www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 8 August, 2013 Page No. 2448-2453

HOP-COUNT ANALYSIS IN UNSTRUCTURED P2P NETWORKS

Mrs.R.Sivapriya, Mr.S.Dhanabal ,

**- PG Scholar, Department of Computer Science and Engineering,*
+- Assistant Professor, Department of Computer Science and Engineering,
rpriya2512@gmail.com
PGP College of Engineering and Technology, Namakkal

ABSTRACT

As the peers participating in Unstructured networks interconnect randomly, they rely on flooding query messages to discover objects of interest and thus introduce Remarkable network traffic. Empirical measurement studies indicate that the peers in P2P networks have similar preferences, and have Recently proposed unstructured P2P networks that organize participating peers by exploiting their similarity. The resultant networks may not perform searches efficiently and effectively because existing overlay topology construction algorithms often create Unstructured P2P networks without performance guarantees.A novel overlay formation algorithm for unstructured P2P networks, is unique in that it poses rigorous performance guarantees. Theoretical performance results conclude that in a constant probability, searching an object in this proposed network efficiently takes hops (where c is a small constant), and the search progressively and effectively exploits the similarity of peers. This overlay approach requires a minimal number of communication over the network and provides tunable parameters to maximize performance for various network topologies. It provides a powerful technique for the aggregates of various topologies and data clustering, but comes with limitations based upon a given topologies structure and connectivity. For topologies with very distinct clusters of peers, it becomes increasingly difficult to accurately obtain random samples due to the inability of random-walk process to quickly reach all clusters.

Index Terms—Peer-to-peer systems, unstructured overlay networks, search,hop count,clustering.

I.INTRODUCTION

PEER-TO-PEER (P2P) networks (or overlay networks) have been widely deployed in the Internet, and they provide various services such as file sharing, information retrieval, media streaming, and telephony. P2P applications are popular because they primarily provide low entry barriers and self-scaling.

Centralized: Napster and other similar systems have a constantly-updated directory hosted at central locations (e.g., the Napster web site). Nodes in the P2P network issue queries to the central directory server to find which other nodes hold the desired files. Such centralized approaches do not scale well and have single points of failure.

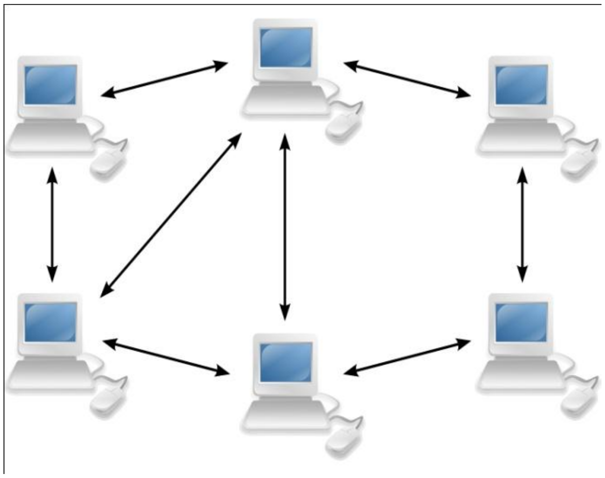

Fig 1. Peer-to-Peer Network

Decentralized and Unstructured: These are systems in which there is neither a centralized directory nor any precise control over the network topology or file placement. Gnutella is an example of such designs. The network is formed by nodes but not based on any knowledge of the topology (as it is in structured designs).

File sharing using P2P networks has gained wide popularity; some reports suggest that P2P traffic is the dominant consumer of bandwidth ahead of Web traffic. This popularity triggered a lot of research activity. While one research trend aims at improving the performance and features, another trend is concerned with analyzing and modeling these networks.

Object search is an essential building block in several P2P applications (e.g., file sharing and information retrieval). Gnutella is a popular P2P search protocol in the mass market. Specifically, because Gnutella networks are unstructured, and the peers participating in networks connect to one another randomly, peers search objects in the networks through message flooding.

To flood a message, an inquiry peer broadcasts the message to its neighbors (by the neighbors of peer i, we mean those peers that have end-to-end connections with i).

The broadcast message is associated with a positive integer time-to-live (TTL) value. Upon receiving a message, the peer decreases the TTL value associated with the message by 1 and then relays the message with the updated TTL value to its neighbors, except the one sending the message to j, if the TTL value remains positive. Aside from forwarding the message to the neighbors, j searches its local store to see if it can provide the objects requested by peer i. Conceptually, if j has the requested objects and is willing to supply them, then j either directly sends i the objects or returns the objects to the overlay path where the query message traverses from i to j.

