An Intelligent Overlay Search in UP2P

D. Sangeetha, M. E.,
Lecturer,
Department of Information Technology,
Sri Muthukumaran Institute of Technology,
Chennai-600 069.
sangeethabaskaran242524@yahoo.com

R. Lalitha, M. E., Ph. D.,
Assistant Professor,
Department of Information Technology,
Sri Muthukumaran Institute of Technology,
Chennai- 600 069.
lalitha.sadha@gmail.com

ABSTRACT

P2P is a fascinating and extraordinary architectural model for developing Internet applications. These systems impose a little overhead in the efficient retrieval of scattered information. Many peer-to-peer networks are overlay networks because they run on top of the Internet. Here we throw insight into the effects of topologically aware overlay construction technique on P2P search algorithms. Here we propose an innovative decentralized Unstructured P2P architecture which comprises of Unstructured Domain Overlay for the construction of overlay network and Sharp Overlay for efficient search and information retrieval on peer-to-peer networks. Moreover to ease the task of search in Unstructured P2P applications like Gnutella, Sharp Overlay uses an efficient overlay query routing with less overhead which uses local knowledge in a decentralized environment. A better latency is achieved by this technique which can be evaluated by the recall rate and aggregate delays. The designed overlay can be deployed in a real time University environment. Here we aim at finding the documents in a unstructured P2P using distributed interestingness search performed on an Unstructured Domain overlay.

In a decentralised view, retrieval and search of documents should be done with less resource consumption and reduced latency rate, when a set of query terms are given.

In this paper we investigate a fully distributed technique for addressing the information retrieval problem in pure P2P networks. We present the following section of proposals as follows:

- An Unstructured Domain Overlay, which is an integrated distributed architecture to perform an efficient interestingness search in an overlay network.
- An overlay Query routing provides an efficient searching mechanism in UP2P using a new approach called Sharp Hunter. It is an interestingness search algorithm which is efficient, scalable yet simple mechanism for improving the information retrieval problem.

Our objective is to construct an efficient Unstructured P2P networks and a simple search mechanism, which improves the latency with accurate results.

In this paper, we can deploy the overlay domain with interestingness search in a real time University environment. Here different domains may be the different Universities. The following sections will explain the construction of overlay domain following the search based on interestingness of the user in the particular University.

1 INTRODUCTION

In a P2P system, all participating peers form a P2P network on top of an underlying physical network. A P2P network is an abstract, logical network called an overlay network [3]. The P2P overlay network consists of all the participating peers as network nodes[9]. An unstructured P2P network is formed when the overlay links are established arbitrarily. In UP2P no global structure is maintained. Messages are sent over multiple hops from one peer to another peer. The peers can join and leave the network dynamically and hence the name Unstructured P2P.

The heterogeneity of UP2P [11], which affects the performance of the information retrieval algorithm, should be improved by using better search algorithm to reduce latency.

Here we aim at finding the documents in a unstructured P2P using distributed interestingness search performed on an Unstructured Domain overlay.

In a decentralised view, retrieval and search of documents should be done with less resource consumption and reduced latency rate, when a set of query terms are given.

In this paper we investigate a fully distributed technique for addressing the information retrieval problem in pure P2P networks. We present the following section of proposals as follows:

- An Unstructured Domain Overlay, which is an integrated distributed architecture to perform an efficient interestingness search in an overlay network.
- An overlay Query routing provides an efficient searching mechanism in UP2P using a new approach called Sharp Hunter. It is an interestingness search algorithm which is efficient, scalable yet simple mechanism for improving the information retrieval problem.

Our objective is to construct an efficient Unstructured P2P networks and a simple search mechanism, which improves the latency with accurate results.

In this paper, we can deploy the overlay domain with interestingness search in a real time University environment. Here different domains may be the different Universities. The following sections will explain the construction of overlay domain following the search based on interestingness of the user in the particular University.

2 PROPOSED SYSTEM

In Unstructured P2P, peers or nodes join and leave the network without affecting system performance. Here, we propose a Unstructured Overlay P2P architecture which is decentralized with no global component. The foresaid construct can be
deployed in a University environment with various Universities. It consists of the following modules:

2. Intelligent Overlay Search

Here the Universities in different countries act as separate domains where the various users in the domain involve in file sharing. We aim at finding documents in such a decentralized environment with dynamic users joining and leaving the University P2P network.

### 2.1 Unstructured Domain Overlay

Our proposed system consists of a core decentralized overlay domain that forms the framework for the construction, routing and search in a unstructured peer-to-peer networks.

In our example, our real time environment consists of various Universities. Here we organize a particular University in a country into separate domains.

In order to build an Unstructured Domain Overlay, we need to consider the following tasks:

- Overlay Domain Construction
- Joining a Domain overlay.
- Active cache

Which is explained in detail in the following sections.

