

Query Indexing on Spatial Data in Peer to Peer Systems

Bhuwan chandra^{#1}, Chandan Pal Singh^{#2} Kamal Deep^{#3}

^{#1}Department of Computer Science, Singhania University, Pachari Bari, Jhunjhunu, Rajasthan, India

^{#2}Computer Science, Jagannath University, Jaipur, Rajasthan, India

^{#3}Computer Science from Jagannath University, Jaipur, Rajasthan, India

¹bcd_hald@yahoo.com

²chandanengg@gmail.com

³kamal.thelotus@gmail.com

Abstract --Peer to peer technology plays an important role in spatial data sharing because of its distributed and decentralized nature. In spatial P2P each peer stores a part of the available spatial information and runs software with equivalent functionality rather than through a central server. Spatial P2P is dynamic in nature where the network structure is always changing. Spatial objects are having distinct properties like identity, location and relationship among them and spatial data is multidimensional. Spatial data sharing process like storing, querying and retrieving large amounts of data is a complex task. It requires the potential of preserving multidimensionality of heterogeneous spatial data without the time complexity and load balancing. It can be simplified if some intelligent techniques will be used. The main objective of this paper is to analyze the process of preserving the locality and dimensionality of space in the network. It maintains the multidimensionality of the spatial data in grid approximation which is done by some intelligent indexing technique and searching .

Keywords-- Spatial data, P2P system

I INTRODUCTION

Spatial data is the geographic data which is related to space. Spatial database management system is used to manage the spatial data. Because of the underlying peer to peer architecture there is no central server. They have direct links with other nodes in the network. Basically in spatial peer to peer system each node has its own database and capable of sharing spatial information of different kind. They are all independent nodes and are maintaining the databases on their own. A peer can join or leave the network whenever there is a need. So the network structure is always changing.

Spatial objects have some specific properties like identity, location, and relationship among them. Spatial relationship is distinguished as topological, direction and metric relationships. Spatial database systems offer the underlying database technology for geographic information systems and other applications. In spatial peer to peer database we not only save one dimensional data but multidimensional also. Spatial data handling is a difficult task because of its multidimensionality. Spatial querying is still a complicated one because of the spatial object properties. Locality stores the dimensional

information and directionality preserves the structure orientation. The corresponding indexing structure explains the locality and directionality. Usually Distributed Hash Tables (DHT) are used for indexing. The One dimensional DHT uses a distance metric to map the multidimensional information into a single dimension. So there is no preservation of multidimensionality. To preserve the multidimensionality and lessen the load balancing a decentralized framework is proposed by Verena Kantere, Spiros Skiadopoulos, and Timos Sellis which is funded by the project PENED 2003 multidimensional spatial data. It manages the changing network structure without the need of load balancing and simplifies the process of storing, querying, indexing and searching of any spatial data with low space and time complexity.

II PEER TO PEER SYSTEMS

Peer to Peer networks are the efficient and effective resource sharing systems. P2P file sharing is a way to share any type of files with someone else anywhere in the world. The peer to peer network may be structured or unstructured. Structured P2P networks uses distributed hash table (DHT) techniques to ensure that any node can efficiently process a search to some peer that has the desired file. DHT contains (key, value) pairs that provides a lookup service for information. Here the data retrieval can be done effectively because the key is associated with the value. Here the peers are responsible for mapping the keys to values to maintain the continual changing in network structure. But in unstructured peer to peer networks exact look up operation is not possible and it causes query flooding. So exact lookup operations are possible in pure peer to peer systems. Some of the systems are Gnutella, and Freenet. Gnutella is purely decentralized so there is no central point of failure. Each peer stores files and index values and routes queries from and to its neighboring peers. It has its own disadvantages like query flooding and reverse path forwarding. It hasn't any intelligent indexing technique to overcome this. This causes problems in maintaining network structure. Napster is the example of the hybrid peer to peer system. Napster has the centralized directory which is used to manage the network structure. It contains every

information about file and hosts in the system. Each peer establishes a connection with the central node when they need the information and can break the connection when the process is over. Here exact lookup is possible. If the central system fails then the whole network will be affected. It is not suitable for dynamic peer to peer network system. So to overcome the high overhead in query processing and central point of failure we are proposing a frame work which contains super nodes to minimize the query flooding and the possibility of errors in order to increase the efficiency in query processing and network maintenance.

III SPATIAL P2P FRAMEWORKS AND OPERATIONS

The proposed spatial peer to peer framework is a grid based spatial data model. The space is partitioned by a grid. Each cell is identified by (x, y) coordinates. The Euclidean distance metric is used to measure the distance between two cells in the grid.

For example consider two cells in the basic grid are A(1,2) and B(3,6), then the distance between them is denoted as $D(A,B) = (4,2)$. The distance is calculated by the maximum and minimum coordinates of the cells.

$$D(C,C1) = (d1,d2)$$

Where

$$d1 = \max\{|cx-cx1|, |cy-cy1|\}$$

$$d2 = \min\{|cx-cx1|, |cy-cy1|\}$$

7								
6				B				
5								
4								
3								
2		A						
1								
0								
y/x	0	1	2	3	4	5	6	7

Table 1. Euclidean distance metric

In spatial P2P, each peer maintains links with successors and indexed peers. They guarantee the connectivity of the network. Thus, successors assure that there is always a path between two arbitrary peers in the network. Moreover, successors preserve the locality and the directionality of the 2D space in the P2P network. This allows the application of traditional spatial query evaluation techniques [21]. To speed up searching, a peer also stores the location of additional peers. Such peers are called indexed peers. Here we require that each peer p should have 4 successors s1,s2,s3,s4 where each successor lies in a different quadrant.

