A Selective Neighbor Search Algorithm in Unstructured Peer-to-Peer Networks

Sujeeth .T
Dept. of Computer Science &Engineering
Sri Venkateswara University,Tirupati
Andhra Pradesh ,India

Mahesh Kumar KondReddy
Dept. of Computer Science &Engineering
Sri Venkateswara University,Tirupati
Andhra Pradesh, India

Kiran Kumar Reddy.D
Dept of computer Science &Engineering
Sree Venkatesa Perumal College of Engineering & Technology
Andhra Pradesh, India

Abstract –Searching for a resource in an unstructured P2P Network is a challenging task and well studied problem in the literature. Many search algorithms are proposed based on Flooding and Random Walk strategies. Flooding based algorithms search aggressively by forwarding query to all the neighbor nodes while Random Walk forwards query selectively to one of its neighbors. Both approaches, suffer from drawbacks of large query hits , search delays respectively. In this paper we propose a novel approach based on flooding, which forwards the query request only to its effective nodes. Whenever a node joins/leave the network the effective nodes are calculated and for its neighbors which are maintained in a table data structure internally.

Keywords: Peer-to-Peer, search, performance, unstructured

I .INTRODUCTION

In order to meet the needs of large-scale network applications, integrate and collaborate with the rich resources, the model of the Internet system have changed from the Client/Server to Peer-to-Peer(P2p)[6]. A peer-to-peer (P2P) network is a distributed system in which peers employ distributed resources to perform a critical function in a decentralized fashion. Nodes in a P2P network normally play equal roles; therefore, these nodes are also called peers. A distributed architecture consisting of a collection of resources (computing power, data, meta-data, and network bandwidth) performing a distributed function is called a peer to peer architecture. P2P computing is the sharing of computer resources and services by direct exchange between systems P2P network is built on the overlay of internet. The huge popularity of recent peer-to-peer (P2P) file sharing systems has been mainly driven by the scalability of their architectures and the flexibility of their search facilities. Such systems are usually designed as unstructured P2P networks, because they impose few constraints on topology and data placement. Each node in P2P network is equal logically, which is both a provider and recipient of resources.

In an unstructured P2P network[1] such as Gnutella[9], no rule exists which defines where data is stored and the network topology is arbitrary. In a loosely structured network such as Free net and Symphony the overlay structure and the data location are not precisely determined. In Free net, both the overlay topology and the data location are determined based on hints. The network topology eventually evolves into some intended structure. In Symphony, the overlay topology is determined probabilistically but the data location is defined precisely. In a highly structured P2P network such as Chord, both the network architecture and the data placement are precisely specified. The neighbors of a node are well-defined. The data is stored in a well defined location.

In unstructured peer-to-peer (P2P) networks[5],[10], each node does not have global information about the whole topology and the location of queried resources. P2P participants join or leave the P2P system frequently, hence,P2P networks are dynamic in nature. P2P networks are overlay networks, where nodes are end systems in

the Internet and maintain information about a set of other nodes (called neighbors) in the P2P layer. These nodes form a virtual overlay network on top of the Internet. Each link in a P2P overlay corresponds to a sequence of physical links in the underlying network. Unstructured p2p Networks are more flexible in that there is no need to maintain special network structure, and they can easily support complex queries like keyword/full text search[7]. In unstructured networks such as nodes are interconnected in a random manner, which offers high resilience for the whole network to be tolerable to frequent node leaves/joins/failure[4].Its well-known characteristics offer enormous potential benefits for search capabilities powerful in terms of scalability, efficiency and dynamics. The routing efficiency in a unstructured Peer-to-Peer Network is low because a large number of peer nodes have to be visited during the search process. Also, query requests are repeated across the nodes with loops.

