

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

## An Efficient Weighing Algorithm for Information Retrieval in Unstructured P2P Networks

AR.Arunachalam, Sundararajan.M, Arulselvi S

Asst.Professor, Department of CSE, Bharath University, Chennai, India

Director, Research Center for Computing And Communication, Bharath University, Chennai, India

Co-Director, Research Center for Computing And Communication, Bharath University, Chennai, India

**ABSTRACT:** In this paper we propose a generalized searching strategy for judging the best route to retrieve the results in unstructured peer-to-peer (P2P) networks. In our algorithm the query is sent to index table first. The index table will contain all the files names as index in the network. So that the peer will check its index table and revert the query back to the original node which has the resultant data in it. This process is efficient in retrieving the result within the time specified. We use a weighing algorithm for finding the best information among the peers.

**KEYWORDS:** Peer-to Peer, Weighing Algorithm, Generalized search algorithm

### I. INTRODUCTION

Peer-to-peer (P2P) paradigm has emerged as new model for distributed networked services and applications. P2P applications have been deployed in many different areas, such as distributed grid computing, storage, web cache, Internet telephony, streaming, conferencing, content distribution, and so on. But file sharing applications are perhaps the most popular P2P applications: many different file sharing systems, such as Gnutella, Kazaa, Edonkey, Emule, BitTorrent, exist and collect millions of users. These types of applications are characterizing a great fraction of the Internet traffic nowadays as witnessed by several studies that recently put in evidence that P2P traffic is starting to dominate the bandwidth in certain segments of the Internet. In a these applications the participants (termed as peers) organize themselves in an overlay (logical) network on top of the physical network. Each peer establishes application level connections only to a subset of known peers (its neighbors).

When a peer needs to locate a resource it sends out request messages (queries) to its neighbors. There are two main approaches for locating a resource in unstructured decentralized P2P networks: flooding and random walk. In random walk based search strategies peers forward a query message (termed as walker) to one randomly chosen neighbor at each step although several walkers can be employed in parallel to increase the probability of successfully locating a resource (hit probability). In flooding based search strategies, when a peer requests a resource it sends queries to all its neighbors. This collection of neighbors may then forward the query to their neighbors (excluding, of course, the neighbor that sent the original request). These neighbors may then propagate the query to their neighbors and so on up to a certain predefined maximum level. In the Existing System of implementation, we were using two Principles

1. Random Walk
2. Flooding

In Random Walk, the query is send to the Route in a random way, route is selected randomly. So, the Time taken for retrieving the data is prolonged as the Result retrieval may not selected at the first route itself. In Flooding, the query is sent to all the Routes, the cost to retrieving the data is more. We are proposing the generalized search algorithm for judging the best route to retrieve the results. The query sent to index file first. The index file will contain all the files names as index in the network. So that the peer will check its index file and revert the query back to the original node

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

which has the resultant data in it. So the resultant node will retrieve the result to the user. User will specify the Time to Live (TTL) for the query raised. This process is efficient in retrieving the result within the time specified.

