EFFICIENT APPROACH FOR IMPLEMENTING SPATIAL DATA IN PEER TO PEER (P2P) SYSTEMS

Kiruthika S[1], Pranitha S [2]

[1] Assistant Professor , Dept. of CSE , Rajiv Gandhi Institute of Technology, Karnataka, India

[2] Assistant Professor , Dept. of CSE , Rajiv Gandhi Institute of Technology, Karnataka, India

ABSTRACT

Spatial data that identifies the geographic information and boundaries on Earth, such as natural or constructed features, oceans, maps etc. P2P systems provide a distributed platform that is very scalable. Peer-to-peer (P2P) computing or networking is a distributed application architecture that partitions tasks or workloads between peers. In a computer network, the P2P systems are very useful in sharing the one dimensional information. But nowadays, the need for P2P applications with multidimensional data has risen, inspiring research on P2P systems that can handle such type of data. Now the point is to share Spatial information and found the area and directionality of the multidimensional information. The peer-to-peer (P2P) paradigm has become very popular for storing and sharing information in a totally decentralized manner. At first, research focused on P2P systems that host 1D data. Nowadays, the need for P2P applications with multidimensional data has emerged, motivating research on P2P systems that manage such data. The majority of the proposed techniques are based either on the distribution of centralized indexes or on the reduction of multidimensional data to one dimension.

Keyword :- *Peer-to-peer, structured overlays, distributed hash tables, Spatial data.*

1. INTRODUCTION

PEER-TO-PEER (P2P) networks are self-configuring networks with minimal or no central control. P2P networks are more vulnerable to dissemination of malicious or spurious content, malicious code, viruses, worms, and Trojans than the traditional client-server networks, due to their unregulated and unmanaged nature. The peers in the P2P network have to be discouraged from leeching on the network. Policing these networks is extremely difficult due to the decentralized and ad hoc nature of these networks. P2P networks can be categorized into structured and unstructured P2P networks .Besides, P2P networks, like the Internet, are physically spread across geographic boundaries .The traditional mechanisms for generating trust and protecting client-server networks cannot be used for pure1 P2P networks. This is because the trusted central authority used in the traditional client-server networks is absent in P2P networks. P2P system is a distributed environment formed by autonomous peers that operate in an independent manner. Each peer stores a part of the available information and maintains links (indexes) to other peers. P2P systems provide a method to distribute the available information to peers, guarantee the retrieval of any information that exists in the system, achieve a reasonable index size for all peers, achieve a reasonable search path for any search performed in the system, maintain a low cost for updating peer indexes when peers join or leave, and achieve data and search load balancing, i.e., there are no peers overloaded with stored data and there are no traffic bottleneck in the network. Until recently, research has focused mostly on P2P systems that handle 1D data such as strings and numbers. However, the need for P2P applications that manage multidimensional data has emerged. These systems pose additional requirements that stem from the particularities of such data.

In centralized multidimensional applications, information is stored according to its multidimensional extent using an indexing structure. Typically, these structures preserve the locality and the directionality of multidimensional information. Intuitively, locality implies that neighboring multidimensional information is stored in neighboring nodes, while directionality implies that the index structure preserves orientation. Typically, these structures preserve the locality and the directionality of multidimensional information. Intuitively, locality implies that neighboring multidimensional information is stored in neighboring nodes, while directionality implies that the index structure preserves orientation. The notions of locality and directionality are very important. If an index structure preserves these properties then searching in the index corresponds to searching in the multidimensional space which can highly improve query evaluation cost In the literature, there exist two different paradigms for handling multidimensional data in P2P systems. The first paradigm proposes the use of a distributed version of some centralized multidimensional index. The main challenge of these approaches is to avoid the bottleneck .Occurring at the peer storing the root of the tree (since every search has to pass through this peer). The work in alleviates the traffic burden of the root by initiating the search at a prespecified level below the root. Moreover, the work in proposes sideway index links that help to avoid searching the index from the root. The second paradigm maps the multidimensional data into a single dimension and uses a Distributed Hash Table. Briefly, 1D DHT techniques use distance metric to define the locality of the 1D data. The main challenge of these approaches is to preserve the important properties of multidimensional space (i.e., locality and directionality).

