International Journal of Advances in Computer Networks and its Security – IJCNS
Volume 4 : Issue 1 *[ISSN 2250 – 3757]*
Publication Date : 09 January 2014

An Efficient and Scalable Search Mechanism in Unstructured Peer to Peer Network

Md. Sohrab Mahmud, S.M.G. Mortuza Azam, Md. Nasim Akhtar
Department of Computer Science & Engineering
Dhaka University of Engineering & Technology, Gazipur
Gazipur, Bangladesh
sohrabmahmud@gmail.com, sazamduet@gmail.com, drnasim@duet.ac.bd

***Abstract*—Peer-to-peer (P2P) network systems gain a huge popularity due to their scalability and reliability in architectures and search facilities. Basically, most of the real world P2P network is unstructured. Due to their unstructured nature it is often impossible to pre-define the searching criteria. As a solution, flooding scheme is used in most cases. But one major limitation of flooding is its query overhead and unnecessary use of bandwidth. In this paper, we propose a novel mechanism to improve the search efficiency in unstructured P2P networks. The method is based on feedback biased walk. Instead of keeping only the feedback report a peer keeps measuring its rank among the neighbors, this algorithm maintains another rank index for a peer that indicates the reliability to make a successful query hit through that peer. Cumulative Feedback-biased Walk (CF-walk) removes the limitations of unstructured P2P networks based on flooding point and cache based searching, improves the hit rate in case of search resources and reduce the query delay. Extensive simulations show that it improves the workload and reduce query traffic.**

***Keywords*—P2P Networks, Unstructured P2P, Feedback-biased walk, Scalable, Rank based.**

I. Introduction

At the present time P2P network is termed as so much essential part of searching approach in network. It is also now emerges as the dominant model for the networks of the future. The P2P paradigm dictates a fully distributed, cooperative network design, where nodes collectively form a system without any supervision. For enhancing the search performance, social networks have been utilized to develop P2P applications with the cooperation between peers[2] and the P2P structures based on social networks [4], [5]. Social relationship is the part of sociology, which is studied by sociologist from many years ago [6], [7]. From the increasing of the popularity of the internet as well as increasing the number of user of computer networks the social behaviors [10], [11] as well as complex interpersonal on-line relationship has been created with the assistance of various kinds of social network applications , like: Facebook [8], Myspace [9], Orkut etc. People share same interest to each other as well as talk to each other via network. P2P make use of various kinds of data sharing purposes. According to the data sharing P2P can be divided into three categories: mixed, structured and unstructured P2P system [12].

Structured P2P data sharing systems such as Chord [13], in this system is strictly controlled, as with the distribution Hash table, the data strictly mapped to the node, such systems have more efficient query routing, it ensure that the routing hop of querying data objects no more than O(log N). On the other hand, because of the topology maintenance costs is expensive and it does not support complex and fuzzy retrieval, resulting in its application is still not very widespread at the present stage.

Gnutella [14] and KaZaA [27], are the example of unstructured P2P systems. There is neither any centralized system nor was there any the data to place and control mechanisms. The practicality of P2P networks relies on the data location and extraction efficiency. How to efficiently search for resources on the P2P network is one of the most significant issues. From the user point of view, the algorithm is good or bad will directly affect the quality of services (such as time delay, the satisfaction of the results, etc.), as well as the user experience.

In P2P networks the resources are not stored in a single server but scattered in all of the other nodes. As well as although the ideal of P2P network each node is equal in practice their ability to provide resources to different [15] nodes heterogeneity exists, the search methods need to take into account whether the nodes should be given different priorities and approach. In this paper, we will discuss how to remove the limitations of unstructured P2P networks based on flooding point, cache based searching as well as how to improve the hit rate in case of search resources and reduce the query delay, improve the workload, reduce query traffic, ensure user flexibilities and give service according to the priority of the user is the goal of this paper. The rest of the part of this paper is organized as follows: related work, proposed method, performance analysis, simulation results and conclusion.

Section II briefly introduces different kinds of searching algorithms in peer-to-peer networks and the recent progress in social networks studies. Detailed descriptions of the proposed model and assumptions are presented in Section III. Our mathematical and analytical results are provided in Section IV. Simulation results are presented in Section V and Section VI presents the conclusion as well as the scope for future work.

II. Related Work

A Peer-to-Peer (P2P) network consists of a collection of nodes that communicate with each other through direct connection. In P2P networks, nodes are not assigned a fixed role and so can act as both client and server. Nodes are capable to provide services for other nodes. For instance, nodes in P2P file sharing networks provide files to each other through direct connections among the nodes. A direct connection between two nodes is represented by a logical network link as well as a structure formed by logical links and nodes, that is, a P2P network topology, can be formed freely [1].

