RESEARCH ARTICLE OPEN ACCESS

A Survey on Handling Multi-Attribute Range Queries in Peer2Peer Grid Systems

G ShanmugaPriya[1], G Saranya[2]
ME, PG Scholar[1&2]
Department of Computer Science and Engineering
SVS College of Engineering[1]
Dhanalakhmi Srinivasan College of Engineering[2]
Coimbatore
Tamil Nadu-India

ABSTRACT

This paper focuses on how the queries are handled in distributed peer to peer systems. The queries are generated for different kinds of resource attributes by the node and it is routed towards the nodes in the network to find the appropriate results using different search algorithms.varoius search algorithms are proposed for grid systems by considering different parameters such as node heterogeneity, efficiency, scalability .There are different kinds of centralized as well as the decentralized approaches like the gianduia, mercury, MAAN .The proposed algorithms are discussed here and a comparison is made to find the effectiveness of these algorithms.

***Keywords*:-** Peer to peer grid systems, distributed hash table, scalability, multi attribute range queries, node heterogeneity.

I. INTRODUCTION

Grid computing is one of the novel approaches for solving large scale problems in science, commerce, engineering [1].It has the capability to merge the computers, clusters, instruments and storage systems. The resources in the grid system are needed by applications and so it is described by the multi attribute range queries. Hence the query constitute a set of attributes such as the available computing power, bandwidth memory .The ranges of each attribute is also specified. According to the attribute inputs the resource discovery will locate the resources. For large scale and to support dynamic resource discovery, an architecture was introduced called the distributed hash table or simply DHT [2], compared to the previous peer to peer systems it provides good scalability and enhanced dynamism resilience features .The information about the node resources are stored by the DHT server, and also receiving the resource queries and sending the requested resource information are done by the server. The single DHT decentralized approach support the range queries.

Hence different approaches has been proposed to handle the query searching are discussed as follows,

The first category proposed the gianduia algorithm that makes four simple changes to the previously proposed gnutella algorithm to achieve scalability, since gnutella faces some problems like it cannot able to handle high aggregate query rate where the nodes become overloaded.[3] When the system size get increased, the problem become more worse. The changes to the algorithm included are topology adaption, flow control, one-hop replication and node heterogeneity. Hence these contributions make improvement in the total capacity of the

System .The increased capacity make the searching much more scalable.

The second category called mercury [4] is a scalable protocol that focuses on supporting multi attribute range based searches as well as explicitly performs well in load balancing. It provides efficient routing by using novel light-weight sampling mechanisms. The random sampling in this method performs node count estimation and query selectivity estimation. Some of the existing systems supports only range based queries [5][6][7],where the mercury supports multi attribute range based queries. The algorithms included are low overhead random sampling algorithm, load balancing algorithm.

The third category called multi –attribute addressable network (MAAN) was proposed to extend the DHT based protocol called CHORD [8] where initially it perform well only for the single-key based -registration and lookup services for decentralized resources. The extension of proposed CHORD supports range queries as well as the multi-attribute based lookup. The algorithms included are locality Preserving hashing and recursive multi-dimensional query resolution mechanism.

Hence all the proposed algorithms aimed at improving the query searching more convenient across the nodes for multiple attributes but with performance variance.

The rest of the paper is structured as follows; Section 2 describes the related work. Section 3 describes the proposed algorithms. Section 4 presents the comparison among the proposed algorithms and section 5 presents the conclusion.

II. RELATED WORK

A. *Research study*

The researchers have done wide-ranging measurement studies of P2P infrastructures .Saroiu et al [8] have studied the node availability, latency, bandwidth in the protocols like Gnutella,Napster.From the study they found the existence of the node heterogeneity in Gnutella as well as Napster.Napster is a centralized search algorithm where gnutella is a decentralized search algorithm. Initially Napster was the system which found that most popular content need not to be sent to the central server instead of that it can be handled by more no of the nodes or peers that already holds the required content. It made centralized search facility which is not so effective. so the decentralized approach called gnutella Was introduced which can able to distribute both the search and download capabilities. The mercury protocol creates a routing hub to handle multi-attribute queries for each of the attribute. Every routing hub is a logical collection of nodes Queries are passed to the corresponding hub for each attributes. so this conclude that queries can able to retrieve the corresponding data items present in the system.

The range queries in MAAN extends chord that assigns each node and key as m-bit identifier using SHA1 hashing function for mapping keys to the nodes

The number of routing hops to resolve a query is,

$$O(\log N + N \times S_{min}) \qquad (1)$$

Where s_{min} is the minimum range selectivity on all attributes, and also it can able to scale well in the number of attributes

$$s_{min} = \varepsilon \qquad (2)$$

The number of routing hops is logarithmic to the number of nodes.