II. LITERATURE SURVEY

Flooding and Random Walk are two typical examples of blind search algorithms by which query messages are forwarded to neighbors without any knowledge of the possible locations of the queried resources or any preference for the directions to send. Modified versions of Blind Flooding[10],[8] are proposed in the literature where the algorithms try to modify the operation of flooding to improve the efficiency by reducing the redundant messages . Efficient search strategies by searching the content based on its type of content is calculated using a correlation or similarity index . Some examples are adaptive probabilistic search (APS), biased RW routing index (RI) local indices, and intelligent search . APS builds the knowledge with respect to each file based on the past experiences. RI classifies each document into some thematic categories and forwards query messages more intelligently based on the categories. Each node collects the file indices of peers within its predefined radius. If a search request is out of a node's knowledge, this node would perform a flooding search. The intelligent search uses a function to compute the similarity between a search query and recently answered requests. Nodes relay query messages based on the similarity. There are some other research works that focus on replicating a reference pointer to queried resources in order to improve the search time .

A k-walker random walk algorithm[4] query messages are called walker and this message is forwarded node to node. Each node propagates the query to K randomly selected neighbor nodes. Each walker periodically checks the original requester, if query is satisfied then terminate searching or until the query propagates a predefined, maximum number of hops away from its source (Time-To-Live or TTL). The approach obviously can reduce the number of messages required for each query at the cost of reduced network coverage. Adaptive Probabilistic Search (APS)[3] is based on the kwalk for object search, but the forwarding process is probabilistic instead of random. In APS each peer effectively direct walkers using feedback from previous searches, while keeping information only about their neighbors.

III. PROPOSED ALGORITHM

The proposed algorithm searches for the queried resource by forwarding the query requests only to its effective neighbors which are calculated and maintained in a table data structure by each node . The effective neighbours of a node are calculated w.r.t the node that have been forwarded the query request. Each entry in the table will have a form <Query Originating Node, list of effective nodes>.When search query is received ,the source node is matched with the Query Originating Node in the table and the query is forwarded to list of effective nodes specified in the matched entry.

Let N(x) denote the set of neighbors of given node x , N(x) ={x1, x2, x3....xn} Effective Nodes of xi is obtained by calculating N(xi) - U j=i+1.to .n N(xj).

x1 is called as don't care node for Node x if, N(x1) - U i=2..n N(xi)=

The below procedure is called whenever any node in the network either joins or leave the network , the entries of effective node tables are updated in the respective nodes accordingly.

Selective Neighbor Forward Flooding Algorithm(SNF)

In Flooding method when a query is received the query is forwarded to forwarded to all its neighbors except to the originating code which causes overwhelming traffic in the network. In the proposed method query is forwarded to only selected nodes at each hop. There are two types of query messages used while searching a resource

Query-Resource This message only queries the node for the existence of resource. If the node contains the requested resource it simply the resource to the queried node. If it doesn't contain the resource then it ignores the search query.

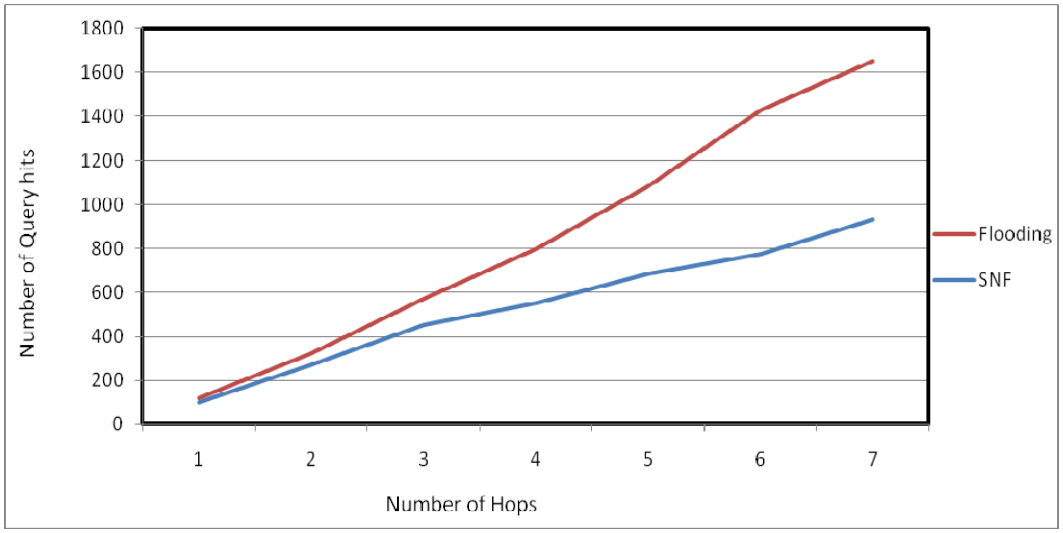
Query-Resource-Forward: This message queries the node for the existence of a resource. If the node doesn't contain the requested resource it simply forwards the Query-Request to its neighbors .