Successor s1 is the closest peer to p with respect to the D distance that lies in Group 1 or 2 of p. Successor s2 is the closest peer to p with respect to the D distance that lies in Group 3 or 4 of p. Successor s3 is the closest peer to p with respect to the D distance that lies in Group 5 or 6 of

p. Successor s4 is the closest peer to p with respect to the D distance that lies in Group 7 or 8 of p.

Ties between peers are solved similar to hashing. If there are no peers in one of the above pairs of groups, the corresponding successor does not exist. If there are no peers in Groups 1 and 2 of p, then s1 does not exist, if there are no peers in Groups 3 and 4 of p, then s2 does not exist, if there are no peers in Groups 5 and 6 of p, then s3 does not exist, and if there are no peers in Groups 7 and 8 of p, then s4 does not exist.

When a peer p is seeking information about cell A then p checks if this information is stored in p. If it is not there then p checks whether the information is stored in its successors or its indexes peers. If the requested information cannot be found, p forwards the search to a peer that is closer to A. The time complexity is $O(\log n)$, where n is the number of peers in the network.

IV SPATIAL P2P QUERY PROCESSING

4.1. NN-Queries

Nearest neighbor querying involves finding the spatial objects that are nearest to a query location. Here we are using a technique called peer tree to do this. We frequently need to find the spatial objects closest to the query point. In NN queries, we are finding the nearest spatial objects using the Euclidean distance between the object and the query point. NN query is first initiated on a single peer in the P2P network. In traditional database systems, NN queries are executed always starting from the root of that tree, i.e., the top-most level clustering; in contrast, in a peer to-peer environment any node should have the capability of initiating the query. In our peer-tree structure, data is stored only at the physical nodes, i.e., nodes that are at level 0 of the hierarchical clustering. In our peer-to-peer NN query execution model, we are able to limit an NN query only to the relevant nodes and can prevent it to propagate till up the root (i.e., topmost) partition. If the query can be answered locally, there is no need to go up till the root partition and query the root. By limiting the query to only the relevant nodes, we conserve energy.

4.2. Range Queries

Here the data objects that fall into a query region is asked. Similar to the NN query, the query can be initiated from any node of the peer-tree and the processing needs to be developed in a distributed manner.

4.3. Constrained Nearest Neighbour Query (CNN)

CNN queries are recently formulated in the context of database systems. This type of query is targeted towards users who are particularly interested in the nearest neighbor(s) in a region bounded by certain spatial conditions, rather than in searching for nearest neighbors in the entire data space. This can be seen as performing nearest neighbor and range queries in a single query.

4.4. Multi-dimensional Queries

Besides the location-based queries, the proposed framework can be utilized for other multi-dimensional

queries over sensor networks. One such example could be environmental monitoring by running continuous queries asking temperature, etc. Each of the sensed attributes can be defined as a dimension in multi-dimensional space and the peer-tree can be constructed as a multi-dimensional index structure. The dynamic nature of the system should be carefully addressed while building such an index structure.

V CONCLUSION

In this paper the framework SPATIALP2P which they have proposed maintains the multidimensionality of the spatial data. SPATIALP2P supports dynamic insertion and deletion of spatial information of various sizes and dynamic joining and leaving of peers. The peer tree indexing structure performs the search process very efficiently without the need of load balancing and provides efficient storing, indexing, and searching services by preserving locality and directionality. As a result, SPATIALP2P performs exceptionally well for point and range query operations.

REFERENCES

- [1]. Verena Kantere, Spiros Skiadopoulos, and Timos Sellis. Storing and Indexing Spatial Data in P2P Systems. *IEEE Transactions on Knowledge and data engineering Vol 21, No 2 February 2009*.
- [2]. N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R* tree: An efficient and robust access method for points and rectangles. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, 1990.
- [3]. K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is "nearest neighbor" meaningful. In *Int. Conf. on Database Theory*, pages 217–225, Jerusalem, Israel, January 1999.
- [4]. N. Bulusu, J. Heidemann, and D. Estrin. Adaptive beacon placement. In *Proceedings of the 21st International Conference on Distributed Computing Systems*, pages 489–498, 2001.
- [5]. X. Cheng, R. Dolin, M. Neary, S. Prabhakar, K. Ravikanth, D. Wu, D. Agrawal, A. El Abbadi, M. Freeston, A. Singh, T. Smith, and J. Su. Scalable access within the context of digital libraries. In *IEEE Proceedings of the International Conference on Advances in Digital Libraries, ADL*, pages 70–81, Washington, D.C., 1997.
- [6]. Y. Choi, M. Gouda, M. C. Kim, and A. Arora. The mote connectivity protocol. Technical Report TR03-08, Department of Computer Sciences, The University of Texas at Austin, 2003.
- [7]. A. Cournier, A.K. Datta, F. Petit, and V. Villain. Self-stabilizing PIF algorithms in arbitrary networks.