efficiency. Have The Third class of methods file location efficiency is to cluster nodes with similar interests, to reduce the file Stelle, to improve latency.Although numerous neighborhood-based and interest-based super-peer topologies have been proposed with different features, are some methods capable of peers according to both proximity cluster and interest. Moreover most of synthesis methods on unstructured P2P systems that no strict policies have topology structure. You cannot provide general DHTs be directly despite their higher file location efficiency. This paper presents a near-aware and interest clustered P2P file-sharing system (ETAFS) on a structured P2Psystem. It forms physically-close-node to a cluster and further groups physically close and of common interest nodes into a sub-cluster. It also files with the same interests together and make it through the DHT Lookup () routing function accessible. More importantly, it keeps the all advantages DHTs about unstructured P2P. Relying on DHT Lookup policy rather than broadcasting, construction consumes the ETAFS much less cost in mapping node to clusters and mapping cluster that interest sub clusters.

ETAFS uses an intelligent file replication algorithm to file further enhance lookup efficiency. It creates copies files requested that are frequently by a group of physically close nodes in their position. In addition, the ETAFS intra sub cluster file improves search through multiple approaches. First, they continue it classifies the interest of a sub-cluster at a lower Number Of interests and common cluster sub interest nodes into a group for file sharing. Secondly ETAFS builds a overlay for each group, the lower capacity nodes higher capacity connect to the node for distributed file search during avoiding node load. Third, to reduce file search delay, ETAFS uses proactive file information collection, so that a file requester can know whether its requested file in its nearby nodes. Fourth, information collections reduce the overhead of file ETAFS flower used filter based file information collection and corresponding Distributed File Search. Fifthly, to enhance file sharing efficiency, ETAFS arranges the bloom filter results in order. Sixth, if you consider that a recently visited file tends to be revisited, the bloom is filter-based approach only the verification of newly added bloom filter information added, to reduce file search delay.Note that although this work is for P2P file sharing systems, the techniques that can be used in this proposal paper fit many current applications such as content delivery networks, P2P video-on-demand systems and data sharing in online social networks animals. Since the architecture of ETAFS is based on a structured P2P system, its architecture cannot be for unstructured P2P systems are used. However ETAFS's techniques for improving efficiency of the inner sub-cluster file search can for unstructured P2P system since node in an intra-sub-cluster are used connected in an unstructured [4]. The remainder of this paper is structured as follows. Section 2 presents a concise review of representative approaches to fi le location efficiency improvement in P2P systems. Section 3 describes ETAFS, referring to its structure construction and file searching algorithms. Section 4 describes the approaches that improve ETAFS the intra-cluster file search. Section 5presents-tracking test results on the effectiveness and efficiency of ETAFS comparison with other systems in static and dynamic environment Section 6 presents the conclusion of this paper.

Related Work: We discuss the related works most relevant to ETAFS in three groups: super-peer topology, proximity-awareness and interest-based ?les sharing. Super-peer topology. Fast Track and Morpheus use super-peer topology. The super-peer network is for efficient and scalable ?les consistency maintenance in structured P2P systems. Our previous work built a super-peer network for load balancing. Garbacki et all proposed a self-organizing super-peer network architecture that solves four issues in a fully decentralized manner: how client peers are related to super-peers, how super-peers locate ?les, how the load is balanced among the super-peers and how the system deals with node failures. Mistreat all. Developed an analytical framework, which explains the emergence of super-peer networks on execution of the commercial P2P bootstrapping protocols by incoming nodes. Chordella is a P2P system that is particularly designed for heterogeneous environments such as wireless networks. Sachez-Artigaz et all 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 all proposed a hierarchical secure load balancing scheme in a P2P cloud system. It ?rst balances the load among super nodes and then depends on each super node to balance the load among nodes under its management [5].

Garbacki etal. proposed self-organizing super node architecture to facilitate ?les querying. Each super node caches the ?les recently requested by its children and other peers send requests to the super nodes that can solve most of their requests. 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 CAN. In the proximity routing method, the logical overlay is constructed without considering the underlying physical topology. In a routing, the node with the closest physical distance to the object key is chosen among the next hop candidates in the routing table. The entries of a routing table are selected based on a proximity metric among all the nodes that satisfy the constraint of the logical overlay [6].

Genaud and Rattanapoka proposed a P2P-based middleware for locality-aware resource discovery. The works measured the inter-ISP traf?c in Bit Torrent and indicated the importance of locality-awareness traf?c in reducing the traf?c over long-distance connections. Liu et al. examined traf?c locality in PPLive and revealed that PPLive achieves high ISP level traf?c locality.

