Reducing Redundancy of Random Walk Search on Grid Topology for Unstructured P2P Network

Rakesh Rathi
Assistant Professor
Department of Computer Science
Govt. Engineering College, Ajmer, (Raj)

Sujata Salvi
Department of Computer Science
Govt. Engineering College, Ajmer, (Raj)

ABSTRACT
Unstructured systems are more supportive for present mass market where demands are changing day by day. It is always a challenge to design a more efficient search algorithm for unstructured file sharing systems. Flooding and random walk are commonly used techniques but have many limitations. Flooding has poor granularity and very high search cost due to its dependency on ttl. On the other hand random walk is more efficient with its random neighbor selection but it requires high search time. In this paper, we propose an approach to reduce random walks search time by reducing redundancy in algorithm. We have evaluated our changes in random walk on grid topology and our simulation results prove that by working on same network with same parameters, search performance of our approach is 3 times better than simple random walk.

General Terms
Peer to Peer network, Random Walk.

Keywords
Search Performance, Unstructured P2P network, Random Walk.

1. INTRODUCTION
P2P network [1] is a distributed network composed of a large number of distributed, heterogeneous, autonomous, and highly dynamic peers. Today p2p systems are widely being used with increasing speed of communication and computational power of nodes. P2P systems mainly distinguish according to placement of data and resources in system. Structured system guarantee to find existing data but suffers from high overhead due to frequent change in structure. Most structured systems are very efficient as it do not suffer from false negatives. On the other hand unstructured systems are more flexible about resource location so easily support complex queries and structure changes. The only drawback is their routing efficiency is low as number of peers joins and leaves during search process.
In unstructured system, where there is no control over resource location search algorithm play an important role to make system efficient. System performs searches by keywords instead of exact matching performed in structured systems. Search techniques in unstructured network can be classified as informed search and uninformed search. Informed searches like APS [2], routing based search [1], local indices based search [1] etc are more efficient but difficult to implement and maintain.

Flooding is commonly used technique to locate resource in unstructured system, it floods query in system with specific TTL and wait for result. Hence flooding performance highly depends on TTL value, in case of large value system suffers from scalability problem. Flooding has poor granularity [3] and high search cost. To overcome this random walk is applied, in which k number of query messages are sent to k random neighbors, if any node have the queried resource it reply success else keeps on forwarding to k random neighbors. It reduces search cost but requires long search time and less coverage than flooding. Search time will improve by increasing number of walkers but the effect is limited due to redundant path. So, if redundancy will remove in random walk then this highly preferable search technique will result with better search efficiency and high success rate.

Main objective of this paper is to improve random walk in unstructured peer to peer system as random walk search technique is easy to implement and highly preferable for search in unstructured systems. Redundancy will remove by selecting only those neighbors which have higher hit count. We also present a comparative analysis of random walk and this scheme by evaluating search efficiency in grid like network.

The rest of the paper is organized as follows: section 2 presents work done related to our approach. Section 3 briefly discusses about our approach, section 4 is all about the experimental evaluation. Finally, in section 5 we conclude and present future work that can be done.

2. RELATED WORK
Many schemes have been designed to improve search performance of unstructured p2p network. Basically, every new technique is inspired from either flooding or random walk. Mainly they are distinguishing as techniques based on selection and techniques based on underlying topology changes [4][5]. Selection based techniques differ by their selection criteria i.e selecting type of content [6], selecting location [7], selecting fie index [8] etc.

Most of the improvement work is based on flooding, random walk or both like random walk with look ahead [3], dynamic search [9], supervised random walk [10], adaptive random walk [11], random walk on embedded tree structure [8] etc. In [3] author introduced techniques based on locally mentioned network information selection. Various hybrid searches are designed on the basis of random walk and flooding for regular topologies. Random with look ahead, technique works for shorter walk. Our work do not take network coverage in consideration, it works for right direction of walker. In [9] dynamic search technique is based on combination of flooding

and random walk, supervised random walk is based on link selection, it is quite similar as our approach but it is based appropriate neighbor link selection by assigning weight to network link, we are only considering node attributes instead of link.