Unfortunately, space-filling curves do not always preserve locality and directionality. For instance, two multidimensional regions that are close in the original space are not necessarily close in the z-ordering curve. The searching strategy of the P2P network inherits these locality problems of the 1D ordering. The proposed SPATIALP2P, a totally decentralized framework for Spatial (i.e., 2D) data that conforms to the autonomy principle of P2P networks.SPATIALP2P provides storing, indexing, and searching services for Spatial data in a P2P network. SPATIALP2P exploits existing experience in the field of DHTs and can be built on top of any 1D DHT. Thus, SPATIALP2P distributes the spatial information to peers and guarantees the retrieval of any Spatial area that exists in the system with low space and time complexity. Additionally, SPATIALP2P efficiently handles changes in the Spatial information and in the network structure caused by joining or leaving peers without the need of load balancing or restructuring. Finally, contrary to existing DHT-based approaches, SPATIALP2P captures the locality and directionality of the 2D space. To achieve these goals, SPATIALP2P undertakes a different approach and does not use a global ordering like1D DHT methods. In SPATIALP2P each peer stores data and indexes peers that are close in the 2D space.

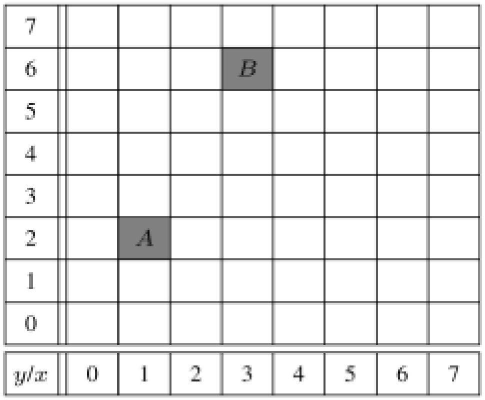

Fig-1: **SPATIAL partition the 2D spa**ce

2. THE SPATIAL P2P FRAMEWORK

SPATIAL P2P uses a grid to partition the space and defines an appropriate distance. Moreover, we discuss how Spatial information is assigned to peers, how we search for Spatial information, how new peers join the system, and how peers leave the system. Hashing in SPATIAL P2P requires that a cell A for which there is no peer with the same identifier is stored in a peer with the closest (to cell A) identifier with respect to the D distance. Ties (i.e., cases where cell A has the same D distance from more than one peers) are solved according to the group ordering. If cell A has the same D distance with more than one peer, we assign it to the peer that lies in the group of A with the smallest number. RANGE SEARCH algorithm can be further optimized as follows: When visiting a successor or an indexed peer, we gather all information concerning C (and not only information of the requested cell). By doing so, we do not have to visit this peer multiple times. The time complexity of RANGE SEARCH algorithm. where k is the number of cells of the search area and n is the number of peers. Notice that this is a worst case bound that corresponds to very rare configurations. In practice, the optimizations of RANGES EARCH algorithm result in a much better average performance. This is also verified by the experimental results of Section 3.Updating peer indexes. In several situation, a peer needs to update its links to other peers (indexed peers). To safely leave the network, a peer p has to assign its stored data to appropriate peers. To this end, peer p communicates with peers in its neighbourhood. Notice that p does not have to notify other peers that it is leaving the network (not even peers having p in their successors or index peers list). For instance, assume that p leaves the network and it is a successor of p host. The first time that p host will use the link to p (due to a search), it will discover that p is not available. Thus, post will initiate the index update operation to find an appropriate replacement forward the request to. Replication and node failure. Let us consider a peer p that has exceeded its processing limits. To share the processing cost, we may replicate the information stored in p to other peers, called p's replicas, which can be used instead of p. In other words, any peer that refers to p, it may use any of p's replicas. Replica peers index each other so that if the chosen replica is overloaded, it forwards (using its links) the request to another replica. Therefore, the processing cost is amortized among replicas. Node failure is handled in a straightforward way in the SPATIAL P2P framework. Periodically, every peer p exchanges messages with its successors and its indexed peers to reassure the presence of each other. When a peer is not responding to these messages within a certain amount of time, we assume that this peer is disconnected from the network. Thus, p initiates a stabilization procedure to replace the disconnected peer.In order to hash areas to peers, the space of peer identifiers must be the same as the space of area identifiers, i.e., a peer is also identified by a quadruple. Thus, we may treat peers as areas. Similarly to the basic grid, hashing in SPATIALP2P requires that an area A for which there is no peer with the same identifier is stored in a peer p.

