An Intelligent Approach to Query Processing in Peer to Peer Networks

Vishnu V[1], Senthilkumar N C[2]

[1] School of Information Technology and Engineering, VIT University, Vellore, Tamil Nadu 632014, India

[2] School of Information Technology and Engineering, VIT University, Assistant Professor (senior) Vellore, Tamil Nadu 632014, India

Abstract

Peer-to-peer systems are mainly used for object sharing and file sharing, but they can provide the infrastructure for many other applications. These systems need an algorithm to evaluate the queries efficiently so that the best of the results can be achieved in data sharing and message passing. The delay in the query can be caused by scale of the system and the size of the query space. In this project an efficient query processing scheme is proposed to limit the search space of the given query by a domain based query routing mechanism which limits the search to related peers. Therefore, the mechanism consists of domain based architecture for grouping peers in the network into domains based on their contents. In each domain, interconnected groups of peers are constructed based on attributes. To perform the query routing, this mechanism provides the required methods for specifying the interaction between the peers in the same or different groups.

Keywords: *Peer to Peer network (P2P), domain based content mechanism (DBCM), cross domain content algorithm (CDCA), intra domain content algorithm (IDCA)*

1. Introduction

As a fact more than 30% of the internet traffic is caused by peer to peer networks (P2P). The P2P network model is quickly becoming the preferred medium for file sharing and distributing data over the Internet. A P2P network consists of numerous peer nodes that share data and resources with other peers on an equal basis. Unlike traditional client-server models, no central coordination exists in a P2P system; thus, there is no central point of failure. P2P networks are scalable, fault tolerant, and dynamic, and nodes can join and depart the network with ease. The most compelling applications on P2P systems to date have been file sharing and retrieval. It is very necessary to see that many factors like file sharing, file exchange and various search algorithms involved in peer to peer networks can be made more efficient by proper query processing. Hence to contribute to the research of query processing in peer to peer networks an efficient algorithm is proposed to increase the efficiency of the existing mechanisms.

In this paper the following factors are dealt with. Decreasing bounded delay of query results. Flooding, DHT based techniques of search have flaws other more efficient methods are proposed .To find the optimal search path of the query results within the network.

2. Related Works

The researches in the area of query processing in peer to peer networks have started to increase the efficiency of retrieving and exchanging information between peers and file sharing. The following were the areas in which researches took place. Chord is a famous DHT of O(log N) degree. A. Gupta, D. Agrawal [1] proposed a probability scheme to support single-attribute range queries on Chord. They have developed a peer to peer architecture for computing approximate answers for complex queries. However, it can only return approximate results.Dongsheng Li, Jiannong Cao [2], have proposed a delay-bounded general range query scheme called Armanda. An efficient range query processing scheme to support delay bounded single and multiple attribute range queries .The maximum delay is less than 2 log N which is efficient but can be improved. Since DHT is a structured topology, when peers join and leave the network. The whole structure should again be modified. Anis ismail & Mohamed quafafou[3] proposed a super peer search algorithm to discover the domains on the fly. which is required for the domain direct search. It reduces the response time by 35%. The advantage of this technique is that they provide robustness in queries routing. The advantage of this process is that, it avoids to making semantic mapping, between heterogeneous data sources owned by super peers, each time the system decides to route query to other super peers. It improves the robustness in queries routing mechanism and scalability in P2P Network. Ghada Hassan, Hamidah Ibrahim [4] have proposed a query search algorithm which keeps the search costs low. The peers are divided based on their content present in the different peers. All peers having similar data come under one domain. This minimizes the search of the requested query. Taizo Yamada', Kenro Aiharat, Atsuhiro Takasut[5] have proposed an indexing method for query processing in P2P systems. The proposed index stores the usefulness of peers and prevents peers from forwarding messages to less informative peers. As a result, the proposed method reduces the bandwidth consumption and realizes efficient query processing. Christos Gkantsidis, Milena Mihail [6] have proposed random walks on peer to peer system and justified their superiority over the flooding techniques. They are many problems regarding the random walk methods. It is weakly decentralized and there are too many parameters which are considered. The query is randomly forwarded to the peers with respected to the time to live value and the 'k' value which is the number of peers to be sent. Generally this mechanism reduces the network load but increases the search latency compared with the flooding mechanism.

3. Flooding mechanism

The flooding mechanism can be created for a small network. The method of flooding is that the queries are sent to the immediate neighbors and in turn passed on to the entire network. It will have low maintenance cost, easy to handle membership, data content changes. Therefore in this paper the flooding mechanism is compared with an intelligent query routing. And both of the mechanisms are compared.

