

# TRUST BASED SEARCH TIME REDUCTION IN PEER TO PEER SYSTEMS

S.Shanmathi, M.Sangeetha, S.Rohini

**Abstract:** Peer to peer is a decentralized network. In unstructured peer to peer network, the local information about the peers will be known to the neighboring nodes. DHT (Distributed Hash Table) search engine, based on distributed indexing and Federated search engine, based on flooding through data replication leads to poor full text retrieval and raising possibly unacceptable communication and storage costs respectively. To overcome, Bloom Cast mechanism is proposed to increase the efficiency of full text retrieval, based on data collection and query logs. To reduce the cost of replication, Bloom Filter encoding is used, which serve as an intermediate between the peers. Security will be enabled using self – certification, which is created by the peer itself by means of their unique identity given to them, while entering into the network. Thus it is implemented to enhance the efficiency of full text retrieval and get better security in the unstructured peer to peer network.

**Index Terms:** Bloom Cast, Peer-to-Peer Systems, Reputation, Security, Trust Management.

## I. INTRODUCTION

The emergence of peer-to-peer (P2P) file sharing applications, such as Napster and Gnutella, helps millions of users have used P2P systems to search desired data.

Existing P2P full-text search schemes can be divided into two types: DHT -based global index and federated search engine over unstructured protocols. DHT-based searching will be done using distributed indexes, which partition a logically global inverted index in a physically distributed manner.

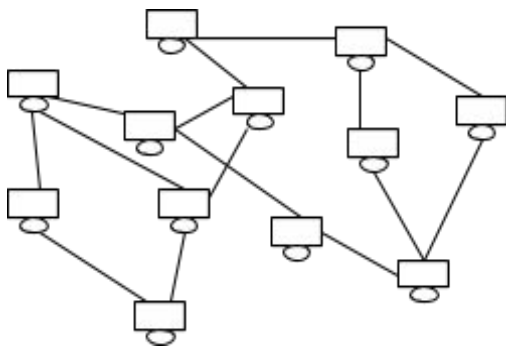


Figure.1 Unstructured P2P network

In federated search engines, queries are processed based on flooding. These unstructured networks are commonly believed to be the best candidate for supporting

full-text retrieval because the query evaluation operations can be handled at the nodes that store the relevant documents.

However, search recall is not guaranteed with acceptable communication cost using a flooding-based scheme. Replication strategies are extensively utilized to improve search performance in unstructured P2Ps. The existing strategies assume that the access frequencies of the items are known and the number of replicas is determined by the query's popularity.

## II. RELATED WORKS

The content retrieval scheme is an important issue in the distributed P2P information sharing systems. There are two content searching schemes in the existing P2P systems. For structured P2P networks, DHT-based distributed global inverted index is used. For unstructured P2P networks, federated search engines are used.

### A. Pure P2P Systems

A pure P2P network does not have the notion of clients or servers but only equal peer nodes that simultaneously function as both "clients" and "servers" to the other nodes on the network. This network model differs from the client-server model. File Transfer Protocol (FTP) is one of the best examples of a file transfer that does not use the P2P model. Here the clients used to open the transfer, and the servers used to continue with those requests. If there exists a link between any two nodes, then there forms a directed edge between the former node and the latter in the overlay network. Based on the linking between the various nodes the P2P networks were classified as structured or unstructured.

### B. Searching In Structured Networks

Structured P2P networks employ a globally consistent protocol to ensure that any node can efficiently route a search to some peer that has the desired resource or data, even if it is a rare one. This needs more structured pattern of overlay links. Most commonly, distributed hash table (DHT) is used to implement in structured P2P network, used to assign ownership of files to that particular peer. The term DHT is generally used to be implemented on top of a structured overlay.

### C. Searching In Unstructured Networks

Unstructured P2P networks are formed when the overlay links are established randomly. In this type of P2P network, if a peer wants to find out any data in the network, a query will be carried out through the network, which may finds many shared data. If any popular content is searched,

S. Shanmathi is an M.E. student in the Department of Computer Science and Engineering, SriGuru Institute of Technology, Coimbatore, India

M. Sangeetha is an M.E. student in the Department of Computer Science and Engineering, SriGuru Institute of Technology, Coimbatore, India.

S.Rohini is an M.E. student in the Department of Computer Science and Engineering, SriGuru Institute of Technology, Coimbatore, India.

then the available peers and the peer searching for it are likely to find the same thing. If a peer is looking for rare data, then it is highly improbable that search will be successful. Since the peer and content management are independent of each other, there is no guarantee that flooding will find the desired data and also causes a high amount of signalling traffic in the network. These networks typically have very poor search efficiency.

