Task allocation in Distributed computing VS distributed database systems : A Comparative study

***Dr. Suchita Upadhyaya ** and Suman Lata**
* Reader : Dept. of Computer Science K. U.K
** Ph.D.Research Scholar K.U.K / Scientist(SS) Directorate of Wheat Research

Summary
Task allocation in Distributed computing systems (DCS) is an important research problem . When resource to be shared in DCS is a database that system is classified as Distributed database system(DDBS) . In DDBS systems Data & operation allocation are both closely interrelated & highly dependent on each other. Here it is represented along with model of allocation and development of such a model in general .DCS & DDBS are compared in this paper with reference to task allocation Models , Algorithms, Issues and Tools . General models and objective function explained in this paper can be treated as basic platform for research in this area of task allocation. Major issues in DCS have been explored by research in this area so far, while in DDBS the main issues are high lighted in this paper. The characteristics of DDBS like distributed data, distributed operations from query tree and result file are mentioned as tools to be taken in this field of research. An objective function can be derived by modifying the terms present in general model ,which inturn depend on characteristics of the system concerned ex. Distributed computing system ,distributed database system ,parallel system & multiprocessors etc.

Key words:
Distributed computing systems, Task allocation issues ,Task preference, Task exclusion, Distributed database systems, data & operation allocation, task allocation model, Data fragmentation, data replication, Task allocation tools ,task allocation algorithm.

1. Introduction

The advancement in new technologies in communication and information lead to the development of distributed systems and parallel systems . Distributed computing systems such as a network of heterogeneous workstations or PCS become an attractive alternative to expensive ,massively parallel machines . But to exploit effective parallelism or distributed system , the tasks (more processors have more tasks) must be properly allocated to the processors. Multiple tasks if not managed properly would lead to the degradation of overall system .Thus task allocation is an important research problem.

DCS & DDBS are the outcome of computer & communication technology advancements. and are used for their higher system throughput & improved availability. DCS and DDBS differ in the resources to be shared. DCS share hard disks and printer etc. while DDBS share distributed databases ,where the data as well as operations on these data items are equally important. Importance of distributed databases and more specifically task (operation) based distributed database designs in design development is gaining importance. The DDBS defines and makes it possible to identify the tasks handled as well as distributed nodes for executing these tasks . In distributed computers system the computers may be homogenous or heterogeneous, connected together for some common application execution .

A task is a set of modules and module (in task)executes on one of the processing nodes (therefore executes on every one of the processors at that node) and communicates with some other modules of the task by inter module communication(IMC).A task is a program or a part of a program in execution. It is one of the important computational activities that takes place at the same time and /or at different locations In contrast to a task , a job is a whole, usually sequential ,program in execution.

Task allocation is an essential phase in distributed system & distributed software design. In DCS the software application is called a task and is a set of cooperating modules . For achieving a fast response time from such systems , an efficient assignment of the application tasks to the processors is imperative.

The general assignment problem is known to be NP – hard ,except in a few special cases with strict assumptions. A task is represented by a task graph. As the position of modules in task graph represents its precedence of execution ,it is called a task precedence graph(TPG).The communication among its modules can also be shown by a graph known as task interaction graph(TIG).These two graphs can be represented separately or can be combined into one. The network of nodes in DCS is represented by another graph known as processor graph PG. So task allocation is solved by mapping the two graphs.TG → Mapping → PG

Manuscript received March 5, 2008
Manuscript revised March 20, 2008

2. Task allocation in DCS & DDBS

2.1 A DCS (distributed computing system)

consists of multiple heterogeneous computing nodes with their memories and clocks, that communicate with each other by message passing mechanism. Over last several years, these have become popular for high performance computing and information processing .The main incentives for choosing DCS are higher system throughput and improved availability. A distributed processing system has the application software portioned into a set of program modules , allocation of these modules to the processors is an important research problem.