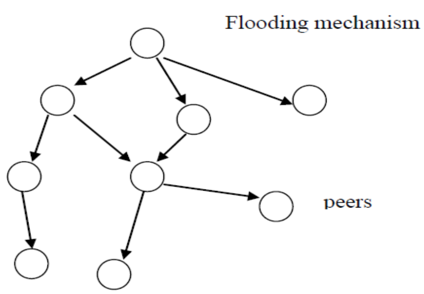
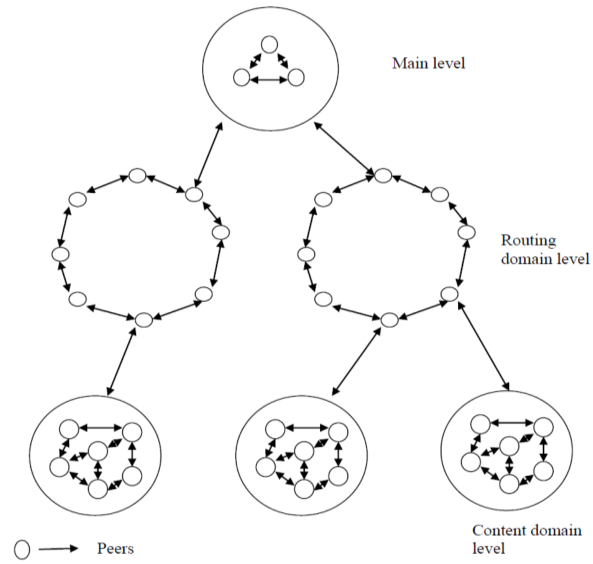


Fig. 1 Flooding Mechanism

4. Domain based content mechanisms

It has three components content domain level, routing domain level, main level



4.1 Content domain level

This level is composed of normal nodes in which the data in the network is distributed. There are logical grouping in the level where domains are formed by logical grouping of the nodes. Each domain consists of a header node which is present in the routing domain level. The header of each domain is called the super peer.

4.2 Routing domain level

In this level, the headers of different domains are grouped into super domains which form the routing domain level. The nodes are connected in ring topology

4.3 Main level

This level contains the main peers. The different rings of the routing domain level are linked together through the main peer. It acts as a intermediate between them. The main level is connected to user interface where the user query is given. The main peers are elected by bully algorithm.

5. Intra domain content algorithm (IDCA)

In this algorithm the search for the requested object is done within the respective domain. Each domain has its own schema. And the schema is represented by its attributes. When the request is send the domain header pass the query within the domain. If the required result is not achieved the query is passed on to the domain header which passes it onto the other domains.

If A1…A10 are attributes then these attributes can be further classified into groups. As shown in the figure below. Each group has a group coordinator which connects to the domain header. So the peers in the same domain can be in different groups so that the query routing takes place with reference to the groups. Based on the information deduced by the peers the results are returned. The group coordinator is elected through bully algorithm

Algorithm:

The query received by the peer is tested with its respective contents or attributes.

- If the answer is present in the same domain

Then

- It finds the attribute groups for which the query is to be executed.
- It forwards the query to the corresponding group headers.

Else

- It returns the query to the domain headers.

5. Cross domain content algorithm

In this algorithm the search for the requested object is done in super domain level. The headers of all domains are present in the super domain level. All the headers are arranged in a ring topology. When the required object is not obtained within certain domain this algorithm takes place. The header peers arranged in ring topology should not exceed NN, where 1<NN<n. and n is total number of headers in each ring. The headers in the super domain level have the knowledge of only the direct neighbors, the successor and the predecessor. It has two types of links, the direct link and the partner link. The direct link links its direct neighbors while the partner link links the header with its partner header. The message passing between these links takes place in opposite direction.

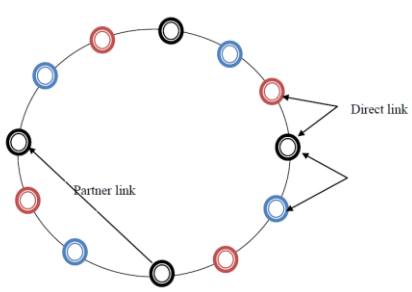


Figure 2 cross domain routing

The notation for direct successor (DS) and the predecessor (PS) are calculated using this notation. The partner link (PL) is also defined:

DS=NN+1mod n, 1<NN<n

DP=NN-1mod n, 1<NN<n

PL=NN+3mod n

Maximum number of hops in this level (H) = trunck((n-2)/3), where n is number of nodes in the super-domain ring. To ensure that the query forwarding does not exceed „H" an descending factor (DF) is proposed. Both the values of „H" and DF are made equal. The value of DF is decreased by 1 after every forwarding through the ring. therefore if DF>0 the query can be forwarded to the next partner. If DF<=0 the query should be returned to the main peer.