III. LITERATURE REVIEW

The proposed algorithms are reviewed to identify the ultimate aim of each of the algorithms .The algorithms considered are dynamic topology adaption algorithm, flow control algorithm, one-hop replication algorithm, search algorithm, low-overhead random sampling algorithm, load balancing algorithm, locality preserving hashing and recursive multi-dimensional query resolution algorithm.

B. *Dynamic Topology Adaption*

The aim of the topology adaption algorithm is to ensures that the nodes holding higher degree based on their capacity handles a large portion of the queries and the low capacity nodes are within short reach of the higher capacity nodes .Hence each node computes individually *a level of satisfaction*, say (**S**).The quantity is taken between 0 and 1.It represents the node satisfaction within its current set of neighbours

There are two conditions considered,

- A value S=0 represents a node is fully dissatisfied.
- A value S=1 represents a Node's full satisfaction with its current neighbours.

C. *Flow Control*

The goal of the flow control algorithm is to avoid overloading any of the nodes within in the system. A sender node will direct the queries to the neighbour only if the corresponding node is agreeing to receive the queries. To achieve a better flow control client node periodically assigns the flow control tokens to the neighbours based on the available capacity. The token represents one single query that a node is agreeing to accept; hence a node can send a query to its neighbour node only if it has received the token from the neighbour node which helps in avoid being overloaded.

D. *One-Hop Replication*

This aimed at increasing the searching process. every node in the system maintains an index of the content of each of its neighbours, The index maintained by every node is shared among them when they get connection between them and the changes occurred are periodically updated, so that not only the node respond to the match with its own relevant content but also the match from the content offered by the other nodes based on index updation.This method ensures the searching is consistent throughout the life time of the network.

E. *Search Algorithm*

This algorithm ensures the combination of both dynamic topology adaption algorithm and one hop replication where the high capacity nodes provide good response to the large proportion of the queries where the search algorithm uses the random walk, It aims at rather than forwarding queries to the randomly chosen neighbours, it chooses the one with the highest capacity node with its flow control token.

F. *Low overhead Random Sampling*

The random sampling algorithm estimate a system wide load distribution to reduce the overhead .It aims at sending a sample-request message with a Time-To-Live (TTL).every node selects a random neighbour link and forward it and decrement the TTL.when the TTL expires it send back the sample.only the local information is utilized by this algorithm but can easily adapt to high dynamic distributed overlay. Hence the messages are piggy-backed on any of the existing keep-alive traffic between the neighbours to minimize the overhead.

G. *Load Balancing*

In this algorithm the average load existing in the system can be determined by the node by using histograms, where the histograms contain the information about the heavily loaded nodes and lightly loaded nodes .The load balancing is effectively performed well in the network by making the lightly loaded node to leave the network and rejoin in the near location of the heavily loaded node to partition the load to the nearby neighbour nodes.

H. Uniform Locality Preserving Hashing

This algorithm can produce uniform distribution of hashing values. The input attribute's distribution function is increasing and the value is continuous, where always it is known in advance .It is satisfied by the Gaussian pareto and Exponential distributions.

The uniform locality preserving hashing function can be designed as ,

$$H(v) = D(v) \times (2^m - 1) \qquad (3)$$

Where, H (v) denotes hash function,
D (v) denotes distribution function

I. Multi-Dimensional Query Resolution

In this algorithm each resource will register its information (attribute, value pairs) at node n_i = *successor (H (v_i))* for each of the attribute value *v_i*, Where *1<=i<=M*.each node categorizes the indices of <attribute-value, resource info> pairs by different kinds of attributes. When a node needs to search for the required resources, it generates a multi-attribute range query, where it is a combination of multiple sub queries on the each dimension.

To resolve the M-attribute queries it takes,

$$\boldsymbol{O}\left(\sum_{i=1}^{M} \log N + K_j\right) \qquad \textbf{(4)}$$

routing hops.

Where K_i is the number of nodes intersects the query range on attribute a_i and N denotes all the nodes.