Task allocation criteria :- Several task allocation algorithms for distributed computer system have been reported in literature[1-5].These algorithms consider execution time of the different modules of a task, executing on different processing nodes. The task assignment problem(NP-hard) tries to maximize the throughput of the system by **minimizing the cost** .This cost may be **in terms of time** that is execution time **or in terms of bytes of data transferred.** The module are allocated onto processing nodes so as to minimize the time taken .This can be done by maximizing and balancing the utilization of resources while minimizing the communication cost between these processors. But both criteria are conflicting , as the load balancing calls for distributing the tasks over different processors while minimizing of interprocessor communication drives the task assignment to assign all the tasks onto a single processor. The task assignment problem in DCS is NP-Complete .Hence satisfactory suboptimal solutions obtainable in a reasonable amount of computation time are generally sought.

2.2 DDBS

DDB system is group of distributed computers on which a database is portioned and saved so that for processing user query all look as single unit. A DDBS system consists of a set of data objects each of which is assigned a value. A data object represents the smallest unit of the database accessible to transactions. All the user requests for access to database are handled by the database management system. The basic units of user activity in database systems are transactions .Each transaction represents the basic machanism for identifying and defining a set of logically associated operations. Each transaction has time factor associated with it.

Data (object) and operation (task) allocation are among the key issues during designing a distributed database. Data allocation defines what data is stored and at what nodes(with replication) while operation allocation states where accessing and processing of operations(select ,join, project etc.)will take place. Each node has its own (NDS) network directory structure which is small and stable relatively giving details(index) of data stored at that node. Both data and operation allocation are interdependent problems and must be solved simultaneously. The optimal set of all the data copies and their optimal allocation depends on processing schedules of all retrieval and updates accessing these data(ie task allocation).However the processing schedules depend on where data sets & their copies are located (ie data allocation)hence it is circular problem and to be effective a DDB design approach must comprehensively treat both data & operation allocation as a unified whole.

3. Task Allocation Models

3.1 Developing DCS allocation model

DCS model developed should be able to satisfy the goals of task allocation which may be any one of the following

1. Balancing the processors load in DCS.
2. Increase the throughput of the system / minimizing the processor turnaround time.
3. Minimize IPC cost.

To find the solution to task allocation problem A (mathematical) model is to be developed . Before that one should have complete knowledge about the tasks and processors. So that policy of assigning these tasks on the given processors may be developed.

The task attribute's details *include*:- 1. Task size 2. Task dependency(represented by task graph)The sequence decides predecessors and successors 3.coupling factor between tasks(unit data transfer between tasks). 4. task execution (in terms of execution time matrix)5. Task communication (in form of m x n matrix)

The processor's details in the network *may include* :- 1. Network topology (architechture of processors connections) which includes
interprocessor's distance.2. Processors attributes like H.W. related may be speed and storage space.

Objective cost function is application and environment dependant.