3. EXPERIMENTAL EVALUATION

The hashing, indexing, and searching algorithms of spatial P2P.Methodology. As discussed there are many methods that handle multidimensional data in distributed and P2P networks. It decided to compare SPATIALP2P with methods that adhere to the same P2P principals as SpatialP2P (e.g., peers are autonomous and free to decide their identifier and the data they store). Such methods rely on 1D DHT techniques that use space-filling curves to identify multidimensional data. Thus, we compare the SPIIALP2P framework against the widely approved Chord method appropriately modified to store Spatial data. Thus, to store Spatial information in z Chord, we have to map the 2D area into one dimension. To this end, we use z-ordering.z-ordering is a continuous fractal which maps each area of a multidimensional space to an integer. In our case, the coordinates of grid cells are mapped to a 1D value. Fig. 11a,for instance, shows the curve for the cells of a 4 _ 4 grid.In chords, z-order data are hashed on a Chord ring [25].Chord is a DHT method, where peers are organized in a ring (or torus) instead of an array. Furthermore, in the second set of experiments, evaluated Spatial P2P for the general case that stores arbitrary sized areas. In this case, range queries may retrieve Spatial areas of any size. It used a similar setting with the first set of experiments (i.e., networks of 5,000 and 10,000 peers and range queries asking for areas using a square window

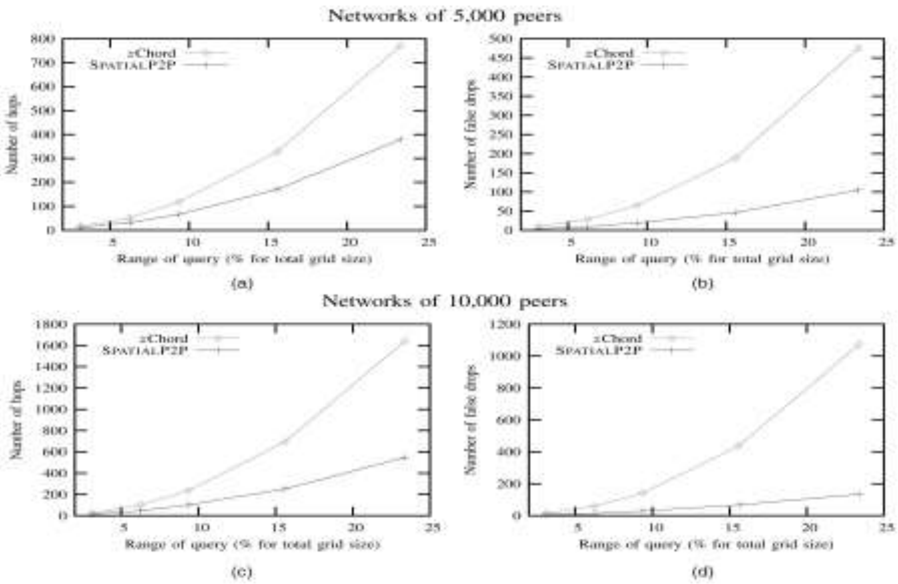

Chart - 1 : Ranges of queries

Individual searches. This is the simplest method. The peer p that wants to update its index list initiates individual searches for each index entry.. Inherit from four neighbours. In this method, the peer p that wants to update its index list inherits the index entries from its four successors. These peers serve as estimations to the actual peers and may be used to significantly prune the search space.. Inherit from eight neighbours. This method is similar to the above method, but this time the peer p that wants to update its index list, inherits the index entries from eight neighbours (instead of four). These neighbours are formed by p's four successors and four peers that each one is a successor of one of p's successors. Spatial information often spans into more than one cell of the grid. Thus, areas that are formed by groups of cells must be represented. Spatial P2P stores information about any group of cells forming a rectangular area in the grid. Information about areas of arbitrary size is stored by their minimum bounding rectangle. A rectangular area A in a grid Individual searches. This is the simplest method. The peer p that wants to update its index list initiates individual searches for each index entry. Inherit from four neighbours. In this method, the peer p that wants to update its index list inherits the index entries from its four successors. These peers serve as estimations to the actual peers and may be used to significantly prune the search space. DHTs characteristically emphasize the following properties,

