IJDCST @ Nov-Dec, Issue- V-3, I-1, SW-063
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Full Text Retrieval in P2P Networks Using Compressed Bloom Filters

Talatoti Ratna Kumar[1], T.Sudhir[2]

[1] **M.Tech(CSE), VVIT Guntur, A.P., India.**

[2]**Asso. Professor, VVIT Guntur, A.P., India, A.P., India.**

Abstract: Now a day millions of users to search and download desired datasuch as Napster and Gnutella as they are Peer-to-Peer (P2P) files sharing applications. In order to improve performance in unstructured P2Ps replication strategies are used.Efficient and effective full-text retrieval over unstructured p2p networks was developed in order to address the problems of the query popularity independent replication strategies, previously a novel strategy, called Bloom Cast, which implements Bloom Filters in WP (With Pointers) scheme. In order to support random node sampling and network size estimationbloom Cast hybridizes a lightweight DHT with an unstructured P2P overlay.However thesewell-organized techniques are executed irrespective of topologies and network size concerns. So, to overcome this problem we propose to use Compression bloom Filter, Long random walk, and Short Random Walk with local flooding schemes in accordance with variable p2p topologies and network sizes. Bloom filters these Hybrid Query Propagation schemes offers best performance over unstructured p2p networks and a practical executionauthenticates the claim.

Key Words:Peer-to-Peer,CompressionBloom Filters,Long random walk,General search scheme,Short Random Walk.

I. INTRODUCTION

A peer-to-peer (P2P) network is a type of decentralized and distributed network architecture in which individual nodes in the network (called "peers") act as both suppliers and consumers of resources, in contrast to the centralized client–server model where client nodes request access to resources provided by central servers.

In a peer-to-peer network, tasks (such as searching for files or streaming audio/video) are shared amongst multiple interconnected peers who each make a portion of their resources (such as processing power, disk storage or network bandwidth) directly available to other network participants, without the need for centralized coordination by servers.[1]

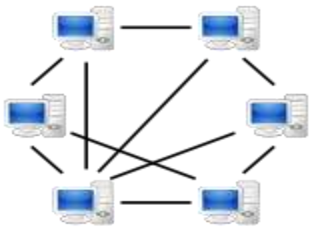

Figure 1: A peer-to-peer (P2P) network in which interconnected nodes.

A peer-to-peer network is designed around the notion of equal peer nodes simultaneously functioning as both "clients" and "servers" to the other nodes on the network. This model of network arrangement differs from the client–server model where communication is usually to and from a central server. A typical example of a file transfer that uses the client-server model is the File Transfer Protocol (FTP) service in which the client and server programs are distinct: the clients initiate the transfer, and the servers satisfy these requests.

Routing and resource discovery:

Peer-to-peer networks generally implement some form of virtual overlay network on top of the physical network topology, where the nodes in the overlay form a subset of the nodes in the physical network. Based on how the nodes are linked to each other within the overlay network, and how resources are indexed and located, we can classify networks as unstructured or structured (or as a hybrid between the two).[2][3][4]

Unstructured networks:

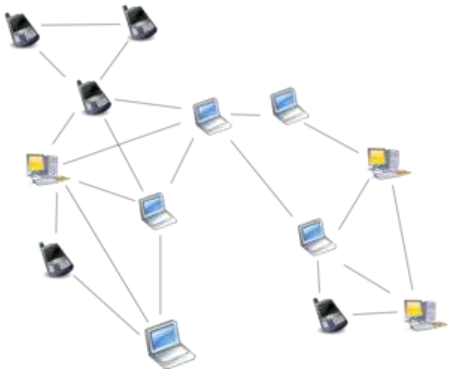

Figure 2: Overlay network diagram for an unstructured P2P network, illustrating the ad hoc nature of the connections between nodes.

Unstructured peer-to-peer networks do not impose a particular structure on the overlay network by design, but rather are formed by nodes that randomly form connections to each other. [5] (Gnutella, Gossip, and Kazaa are examples of unstructured P2P protocols.[6])

However the primary limitations of unstructured networks also arise from this lack of structure. In particular, when a peer wants to find a desired piece of data in the network, the search query must be flooded through the network to find as many peers as possible that share the data. Flooding causes a very high amount of signaling traffic in the network, uses more CPU/memory (by requiring every peer to process all search queries), and does not ensure that search queries will always be resolved. Furthermore, since there is no correlation between a peer and the content managed by it, there is no guarantee that flooding will find a peer that has the desired data. Popular content is likely to be available at several peers and any peer searching for it is likely to find the same thing. But if a peer is looking for rare data shared by only a few other peers, then it is highly unlikely that search will be successful.[10]

Structured Networks:

In structured peer-to-peer networks the overlay is organized into a specific topology, and the protocol ensures that any node can efficiently search the network for a file/resource, even if the resource is extremely rare.