To find a file, a node queries its neighbors. The most typical query method is flooding, where the query is propagated to all neighbors within a certain radius. These unstructured designs are extremely resilient to nodes entering and leaving the system. However, the current search mechanisms are extremely unscalable, generating large loads on the network participants.

In this study,optimizing the search performance in Gnutella-like unstructured P2P networks.Existing techniques in the literature for improving search performa nce in unstructured P2P networks include indexing, replications , superpeer architectures, and overlay topologies , among others. In this paper, we primarily study the overlay topology formation technique for unstructured P2P networks, aiming to enhance search efficiency and effectiveness. In particular, as recent measurement studies, show that peers are
likely to resolve the queries issued by the peers sharing the common preferences, our study intends to organize the participating peers to exploit their similarity

II. RELATED WORKS

The significance of small-world patterns is twofold: it provides a rigorous support to intuition and it suggests the potential to exploit these naturally emerging patterns. As a proof of concept, Design and evaluate an information dissemination system that exploits the small-world interest-sharing graphs by building an interest-aware network overlay. This approach leads to improved information dissemination performance. As a result, the efficiency of increasing the cache size is not linear: caching is most efficient for the most popular items, and there are diminishing returns from increasing the cache to provision for less popular items.

Peer-To-Peer (P2P) networks, such as Napster, Gnutella, and BitTorrent, have become essential media for information dissemination and sharing over the Internet. Concerns about privacy, however, have grown with the rapid development of P2P systems. In distributed and decentralized P2P environments, the individual users cannot rely on a trusted and centralized authority, for example, a Certificate Authority (CA) center, for protecting their privacy. Without such trustworthy entities, the P2P users have to hide their identities and behaviors by themselves. Hence, the requirement for anonymity has become increasingly critical for both content requesters and providers.(2)

Peer-to-peer file sharing systems now generate a significant portion of Internet traffic. A good understanding of their workloads is crucial in order to improve their scalability, robustness and performance. Take different measurements: instead of passively recording requests, probe peers to get their cache contents information. This provides with a map of contents that is use to evaluate the degree of clustering in the system, and that could be exploited to improve significantly the search process. (3)

Building upon prior research, we propose several modifications to Gnutella's design that dynamically adapt the overlay topology and the search algorithms in order to accommodate the natural heterogeneity present in most peer-to-peer systems. We test our design through simulations and the results show three to five orders of magnitude improvement in total system capacity. We also report on a prototype implementation and its deployment on a testbed. The peer-to-peer file-sharing revolution started with the introduction of Napster in 1999.(4)

Search is blind in that it is independent of the query and is thus not more effective than probing randomly chosen peers. One technique to improve the effectiveness of blind search is to proactively replicate data. We evaluate and compare different replication strategies and reveal interesting structure: Two very common but very different replication strategies uniform and proportional yield the same average performance on successful queries, and are in fact worse than any replication strategy which lies between them. The optimal strategy lies between the two and can be achieved by simple distributed algorithms. (5)

Centralized-search systems use specialized nodes that maintain an index of the documents available in the P2P system. To find a document, the user queries an index node to identify nodes having documents with the content of interest. These central indices may be built with the cooperation of the nodes or by crawling the P2P network (as in a web search engine). The advantage of a centralized-search mechanism is efficiency (just a single message is needed to resolve a query). However, a centralized system is vulnerable to attack (e.g., index sites can be shut down by a court order or a hacker attack) and it is difficult to keep the indices up-to-date.(6)

To enhance the search performance in unstructured P2P networks through exploiting users' common interest patterns captured within a probability-theoretic frame

work termed the User Interest Model (UIM). A search protocol and a routing table updating protocol are further proposed in order to expedite the search process through self organizing the P2P network into a small world. (7)

To enhance the search performance in unstructured P2P networks through exploiting users' common interest patterns captured within a probability-theoretic framework termed the user interest model (UIM). A search protocol and a routing table updating protocol are further proposed in order to expedite the search process through self organizing the P2P network into a small world. (8)