#### D. Overlay Networks

An overlay network is a type of the computer network which is built on the top of another existing network. Nodes that are present in the overlay network can be thought of as being connected as virtual links or logical links, each one of which is to be in contact with an appropriate path, connected in the existing network. The most important applications of overlay networks are cloud computing, peer-to-peer systems, and client-server systems. Initially the internet was built as an overlay network upon the telephone network whereas nowadays with the invention of VoIP, the telephone network is turning into an overlay network that is built on top of the Internet. The area in which the overlay networks used is telecommunication and internet applications.

### III. PEER CREATION

**Peer-to-peer (P2P)** computing or networking is a distributed application architecture that partitions tasks among peers. Peers are equally privileged in the application to form a peer-to-peer network.

The peer-to-peer application structure was made simpler by the file sharing systems like Napster. The concept has stimulated new structures in many areas of human interaction.

### IV. CONTENT SEARCHING AND REPUTATION MAINTAINANCE

In this section we are going to discuss detail about the concepts of content searching and reputation maintenance.

#### A. Content Searching

In a content search function, the input is a set of keywords representing a user's interests and the output is a set of resources containing these keywords. In the content search context, resources represent text documents or metadata of general resources. Content search is useful when a user does not know the exact resource names of interests, this case is common in P2P-based searches as well as in web searches. Flooding is the basic method for content searching in unstructured P2P networks. However, unnecessary traffic will be seen in blind flooding based search mechanism. This greatly limits the performance of P2P systems. This unwanted traffic is dividable and can be avoided while searching in P2P networks. The bloom hash table is used to store the resources which help in effective searching of resources with desired capabilities. It also provides more

information about the path-name. This design enables resource discovery without knowledge of where the corresponding data items are stored.

#### B. Reputation Maintenance

The reputation systems based on the client-server model, the server here provides pseudonyms (identities) to users and inducts them into the system. Once a peer has logged into the system, the requester (client) will select a service provider (server) (from other users) for the given service, based on the previous reputation of the service provider. Then the requester will receive the requested service from the appropriate provider. Once the transaction of data is completed, the requester gives recommendation to the server based on its satisfaction level from the transaction. In P2P networks, there is no way to determine the distinctness of a peer in the absence of a central agency or without using external means.

### V. BLOOMCAST

In Unstructured P2P networks, Bloom Cast is an effective and efficient full text retrieval scheme. By the advantage of the hybrid P2P protocol, it replicates the data items uniformly at random across the P2P networks. Bloom Cast hybridizes a lightweight DHT with an unstructured P2P overlay to support random node sampling and network size estimation. It is an inter-domain protocol for routing between the source and the border routers of the ASes hosting the receivers.

#### A. Bloom Filter

Bloom Filters were used to encode the transferred lists while recursively intersecting the matching document set. Bloom Casting is a secure source specific multicast technique, which reassigns the membership control per group forwarding state from the multicast routers to the source. It uses in-packet Bloom filter (iBF) to encode the forwarding tree. Bloom Casting separates multicast group management and multicast forwarding. It sends a Bloom Cast Join (BC JOIN) message towards the source AS. Initially the ASes message contains an empty Bloom filter. While the message moves upstream towards the source, these AS records the convey information in the control packet by inserting the corresponding link mask into a collector. After this, it performs a bit permutation on the collector. The Bloom Cast architecture that is designed to interconnect source specific multicast protocols. Unlike conventional IP multicast method, the onward information is organized in routers on the delivery tree, in Bloom Cast, the transit routers will not store any group-specific state.

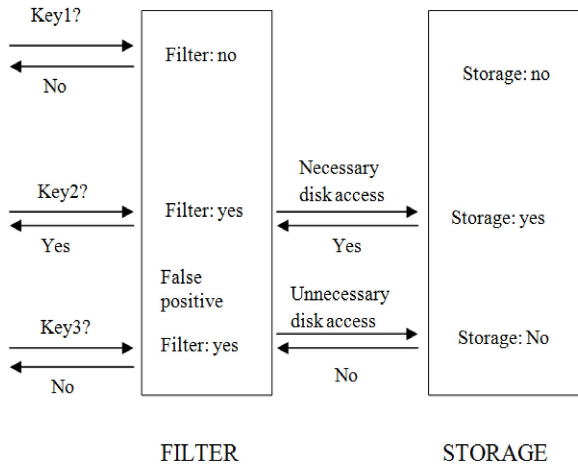


Figure.2 Bloom Filter

**Algorithm:**

Require: EstimatedNetworkSize= $N$  is achieved

**for all** documents in local collection **do**

create an empty bit vector with  $m$  bits for document  $x$ ,  $BF_x$

**for all** terms in a document **do**

insert terms  $t$  into  $BF_x$  by setting the  $h_j(t)$ th bits of  $BF_x$  to 1, where  $\{h_j(\cdot), 1 \leq j \leq k\}$  hash functions used by  $BF_x$

**end for**

**end for**

sample an optimal number of  $r = \frac{c}{\sqrt{\frac{s_q}{s_d}}} \sqrt{N \cdot \frac{s_q}{s_d}}$  random peers in the network by the lightweight DHT.