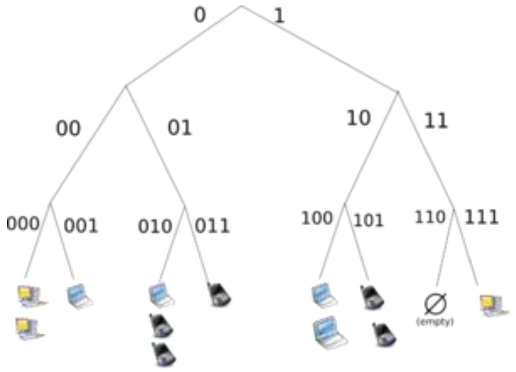

Figure 3: Overlay network diagram for a structured P2P network, using a distributed hash table (DHT) to identify and locate nodes/resources.

The most common type of structured P2P networks implement a distributed hash table (DHT), in which a variant of consistent hashing is used to assign ownership of each file to a particular peer [11].This enables peers to search for resources on the network using a hash table: that is, (key, value) pairs are stored in the DHT, and any participating node can efficiently retrieve the value associated with a given key.

II. RELATED WORK

Dongsheng Li states that with the increasing popularity of the peer-to-peer (P2P) computing paradigm, many general range query schemes fordistributed hash table (DHT)-based P2P systems have been proposed in recent years. Although those schemes can support rangequery without modifying the underlying DHTs, they cannot guarantee to return the query results with bounded delay. The query delay in these schemes depends on both the scale of the system and the size of the query space or the specific query. In this paper, wepropose Armada, an efficient range query processing scheme to support delay-bounded single-attribute and multiple-attribute rangequeries.

Christos Gkantsidis states that we quantify the effectiveness of randomwalks for searching and construction of unstructuredpeer-to-peer (P2P) networks. For searching, we arguethat random walks achieve improvement over floodingin the case of clustered overlay topologies and in thecase of re-issuing the same request several times. Forconstruction, we argue that an expander can be maintaineddynamically with constant operations per addition. The key technical ingredient of our approach is a deep resultof stochastic processes indicating that samples taken fromconsecutive steps of a random walk can achieve statisticalproperties similar to independent sampling (if the secondeigenvalue of the transition matrix is bounded away from1, which translates to good expansion of the network;such connectivity is desired, and believed to hold, in everyreasonable network and network model). This property hasbeen previously used in complexity theory for constructionof pseudorandom number generators. We reveal anotherfacet of this theory and translate savings in random bitsto savings in processing overhead.

We evaluate and compare different replication strategies and reveal interesting structure: Two very common but very different replication strategies uniform and proportional yield the same average performance on successful queries ,and are in fact worse than any replication strategy which lies between them. The optimal strategy lies between the two and can be achieved by simple distributed algorithms. These fundamental results order a new understanding of replication and show that currently deployed replication strategies are far from optimal and that optimal replication is attainable by protocols that resemble existing ones in simplicity and operation.

III. EXISTING SYSTEM

Traditionally evaluate the performance of BloomCast design usingtrace-driven simulations. In this section, we describe thesimulation setup. First, we introduce the Gnutella traces wecollected. We then describe the data used for evaluationincluding the WT10G data collection from NIST and thequery logs. Finally, we present the metrics used forperformance evaluation

The topology ofa small-world network has the properties of sparseness, short global separation, and high-local clustering of nodeswhile power law denotes the property of the node degreedistribution.

BRITE is a topology generation tool that provides theoption of generating topologies based on the AS model.Using BRITE, we generate a physical topology with 100,000 nodes. Using the physical topology generated by BRITE, we can simulate the underlying Internet with rich configurationinformation, including bandwidth configuration, latency,and so forth.

Using BRITE, we configure the upload bandwidth of apeer according to the measurement study on MSN fromMicrosoft [21] in 2007. The study has shown that 97.2 percentMSN video users have upstream bandwidth higher than128 Kbps (16 KBps).On one hand, this conservativeconfiguration about peer bandwidth capacity indeedpushes the system performance examination close to thesystem limits.