## II. RELATED WORK

Until today, the development of analytical models to gain insight of the behavior of P2P-based applications has mainly focused on homogeneous assumptions on the behavior and characteristics of the application participants. Two noticeable exceptions are the works in [28] and [29] that take into account the bandwidth diversity problem in BitTorrent-like file-sharing applications. These studies deeply differ from ours since in the type of systems they consider there is no need to locate a resource (the location of a tracker process suffices to be able to start downloading a file). Furthermore, none of the following papers has dealt with the limited processing capacities of peers that result in the possibility for a peer to be overloaded and thus not able to forward other peer queries. Several papers have analyzed the behavior of P2P-based applications by means of analytical models. The paper explores alternatives (expanding rings and random walks) to the classical flooding search strategies. The author also evaluates different network topologies and resource replication strategies as to provide better performance. When alternatives to the classical flooding are analyzed uniform resource distribution is considered and peer availability is neglected. The work in [26] exploits generalized random graphs to represent overlay networks but only focuses on simple homogeneous scenarios where peers are always available, resources are uniformly distributed among peers, and flooding-based search strategies are probabilistic but uniform across peers. This work has been extended in that contains the design and exploitation of a Gnutella crawler as well as an analysis of the measurements obtained with this tool. The authors use GRGs and the generating function analysis technique to demonstrate the utility of these strategies on the Gnutella peer-to-peer network. This paper differs from ours since it is based on the use of random walk; furthermore, it only exploits degree properties of the query receiver and assumes uniform resource distribution. The work in [27] quantifies the effectiveness of random walks for searching and construction of unstructured P2P networks. It also compares flooding and random walk by simulations on different network topologies where resource are uniformly distributed and peer availability is not considered. The work proposes a query routing mechanism for unstructured P2P networks where the participant peers build probabilistic routing tables, constructed and maintained through an exchange of updates among immediate neighbors in the overlay. The proposed routing mechanism uses the information of these routing tables to forward search queries. Availability of peers is not included in the analysis and the topology considered for the analysis is a regular random graph. Furthermore, validation on real overlay topology is not presented. The paper proposes a service differentiation policy based on the amount of services each peer has provided to the P2P community. In this manner all peers are encouraged to share resources in a P2P application. The resource request and distribution processes are modeled as a competition game among all peers and then the proposed service differentiation based mechanism is analyzed by means of a game theoretical approach. The same theoretical framework has been used in [28] to analyze the behavior of P2P applications in the presence of free-riding. This paper shows that a P2P application can operate effectively in the presence of a significant level of free riders. The client-side behavior is important in a P2P-based file sharing application when multiple peers hold a copy of the requested resource; in this case, the “client” peer may select one or more peers to transfer the resource. The paper addresses the problem of optimal server peer selection for both P2P downloads and P2P streaming. The problem has been studied within the framework of a free-market resource economy in which peers buy and sell resources directly from each other. In this framework, a server peer’s price would depend on parameters such as the specific object being downloaded, the duration of the download, the rate at which the download is to occur, and so on. The paper addresses several peer selection problems and provides theoretical methodologies that could be helpful in solving different variations of peer selection problems. In the goal of the server peer selection is the download time minimization. This paper proposes the use of a methodology of machine learning for the construction of good peer selection strategies from peer’s past experience. The server-side behavior accounts for the possible strategies that a server peer may use for serving the set of client peers. In general, these strategies are associated with incentive mechanisms and propose service differentiation strategies and a game theoretic approach for the evaluation of these strategies in other cases, policies that allow the service rate allocation based on an auction strategy are defined. the (distribution of the) time to transfer a resource in a P2P application is one of most effective

# International Journal of Innovative Research in Science, Engineering and Technology

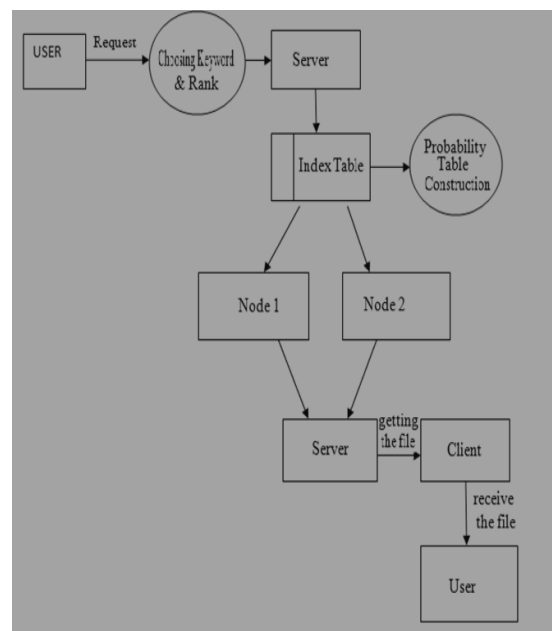
(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

