A COMPARATIVE STUDY ON OPTIMIZATION OF SEARCH IN OVERLAY NETWORKS

C.Priyanka[1], Dr.K.Deeba[2]

[1]*PG Scholar, Department of Computer Science and Engineering, chinnaduraipriyanka@gmail.com*
[2]*Associate Professor, Department of Computer Science and Engineering, deeba.senthil@gmail.com*

Abstract

Overlay Networks both in structured and unstructured environments have a massive growth in today's market providing support of several applications. These networks can be erected very competently without any distinct rules and therefore considered suitable to the Internet environment. But, it is oblivious that searching techniques that are proposed usually perform poorly when the size of network is large network and ignorant to the physical network topology that also introduces wide-area network traffic. To enhance the search in overlay network different approaches of search patterns are provided in prior. These searches are done with various topologies in order to further expedite the search progress through self-organizing the P2P network into a small world. Both theoretical and experimental results on optimization on various search approaches are studied also effectiveness and efficiency of various approaches in comparison are discussed.

Keywords—**Overlay Networks; Optimization; Peer topology; Search**

1. Introduction

   Structured and unstructured overlay networks are attractive for certain applications because they require no centralized directories and no precise control over network topology. This Network provides various services such as file sharing, instantaneous messaging and other popular network applications that rely on P2P technology. This growth of such applications introduces internet traffic up to 20 per cent as in prior studies [1] can have significant impact on the underlying network. However, searching an object is an essential part in several applications where in the popular protocols as Gnutella which are unstructured and peers participating interconnect randomly through message flooding by enquiring a peer and broadcasts the message to its neighbours that have end-to-end connections. The flooding based query algorithm which is used in Gnutella does not scale; where each query generates a large amount of trace and large systems quickly become overawed by the query persuaded load. This type of searching an object takes place in most of the overlay networks.
    In this study we are interested in enhancing the search performance in both structured and unstructured Peer to Peer Networks. Existing methodologies includes indexing [5][6][7], clustering [1], replications [8] and other overlay topologies improves search performance of Unstructured Peer to Peer networks. For unstructured environments such as Semantic Overlay Networks (SONs), peers that are socially similar are *organised* into groups. These while being highly flexible, improve query performance and promise high degree of peer autonomy [3]. Additionally, this method offers better support of semantics due to their ability to provide mechanisms for estimated range, or text queries, and highlight peer independence. Rather than this type of search topology another approach given to the search enhancement was *RefNet* method, which focused on the effect of network clustering that, finds ability of relevant information sources of the decentralized search [4].
    In highly dynamic P2P systems, when a selected peer leaves, the whole path fails. Unfortunately, such a collapse is often difficult to be known by the initiator. Therefore, a randomly-assigned path is very variable, and users have to frequently inquiry the path and retransmit messages, to address this issue a non-path-based

anonymous light weight P2P protocol called Rumor Riding (RR) was anticipated. Where RR provides a high degree of anonymity and outperforms existing approaches in terms of reducing the traffic overhead and processing latency. The overall idea of this paper is that the statistical patterns over locally shared resources of a peer can be explored to guide the distributed resource discovery process and therefore enhance the overall resource discovery performance in unstructured peer to peer networks. In this paper, we primarily focus our study on the overlay topology formation techniques for both structured and unstructured P2P networks, aiming to enhance search efficiency and effectiveness on optimization.

2. Related Study on Structured and Unstructured Networks

In this section we primarily discuss about P2P networks by different generations. In first generation early P2P networks, such as Napster and Gnutella, aim to deploy easily and quickly. They are too simple, but have bad scalability or low query efficiency. In Second generation P2P networks utilizes the DHT technique to obtain better scalability and query efficiency, and provide load balancing and deterministic search guarantees. But the resilience or fault-tolerance is not good, especially under malicious attacks. In recent generation the contribution aims in providing high resilience assuming the node will be down with some failure probability.