Generally most of the searching schemes in unstructured P2P networks are forwarding-based [16], including iterative deepening [17], local indices [17], k-walker random walk [23], modified random BFS [21], two-level k-walker random walk [20], directed BFS [17], intelligent search [20], routing indices based search [19], adaptive probabilistic search [24], and dominating set based search [18] etc. These schemes are different variations of flooding used in Gnutella [14]. They can be classified as deterministic or probabilistic [22]. At first the non-forwarding concept was proposed in GUESS [25]. In case of non-forwarding approach each node fully controls the entire process of its own queries. Each node directly probes its own neighbors in a sequential order until the query is satisfied or until all neighbors have been probed. Every node uses a link accumulation to keep information about its neighbors. There is one entry for each neighbor in the link accumulation. These link cache entries are refreshed through periodic pings. For adding new neighbors into the link accumulation, each node also requests that its neighbors select a certain number of their own link cache entries and return them in the pongs during the periodic pings [16].

In case of cache-based search algorithm when a node of the remaining load capacity is high then it will be the center node as well as create a joint topology area with the nearby nodes together. It is needed for the center node and the ordinary nodes to store the index cache, at the local region the overheating resources will be copied to the local for this approach [26].

Flooding [14] is a popular query scheme to search a data item in fully unstructured P2P networks such as Gnutella. While flooding is simple and vigorous, its communication overhead, that is, the number of messages, increases exponentially with the hop number. In case of search for resources most of the messages visit the node that has been searched in the same query, we can say as duplicity of messages. As a result communication overhead and scalability is the main problem in this approach.

Interest-based shortcut [28] is used to avoid the blindness in random walks. In case of blind search nodes do not store any information regarding object locations. Every search requires contacting several nodes within some distance called time to live (TTL), creating enormous overhead to all nodes involved. By favoring nodes sharing in similar interest with the source, which can be regarded as a variation of Markov random walks, biased towards some specific nodes. Markov random walks may accelerate the query process to some extent in some cases. On the other hand it causes new problems, such as nodes in an interest group have formed a cluster, and query messages can be artificially restricted in this specific cluster. Therefore, the query procedure should shorten the covering time of the whole cluster instead of the hitting time of some specific nodes in it. As a result, Markov random walks work poorer than uniform random walks if both of them can be confined in specific clusters.

III. Proposed Method

In this paper we mainly proposed a novel method that enhanced the performance of Feedback biased walk. In Feedback biased walk a walker selects the neighbor, to forward the query, with the highest probability if that particular neighbor has replied a feedback with many results records before among all the neighbors of the peer that initiates a query or currently the forwarder. The selection for the best neighbor is based on a metric that indicates how many times it makes a query hit. A query hit is called when a peer makes a successful reply against a query. When the system is initialized all peers are treated as same.

In cumulative feedback-biased walk each peer in the network maintains two indices of feedback. It maintains one index to keep the ratio of query hit it induces in respect to query requests. Another index it keeps refers to the ratio of the collective feedback of its neighbors those pass a query hit reply through this particular peer. For example node A will maintain a rank that indicates the performance of itself that is

$$P_A = Q_{hit}/Q_{req} \quad (1)$$

Where Q_{req} is the total number of query requests A receives and Q_{hit} is the total numbers it makes a query hit. The value of P_A denotes that how likely node A is responsive to a particular query request. The cumulative feedback index maintains by a node is

$$CP_A = Q_{rep}/Q_{req} \quad (2)$$

Where Q_{rep} is the total number a query hit reply goes through node A. The CP_A denotes the path reliability through node A as it collects the total successful responses through this path. To get better performance a node uses both the ranks on demand. In this paper we propose an efficient approach to search in unstructured P2P network. When a node A initiates or forwards a query it chooses the node that has the highest CP_A value among the neighbors those connected directly to node A. If the TTL value is 1 then node A will choose a node based on the value of P that is its individual performance metric.

IV. Performance Analysis

In figure 1 we studied a case where 7 peers form a network with different degree of connectivity with each other. Node A issues a query with TTL value set to 3 and hence is called as the query issuer. To start a query search over the network it first searches the indices of its vicinity. As it is a query issuer so it sends the message to the neighbor whose Cumulative Feedback is better. From the rank table A finds B to be the suitable route to go deep as it can yet travel 2 more hops. In node B, it will search the local contents if it can match the search. If thus then it become the query hit node. But in this case it fails to have the desired content and decides to forward the message. Again it checks the rank table and determines D to be the best node to forward the message. Before forwarding the message to D, B decreases the TTL value by 1. Hence the message reaches to node D with TTL = 1. In this case D will be interested to consider the individual performance rather than cumulative one as the message can move only one more hop if necessary. D again fails to make a query hit and thus checks the rank table to forward the message to node G which makes a query hit and reply on the same path. During the search the related nodes update both the feedback indices in rank table and with the query message each node informs another participating in the search procedure through the packet header.