IV. PROPOSED SYSTEM

The method of random walk has been proposed as a practicalalternative to implement uniform sampling [1], [3]. In particular,in several random graph models, the so-called mixing timeof the random walk, which is the number of simulation stepsin order for the random walk to reach a distribution close (forsampling purposes) to uniform, is O(log n). This means thatwe may simulate k uniform samples with O(log n) randomwalk steps for each uniform sample. Since the random walkscan be simulated in parallel, and assuming that the responsedelay of a random walk is proportional to the number ofsimulation steps of the walk, we get maximum response timeO(log n), overhead at most O(k log n), while achieving performancesimilar to uniform sampling. The drawback of thisapproach is the network overhead which scales as O(log n).On the positive side, the theory of cover times [15] [16],complexity theory [17], [18] and extensive experimentation[1] suggest that this overhead can be reduced to a constantby taking O(log n) steps to randomize and then using ksuccessive steps of the random walk in the place of independentsamples. The drawback however is that the approach isinherently sequential and hence introduces maximum response time at least k. The behavior of

V. EXPERIMENTAL SETUP

In this section we describe the analysis of each processing technique in sharing information from one other networks presented data efficiency process generation.

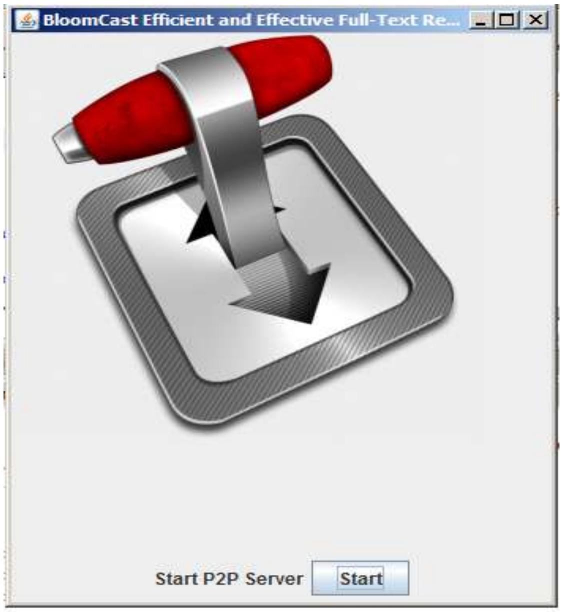

Figure 4: Server setup operations in real time torrent application development.

The above figure show efficient process generation in construction of client and server authentication for accessing services in dependent process of one peer to other peer present in the network progress environment specification in real time network application development. Client sendsa requires to server then server provides efficient process generation in commercial event management in real time application server and client communication for accessing services.

VI. PERFORMANCE RESULTS

In this section we describe process generation in real time application development features in real time data sharing application development. In this paper we propose to develop Random Work Propagation and Compressed Bloom filter.

ID	Name	Size	Status	Speed	Progress	Location	Limit
	A Fast Clustering Based...	10184				C:\Users\SIVAYAM...	
	CL.docx	5985				C:\Users\SIVAYAM...	
	Cross-site application do...	10057				C:\Users\SIVAYAM...	
	Project Installation Steps...	12808				C:\Users\SIVAYAM...	
	User histories Modules d...	12425				C:\Users\SIVAYAM...	

Figure 5: File sharing operations in real time application development.

As shown in the figure 5 we upload different files from subscribed event management process between each client. Each client maintains efficient processing in commercial work load between each client. If we are downloading a file from other files then we want to access services from server and then we process all the integrated operations in real time torrent application environment progression. These services are processed when retrieving all the uploaded files from other peers present in the network process. These results are accessed in commercial event management in real time application present in each peer.

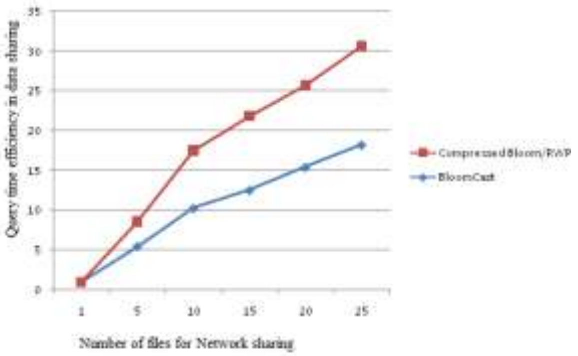

Figure 6: Experimental results for query processing time and data retrieving.

Data required peer searching files list from other peers, even they are present in the overall network progression. These network applications are efficient in searching files in less time with sufficient file sharing. Then server give all file list to searching peer then peer verify each file, if it was required then download those files with relevant application process. In this way we are sharing information from one to other peers network progression with commercial process generation in each data management with commercial searching process in each peer.