measures to evaluate the QoS perceived by the user of a P2P-based file sharing application. The modeling technique accounts for several different aspects of P2P-based file sharing applications; in particular, the impact of the resource popularity, the server peer selection policies, and bandwidth characteristics, concurrent downloads and uploads, and in a simplified manner some issues related with cooperation level among peers. The paper is based on the use of stochastic fluid process where a fluid process is modulated by a finite homogeneous Markov process. The paper addresses the same issue, i.e., the download time evaluation, but it is developed by means of different modeling techniques and is based on different assumptions simplifications. In particular, this paper evaluates the expected resource download time and tries to capture the impact of various network and peer level characteristics. To address the above concerns, the modeling approach is proposed which decomposes the time required to download a resource into two contributions: network latency time and peer level latency time.

## III. SYSTEM DESIGN

We are proposing the Generalized Search algorithm for judging the best route to retrieve the results. The query is sent to index table first. The index table will contain all the files names as index in the network. So that the peer will check its index table and revert the query back to the original node which has the resultant data in it. So the resultant node will retrieve the result to the user. User will specify the Time to Live (TTL) for the query raised. This process is efficient in retrieving the result within the time specified.



Data Flow Diagram

### 3.1 Random Walk

In random walk the system selects a neighbor randomly to send request query for searching data. And then the node will check that the searching data is available or not. If it is available then it will reply to the sender otherwise it will send to its own neighbor. Like that request query will process till they require data is not found. Actually mechanism uses the depth search algorithm to search the require data. Time Consumption is Very High.

### 3.2 Flooding Search

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

In flooding search the request query for searching data is send to all the neighbors at the same time and if the required data if available at any of the neighbor, it will give response back to source. Otherwise these neighbors will send the query to their own neighbors again. This process will be continued until the required data is found. Actually this mechanism uses the breadth search algorithm to search the require data. The Network Traffic is increased to the Maximum extent.

### 3.3 Generalized Search

In this algorithm, the search is made to the Index Table, to fetch the Best Data, which implicitly specifies the Node and the Path to reach the Node. This Module identifies the Node which has the data through searching with the Query made by the user. This process avoids Time loss, reduces Band width and also fetches the data very effectively.

### 3.4 Index Table Construction

We construct a index table based on generalized search .The information's are stored as in the index table, so that the data search is made to the index table to find the Best Data Relevant to the query. We can increase the performance by reducing number of query messages and time taken with the help of previous experience.

### 3.5 Weighing Algorithm

Weighing algorithm is applied to calculate to find the best book in the entire peer group. Once the query is provided to the index table, the index table will consider the index file stored in it & process the query to all the peers. Query key word is called as "Term", after the streaming algorithm is applied to fetch the Key words in each & every file which is stored in all the peers. Once the peer receives the query, the index table calculates the Term Frequency which is achieved by counting the number of occurrence of the Term which the user has queried in each file. So the ratio is calculated by comparing Term Frequency with the total number of Key words. Consider a book containing 100 words wherein the keyword appears 3 times. Following the previously defined formulas, the term frequency (TF) for keyword is then  $(3 / 100) = 0.03$ . Now, assume we have 10 million books and keyword appears in one thousand of these. Then, the inverse document frequency is calculated as  $\log(10\ 000\ 000 / 1\ 000) = 4$ . The  $tf*idf$  score is the product of these quantities:  $0.03 \times 4 = 0.12$ . We land up with a value for every document which is stored in the index table. This value is called as Weight of Term in a Book of the particular Peer. All the values are tabled in an index table.

### 3.6 Ranking Algorithm Module

We will take an example of a book file sharing p2p application. After getting the corresponding Weight of the term in a book of the particular peer, all the values are tabled for further processing. The scores are compared with each other, to fetch the Top values and it is made in the ascending order. The top valued books are kept in the order. For example Weight of Book 1 in Peer 4 : **0.75**, Book 2 in Peer 1 : **0.70**, Book 3 in Peer 2 : **0.65**, Book 4 in Peer 4 : **0.60** and so on.