3. SEARCH APPROACH

This section is divided into two parts. First part describes the proposed scheme and second about the content distribution in network

3.1 Proposed Scheme

Our scheme represents slight improvement in random walk search. In K-random walk search technique query messages are forwarded to randomly selected neighbors as represented in

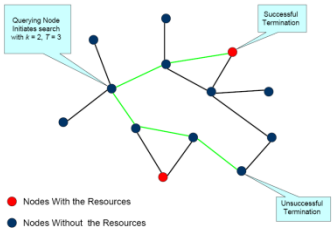

Fig 1: Simple K-Random Walk

Where, initiating node with K=2, forwards query to two random neighbor, if node is having resource query terminates else again forwarded to random neighbor, at each level selection of neighbor is done randomly without any hint. It takes long search time.

Most of the times same node is selected randomly and this will result in redundancy. To reduce redundancy, our scheme suggests selecting that neighbor which has not been selected previously. Unlike random walk, searching node forwards K query messages to most "eligible" neighbor i.e node which has higher probability of being successful node than those which were selected previously and proved as unsuccessful.

Our approach is simply based on information maintained on node. Each node maintain data structure with six type of information: message table that count of number of times current node has seen this packet, hit table that stores packets for which current node reports query hit, list that buffers incoming packets, current nodes view (connected links), key storage that represents frequency of key, tree set represents cycle in which query is scheduled. The main attribute which is used to mark a node as eligible is first one; it gives information about previous visits. By adding these simple parameters in query table, our technique results with better success rate.

3.2 Environment Definition

Mainly two factors are important to simulate p2p environment, network topology and content distribution on peers.

3.2.1. Network Topology

Network topology represents peer to peer connection manner of network. Initially we have selected grid topology for our simulation where each node has either 4 or 8 neighbor.

3.2.2 Content Distribution

Each peer hosts some data which are searched during query. We have assumed data simply in bit string. Each peer hosts a 10 bit string, which is also referred as token. Query is also in form of tokens. There is 1024 different type of data in presented network. Each peer data is classified as data to be shared and data to be queried. Data and query distribution is based on zipf law [12].

4. PERFORMANCE EVALUATION

In this section, we studied gains of our scheme and presented its improvement by comparing it with simple random walk. Below, we first describe our experimental setup and then results.

4.1 Experimental Setup

In order to test our improvement, we have evaluated its performance on peersim [13][14][15][16] simulator, which is specifically designed for p2p protocols. Using peersim, we generate grid network containing 10000 peers. We have performed coverage experiment on stable conditions of network i.e the search operation is performed until the message packets cover entire network in which no peer join or leave during experiment. In peersim, coverage experiment varies due to parameters mentioned in configuration file. We had given same parameters for both random walk and our improved walk which is ttl value 50 and walkers 20, data and query distribution 1.0 and for same number of cycles and queries. An efficient algorithm generates less number of messages and at the same time produce high hit rate. To evaluate efficiency of algorithm mainly three metrics [17] query efficiency, search responsiveness and search efficiency is calculated. All of the three are based on higher hit rate.

4.2 Experimental Result and Analysis

The performance of our scheme is compared with simple random walk by setting same parameters during simulation. On network of 10000 nodes, search is performed for maximum 100 queries. The result observed in following form: query id, message ttl, nodes covered on network, query hits, message packets used. As in our simulation both have same value of walkers, so both generate same number of message packets, but on comparing performance our scheme represent remarkable improvement. With same parameters, nodes coverage on network and number of hits total value of our improvement is three times more than simple random walk. Hence, it has better query efficiency, search responsiveness and search efficiency due to its high success rate.