IV. COMPARISON OF SEARCH ALGORITHMS

TABLE.1
QUERY SEARCH ALGORITHMS

S. NO	***Query searching Algorithm***			
	protocols	***goal***	***benefits***	***limitations***
1	Dynamic Topology Adaption (DTA)	To Achieve nodes Level of satisfaction	Gathers more neighbours to increase satisfaction level	Degree is to be often update if a new node is
	Flow Control (FC)	To Aviod hot-spot or node overloading	Continous query propagation	No explicit feedback mechanism
	One-Hop Replication (OHR)	To improve efficiency of search process	Periodical updation of incremental changes	when a node leaves the system ,the neighbour node information is lost
	Search Algorithm (SR)	Takes baised random-walk to avoid redundany path	Query traversing more than one time is avoided	Multiple Query redundant response is not produced
2	Low-Overhead Random Sampling (LRS)	Estimation of load distribution	random subset information is delivered	It does not influence the random accuracy
	Load Balancing Algorithm (LB)	To ensure uniform distribution of load	Only lightly loaded nodes are considered	To re-join the network the node need to find long link
3	Uniform locality Preserving Hashing (ULPH)	To produce uniform distribution of hashing values	Monotonious attribute values are known in advance	Need to cover m-bit identifier space
	Multi-Dimensional Query Resolution (MQR)	Searching resources for multi attributes	Routing hop for searching increase linearly.	No of routing hops is dependent on number of attributes

V. EVALUATION RESULTS

The evaluation is done among different algorithms to know their performance variance. The number of the routing hops is very low for the single attribute query range that is topology adaption, flow control and one-hop replication, so the searching becomes more complex. It provides only less query transformation across the network. The low overhead random sampling provides better searching way within the network .Apart from this the uniform locality preserving hashing provides even more reliable query searching than other algorithms.

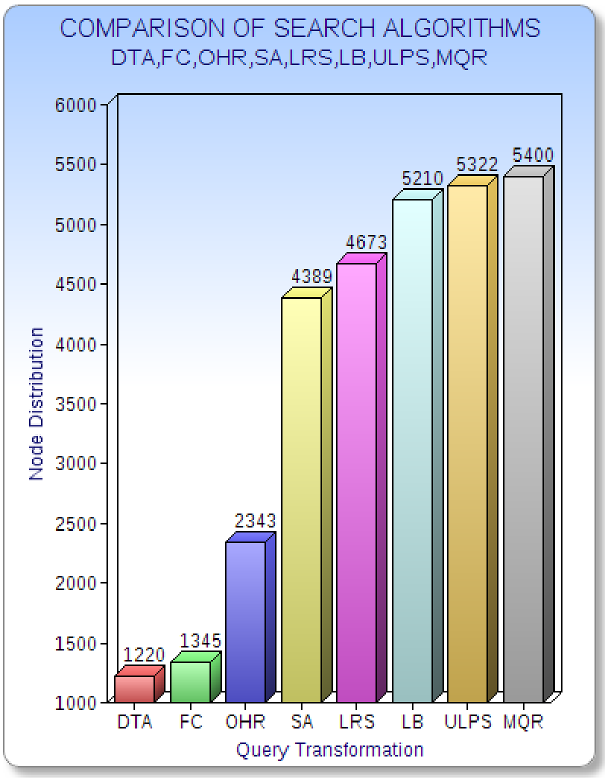

Fig.1 comparison of search algorithms

VI. CONCLUSION

In this survey we have focused on how queries are handled, how they transformed within the network and how the node accepts the queries to improve the heterogeneity of the node. Hence it improves the node capacity as well as makes the system a scalable one.

REFERENCES

[1] A. Abbas - Grid Computing: A Practical Guide to Technology and Applications, Editura Charles River Media, 2005

[2] Dhiraj .M . Bochare,Dr.A..S.Alvi -Distributed hash table in peer to peer (P2P) System.

[3] Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S.Shenker, "Making Gnutella-Like P2P Systems Scalable," Proc.ACM SIGCOMM, 2003.

[4] A.R. Bharambe, M. Agrawal, and S. Seshan, "Mercury: SupportingScalableMultiAttributeRange Queries," Proc. ACM SIGCOMM, 2004.

[5] Harvey, N. J. A., Jones, M. B., Saroiu, S.,Theimer, M., and Wolman, A. Skipnet: A scalable overlay network with practical locality properties. In Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems (Seattle, WA, Mar.2003).

[6] Heubsch, R., Hellerstein, J., Lanhan, N., Loo,B. T., Shenker, S., and Stoica, I. Querying the Internet with PIER. In Proceedings of the 29th International Conference on Very Large DataBases,Sept. 2003).

[7] Li, X., Kim, Y.-J., Govindan, R., and Hong W. Multi-dimensional range queries in sensor networks. In Proceedings of the ACM Sensys 2003 (Nov. 2003).

[8] M. Cai, M. Frank, and P. Szekely, "MAAN: A Multi-Attribut Addressable Network for Grid Information Services," Proc. Fourth Int'l Workshop Grid Computing, 2004