## IV. CONCLUSION

In this paper we analyzed the impact of heterogeneity in P2P-based applications on the number of queries sent throughout the network by peers that request a resource and on the hit probability for the search process. We also analyzed the congestion of the network when the search algorithm is able to overload peers with limited processing capacities. To this end, we developed an analytical model exploiting generalized random graphs to represent the overlay network and incorporating the dependence of peer availability and non-uniform resource distribution by considering probabilities that depend on the peer degree. We thoroughly validated the model that showed good agreement with the predictions obtained by simulations on real overlay networks obtained from crawling popular P2P-based file-sharing applications. To this end, we developed a distributed crawler inspired to previous work on this subject that is able to gather Gnutella snapshot in a few minutes. We observed interesting behaviour of a simple probabilistic flooding algorithm that leads the network to congestion. We also showed that neglecting heterogeneity

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

leads to rather different results even in these simple settings. Furthermore, we provided an example of definition of a complex search algorithm that could be easily analyzed by means of our techniques to find optimal parameters setting.

## REFERENCES

- [1] "Entropy," Tech. Rep., <http://www.entropy.com>.
- [2] Kiran Kumar T.V.U., Karthik B., 'Improving network life time using static cluster routing for wireless sensor networks', Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S5) (2013) pp.4642-4647.
- [3] F. Dabek, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, "Wide-Area Cooperative Storage with CFS," in *Proc. of the ACM Symposium on Operating Systems Principles (SOSP)*, 2001.
- [4] Langeswaran K., Revathy R., Kumar S.G., Vijayaprakash S., Balasubramanian M.P., "Kaempferol ameliorates aflatoxin B1 (AFB1) induced hepatocellular carcinoma through modifying metabolizing enzymes, membrane bound ATPases and mitochondrial TCA cycle enzymes", Asian Pacific Journal of Tropical Biomedicine, ISSN : 2221-1691, 2(S3)(2012) pp.S1653-S1659.
- [5] S. Iyer, P. Rowstron, and P. Druschel, "Squirrel: a Decentralized Peer-to-Peer Web Cache," in *Proc. of the ACM Symposium on Principles of Distributed Computing (PODC)*, 2002.
- [6] Jayalakshmi T., Krishnamoorthy P., Kumar G.R., Sivamani P., "The microbiological quality of fruit containing soft drinks from Chennai", Journal of Chemical and Pharmaceutical Research, ISSN : 0975 - 7384, 3(6) (2011) pp. 626-630.
- [7] "Global Index (GI)," Tech. Rep., URL [http://www.skype.com/skype\\_p2pexplained.htm](http://www.skype.com/skype_p2pexplained.htm).
- [8] Kulanthaivel L., Srinivasan P., Shanmugam V., Periyasamy B.M., "Therapeutic efficacy of kaempferol against AFB1 induced experimental hepatocarcinogenesis with reference to lipid peroxidation, antioxidants and biotransformation enzymes", Biomedicine and Preventive Nutrition, ISSN : 2210-5239, 2(4) (2012) pp.252-259.
- [9] S. Iyer, P. Rowstron, and P. Druschel, "SplitStream: High- Bandwidth Multicast in Cooperative Environments," in *Proc. Of the ACM Symposium on Operating Systems Principles (SOSP)*, 2003.
- [10] Kulanthaivel L., Srinivasan P., Shanmugam V., Periyasamy B.M., "Therapeutic efficacy of kaempferol against AFB1 induced experimental hepatocarcinogenesis with reference to lipid peroxidation, antioxidants and biotransformation enzymes", Biomedicine and Preventive Nutrition, ISSN : 2210-5239, 2(4) (2012) pp.252-259.