V. Simulation Results

We conducted the simulation on a graph with 1000 nodes. Degrees of each node vary within the range of [3,100] with an average value of 35. The TTL value was considered to be 10 initially. For a better comparison between max-feedback biased walk and our proposed mechanism cumulative feedback biased walk we conducted a simulation for a certain period with an increasing rate of TTL values. We figured out the comparison by setting up the TTL values within a varying range of 1 to 100. We consider the network remain almost stable and didn't change the topology frequently. Figure 2 shows the performance comparison of success rate between max-feedback biased walk and Cumulative feedback biased walk.

Table 1. Individual And Cumulative Feedback Rank For Each Peer (Rank Table)

Peers	Individual Rank	Cumulative Rank
A	5/12	7/12
B	3/15	8/11
C	6/16	9/13
D	4/18	9/15
E	8/14	10/13
F	4/12	5/10
G	9/17	6/14

Compared with the max-feedback biased walk our proposed method (CF-Walk) performs better in terms of success rate against increasing TTL values. The simulation shows the expected results as with increasing TTL CF-Walk has a significant chance to find the appropriate route that can return a query hit. As a successful route contains the nodes or peers that forward the query it is like as max-feedback biased walk. For this reason we noted CF-Walk as a method that have knowledge of peers those resides at two hop distance and even more.

VI. Conclusion

In this paper, we present the design and evaluation of cumulative feedback biased walk, a rank based search method in Gnutella like peer to peer networks. Comparisons show that Cumulative Feedback biased Walk (CF-walk) achieves high success rate and fast response time while reduces traffic volume and message overhead at a significant rate. So far we have studied the methods that worked on neighboring information to forward a query, have the information of one hop distance peer. Using CF-walk we can have peer information of 2 hops distance and even more. In this paper, we implement CF-walk with one walker. In near future we will implement the method with k-walker to analyze the performance.

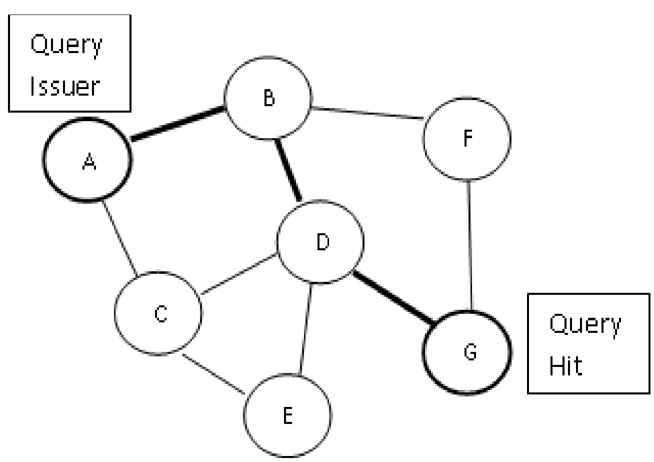

Figure 1. Cumulative Feedback biased Walk procedure.

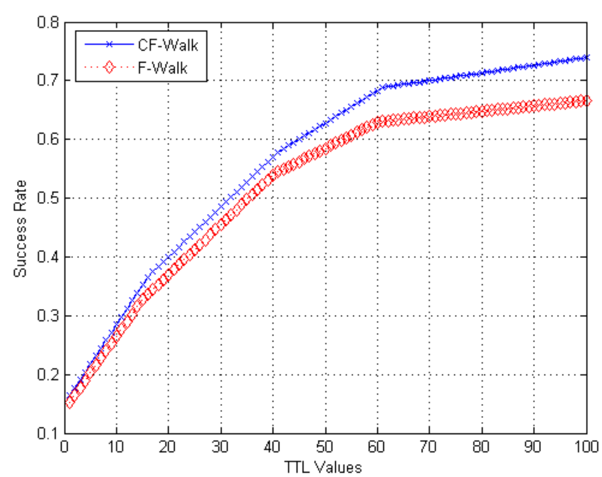

Figure 2. Performance comparison of Cumulative Feedback biased Walk and Feedback biased Walk.

SSACKNOWLEDGMENT

This work is partially supported by Dhaka University of Engineering & Technology, Gazipur, Bangladesh.

REFERENCES

[1] Kei Ohnishi, Yuji Oie, "Parallel Evolutionary P2P Networking for Realizing Adaptive Large-scale Networks", 2011, IEEE/IPSJ International Symposium on Applications and the Internet.