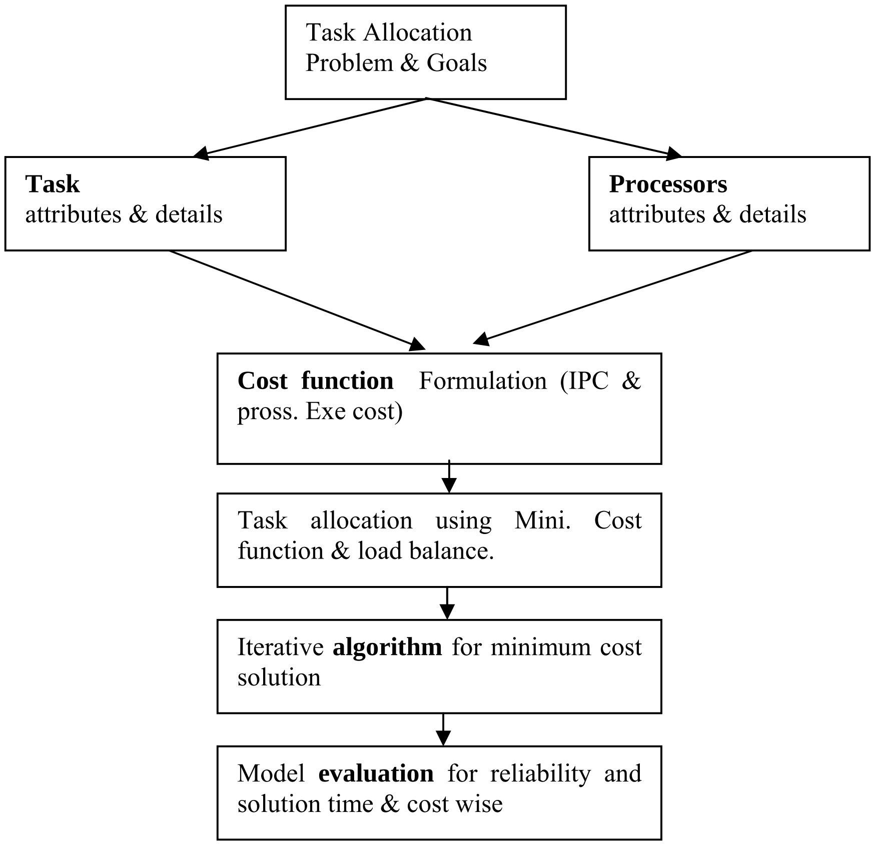
Total processing cost = IPC(Inter processor communication) cost + Processing cost

Objective function development

Total processing cost =

$$\sum_{i} \underbrace{\sum_{k} (\mathbf{WQ}_{ik}\ \mathbf{X}_{ik}}_{\text{Processing cost}} + \sum_{i} \sum_{j} \underbrace{(\mathbf{C}_{ij} * \mathbf{d}_{kl})\ \mathbf{X}_{ik}\ \mathbf{X}_{jl})}_{\text{IPC} \qquad \text{cost}}$$

Developing DCS Task allocation model


Task Allocation Problem & Goals
Task attributes & details
Processors attributes & details
Cost function Formulation (IPC & pross. Exe cost)
Task allocation using Mini. Cost function & load balance.
Iterative **algorithm** for minimum cost solution
Model **evaluation** for reliability and solution time & cost wise

Objective function for DCS [1]
C_{ij} = No. of data units transferred from task i to j
D_{kl} = distance related comm.. cost for one unit data trans.from processor k to l
X_{ik} = assignment variable (task i is assigned to processor k (0,1)
Q_{ik} = Processing cost of task (i) on processor k bjective function for task allocation model may be cost model [1] and may include other parameters like *Precedence Relation(*PR) among tasks or it may be common probability formula in case of *reliability* being main objective of task allocation. In case of task allocations considering *load balance* as main goal .The load on a processor is the combination of all the execution and communication costs associated with tasks assigned to it.

In most of distributed systems task allocation objective function is f(v) = g(v) + h(v) where g(v) is search path from source node and h(v) is lower bound from current node to goal node.

In optimal task assignment goal is to minimize the maximum process turn around time .

Turn around time = time for module execution + time for IPC cost .