Other popular methods on optimizing the search performance where pSearch[6] and SSW[7] which are content based P2P networks providing semantic search. Similar to most overlay networks based on distributed hash tables (e.g., Chord[8]) in pSearch and SSW, the peers those participating need to maintain foreign indices, that is, the indices of the object stored in the distant peers. To locate an object, a requesting peer routes a message toward peer responsible for the key subspace where the object is indexed. In contrast, we can construct unstructured peer to peer networks where the participating peers need not to organize themselves into a secure, deterministic topology group, considerably reducing the maintenance overhead of overlay topology. Unlike pSearch [6] and SSW[7], the peers in the network hosts the objects of interest and maintain no foreign index, eliminating storage bandwidth overheads for publishing and managing such index. However, due to space complexity, we focus on optimizing searching protocol. While maximizing the coverage of a query message may not guarantee the efficiency and efficiency of searches.

In an unstructured peer to peer network environment, the concept of anonymity, many approaches have been proposed to support anonymous communications. The prior methods that are proposed in second generation such as Onion Routing, Tor, APFS, and Shortcut Protocol provides P2P mutual anonymity are with a reduced response delay [2]. When receiving a packet, a peer has two options as such forwarding the packet to a randomly chosen peer or directly sending it to the destination peer. All these above methods have different mechanisms focusing on the peer search improvement and managing the overlay topology by different methods.

3. Object Search Algorithms

Searching an object in Peer networks is the primary key component where it nearly determines the total performance of network. The searching algorithms based on query search are of random and deterministic, flooding and non-flooding, centralized and de-centralized. In random method based on unstructured as *Gnutella* queries are forwarded to a randomly selected neighbour node improve the query resilience compared to the deterministic unique query, but it will have long query path length. Compared to random query, non-random query choose the next-hop deterministically, which is usually used in structured P2P networks such as 2G DHT-based P2P networks [16]. In this method the query can also be classified into unique-query and multi-query, flooding-based query is a special form of multiple-query. Prior efforts in developing distributed overlay formation algorithms for semantic small-world P2P networks where any peer in these proposals often selects its

neighbours heuristically. All these approaches are practical for large-scale distributed environments but vary in their performance investigation. In recent we the focus is similarity based clustering methods where highly suited for unstructured environments.

Consider any given unstructured P2P network G = (V, E), where V is the set of peers that are participating, and E is the set of overlay connections linking the peers in V. The peers in G may be interrelated randomly. Where G should satisfy the following strategies to exploit the peer similarity such as:

a. *High Clustering*:
   Any peer in V should connect *max* peer in V and the neighbours are selected in top most cadets of V.
b. *Low Diameter*:
   Consider two different peers in V, such that both peers should have at least one overlay path existing between them, with the hop count almost small as possible.
c. *Progressive*:
   There should be a path existing between the peer who issues the query and the peer which resolves the query.

The above a, b, c should be maintained effectively in the following overlay search topologies. Exploiting the peer similarity is taken by;

A. *Definition:*

Let V be the set of peers participating in an Overlay network. The peer semblance function measures the degree of the similarity between any two peers, u$\in$V and v$\in$V in the system where u and v is any two peers that participate.

$$F: V \ X \ V \rightarrow IR()+$$

{a,b,c}   {b,c}
3 ——2/3—— 4

1/3   1/2

1 ——1/2—— 2
{a}   {a,c}

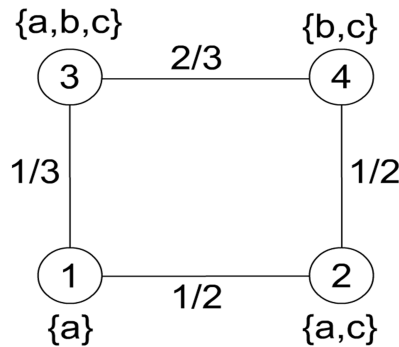**Fig. 1 Peer Similarity Graph**

Consider Figure 1 as Graph G = (V, E), where V={1,2,3,4}. Peers 1,2,3 and 4, respectively, host set of objects O1={a}, O2={a,c}, O3={a,b,c}, O4={b,c}. Any two peers u and v have an edge in E if both peers share at least one common object. That is, $F(u,v) = \left| O_u \cap O_v / O_u \cup O_v \right|$, the value nearby an edge (u,v) indicates F(u,v). F(u ,v) is defined as the inverse of the cosine angle of two summarized latent semantic vectors representing any two peers u and v in a P2P network, where every element in a summarized vector for any peer say i calculates the total frequency of the corresponding keyword appearing in the data items stored in i. The following table debits the notations that are frequently used in exploiting the peer similarity.