Shen and Hwang proposed a locality-aware architecture with resource clustering and discovery algorithms for ef?cient and robust resource location in wide-area distributed grid systems. Lehrieder studied locality awareness in scenarios with real-life, skewed peer distributions and heterogeneous access bandwidths of peers[7].

Manzillo proposed the collaborative locality-aware overlay service, which reduces the transmit cost of ISPs by switching to the source inside the same ISP with the requester. Interest-base ?le sharing. One category of interest-base ?le sharing networks is called schema based networks. They use explicit schemas to describe peers' contents based on semantic description and allow the aggregation and integration of data from distributed data sources. Hang and Sia proposed a method for clustering peers that share similar properties together and a new intelligent query routing strategy. Crespo and Garcia-Molina proposed a semantic overlay network (SON) based on the semantic relations among peers. Ruffo and Schifanella studied the spontaneous communities of users in P2P ?les sharing applications and found that a family of structures show self-organized interest-based clusters. The works in consider node interest for publish and subscribe. Iamnitchieal. found the small world pattern in the interest-sharing community graphs and suggested clustering common-interest nodes to improve ?les searching efficiency. Some works leverage the social network common interest property for efficient ?les searching.

Proposed Work: We provide a new concept of a resource for file replication, which holds both node storage and meeting frequency. We theoretically investigate the impact of resource allocation on the average query delay and from a resource allocation rule to minimize the average query delay. We a distributed file replication protocol further propose to implement the proposed scheme. Extensive track-driven experiments with synthesized and real traces show that our protocol can achieve shorter average polling delay at a lower price than the current replication protocols. File Replication is an effective way to increase the availability and reduce file polling delay. It creates replicas of a file, its probability encountered by requests to improve. Unfortunately, it is impractical and inefficient to allow all nodes that hold replicas of all files in the system, despite limited node resources. Also file is querying delay is always a major concern in a file-sharing system. Users often wish to provide their requested files received quickly, regardless of whether the files are very popular or not. Thus, a critical issue for further investigation is applicable protocols, each node replicates files often interrogates or a group of nodes to create a replica for each file they often query. In the former redundant copies are created any problems in the system, whereby the waste of resources. In the latter, albeit redundant copies of group on the basis of cooperation reduces neighbor nodes can be separated by node mobility, resulting in large polling delay. There are also some works addressing content caching in separate networks for efficient data recovery or message routing. They basically cache data, which frequently queried on places that are frequently visited by the mobile node. Both the two categories of replication methods fail thoroughly to consider that the mobility of a node affects the availability of the files.

Despite the efforts, current file replication protocols lacks a rule for limited resources in order to allocate files for replica creation, i.e. to achieve global search efficiency optimization under limited resources, the minimum average query delay. Simply hold memory as a resource for the replicas, but neglect the frequency of a node other nodes (meeting briefly ability) to make also affects the availability of the files.Files in a node with higher capacity meetings has higher availability. It allows file sharing if no base stations available. With the P2P architecture, the congestion can be avoided on overloaded server in the current client-server based file sharing systems. He wasted otherwise uses peer-to-peer communications to mobile nodes on. Free and unobtrusive node access and share files in distributed.

Method

Analysis of Globally Optimal File Replication: We present the general process to model the expected file querying delay with file replication. We let m^i be the probability that a node's newly met node in the coming time interval is node i, which reflects the meeting ability of the files on node i. We also use Xij to denote whether node i owns file jor its replication.

Meeting Ability Distribution in Real Traces: For each trace, we measured the meeting abilities of all nodes and ranked them in decreasing order, as shown in Figs. 1a and 1b. We see that in all three traces, node meeting ability is distributed in a wide range. This matches with our previous claim that nodes usually have different meeting abilities. Also, it verifies the necessity of considering node meeting ability as a resource in file replication since if all nodes have similar meeting ability, replicas on different nodes have similar probability to meet requesters and hence there is no need to consider meeting ability in resource allocation.

Node Movement Models: The community-based mobility model has been used in content dissemination or routing algorithms for disconnected MANETs/DTNs to depict node mobility [11]. In this model, the entire test area is split into different sub-areas, denoted as caves. Each cave holds one community. A node belongs to one or more communities (i.e., home community). The routines and (or) social relationships of a node tend to decide its mobility pattern. When moving, a node has probability Pin to stay in the home community and probability 1 _ Pin to visit a foreign community [9]. A node moves within its home communities for most of the time.