$$\mathbf{Tp\ (A) = t_p{}^e(A) + t_p{}^c(\ A\ )}$$

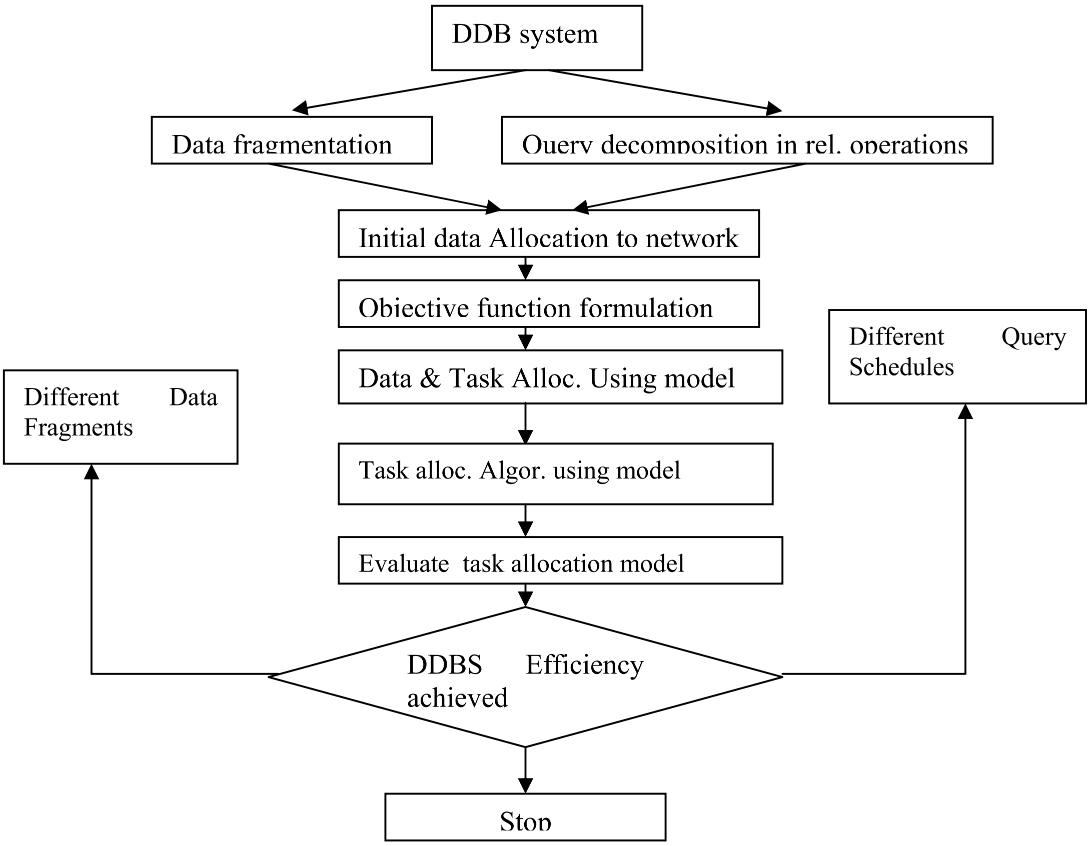
A = task assigned P = processor.

After deciding about the objective function & knowing the details about the tasks & processors of the particular Task allocation problem there are some *restrictions* like of processors H.W. capabilities, network architechture and of homogenous or heterogeneous processors in the same way tasks have their nature like interdependent ,tasks sizes etc. Under the current model situation ,attributes are the collection of constraints from application tasks and given network topology ex. Memory of processor is limited.. Some of the attributes are to be taken as constraints. It is essential to incorporate them in Allocation model to achieve load balance and for application requirements.

Next step after all this is to decide about the iterative *algorithm* that will use the objective function developed and lead to solution or optimal soluition .Again this algorithm selection depends on application and may use branch and bound ,A* best search algorithm of AI(state space search), greedy algorithm. Some of the techniques used are graph theoretic, network flow technique and clustering technique. Using anyone of the above methods the repetitive algorithm finds partial schedule (task assignment)in each step of algorithm execution and finally find the complete path to goal node. Further refinements in the selected technique's based algorithm are achieved by using tasks dependent information as heuristic to reduce the search efforts in finding an optimal path to the goal node .But main question is to decide about what heuristic is appropriate for a given situation.

Evaluation of task allocation algorithm: Task allocation algorithm is evaluated either with actual implimentation or with simulation .More flexible & powerful simulations are needed for this purpose .Usually task allocation algorithm are implemented using programs on some system. By applying several sets of input data to this program the time complexity of algorithm is tested and algorithms performance in terms of cost & time is observed and reliability of algorithm under different situation can also be tested for different sets of data.