- [11] X. Zhang, J. Liu, B. Li, and T. P. Yum, "DONet/CoolStreaming: A Data-driven Overlay Network for Live Media Streaming," in *Proc. of IEEE Infocom*, 2005.
- [12] Y.-H. Chu, S. Rao, , and H. Zhang, "A case for end system multicast," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, 2002, special Issue on Networking Support for Multicast.
- [13] J. Li, P. A. Chou, and C. Zhang, "MutualCast: an efficient mechanism for one-to-many content distribution," in *Proc. of the ACM Sigcomm ASIA Workshop*, 2005.
- [14] B. Cohen, "BitTorrent protocol specification," in *First Workshop on Economics of Peer-to-Peer Systems (P2P)*, 2003.
- [15] Z. Ge, D. R. Figueiredo, S. Jaiswal, J. Kurose, and D. Towsley, "Modeling Peer-Peer File Sharing System," in *Proc. of IEEE Infocom*, 2003.
- [16] D. Stutzbach, R. Rejaie, and S. Sen, "Characterizing Unstructured Overlay Topologies in Modern P2P File-Sharing Systems," in *Proc. of the ACM Sigcomm Internet Measurement Conference*, 2005.
- [17] S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 2, pp. 219-232, Apr 2004.
- [18] P. K. Gummadi, R. J. Dunn, S. Saroiu, S. D. Gribble, H. M. Levy, and J. Zahorjan, "Measurement, Modeling, and Analysis of a Peerto-Peer File-Sharing Workload," in *Proc. of the ACM Symposium on Operating Systems Principles (SOSP)*, 2003.
- [19] S. Saroiu, P. K. Gummadi, and S. D. Gribble, "Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts," *Multimedia Systems Journal*, vol. 8, no. 5, Nov 2002.
- [20] R. Bolla, R. Gaeta, A. Magonetto, M. Sciuto, and M. Sereno, "A Measurement Study Supporting P2P File-Sharing Community Models," *Computer Networks*, vol. 53, no. 4, pp. 485 - 500, 2009.
- [21] M. Ripeanu, "Peer-to-Peer Architecture Case Study: Gnutella Network," in *Proc. of IEEE 1st International Conference on Peer-to-peer Computing (P2P)*, 2001.
- [22] S. Zhao, D. Stutzbach, and R. Rejaie, "Characterizing Files in the Modern Gnutella Network: A Measurement Study," in *Proc. Of SPIE/ACM Multimedia Computing and Networking*, 2006.
- [23] T. Klingberg and R. Manfredi, "Gnutella 0.6," <http://rfcgnutella.sourceforge.net/src/rfc-06-draft.html>, Tech. Rep.
- [24] M. E. J. Newman, "The spread of epidemic disease on networks," *Physical Review E*, vol. 66, 2002.
- [25] J.Arul Hency Sheela, GC-MS Studies of the Plant Clematis Gouriana , International Journal of Innovative Research in Science, Engineering and Technology , ISSN: 2319-8753 ,pp 13514-13519 ,Vol. 3, Issue 6, June 2014.
- [26] Jemima Daniel, Usage of Language, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, pp 7073-7075, Vol. 2, Issue 12, December 2013.
- [27] Jemima Daniel, The Duality of Human Nature, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, pp 511-512, Vol. 2, Issue 2, February 2013.
- [28] Jemima Daniel, Impact of E-Mail communication, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, pp 527-528, Vol. 2, Issue 2, February 2013.
- [29] Jemima Daniel, The Enchanting World In Karnas`S Plays, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, pp 757-758, Vol. 2, Issue 3, March 2013.



ISSN(Online): 2319 - 8753  
ISSN (Print) :2347 - 6710

# **International Journal of Innovative Research in Science, Engineering and Technology**

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 1, January 2015**

[30] Jemima Daniel, Myth in Indian English Dramas, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, pp 1551-1555, Vol. 2, Issue 5, May 2013.