International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 3, Issue 3 (October 2013) PP: 15-18

A Comparison between Flooding and Bloom Filter Based Multikeyword Search in Peer-To-Peer Network

Amulia P.M[1], Jisha G[2]
[1]*Department of Information Technology, RSET, Kakkanad,*
[2]*Department of Information Technology, RSET, Kakkanad*

ABSTRACT: *P2P networks are popular in multikeyword searching systems. There are centralized and decentralized P2P networks. P2P systems can also be structured or unstructured. The common technique like flooding used in unstructured network for keyword search incurs large amount of unnecessary traffic. The bloom filter technique used in keyword search reduces unnecessary traffic in the network. With user demand becoming complex and broad, multikeyword search is becoming popular. This paper gives a comparison between flooding and bloom filter based multikeyword search techniques.*

Keywords: *P2P,flooding, TTL, query, bloom filter*

I. INTRODUCTION

Peer-to-peer (P2P) network is a popular information sharing tool where data is located mostly in geographically separate locations. Peer-to-peer network could be structured or unstructured. Search process includes query forwarding, identifying the set of nodes that should be contacted for the result, local processing of the query, local indices that are stored in each peer and its updating. The search efficiency depends greatly on the time taken to provide the result. The search techniques [1] used in unstructured networks is blind or informed search. The common blind search technique [2] is flooding. Flooding uses the basic technique to flood the network with query. Unnecessary nodes are contacted in flooding based search technique. The single keyword search requires the result of the query to be received with minimum delay and lesser bandwidth wastage. The bloom filters are probabilistic data structures that specify the membership of an element in a set. It is a bit vector that specifies "is an element present in the set". Using encoded filters the amount of data transmitted through the network is reduced. With user query becoming broad and complex multikeyword search is becoming popular.

In a query that includes "cloud computing" the query is separated into individual keywords "cloud" and "computing". The traditional method includes that each keyword be separately searched and the results to be merged at the selected peer node. The keywords "cloud" and "computing" are separately searched. If this multi keyword search requires flooding technique the amount of traffic and the number of unnecessary peer nodes contacted would be numerous. Consider a Google web search that includes contacting each and every node in the network. This could cause wastage of resources. The search will be time consuming leading to user frustration. The benefit of flooding technique is guaranteed search result. There are different ways to improve flooding technique [3]. The informed search technique, bloom filter discussed in the paper reduces the unnecessary overhead data in the network.

The rest of this paper is organized as follows. Section 2 provides an idea of multikeyword search using intersection and union operations. Section 3 gives a comparison of flooding and bloom filter based multikeyword search and a conclusion is given in section 4

II. MULTIKEYWORD SEARCH USING INTERSECTION AND UNION OPERATION

In multikeyword search the search query is separated into individual keywords and each keyword is separately searched. The result is obtained after the distributed intersection of each keyword results if the required operation is AND. Given an example for a two-keyword (keyword1, keyword2) search, the keyword1 is searched and the result identifiers x_id are obtained. The keyword2 is also propagated in the similar manner and the result identifiers y_id are obtained. The identifiers x_id contains the set of result identifiers that contains keyword1 and identifiers y_id contains the set of result identifiers that contains keyword2. The resultant intersection [9, 10] operation is found using x_id ∩y_id. The result is sent to the client node that requested the search result for the query.

In some queries the user requires the results containing any of the keywords, OR operation. The technique used is union of results. The multikeyword query containing keywords (keyword1, keyword2) are separated into individual keywords and searched separately. The search results x_id contains the result identifiers for all the documents containing keyword1 and y_id contains the result identifiers for all the

documents containing keyword2. Finally the result (x_id, y_id) are merged [9, 10], x_id U y_id and the result sent to the client. Depending on the popularity of the query the result can be cached for a time period at the query node.

III. COMPARISON OF FLOODING AND BLOOM FILTER BASED MULTIKEYWORD SEARCH

The common search technique used in an unstructured basic Gnutella [4, 5] network is flooding. Gnutella is the first decentralized peer-to-peer network of its kind. The flooding technique uses simple broadcasting method where each node contacts all its neighbors. Time to live (TTL) or hop count can be used to reduce the number of nodes contacted. Flooding comes under the category of blind search. In blind search there is no information about the route to follow to get to the result. The queries are broad casted or flooded in the same manner with the results of the query directed back to the unique node. The unnecessary overload on the network can be reduced with time to live (TTL) value. Choosing the appropriate TTL is not easy. If the TTL is too low the required search result will not be obtained. As the value of TTL increases, the overhead of flooding increases. Ultra peer is a concept that was introduced in Gnutella 0.6. These super peers have high capacity, processing power, memory etc. This architecture saves the client nodes from the burden of extensive message routing but limits scalability of the network. In multikeyword search in P2P network [7] the query is split into individual keywords. The result of the individual keyword is either intersected or merged based on the operation. There are AND or OR operations in a multikeyword search. Flooding technique used in multikeyword search involves broadcasting of each individual keyword throughout the network. Although the required search result is provided, this technique overloads the network with unnecessary data. The scalability of the network is thoroughly affected. Another problem with flooding technique is user frustration since large amount of time is needed to get the required search result.