Devoloping DDBS Task allocation model


DDB system
Data fragmentation
Query decomposition in rel. operations
Initial data Allocation to network
Objective function formulation
Data & Task Alloc. Using model
Task alloc. Algor. using model
Evaluate task allocation model
DDBS Efficiency achieved
Different Data Fragments
Different Query Schedules
Stop

3.2 DDBS Task Allocation model Development

STEPS OF DDB system Design : The DDBS design goes as follows:-

A. Generate the units of data from

1. logical data model or conceptual schema

2. . characterization of retrieval & update activity

B. Decompose retrieval and update activities into relational operations

On fragments (with reference to A),data fragments made in A.

Now next step should be

C. Allocate both data fragment (A) and operations from(B) to the nodes in network using any model ex Mathematical)

Model has three components of decision ***variables, constraints*** and ***evaluation*** measures.

***Optimized model of DDBdesign*:**

Decision variables includes following

1. X_{ij} data file allocation for i^{t}h data fragment on j^{th} node

2. Z $_{kit}$ for data replication of i^{th} fragment on t th node k =0,1 if used or not.

3. Y_{kmt} for operation m of query K is done at node

all these lead to 3 matrix

Objective function :- It is to minimize the cost of (Retrieval +update+storage). Objective function selected to this problem is minimization of system operating cost including communication cost ,disk IO,CPU processing and storage costs. [18]

Mini cost = $\sum_k \sum_j f(k,j) \sum_m$ (COM(k,j,m) +IO (k,j,m) +CPU (k,j,m) + $\sum_t STO_{(t)}$

f(k,j) is frequency of execution of query k at node j ,other are com,IO,CPU cost for m step of query k at node j and $STO_{(t)}$ is node t's storage cost.

*Constraints_*:- Two different types of constraints are intrinsic problem constraints and resource constraints .Resource constraints are CPU processing and storage capacity of nodes and communication capacity of each link .Which are limited

Intrinstic problem constraints includes :-

$\sum_t X_{it} \geq 1$ all file fragments 1 = 1,2 …… must be stored at one or more nodes.

$Z_{kit} \leq X_{it}$ for all the queries k for all file fragments cannot be accessed from a node
unless it is stored at that node.

$\sum_t Y_{kmt} = 1$ all steps m of all queries must be processed at some node.

After developing such mathematical formulation an algorithm can be developed to solve it. One such algorithm was tested .ie. a genetic algorithm and found an efficient solution of data and operation allocation in reasonable amount of time.

3.3 DDBS model development is two times complex than DCS model as data and operations not only are to be allocated on network of computers but their interdependency is also to be taken care of. This interdependency may lead to reallocation of data and then to again the operation allocation with reference to newly allocated data fragments. Refinement in each iteration is possible as by reallocation of data and operation during different queries & applications and situations which leads to minimum cost of model function.

From Fig. 1 & 2 it is clear that DCS considers only task with reference to the processors on which they are to be executed. While Fig. 2 shows tasks as well as data are simultaneously considered starting with an initial allocation . In DDBS refinements are done at two levels .It is done by changing the fragment allocations and considering different schedules of query execution in each iteration then choosing the good. While in Fig. 2 refinement in each iteration considers different tasks to processor allocations.