A Query-Resource-Forward message is sent only to its effective neighbors. A Query-Resource message is sent to its don't care nodes.

IV. EXPERIMENT RESULTS

Simulation environment consists of 1000 nodes. Each node is connected to 2-15 nodes with an average degree of 5 for each node. A randomly chosen peer will generate search query for a TTL value of 7.

```
Procedure configure Network(Node:N)
begin
    for each neighbour X of N
     begin
        T ← number of neighbours of X
          for   i←0 to  T
            begin
                for   j←i+1 to T
                 begin
                   node_List← node_List U{N(x_i)-N(x_j)}
                  end
                   node_List←node_List - {Neighbours of X}
           Add Entry  <  N(x_i),node_List  >  in the Effective Node Table of  X_i
        end
end
```



The number of query hits generated at each node at each hop is plotted in the below .The proposed Selective Neighbor Forward Flooding method generates less number of query hits compared with flooding, especially when the number of hops have been increased.

V. CONCLUSION

In this paper we propose a novel search algorithm based on Flooding which forwards the query requests to selective neighbor nodes only. The proposed algorithm reduces the total search hits, redundant query messages and search time compared to Flooding and Random Walk search methods. The proposed method performs better even when the number of number of hops increases during the search.

REFERENCES

[1] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. "Search and Replication in Unstructured Peer-to- Peer Networks", In Proceedings of ACM ICS'02, NY,June 2002, pp.84-95.

[2] S. Jiang, L. Guo, X. Zhang, and H. Wang, "Light Flood: Minimizing Redundant Messages and Maximizing Scope of Peer-to-Peer Search," IEEE Trans. Parallel and Distributed Systems, vol. 19,no. 5, pp. 601-614, May 2008.

[3] Tsoumakos D., Roussopoulos N, "Adaptive probabilistic search (APS) for peer-to-peer network's", Technical Report, CS-TR-4451, University of Maryland, 2003.

[4] C. Gkantsidis, M. Mihail, and A. Saberi, "Random Walks in Peer-to-Peer Networks," Proc. IEEE INFOCOM '04, pp. 120-130, 2004.

[5] C. Gkantsidis, M. Mihail, and A. Saberi, "Hybrid Search Schemes for Unstructured Peer-to-Peer Networks," Proc. IEEE INFOCOM '05, pp. 1526-1537, 2005.

[6] A. Crespo and H. Garcia-Molina, "Routing Indices for Peer-to-Peer Systems," Proc. 22nd Int'l Conf. Distributed Computing Systems (ICDCS '02), pp. 23-32, July 2002.

[7] B. Yang and H. Garcia-Molina, "Improving Search in Peer-to-Peer Networks," Proc. 22nd Int'l Conf. Distributed Computing Systems (ICDCS '02), pp. 5-14, July 2002.

[8] P. Lin, T. Lin, and H. Wang, "Dynamic Search Algorithm in Unstructured Peer-to-Peer Networks," Proc. Global Telecomm. Conf.(GLOBECOM '06), Nov. 2006.

[9] K. Sripanidkulchai, The Popularity of Gnutella Queries and Its Implications on Scalability, white paper, Carnegie Mellon Univ.,Feb. 2001.

[10] Z. Zhuang, Y. Liu, L. Xiao, and L. M. Ni, "Hybrid periodical flooding in unstructured peer-to-peer networks," in *Proceedings of ICPP'03*, October 2003, pp. 171–178.