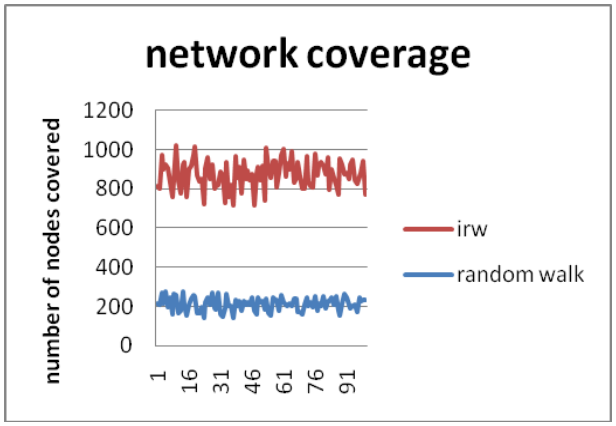

Figure 2: Network coverage on grid topology

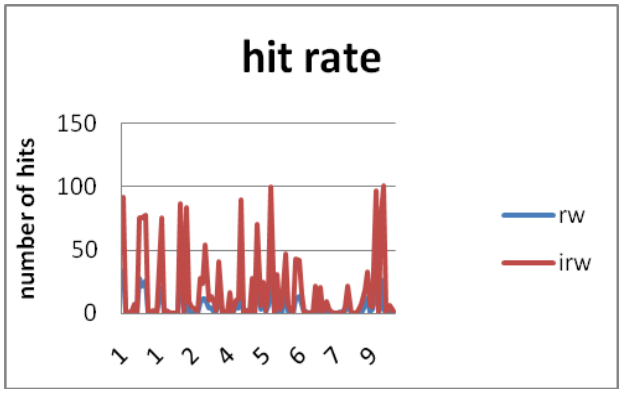

Figure 3: Hit rate of random walk and improved random walk

5. CONCLUSION AND FUTURE WORK

This paper proposes an improved search by doing simple modification in random walk approach. Our scheme simply reduces redundancy in random walk by selecting most eligible node based on previous search data. By this simple improvement our approach performs three times better than simple random walk. We have worked on simple grid topology and with stable conditions of network. In future the work can be analyzed with other topologies and data can be more realistic. Our approach is compared with only random walk, in future it can be compared with other search techniques.

6. REFERENCES

[1] Xiuqi Li and Jie Wu, "Searching Techniques in Peer-to-Peer Networks" Department of Computer Science and Engineering, Florida Atlantic University.

[2] Dimitrios Tsoumakos and Nick Roussopoulos, "Adaptive Probabilistic Search for Peer-to-Peer Networks", Proceedings of the Third International Conference on Peer-to-Peer Computing (P2P03), IEEE 2003.

[3] Christos Gkantsidis, Milena Mihail and Amin Saberi, "Hybrid Search Schemes for Unstructured Peer-to-Peer Networks," Proc .IEEE INFOCOM'05, pp.1526-1537, 2005.

[4] VicentCholvi, AntonioFern´ andez,LuisL´ opez,LuisRodero-Merino," Using Random Walks to Find Resources in Unstructured Self-Organized P2P Networks", Dasson 2007.

[5] William Acosta Surendar Chandra, "Unstructured Peer-to-Peer Networks - Next Generation of Performance and Reliability", IEEE Infocom 2005.

[6] Chinmay R Totekar,Vani M, Palavalli Sharath R, "Search Improvement In Unstructured P2P Network Considering Type of Content", IEEE 2008.

[7] ManalElDick, EstherPacitti, "Locaware:Index Caching in Unstructured P2P-file Sharing Systems", 2nd International Workshop on Data Management in Peer-to-peer systems (DAMAP) (2009).

[8] Phouvieng Hieungmany and Shigeo Shioda,"Characteristics of Random Walk Search on Embedded Tree Structure for Unstructured P2Ps", 16th International Conference on Parallel and Distributed Systems. IEEE 2010.

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

[10] Lars Backstrom and Jure Leskovec," Supervised Random Walks:Predicting and Recommending Links in Social Networks",WSDM'11, HongKong, China, , February9–12,2011.

[11] Nabhendra Bisnik and Alhussein Abouzeid, "Modeling and Analysis of Random Walk Search Algorithms in P2P Networks",ICC'09, IEEE, 2009.

[12] Lada A. Adamic and Bernardo A. Huberman, "Zipf's law and the Internet", Glottometrics, Vol. 3 (2002), pp. 143-150, 2002.

[13] Alberto Montresor and M´ark Jelasity,"PeerSim:A Scalable P2P Simulator", In Proc. of the 9th Int. Conference on Peer-to-Peer (P2P'09), pages 99–100. Seattle, WA, September 2009.

[14] Gian Paolo Jes, "PeerSim HOW TO: Build a new protocol for the PeerSim 1.0 simulator", December 24, 2005.

[15] Moreno Marzolla, "Simulating overlay network with Peersim",http//www.moreno.marzolla.name.

[16] Bruno DEFUDE, "P2P simulation with PeerSim" ,Feburary2007.

[17] Tsungnan Lin, Hsinping Wang, "Search Performance Analysis In Peer-to-Peer Networks",Proceedings of the Third International Conference on Peer-to-Peer Computing, IEEE 2003.