Design of the File Replication Protocol: We propose a distributed file replication protocol that can approximately realize the optimal file replication rule with the two mobility models in a distributed manner. Since the OFRR in the two mobility models present the protocol in this section without indicating the specific mobility model. We first introduce the challenges to realize the OFRR and our solutions. We then propose a replication protocol to realize OFRR and analyze the effect of the protocol.

Performance Evaluation in Disconnected Manets with the Community-Based Mobility Model: We evaluated the performance of PCS in comparison with DCG, CACHE-DTN, OPTM and Random. CACHE-DTN caches each file on the central node of each network center location (NCL). If a central node is full, its replicas are stored in its neighbor nodes according to their popularities [10]. A more popular replica is stored closer to the central node. The experiment settings and metrics are the same as in Section 5 unless otherwise specified in below. The total number of queries was set to 6000 Rp and Rp is the query rate and was varied in the range. In the experiment with the MIT Reality trace and the Haggle trace, queries were generated evenly in the time period of [0.3 Ms and 2.3 Ms] and [0.05 Ms and 0.25 Ms] was set to 0.3 Ms and 0.04 MS and the TTL of each query was set to 0.3 Ms and 0.04 Ms, respectively. We again set the confidence interval to 95 percent when handling experiment.

Implementation

Analysis of Globally Optimal File Replication: We present the general process to model the expected file querying delay with file replication. We let m^i be the probability that a node's newly met node in the coming time interval is node i, which reflects the meeting ability of the files on node i. We also use Xij to denote whether node i owns file jor its replication.

Meeting Ability Distribution in Real Traces: For each trace, we measured the meeting abilities of all nodes and ranked them in decreasing order, as shown in Figs. 1a and 1b. We see that in all three traces, node meeting ability is distributed in a wide range. This matches with our previous claim that nodes usually have different meeting abilities. Also, it verifies the necessity of considering node meeting ability as a resource in file replication since if all nodes have similar meeting ability, replicas on different nodes have similar probability to meet requesters and hence there is no need to consider meeting ability in resource allocation.

Node Movement Models: The community-based mobility model has been used in content dissemination or routing algorithms for disconnected MANETs/DTNs to depict node mobility. In this model, the entire test area is split into different sub-areas, denoted as caves. Each cave holds one community. A node belongs to one or more communities (i.e., home community). The routines and (or) social relationships of a node tend to decide its mobility pattern. When moving, a node has probability Pin to stay in the home community and probability 1 Pin to visit a foreign community[12]. A node moves within its home communities for most of the time.

Design of the File Replication Protocol: We propose a distributed file replication protocol that can approximately realize the optimal file replication rule with the two mobility models in a distributed manner [15]. Since the OFRR in the two mobility models present the protocol in this section without indicating the specific mobility model. We first introduce the challenges to realize the OFRR and our solutions. We then propose a replication protocol to realize OFRR and analyze the effect of the protocol.

CONCLUSION

Expand In recent years, to file location efficiency in P2Psystems that interest clustered super-peer networks and proximity clustered super-peer networks have been proposed.Although both strategies improve the performance of P2Psystems, few works cluster colleague based on both peer intere are physical proximity at the same time. Moreover, it is more difficult to realize in structured P2P systems because of their strictly defined topologies, although they. A high efficiency of file location as unstructured P2P In this paper, we present a proximity-aware and interest-clusteredP2P file-sharing system, based on a structured P2P. It groups colleagues based on both interest and close through the use of a hierarchical structure of a structuredP2P. ETAFS uses an intelligent file replication algorithm that often physically prompted a file close nodes near their physical location to improve the replicated file lookup efficiency [13].Finally ETAFS improved the file search efficiency among the proximity and common interest close nodes by a number of approaches. The trace-driven experimental results to show the efficiency of the planet Lab ETAFS compared to other P2P file sharing systems. It reduces the overhead and leads to significant improvements in file location efficiency in dynamic node [14]. The experimental results demonstrate the effectiveness f approaches to improving file search efficiency among close proximity and common interest knots.