Replicate  $BF_x$  together with  $url_x$ , the URL of document  $x$ , to the set of randomly sampled nodes;

**return**

## VI. CONCLUSION

All peers in the unstructured P2P network were identified by means of their unique certificates created using their aka identity. The reputation of each peer is attached to its identity after each transaction. The identity certificates are generated using self-certification, and all peers keep up their own certificate authority which issues the identity certificate to the peers.

Bloom Cast is an efficient and effective full-text retrieval scheme. Bloom Cast is effective because it certifies the call back with high probability. It is efficient as the entire communication cost of full-text search is reduced below a conventional bound. Moreover, by replicating Bloom Filters instead of the unrefined documents across the network, Bloom Cast significantly reduces the communication cost for replication. The power of Bloom Cast design has been verified through query logs from a real world search engine.

## ACKNOWLEDGEMENT

The authors would like to thank the staff and students of SriGuru Institute of technology, friends and family members for their support and guidance in bringing this research article. The authors would also like to thank them for their valuable support.

## REFERENCES

- [1] Ahmet Burak Can, "SORT: A Self-Organizing Trust Model for Peer-to-Peer Systems", IEEE Transactions on Dependable And Secure Computing, volume. 10, no. 1, February 2013.
- [2] K. Aberer and Z. Despotovic, "Managing Trust in a Peer-2-Peer Information System", Processing 10th International Conference Information and Knowledge Management, 2001.
- [3] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design", IEEE Internet Computing, volume. 6, no. 1, pp. 50-57, January 2002.
- [4] S. Kamvar, M. Schlosser, and H. Garcia-Molina, "The Eigen trust Algorithm for Reputation Management in P2P Networks", Processing 12th World Wide Web Conference, 2003.
- [5] Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H.Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications", Processing ACM SIGCOMM 2001, pp. 149- 160, 2001.
- [6] E. Cohen and S. Shenker, "Replication Strategies in Unstructured Peer-to-Peer Networks", Processing ACM SIGCOMM 2002, pp. 177-190, 2002.
- [7] R.A. Ferreira, M.K. Ramanathan, A. Awan, A. Grama, and S. Jagannathan, "Search with Probabilistic Guarantees in Unstructured Peer-to-Peer Networks", Processing IEEE Fifth International Conference Peer to Peer Computing (P2P 2005), pp. 165-172, 2005.
- [8] Priya Ponnusamy.P, Yoha Lakshmi.M, "Content Based Retrieval in Unstructured P2P Overlay Networks" The International Journal of Engineering And Science (IJES), Volume 1, Issue 2, Pages 318-323, 2012.
- [9] H. Song, S. Dharmapurikar, J. Turner, and J. Lockwood, "Fast Hash Table Lookup Using Extended Bloom Filter: An Aid to Network Processing", Processing ACM SIGCOMM, 2005.



**S. Shanmathi** was born in Tiruchengode on 8<sup>th</sup> August 1991. She received her B.E (CSE) degree from Sengundhar College of Engineering Tiruchengode, Tamil Nadu in 2012. She is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu. She has presented papers in various national conferences. Her areas of interest are Network Security, Data structures and web designing.



**M Sangeetha** was born in Theni on 26<sup>th</sup> December 1990. She received her B.Tech.(IT) degree from Periyar Maniammai University, Thanjavur, Tamil Nadu in 2012. She is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu. She has published research papers and articles in various international journals. She has presented papers in various national and international conferences. She is interested in Secure Computing, Data mining and Audio mining.



**S. Rohini** was born in udumalpet on 13<sup>th</sup> March 1989. She received her B.E (CSE) degree from Sengundhar College of Engineering Tiruchengode, Tamil Nadu in 2012. She is currently pursuing M.E. (CSE) degree in SriGuru Institute of Technology, Coimbatore, Tamil Nadu. She has presented papers in various national conferences. Her areas of interest are Cloud computing, Green computing.