4. Task Allocation Algorithms

4.1 A large number of task assignment algorithms have been proposed using various techniques such as network flow, state space search, clustering, bin packing and probabilistic and randomized optimization. [4] Most of the earlier work of task assignment in DCS by Harold Stone and Bokhari includes networkflow algorithms on dual processor system only ,later on researchers took more than two processors into consideration .The task assignment problem solving approaches can be roughly classified into four categories ,graph theoretical, mathematical programming ,heuristic techniques and probabilistic approaches such as simulated annealing based techniques ,mean yield annealing and genetic algorithms. Because of the intractable nature of the task assignment problem and its importance there is ever increasing demand for distributed computing . An efficient algorithm for the task assignment problems is desirable to get the best possible solution within acceptable CPU times

4.2 DDBSTaskAssignment Algorithms:

Design method has 3 steps

1. set of queries are analysed to define set of file fragments for allocation and may give idea about horizontal and vertical fragmentation of data files.
2. Each query is decomposed into query steps each of which references to file fragments if the referenced data is fragmented additional join operations may be needed.
3. Resulting data fragments and query steps are used as input to an optimized model

In DDBS the Greedy algorithm, AI' A* algorithm and search techniques such as Branch & bound and state space search are found to be good when tested on DDBS task allocation [18]. In a Greedy Algorithm from sequence of choices it makes a locally optimal choice in that which leads to a globally optimal solution .Search space tree is a pool of all possible solution sites which can participate in finding the optimal path of solution .In this process the developed model helps in giving weightage to current site under consideration .The weight is a cost function and is a parameter to be compared with other site and select the better (with minimum cost).

A search space tree is used to represent the large solution space for determining the data allocation to minimize the total data transmission cost .And Branch & bound search technique is used to make decision tree. During search for an optimal data allocation the decision tree constructed so far partitions the space of completely specified allocations into subsets that belong to leaves. A subset belonging to leaf of decision tree contains all completely specifies allocations .The partially specified allocations use a cost estimator which decides which of partial allocations will become completely specified allocation finally.

A* algorithm belongs to Artificial Intelligence applications and is well suited to search through the solution space of DDBS task allocations. Refinement in iteration of the choosen algorithm is possible as by reallocation of data and operation.

4.3 Not much work has been done on DDBS task allocation but work needs consideration in graph theoretic technique ,task clustering & state space search technique.These techniques can be tried as they match the basic nature of tasks represented as graphs. while a lot of work in DCS has been done on task allocation such that the algorithms are broadly classified in to two categories of optimal & suboptimal classes.

5. Task Assignment Issues

5.1 Issues in DCS :-

Depending upon the applications of DCS the scope_and horizon of task allocation fluctuates but total space of tasks ,processors and links includes following:-

1.Task interdependency :- The DCS must consider execution of modules of various unrelated (independent) tasks also. As modules from same task show interdependency but may not depend on the modules of the other tasks so that process idleness is negligible during execution.

2. Task preference and task exclusion:- The task preference matrix indicates that certain tasks can only be executed in the specified processor .It is represented by m x n matrix

Certain pair of tasks must not be assigned to the same processors. This pair of tasks is said to be mutually exclusive and is represented by E matrix and $E_{ij} = 0,1$.

In heterogeneous DCS the different amount of execution time specify different degrees of preference of module on the processors.

3. Task redundancy :-It may be provided for system reliability .Multiple copies of a task exists. ie if task (i) has redundancy of three ,new tasks (i) +1 and (i) +2 are added to the original set of tasks and have similar properties as (i) and additional restriction is added such that tasks (i),(i +1), (i+2) cannot be assigned to the same processor.

4. Channel and processor's load :-While assigning modules of multiple tasks to the processing nodes, IMC (intermodular communication) plays a vital role and is equally important as the speed and capability of the processing nodes. So during task allocation model development the load on the channels must be considered which are used for intermodule communications .Heavily loaded nodes may lead to dead lock or even starvation sometimes. The execution time of a particular module on a particular node will depend on the number of modules already executing on that particular node as per some chosen policy. Preference matrix and exclusion matrix when taken as INPUT parameters would also lead to processor load balance.