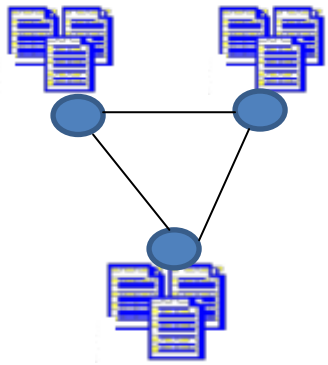

Fig 1 P2P network with documents distributed among peer nodes

In Fig 1 if flooding technique is used for the multikeyword search, the keyword is searched in the document list present in every peer node in the P2P network. For a large network this would be time consuming causing user frustration. The flooding technique greatly affects the efficiency and scalability of a network.

In informed search intelligent choices are made for the query forwarding based on certain criteria. This provides optimistic query forwarding so as to get to the result faster. Informed search techniques [6] include directed breadth first search (DBFS), intelligent search, preferential walk, local indices, routing indices [8] etc. In multikeyword search the bloom filters provide a faster mechanism to specify if the document for the keyword is present in the particular peer node. Bloom filter based search technique gives a structured approach to the search mechanism. The bloom filter [9] technique is also an efficient and improved keyword search mechanism. This includes transmitting documents in encoded form rather than raw data. This reduces traffic in the network. The bloom filters are probabilistic data structures that specify the membership of an element in a set. It is a bit vector that specifies "is an element present in the set".

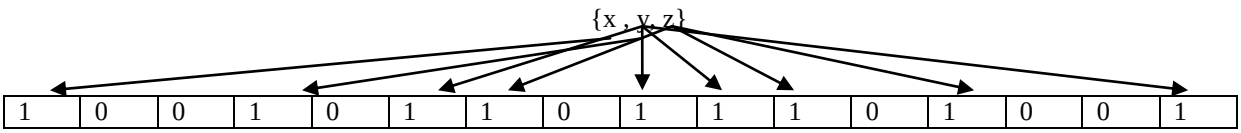

Fig. 2: bloom filter representation of elements {x,y,z}

Fig. 2 shows a bloom filter representation of elements {x,y,z}. The bloom filters shows the bit vector turned on for each element, specifying their membership in the set.

With multikeyword search becoming popular in peer-to-peer network, there should be techniques to guarantee required search results with minimum overhead and delay. This will reduce the communication cost and wastage of resources. Bloom Filters are bit vectors that specify if the result document for the particular keyword is present in the peer node in a P2P network. The problem with using bloom filters is false positive results.

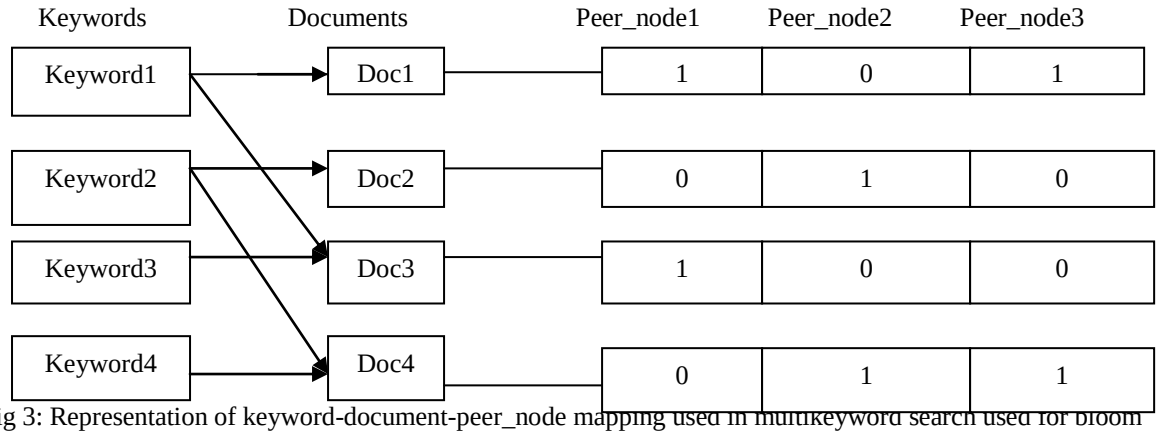

Fig 3: Representation of keyword-document-peer_node mapping used in multikeyword search used for bloom filter creation

Fig 3 shows the keyword-document-peer-node mapping in a multikeyword search used to create bloom filter vectors set. The multikeyword query is split into individual keywords. The bit vector for the peer_nodes in the network that contain the document for that particular keyword is turned on (set to 1). If the document for that keyword is not present in the peer_node the bit vector is set to 0. From the bloom filter set the multikeyword search is made faster as a mapping is obtained about the location of the result document.