Algorithm:

The source of the query is determined by each header.

If query source = main peer,

Then

The header checks the query against its local schema(if true)

The query is forwarded to the underling peers

Else

It checks for answers in the neighbor peers against its local knowledge.(if true)

It directly forwards it to the specific direct neighbor link.

Else

It forwards to the partner link.

If the source query = one of its underlying peers,

Then

The header checks it against its local knowledge about the neighboring schema(if true)

It forwards to the specific direct neighbor

Else

It forwards to the partner link.

If the source of the query = one of its neighbors,

Then

The header checks against its local schema(if true)

The header forwards it to the underlying peers

Else

It checks for answers in the neighbor peers against its local knowledge.(if true)

It directly forwards it to the specific direct neighbor link.

Else

It forwards to the partner link.

6. Handling the crashed peers

In such a connect network if any peer crashes out. They will be trouble in looking for the schema in the respective domains. The peers in the super domain level are very important as it connects both the main level and the super domain level. When a peer in the super domain level crashes out the bully algorithm is performed in the underlying domain level. The bully algorithm is as follows

- It uses time out to detect the crash in the network.
- The election is done based on the processor identifier number.
- When the peer detects the crash of any peer it broadcasts a election message to all the peers with high processor identifier number.
- If it does not receives any acknowledgement. It broadcasts to all the peers as the leader.

7. Implementation

The above proposed algorithms was done using XAMPP control panel where we have distributed the data between the peers. And the data retrieval is done using local host server. Using the above technology we have derived a statistics on the above algorithms

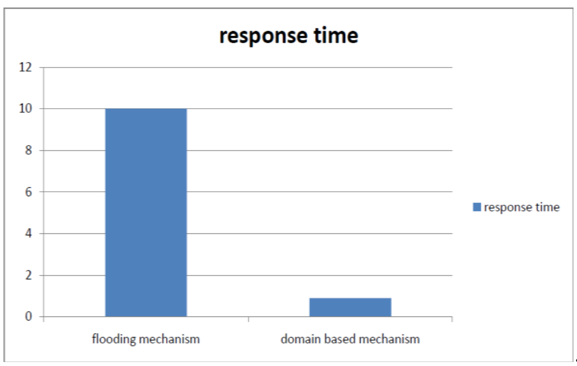


Figure 3 comparisons of algorithms

8. CONCLUSION

In this paper, query response time is compared with the existing flooding mechanism and proposed domain based content mechanism. The response time of the domain based algorithm is decreased which helps in query search in peer to peer networks. This algorithm can applied in peer to peer application.

REFERENCES

[1] A. Gupta, D. Agrawal, and A.E. Abbadi, “Approximate Range Selection Queries in Peer-to-Peer Systems,” Proc. First Biennial Conf. Innovative Data Systems Research (CIDR “03), Jan. 2003.

[2] Dongsheng Li, Jiannong Cao, Senior Member, IEEE, Xicheng Lu, and Keith C.C. Chan “Efficient Range Query Processing in Peer-to-Peer Systems” IEEE transactions on knowledge and data engineering, vol. 21, no. 1, pp. 78 – 90 , January 2009.

[3] A. Ismail, M., Quafafou, G., Nachouki, and M. Hajjar, "Efficient super-peer-based queries routing". MEDES '09. ACM, pp. 91-98, 2009

[4] Ghada Hassan, Hamidah Ibrahim, Md.Nasir Sulaiman, Rasali Yakob, “A Domain-based query routing mechanism for peer-to-peer Networks” The 8th International Conference on Mobile Web Information Systems (MobiWIS-2011) from Elsevier in science direct 2011.

[5] Taizo Yamada', Kenro Aiharat, Atsuhiro Takasut, Jun Adachit,”A Distributed Index System for Efficient Query Processing in Peer-to-Peer Networks”, IEEE transaction on data engineering, pp.139-142, 2003

[6] C. Gkantsidis, M. Mihail, and A. Saberi, Random walks in peer-to-peer networks,Proc IEEE INFOCOM 06, vol. Pp.120-130, 63(3):241-263, 2006.

Author profiles

Vishnu V received B.E degree in computer science and engineering in Anna University, Chennai, India in 2009. Currently he is pursuing M.Tech degree in Information Technology from School of Information Technology and Engineering of VIT University, Vellore, India. His research interests include Network Programming and Database Management.

Senthilkumar N C received ME (CSE) from Vellore Engineering College, Vellore, India. The Author currently works as Assistant Professor in VIT University and doing research work in Spatial Databases.