1. **Decentralization -** The nodes collectively form the system without any central coordination.
2. **Fault tolerance** -The system should be reliable (in some sense) even with nodes continuously Joining, leaving, and failing.
3. **Scalability** - The system should function efficiently even with thousands of nodes.

In Indexing, speed up searching each peer p(x,y) stores the location of a set of indexed peers having progressively increasing distance from p. If an indexed peer q is not present in the network, then p indexes the closest peer with respect to Dz distance instead. Ties between peers are solved similarly to hashing.Consider a peer p seeking information about cell A(ax,ay). Initially, p checks if this information is stored in p. Then, p checks if information regarding A is stored in its successors or its indexes peers. If the requested information cannot be found, p forwards the search to a peer (from its successors or indexed peers) that is closer to A.

4. CONCLUSIONS

The techniques that have been proposed until now belong to two broad categories. Techniques in the first category are based on the idea that 1D index can be reused in order to manage multidimensional data, if the dimensionality is reduced to one. This idea was the first to be explored. Techniques in the second category are based on the idea that centralized hierarchical indexes can be reused to manage dispersed multidimensional data, if they are properly distributed. More elaborated solutions have been proposed in this category. However, the restage of existing techniques in the approaches in both categories leads to the maintenance of some fundamental features that oppose to the nature of either the distributedness or the multidimensionality. Our intention is to overcome these shortcomings by creating a technique that manages disperse multidimensional data in an inherently distributed way without altering the dimensionality. It focused on Spatial data presented the SPATIALP2P framework for handling Spatial data in a P2P network. SPATIALP2P 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. SPATIALP2P supports dynamic insertion and deletion of Spatial information of various sizes and dynamic joining and leaving of peers. In the future, we intend to adjust and test the SPATIALP2P framework for data of higher dimension.

5. REFERENCES

[1] F. Banaei-Kashani and C. Shahabi, "SWAM: A Family of Access Methods for Similarity-Search in Peer-to-Peer Data Networks," Proc. 13th ACM Int'l Conf. Information and Knowledge Management (CIKM '04), pp. 304-313, 2004.

[2] A. Bharambe, M. Agrawal, and S. Seshan, "MERCURY: Supporting Scalable Multi-Attribute Range Queries," Proc. ACM SIGCOMM '04, pp. 353-366, 2004.

[3] M. Cai, M. Frank, J. Chen, and P. Szekely, “MAAN: A Multiattribute Addressable Network for Grid Information Services,” Proc. Fourth Int’l Workshop Grid Computing (GRID ’03), pp. 184-\ 191,2003.

[4] W. Cai, S. Zhou, L. Xu, W. Qian, and A. Zhou, “c2: A New Overlay Network Based on CAN and Chord,” Proc. Int’l Conf. Grid and Cooperative Computing (GCC ’03), pp. 42-50, 2003.

[5] M. Demirbas and H. Ferhatosmanoglu, “Peer-to-Peer Spatial Queries in Sensor Networks,” Proc. Third Int’l Conf. Peer-to-Peer Computing (P2P ’03), pp. 32-39, 2003.

[6] P. Druschel and A. Rowstron, “PASTRY: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems,”Proc. 18th IFIP/ACM Int’l Conf. Distributed Systems Platforms(Middleware ’01), pp. 329-350, 2001.

[7] P. Ganesan, B. Yang, and H. Garcia-Molina, “One Torus to Rule Them All: Multi-Dimensional Queries in P2P Systems,” Proc.Seventh Int’l Workshop Web and Databases (WebDB ’04), pp. 19-24,2004.

[8] A. Guttman, “R-Trees: A Dynamic Index Structure for Spatial Searching,” Proc. ACM SIGMOD ’84, pp. 47-57, 1984.pp. .