Developing peer similarity metrics in modern p2p networks is challenging. First, in p2p networks, users do not explicitly rank their preferences, but simply download content,use it and possibly delete it or simply ignore it if they dislike it. This form of implicit ranking makes it difficult to assess whether users “like” or “dislike” the downloaded content. Additionally, there is a large amount of noise that inherently exists in p2p networks, since content is mostly generated and tagged by the users. This results in an abundance of duplicate content with different titles, multiple and even conflicting tagging and spelling mistakes or ambiguities.(9)

III. OUR IDEA

A novel overlay construction algorithm to enhance the efficiency and effectiveness of searches in unstructured P2P networks. This offers the performance guarantees in terms of the hop count of routing a query message and the successful query ratio. On the other hand, while there are studies in matching P2P overlay networks with physical network topologies, its differs from these studies in that we format the P2P topologies such that searches are efficient and effective. Prior studies on maximizing the coverage of a flooding-based search can be found in the literature.while maximizing the covreage of a guery message may not guarantee the efficiency and effectiveness of searches.

In this paper,we first observe that existing P2P file sharing networks(e.g.,eDonkey) exhibit the power-law file sharing pattern. Based on such sharing pattern, we present a novel overlay construction algorithm to enhance the efficiency and effectiveness of searches in unstructured P2P networks with overlay topologies. unique features.

P2P Networks widely deployed in Internet and they provide various services such as File Sharing,Information retrieval,media streaming and telephony.P2P applications are popular because they primarily provide low entry batteries and self scaling.

This proposal has the following properties

- In a constant probability, the search hop count between any two nodes is Oðlnc1 NÞ, where 1 < c1 <2 is a small constant, and N is the number of active peers participating in the network.
- On the search path from the querying peer to the destination peer progressively and effectively exploit their similarities.
- Whereas some prior solutions require centralized servers to help organize the system, our proposal needs no centralized servers to participate in. Unlike most decentralized overlay construction algorithms for enhancing searches in unstructured P2P networks,our solution is mathematically provable and provides performance guarantees.

IV.METRICS

The performance metrics we measure include the following

Query message hop count :

The number of overlay hopcount required for resolving a query. Regarding the hop count, we only measure those queries that can be successfully sent to their resolvers; the hop count of a query that cannot be successfully resolved is 1, otherwise. Possibly, there are multiple overlay paths between the inquirer (peer u) and the resolver (peer v). If so, we measure the shortest one among these paths.This mainly consider the progressive overlay paths.

Successful query ratio :

The percentage of queries that can be successfully routed to its destination in a progressive fashion

The query traffic overhead :

The messages required for successfully resolving a query include

1) messages that are successfully routed to their destination (may be through different overlay paths)

2) those that reach dead ends and cannot be relayed further.

Traffic overhead for maintaining and rewiring the overlay topology :

This denotes the averaged number of messages required for maintaining and rewiring the neighbors of a participating peer.

- Overlay construction algorithms intending to exploit the similarity of peers for enhancing search performance
- Efficient and effective search
- Search performance is guarantee
- Success ratio is 100%
- Can easily manage the traffic

We validate our proposal with simulations. The simulation results reveal that whereas existing overlay formation works that is, the two representative distributed algorithms among, introduce fair traffic overhead to maintain and rewire their overlay topologies, ours clearly outperforms existing method in terms of

1. The query message hop count,
2. The successful ratio of resolving a query,
3. The query traffic overhead, and
4. The overlay maintenance overhead.

Moreover, we find that together with a similarity-aware overlay topology, the search protocol we have suggested in this paper, which takes advantage of the similarity of peers exploited by our overlay network, can considerably reduce the search traffic. Peers participating in a P2P network are often heterogeneous in terms of their network bandwidth, storage space, and/or computational capability. It would be interesting for our future work to investigate how the heterogeneity affects our proposal. Moreover, the overlay formation algorithm presented in this paper is oblivious to the physical network topology, and this may introduce considerable wide-area network traffic. It would be challenging to design an overlay formation algorithm aware of both the similarity of participating peers and the physical network topology.

V. IMPLEMENTATION

Propoed system satisfy the following properties:

C1. High Clustering

Each peer u connects maxu peersin V , and these neighbors, selected among the peers in V , are the top-maxu nodes most similar to v.

K MEANS ALGORITHM

- Initially, the number of clusters must be known, or chosen, to be K say.
- The initial step is the choose a set of K instances as centres of the clusters. Often chosen such that the points are mutually “farthest apart”, in some way.
- Next, the algorithm considers each instance and assigns it to the cluster which is closest.
- The cluster centroids are recalculated either after each instance assignment, or after the whole cycle of re-assignments.