### 2.1.1 Overlay Domain Construction

This module clusters nodes together without any centralized component differing from the existing systems [3]. Clustering close by nodes together forms a domain. Here peers in the same domain are connected to each other by making dual connection parameters. One half of the connections are said to be in low latency links called L connectivity and remaining as R connectivity. L connectivity are low latency connections, connecting nodes in the same domain they follow the connections by making a fixed fraction (0.5). R connectivity is used to make the peer graph to remain connected. The following structure illustrates how the similar domains are clustered together. Each peer maintains its local domain specific information in a structure called Active Cache.

In our example, various Universities are organized into separate domains i.e., University in same country will form a cluster.

### 2.1.2 Joining an Overlay

Let n be the new node that wants to join an overlay network N. The new node joining the overlay has to decide which domain it has to join, in order to participate in the overlay networks. In order to locate the domain, a data structure called active cache, which contains domain specific information, is maintained at each peer in the various domains separately.

After determining the domain, the new node has to decide to which peer it has to establish connection in a specific domain. The following steps are followed:

- The new peer n establishes connections to the d/2 random nodes in a particular domain. Where ‘d’ is the degree of the node.
- If the new peer n is the first peer in that particular domain it can join with a peer, which belongs to another domain and closer to this domain.

In a University environment, when a new user needs to join a particular domain, the user access the active cache to identify Domain and then identifies the particular user or peer in the domain who have high degree of correspondence thus joining the overlay.
2.1.3 ACTIVE CACHE

A cache guides the new peers to join the overlay, which is maintained at each peer. It forms a data structure with the following fields:

1. **Name**: Name of the domain, for example any University.
2. **Correspondence**: No. of friends or degree of association, for example, no. of colleagues in good terms.
3. **Area of interest**: Interestingness measure to indicate the user preference.
4. **IP address**: IP Address of the participating peer.
5. **Port no**: Port of communication of the participating peer.

<table>
<thead>
<tr>
<th>Name</th>
<th>Correspondence</th>
<th>Area of interest</th>
<th>IP address</th>
<th>Port no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leena</td>
<td>Mary.geo</td>
<td>network</td>
<td>192.168.1.22</td>
<td>54</td>
</tr>
<tr>
<td>Sarah</td>
<td>Sita.gita</td>
<td>signals</td>
<td>192.181.24</td>
<td>56</td>
</tr>
</tbody>
</table>

When a node disconnects, the network re-establishes the dropped connection by consulting the Active cache to maintain a resilient network.

3 OVERLAY QUERY ROUTING WITH SEARCH

The existing query routing [17] includes BFS in Gnutella where each query is associated with TTL [9], which determines the maximum number of hops that a query can be forwarded.

The proposed query routing in Unstructured Domain Overlay uses an intelligent yet interesting search called Intelligent overlay search based on the interestingness measure namely the user preferences.

The proposed Intelligent Overlay search uses an innovative search mechanism by including config data and priority detector, which determine the order, or route in which the query for search is routed to find the absolute document. The following figure shows how the search using sharp hunter is done with a profile details of query stored in a structure called config data.

3.1. INTELLIGENT OVERLAY SEARCH

It achieves reduced messaging by having each peer profile the query/query hit activity of its neighboring nodes. It then knowledge to forward queries to the neighbors that are most likely going to reply to a given query. It consists of two components that run locally in each peer.

1. Config data.
2. Priority Detector (PD)

3.1.1. CONFIG DATA

Each node maintains in a repository the T most recent queries and the corresponding query hits along with the number of results. Once the repository is full, the node uses the LRU replacement policy to keep the most recent queries.

3.1.2. PRIORITY DETECTOR (PD)

This is a function used by a node to perform an order of priority of neighbors in order to determine to which ones to forward a query q. To compute the priority order of each peer pi, pk compares q to all queries in the config data structure, for which there is a query hit, and PD allows us to prioritize the peer that returns more results. Based on the priority determined, the query is routed onto those high priority peers and thus a optimized low latency route is formed which aids the query node to the destination. Thus the shortest path is determined on through low latency links and the search is accomplished with fewer overheads.

6. MERITS OF PROPOSED SYSTEM

The proposed P2P Domain overlay and the Intelligent Overlay search has the following merits:

- No centralized component is used.
• Latency overhead is less since the entire environment is decentralized.
• Backup and recovery is feasible.
• The entire P2P architecture is scalable and reliable.
• Since the search follows a priority function to find the next forwarding peer, the best routing path is chosen.
• Config data is maintained locally which plays a vital role in maintaining the local peer information to give good search results.

CONCLUSION
Unstructured P2P architecture with domain specific overlay is constructed in an absolutely decentralized environment with no global component. Active cache maintained at each peer provides a local pool of access in a diversified environment. It helps our example namely the University to be decentralized such that the various domain specific sharable files can be shared and accessed by a participating peer at their venue without the involvement of centralized server. Above all, the overhead due to centralized server where all the queries during search orients on the centralized component is completely eliminated. Moreover the search is based on user interestingness thereby giving good related documents by using a priority function during routing.

FUTURE SCOPE
The overlay query routing along with search can be done in a multilingual environment so that it can be an extended multilingual search. Also the query routing performance can be enhanced to a greater extent so that it can improve the p2p network performance.

REFERENCES