REFERENCES

1. Carchiolo, V., M. Malgeri, G. Mangioni and V. Nicosia, 2010. An Adaptive Overlay Network Inspired by Social Behavior, J. Parallel and Distributed Computing, 70(3): 282-295.
2. Hughes, D., G. Coulson and J. Walker dine, 2005. Free Riding on Gnutella Revisited: The Bell Tolls?" IEEE Distributed Systems Online, 6(6): 81.
3. Clarke, I., O. Sandberg, B. Wiley and T.W. Hong, 2001. Free net: A distributed anonymous information storage and retrieval system," in Proc. Int. Workshop Des. Issues Anonymity Un observability, pp: 46-66.
4. Stoica, I., R. Morris, D. Liben-Nowell, D.R. Karger, M.F. Kaashoek, F. Dabek and H. Bala krishnan, 2003. Chord: A scalable peer-to peer lookup protocol for internet applications, IEEE/ACM Transnet., 11(1): 17-32.
5. Rowstron, A. and P. Druschel, 2001. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems, in Proc. IFIP/ACM Int. Conf. Distrib. Syst. Platforms Heidelberg, pp: 329-350.
6. Zhao, B.Y., L. Huang, J. Stripling, S.C. Rhea, A.D. Joseph and J. Kubiatowicz, 2004. Tapestry: A resilient global-scale overlay for service deployment, IEEE J. Sel. Areas Commun., 22(1): 41-53.

7. Shen, H., C. Xu and G. Chen, 2006. Cycloid: A scalable constant-degreeP2P overlay network, Perform. Eval., 63: 195-216.
8. Li, Z., G. Xie and Z. Li, 2008. Efficient and scalable consistency maintenance for heterogeneous peer-to-peer systems, IEEE Trans. Parallel Distrib. Syst., 19(12): 1695-1708.
9. Shen, H. and C.Z. Xu, 2008. Hash-based proximity clustering for efficient load balancing in heterogeneous DHT networks, J. Parallel Distrib. Comput., 68: 686-702.
10. Seshadri, S. and B. Cooper, 2007. Routing Queries through a Peer-to-Peer InfoBeacons Network Using Information Retrieval Techniques, IEEE Trans. Parallel and Distributed Systems, 18(12): 1754-1765.
11. Lehrieder, F., S. Oechsner, T. Hossfeld, Z. Despotovic, W. Kellerer and M. Michel, 2010. Can P2P-Users Benefit from Locality-Awareness?" Proc. IEEE 10th Int'l Conf. Peer-to-Peer Computing (P2P), pp: 1-9.
12. Choffnes, D.R. and F.E. Bustamante, 2008. Taming the Torrent:A Practical Approach to Reducing Cross-ISP Traffic in P2P Systems, Proc. ACM Special Interest Group Data Comm. (SIGCOMM), pp: 363-374.
13. Seetharaman, S. and M.H. Ammar, 2008. Managing Inter-Domain Traffic in the Presence of BitTorrent File-Sharing, Proc. ACM Special Interest Group Computer Systems Performance Evaluation Community (SIGMETRICS), pp: 453-454.
14. Zhang, H., Z. Shao, M. Chen and K. Ramchandran, 2011. Optimal Neighbour Selection in BitTorrent-like Peer-to-Peer Networks, Proc. ACM Special Interest Group Computer Systems Performance Evaluation Community (SIGMETRICS), pp: 141-142.
15. Iamnitchi, A., M. Ripeanu, E. Santos-Neto and I. Foster, 2011. The Small World of File Sharing, IEEE Trans. Parallel and Distributed Systems, 22(7): 1120-1134.
16. Li, M., W.C. Lee, A. Sivasubramaniam and J. Zhao, 2008. SSW: A Small-World-Based Overlay for Peer-to-Peer Search, IEEE Trans. Parallel and Distributed Systems, 19(6): 735-749.
17. Chen, G., C.P. Low and Z. Yang, 2008. Enhancing Search Performance in Unstructured P2P Networks Based on Users' Common Interest, IEEE Trans. Parallel and Distributed Systems, 19(6): 821-836.
18. Lin, K.C.J., C.P. Wang, C.F. Chou and L. Golubchik, 2010. SocioNet: A Social-Based Multimedia Access System for Unstructured P2P Networks, IEEE Trans. Parallel and Distributed Systems, 21(7): 1027-1041.
19. Cheng, X. and J. Liu, 2009. NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing, Proc. IEEE Conf. Computer Comm. (INFOCOM), pp: 1152-1160
20. Li, Y., L. Shou and K.L. Tan, 2008. CYBER: A Community-Based Search Engine, Proc. 8th Int'l Conf. Peer-to-Peer Computing (P2P), pp: 215-224.