This process is iterated.

C2. Low Diameter

Consider any two distinct peers u and v in V . There should exist at least one overlay path P connecting u and v, and the hop count of P should be as small as possible, enabling a query message to be rapidly propagated from u to v. Here,
the hop count of an overlay path P means the number of overlay links in P.

C3. Progressive

Let s be the peer that issues a query, and d be the peer that can resolve the query. There should exist an overlay path P connecting s and d such that for any two neighboring peers u and v on P, upon receiving a query message, u forward the message to v that is more similar to d than u.

PSEUDOCODE FOR K-MEANS ALGORITHM

X: a set of N data vectors
C_I: Data set (initialized k cluster centroids)
C: Number of clusters.

$P = \{p(i) \mid i = 1, \ldots, N\}$ is the cluster *label of X*

KMEANS(X, C_I) → (C, P)
REPEAT
$C_{previous} \leftarrow C_I$;
FOR all $i \in [1, N]$
DO
Generate new optimal paritions
$p(i) \leftarrow \arg\min d(x_i, c_j)$;
$l \le j \le k$
FOR all $j \in [1, k]$
DO
Generate optimal centroids
$c_j \leftarrow$ Average of x_i, whose $p(i) = j$;
UNTIL $C = C_{previous}$

VI. CONCLUSION

We have presented an unstructured P2P network with rigorous performance guarantees to enhance search efficiency and effectiveness. In a constant probability, a queryin g peer takes O ($\ln^n$ N) hops (where c is a small constant) to reach the destination node capable of resolving the query, whereas the query messages can progressively and effectively exploit the similarity of the peers. The query can be successfully resolved in an approximate probability of 100 percent. Notably, the theoretical analysis further reveals that the competitive decentralized solutions do not perform well as the hop count of routing a query message in such networks, considering the exploitation of the similarity of participating peers, is in the polynomial of system size N.

REFERENCES

[1] Iamnitchi, M. Ripeanu, E. Santos-Neto, and I. Foster, “The Small World of File Sharing,” IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 7, pp. 1120-1134, http://doi.ieeecomputersociety.org/ 10.1109/TPDS.2010.170, July 2011.

[2] Y. Liu, J. Han, and J. Wang, “Rumor Riding: Anonymizing Unstructured Peer-to Peer Systems,” IEEE Trans. Parallel and Distributed Systems, vol. 22, no. 3, pp. 464-475, Mar. 2011

[3] H. Chen, H. Jin, Y. Liu, and L.M. Ni, “Difficulty-Aware Hybrid Search in Peer-to-Peer Networks,” IEEE Trans. Parallel and Distributed Systems, vol. 20, no. 1, pp. 71-82, Jan. 2009.

[4] F.L. Fessant, S.B. Handurukande, A.-M. Kermarrec, and L. Massoulie´, “Clustering in Peer-to-Peer File Sharing Workloads,” Proc. Third Int’l Workshop Peer-to-Peer Systems (IPTPS ’04), pp. 217- 226, Feb. 2004.

[5] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker, “Making Gnutella-like P2P Systems Scalable,” Proc. ACM SIGCOMM ’03, pp. 407-418, Aug. 2003.

[6] E. Cohen and S. Shenker, “Replication Strategies in Unstructured Peer-to-Peer Networks,” Proc. ACM SIGCOMM ’02, pp. 177-190, Aug. 2002.

[7] Crespo and H. Garcia-Molina, “Routing Indices for Peer-to- Peer Systems,” Proc. 22th IEEE Int’l Conf. Distributed Computing Systems (ICDCS ’02), pp. 23-32, July 2002.

[8] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, “Search and Replication in Unstructured Peer-to-Peer Networks,” Proc. ACM Int’l Conf. Supercomputing (ICS ’02), pp. 84-95, June 2002.

[9] Anusuya .R, Dr.Kavitha. V, Mrs.Golden Julie .E ”Enhancing and Analyzing Search Performance in Unstructured Peer to Peer Networks using Enhanced Guided Search Protocol” Journal Of Computing, Volume 2, Issue 6, June 2010, Issn 2151-9617 https://sites.google.com/site/journalofcomputing/ www.journalofcomputing.ORG

[10] Yuval Shavitt, Ela Weinsberg, Udi Weinsberg “Estimating Peer Similarity using Distance of Shared Files” Tel-Aviv University, Israel{shavitt,ela,udiw}@ eng.tau.ac.il