TABLE I

NOTATIONS FREQUENTLY USED

NOTATIONS	DESCRIPTION
V	Set of participating peers
G=(V,E)	Peer similarity graph
F(u,v)	Similarity function measuring the level similarity between vertices u and v

This type above mentioned similarity function is used for search [1].

4. P2P Comparative Study with Enhanced Search Protocol

In this section various methods that are used previously for search protocol with various topologies are discussed. Chord, Napster, Guntella are some of the famous search protocols in market. In *Chord* [8], the problem in peer to peer systems is opposed by efficiently locating the node that stores a particular data item. Chord adapts efficiently as nodes join and leave the system, and can respond to queries even if the system is unceasingly changing. The performance investigation shows that Chord scales well with the number of nodes, recovers from large numbers of instantaneous node failures and joins, and answers most lookups correctly even during recovery. The other popular search protocols such as *Napster* and *Gnutella* are compared on the basis of their characteristics such that, there is significant heterogeneity and lack of cooperation across peers participating in these systems [9].

Flooding is an essential building block of unstructured peer-to-peer (P2P) systems. To improve the performance from flooding a novel method of *Clustella*, a novel semi-structured P2P architecture with bounded peer degree[10]. It decomposes the network into different clusters, allowing peers to quickly find those neighbours which contribute much to their routing efficiency both in static and dynamic environments. Later another method to overcome the flooding was Rumor Riding (RR), a lightweight and non-path-based mutual anonymity protocol for decentralized P2P systems. Peers participating in unstructured networks interconnect randomly, rely on flooding_query messages to discover objects of interest introduces Network traffic uses a sower to distribute and redistribute the query. The performance analysis shows that RR provides a high degree of anonymity and outperforms existing approaches in terms of reducing the traffic overhead and processing latency [2]. In peer to peer next section upon content based retrieval methods like pSearch, SSW where the next problem raised. pSearch distributes document indices through the P2P network based on document semantics generated by Latent Semantic Indexing (LSI) [11].

In order to provide an enhanced support to this type of peer environment for both structured and unstructured environment the balancing of peer search topology can be studied. The major study is on research of structured peer to peer systems where the nodes are often being homogenous and sometimes to be heterogeneous. While typically there are several load balancing schemes have been proposed are typically ad hoc, heuristic based, and localized. For a structured network HiGLOB, for global load balancing in structured P2P systems was proposed. Each node contains a histogram manager which maintains a histogram that reflects a global view of the distribution of the load in the system also load-balancing manager redistributes the load whenever the node becomes overloaded or under loaded. This outperforms *Skip Graph, BATON,* and *Chord* in balancing the peers [15]. *Pastry*[12] another method where it can efficiently support a variety of peer-to-peer applications. Being decentralized it is widely used for large-scale peer-to-peer systems which out performs Napster, Gnutella and also FreeNet.

Most P2P networks based on distributed hash tables where, different problem on balancing load in p2p systems can be solved by providing a DHT abstraction distribute objects among "peer nodes" by choosing random identifiers for the objects. Designing load-balancing algorithms that uses the notion of "virtual servers" is able to balance the load within 80% of the perfect value, while the most complex scheme is able to balance the load within 95% of the optimal value [13]. In an active environment being the nodes to be *heterogeneous,* of the nodes varying rapidly an algorithm was presented. The simulation results show that in face of rapid arrivals and departures of objects of widely changing the load, load balancing for system utilizations as high as 90% while moving only about 8% of the load that arrives into the system [14]. Similarly many such load balancing techniques can be studied such imposes the enhancement of peer to peer network. Those of which simulation

results can be compared with the existing search topologies where comparative analysis of different search method.