[2] Wenyu Wang, Li Zhao, and Ruixi Yuan, "Improving cooperation in peer-to-peer systems using social networks," in IPDPS, 2006.

[3] Victor Ponce, Jie Wu, and Xiuqi Li, "Improve peer cooperation using social networks," in International Conference on Parallel Processing Workshops(ICPPW2007), 2007.

[4] Vincenza Carchiolo, Michele Malegeri, Giuseppe Mangioni, and Vincenzo Nicosia, "Social behaviours in p2p systems: An efficient algorithm for resource organzisation," in Proceedings of 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06), 2006.

[5] Lu Liu, Nick Antonopoulos, and Stephen Mackin, "Social peer-topeer for resource discovery," in Paralle Distributed and Network-Based Processing(PDP'07), 2007.

[6] P.V. Marsden and K.E. Campbell, "Measuring tie strength," in Social Forces 63, 1985, pp. 482C- 501.

[7] Mark.Granovetter, "The strength of weak ties.," American Journal of Sociology, vol. 78, pp. 1360-1380, 1973.

[8] The Facebook, "http://www.facebook.com," 2008.

[9] The Myspace, "http://www.myspace.com," 2008.

[10] Barry Wellman, "For a social network analysis of computer networks: A sociological perspective on collaborative work and virtual community," in SIGPR/SIGMIS, 1996.

[11] Parag Singla and Matthew Richardson, "Yes, there is a correlation-from social networks to personal behavior on the web," in WWW2008, 2008.

[12] Liu, Y., Xiong N., Zhu L., Gao J. (2010). "An effective simulation method for search strategy in unstuctured P2P network".} Simulation Modelling Practice and Theory, 18(4), 456-469.

[13] Stoiea, I., Morris, R., Karger, D., et a1. Chord: "A scalable peer-to-peer lookup service for Internet applications".} In Proceedings of the 2001 ACMSIGCOMM (pp. 149-160). New York: ACM Press (2001).

[14] Y. Chawathe, S. Ratnasamy, and L.Breslau, "Making Gnutella-like P2P Systems Scalable," In Proceedings of ACM SIGCOMM, 2003.

[15] Saroiu, S., Gummadi, P. K., \& Gribble S. D. (2002). "A measurement study of peer-to-peer file sharing systems." In Proceedings of multimedia computing and networking.

[16] B. Yang and H. Garcia-Molina. "Improving search in peer-to-peer networks." In Proceedings of the 22nd IEEE International Conference on Distributed Computing Systems (IEEE ICDCS'02), 2002.

[17] Xiuqi Li and Jie Wu, "A Hybrid Searching Scheme in Unstructured P2P Networks" Proceedings of the 2005 International Conference on Parallel Processing (ICPP'05)

[18] C. Yang and J.Wu. "A dominating-set-based routing in peerto- peer networks." In Proceedings of the 2nd International Workshop on Grid and Cooperative Computing Workshop (GCC'03), 2003.

[19] A. Crespo and H. Garcia-Molina. "Routing indices for peer-to-peer systems." In Proceedings of the 22nd International Conference on Distributed Computing (IEEE ICDCS'02), 2002.

[20] I. Jawhar and J. Wu."A two-level random walk search protocol for peer-to-peer networks." In Proceedings of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics, 2004.

[21] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-yazti. "A local search mechanism for peer-to-peer networks." In Proceedings of the 11th ACM Conference on Information and Knowledge Management (ACM CIKM'02), 2002.

[22] X. Li and J. Wu. "Searching techniques in peer-to-peer networks." in Handbook of Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks, Edited by J. Wu, CRC Press, 2005.

[23] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. "Search and replication in unstructured peer-to-peer networks.} In Proceedings of the 16th ACM International Conference on Supercomputing (ACM ICS'02), 2002.

[24] D. Tsoumakos and N. Roussopoulos. "Adaptive probabilistic search in peer-to-peer networks." In Proceedings of 2nd International Workshop on Peer-to-Peer Systems (IPTPS'03), 2003.

[25] Guess protocol specification v0.1. http://groups.yahoo.com /group /the gdf /files /Proposals /GUESS /guess o1.txt.

[26] Yuhua Liu · Shishun Wu · Naixue Xiong · Jong Hyuk Park · Maoyuan Zhang, "A cache-based search algorithm in unstructured P2P networks".} © Springer Science+Business Media, LLC 2011

[27] KaZaA, "Kazaa file sharing network," 2002, http://www .kazaa.com/.

[28] K. Sripanidkulchai, B. Maggs, and H. Zhang, "Efficient content location using interest-based locality in peer-to-peer systems," in Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03), vol. 3, pp. 2166-2176, San Francisco,Calif,USA,March-April,2003.