VII.CONCLUSION

In a peer-to-peer network, tasks (such as searching for files or streaming audio/video) are shared amongst multiple interconnected peers who each make a portion of their resources (such as processing power, disk storage or network bandwidth) directly available to other network participants, without the need for centralized coordination by servers. For data retrieving traditionally proposed technique is Bloom cast for efficient process generation in event management of data sharing. For doing this data retrieval efficiently, in this paper we propose to develop Random Work Propagation and Compressed Bloom Filters for efficient data sharing between each peer present in network. Further improvement of our proposed we will develop other efficient algorithms for data sharing in peer to peer networks and we also improve different security consideration in confidential data sharing with network coding.

VIII. REFERENCES

[1]. Rüdiger Schollmeier, A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications, Proceedings of the First International Conference on Peer-to-Peer Computing, IEEE (2002).

[2]. Ahson, Syed A.; Ilyas, Mohammad, eds. (2008). SIP Handbook: Services, Technologies, and Security of Session Initiation Protocol. Taylor & Francis. p. 204. ISBN 9781420066043.

[3].Jump up ^ Zhu, Ce; et al., eds. (2010). Streaming Media Architectures: Techniques and Applications: Recent Advances. IGI Global. p. 265. ISBN 9781616928339.

[4].Jump up ^ Kamel, Mina; et al. (2007). "Optimal Topology Design for Overlay Networks". In Akyildiz, Ian F. Networking 2007: Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet: 6th International IFIP-TC6 Networking Conference, Atlanta, GA, USA, May 14-18, 2007 Proceedings. Springer

[5].Filali, Imen; et al. (2011). "A Survey of Structured P2P Systems for RDF Data Storage and Retrieval". In Hameurlain, Abdelkader; et al.

Transactions on Large-Scale Data- and Knowledge-Centered Systems III: Special Issue on Data and Knowledge Management in Grid and PSP Systems. Springer.

[6].Zulhasnine, Mohammed; et al. (2013). "P2P Streaming Over Cellular Networks: Issues, Challenges, and Opportunities". In Pathan; et al. Building Next-Generation Converged Networks: Theory and Practice. CRC Press.p. 99.

[7].Chervenak, Ann; Bharathi, Shishir (2008). "Peer-to-peer Approaches to Grid Resource Discovery". In Danelutto, Marco; et al. Making Grids Work: Proceedings of the CoreGRID Workshop on Programming Models Grid and P2P System Architecture Grid Systems, Tools and Environments 12-13 June 2007, Heraklion, Crete, Greece. Springer.

[8]. Jin, Xing; Chan, S.-H. Gary (2010)."Unstructured Peer-to-Peer Network Architectures".In Shen; et al. Handbook of Peer-to-Peer Networking.Springer.p. 119.ISBN 978-0-387-09750-3.

[9].Jump up to: a b Lv, Qin; et al. (2002). "Can Heterogenity Make Gnutella Stable?". In Druschel, Peter; et al. Peer-to-Peer Systems: First International Workshop, IPTPS 2002, Cambridge, MA, USA, March 7-8, 2002, Revised Papers. Springer.

[10].Shen, Xuemin; Yu, Heather; Buford, John; Akon, Mursalin (2009). Handbook of Peer-to-Peer Networking (1st ed.). New York: Springer.

[11]. Typically approximating O(log N), where N is the number of nodes in the P2P system

[12].D. Li, J. Cao, X. Lu, and K. Chen, "Efficient Range QueryProcessing in Peer-to-Peer Systems," IEEE Trans. Knowledge andData Eng., vol. 21, no. 1, pp. 78-91, Jan. 2008.

[13]. I. Stoica, R. Morris, D. Karger, M.F. Kaashoek, and H.Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Servicefor Internet Applications," Proc. ACM SIGCOMM '01, pp. 149-160, 2001.

[14]. E. Cohen and S. Shenker, "Replication Strategies in UnstructuredPeer-to-Peer Networks," Proc. ACM SIGCOMM '02. pp. 177-190,2002.

[15]. Christos Gkantsidis, Milena Mihail, and Amin Saberi, "Random walksin peer-to-peer networks," in IEEE Infocom, Hong Kong, 2004.

[16].Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker, "Searchand replication in unstructured peer-to-peer networks," in InternationalConference on Supercomputing, New York, New York, USA, 2002.

[17]. Lada A. Adamic, Rajan M. Lukose, Bernardo Huberman, and Amit R.Puniyani, "Search in power-law networks," Physical Review E.

[18] Hanhua Chen, Hai Jin, Xucheng Luo," BloomCast: Efficient and Effective Full-Text Retrieval in Unstructured P2P Networks",IEEE Transactions On Parallel And Distributed Systems, Vol. 23, No. 2, February 2012.