Here certain method of which the load balancing can be done are discussed. The P2P networks can also be based on maintaining the balance information of peers in Routing table. The differences in both tables are such DHT offers security than Routing table. In concern of search enhancement alone the efficiency is better rated in routing table than DHT. A comparative analysis can be done on both efficiency rating between both maintaining balance information on DHT and routing table.

5. Conclusions and Future Work

In this paper the study of Overlay network is been given where different methods that have been proposed in prior with their performance comparison on basis of how efficiently the query for search protocol in both the structured and unstructured environment has been discussed. Peer environments being decentralized are been used in rapid growth of development of various applications. The comparative analysis of different strategies in peer environment would be rather interesting in designing various topologies.

Our Future work is on the contribution of a topology that is entirely applicable for both structured and unstructured environments. It is also interesting in maintaining heterogeneity throughout supporting dynamic peers. Further the wide area network traffic, storage space, and computational capability that occur in various clustering based search technique is also been studied.

References

[1]. Hung-Chang Hsiao and Hong-Wei Su "On Optimizing Overlay Topologies for Search in Unstructured Peer-to-Peer Networks", IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 5, May 2012.

[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]. P. Raftopoulou, E.G.M. Petrakis, and C. Tryfonopoulos, "Rewiring Strategies for Semantic Overlay Networks," Distributed and Parallel Databases, vol. 26, nos. 2/3, pp. 181-205, Dec. 2009.

[4]. W. Ke and J. Mostafa, "Strong Ties versus Weak Ties: Studying the Clustering Paradox for Decentralized Search," Proc. Seventh Int'l Workshop Large-Scale Distributed Systems for Information Retrieval (LSDS-IR '09), pp. 49-56, July 2009.

[5]. S. Voulgaris, M. Steen, and K. Iwanicki, "Proactive Gossip-Based Management of Semantic Overlay Networks," Concurrency and Computation: Practice and Experience, vol. 19, no. 17, pp. 2299-2311, Dec. 2007.

[6]. Tang, Z. Xu, and S. Dwarkadas, "Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks," Proc. ACM SIGCOMM '03, pp. 175-186, Aug. 2003.

[7]. M. Li, W.-C. Lee, A. Sivasubramaniam, and J. Zhao, "SSW: A Small World-Based Overlay for Peer-to-Peer Search," IEEE Trans. Parallel and Distributed Systems, vol. 19, no. 6, pp. 735-749, June 2008.

[8]. 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. Networking, vol.11,no.1,pp.17-21,Feb.2003.

[9]. S. Saroiu, P.K. Gummadi, and S.D. Gribble, "Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts," Multimedia Systems, vol. 9, pp. 170-184, Aug. 2003.

[10].S. Schmid and R. Wattenhofer, "Structuring Unstructured Peer-to- Peer Networks," Proc. Int'l Conf. High Performance Computing (HiPC '07), pp. 432-442, Dec. 2007.

[11].Tang, Z. Xu, and S. Dwarkadas, "Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks," Proc. ACM SIGCOMM '03, pp. 175-186, Aug. 2003.

[12].Rowstron and P. Druschel, "Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems," Proc. IFIP/ACM Int'l Conf. Distributed Systems Platforms, pp. 161-172, Nov. 2001.

[13].Rao, K. Lakshminarayanan, S. Surana, R. Karp, and I. Stoica, "Load Balancing in Structured P2P Systems," Proc. Second Int'l Workshop Peer-to-Peer Systems (IPTPS '02), pp. 68-79, Feb. 2003.

[14].S. Surana, B. Godfrey, K. Lakshminarayanan, R. Karp, and I. Stoica, "Load Balancing in Dynamic Structured P2P Systems," Performance Evaluation, vol. 63, no. 6, pp. 217-240, Mar. 2006.

[15].Q.H. Vu, B.C. Ooi, M. Rinard, and K.-L. Tan, "Histogram-Based Global Load Balancing in Structured Peer-to-Peer Systems," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 4, pp. 595-608, Apr.2009.

[16]. Chonggang Wang Bo Li, " Peer-to-Peer Overlay Networks: A Survey" Department of Computer Science The Hong Kong University of Science and Technology, April 2003.