TABLE1 COMPARISON OF FLOODING AND BLOOM FILTER SEARCH

Search Techniques	Advantages	Disadvantages
Flooding Technique	Guaranteed Search Result	Large network traffic.
	Can be used with Unstructured Networks	Contacting unnecessary peer nodes.
	No information is required about the search.	Wastage of resources.
Bloom Filter Technique	Can be used with Structured networks.	Generation of bloom filters require location information of the search result.
	Structured search approach. Generated using document-location mapping information. Contact only the necessary peer-node for the required document result.	Storage and maintenance of bloom filters required.
	Reduce unnecessary load on the network.	Bloom filter data could include stale information.
	Minimize wastage of resources and network bandwidth.	False positive results are a very important drawback of bloom filters.
	Reduce communication cost.	

IV. CONCLUSION

There are number of search techniques used with different network topologies. Flooding technique is commonly used in an unstructured network. This method is reliable but affects the scalability of the P2P network. Search techniques are selected based on the requirement. The requirement could be, faster search result or reduced search cost. To reduce communication cost the bandwidth utilization has to be minimized. This requires more organized search like bloom filter based multikeyword search. Although bloom filters have advantages over flooding the most important drawback of bloom filters are false positive results. Structured or unstructured networks, the focus should be to optimize search and provide improved and fast search results.

REFERENCES

[1] Efficient Search Techniques in Peer to Peer Networks, Tarunpreet Bhatia, Dr Deepak Garg, Computer Science Department, Thapar University, Patiala . International Journal of Computer Applications(0975- 8887) Volume 36– No.1, December 2011.

[2] Dynamic Search Algorithm used in Unstructured, B.Srikanth, M.Tech Student, K.Venkateswara Rao, Assoc. Prof in CSE, QIS College of Engineering & Technology, Ongole, Prakasam Dist, AP, India. International Journal of Engineering Trends and Technology- Volume2 Issue3- 2011.

[3] Search and Replication in Unstructured Peer-to-Peer Networks, Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. 2002. Search and replication in unstructured peer-to-peer networks. In *Proceedings of the 16th international conference on Supercomputing* (ICS '02). ACM, New York, NY, USA, 84-95. DOI=10.1145/514191.514206. http://doi.acm.org/10.1145/514191.514206.

[4] Survey of search and replication schemes in unstructured P2P networks Sabu M.Thampi, Dept. of Computer Science and Engineering, Rajagiri School of Engineering and Tecnology ,Kerala , Chandra Sekaran K. Department of Computer Engineering, NIT, Karnataka

[5] Peer-to-Peer Architecture Case Study : Gnutella Network, Peer-to-Peer Computing, FirstInternationalConferenceon Aug 2001 Matei, Ripeanu, Department of Computer Science, Chicago University, IL, USA Pages 99-100 Print ISBN : 0-7695-1503-7 INSPEC Accession Number: 7247201 DOI : 10.1109/P2P.2001.990433, matei@cs uchicago.edu K. Elissa, "Title of paper if known," unpublished.

[6] A Comparison of Peer-to-Peer Search Methods; Dimitrios Tsoumakos, Computer Science Department, University of Maryland, dtsouma@cs.umd.edu, Nick Roussopoulos, Computer Science Department, University of Maryland, nick@cs.umd.edu. International Workshop on Web and Databases, Proceedings of the Sixth WebDB Workshop San Diego, California, June , 2003.

[7] Peer-to -Peer Computing : Dejan S. Milojicic, Vana Kalogeraki, Rajan Lukose, Kiran Nagaraja, Jim Pruyne, Bruno Richard, Sami Rollins, Zhichen Xu [dejan, vana, lukose, pruyne, zhichen]@exch.hpl.hp.com, bruno_richard@hp.com, knagaraj@cs.rutgers.edu, srollins@cs.ucsb.edu, HP Laboratories Palo Alto HPL- 2002- 57 (R.1) July 3rd, 2003.

[8] Routing Indices for Peer-to-Peer Systems ; Distributed Computing Systems, 2002. Proceedings of 22nd International Conference on 2002, Arturo Crespo, Hector Garcia-Molina Computer Science Department Stanford University Stanford, CA 94305-2140, USA {crespo,hector}r@db.Stanford.edu.ISSN: 1063-6927 Print ISBN: 0-7695-1585-1 INSPEC Accession Number : 7404204 DOI : 10.1109/ICDCS.2002.1022239

[9] Optimizing Bloom Filter Settings in Peer-to-Peer Multikeyword Searching Hanhua Chen,Member, IEEE, Hai Jin, Senior Member, IEEE, Lei Chen, Member, IEEE Yunhao Liu,Senior Member, IEEE, and Lionel M. Ni, Fellow, IEEE, IEEE Transactions on Knowledge and Data Engineering, VOL. 24, NO. 4, APRIL 2012.

[10] Efficient Multi-keyword Search over P2P Web; Hanhua Chen, Hai Jin , School of Computer Science and Technology Huazhong University of Science and Technology Wuhan, 430073, China {chenhanhua,hjin}@hust.edu.cn Jiliang Wang,Lei Chen,Yunhao Liu,Lionel Ni Department of Computer Science and Engineering Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong {aliang,leichen,liu,ni}@cse.ust.hk. Proceedings of the 17th international conference on World Wide Web,2008, ISBN: 978-1-60558-085-2 DOI: 10.1145/1367497.136763