5. Message passing :- Non identical communication links are used by the processors for message transmission. That is identical amount of message if transmitted through different communication links will require different amounts of transmission time, which specify different degrees of preference of the message on the links. But sometimes the processors in the system need not to be fully connected but the link between any two processors is symmetric ie the time to transmit a certain amount of message from one processor to another is identical to that to transmit the same message in reverse direction.Thus symmetric links and non identical ones are to be considered while treating task allocation problem.

6. Unit value for cost function:- The cost function in task allocation is sum of IPC and Processing cost that are actually different in measurement units inheritly . So weights(**W** in equation below) should be assigned to these two cost factors to reflect the criticality of the casts in particular application. Such a cost function will help in balancing IPC and processor cost and finally lead to load balance in that application ex. Total processing cost = $\sum \sum (\mathbf{W} Q_{ik} X_{ik} + \sum \sum (C_{ij} * d_{kl}) X_{ik} X_{jl})$ [1]

5.2.Issues in DDBS :-

The research in DDBS needs consideration in following areas indicated here.These better task allocation models developed will give direction to good & efficient DDBS development.

1.Data fragmentation and fragment allocation to proper nodes in distributed database system is most critical issue in designing DDBS.Decision about type of fragmentation ie. Horizontal,vertical is application dependent and node set of fragmentation is transaction arrval frequency dependent.

2.Data replication over sites of network in DDBS increases the *data availability*.But proper & exact replication ie what and how much data to replicate (replication degree) ,Where to replicate data(replica placement) and how to update(replica control) replicated data are important issues of concern.

3. Reallocation of data in partitioned database must be taken into account when the *access pattern of data* change frequently.Which is slight deviation from traditional DDBS (static).

4. Changing workloads :While allocating tasks and data to nodes of network the changing workloads should be considered as overloaded sites have processors memory capacity limitation associated with these.This consideration will lead to the load balanced DDBsystems.

5. Cost and performance: Data and operation allocation during distributed transaction execution usually the cost and performance is considered while data availability is an equally important factor to be considered which will give more reliable DDB systems.

6. Unit cost of transaction: In DDBS the site of computer network is decided for data allocation by considering the cost of transaction in that particular concerned path schedule

5.3 In DCS the task is main issue whose interdependency ,preference ,exclusion, redundancy and unit costs are of concern while allocation is being done. In DDBS the data of distributed database is main on which operation allocation completely depends. That is why data's fragmentation ,replication, allocation and reallocation are of main concern. But following two issue(4th and 6th) are common for both.

1. **Load balance** :- In DCS load on channel & processors affects the task execution speed which finally decides system throughput . In DDBS changing access patterns leading to changing workloads on nodes of network are so important that sometime they lead to reallocation of data to have acceptable system throughput.
2. **Unit cost** :- In DCS unit value cost for cost function needs weight(W) for normalization because of their different units inheritly .In DDBS value of unit cost of transaction helps in deciding the path of schedule by selecting low cost nodes for data allocation.

6. Task allocation tools

6.1 Tools in DCS :-

For solving task allocation problem in DCS a task allocation model has to be developed which is a valuable tool for the design of distributed systems and DDB systems.

Tools used to achieve above usually are

1. Use task dependant information as heuristic to reduce search efforts and search space(informed search to save search time and search space). [4]
2. Use precedence relationship as parameter in heuristic algorithm for task allocation. [6]
3. Maintaining GT (global table) showing that all the processors are being utilized properly for execution. Which further leads to load balancing. [3]
4. To use clusters (task and processors) to reduce the communication penality. [12]
5. Using concurrency (by including interference costs of modules) and parallelism to increase speed of assignment thus leading to minimize turn around time.[11]
6. By using redundancy (H.W. /S.W.) not only increases the system reliability but also improves response time of the system. [22]

6.2 DDBS Tool's Characteristics:-

Before solving the problem of Data & operation allocation in DDB system it is important to analyse the general characteristics of this problem which inturn gives a foundation to the development of specialized algorithm.These are Data fragmentation,query tree,data & operations,and finally the result data file. [16]

1. Database distribution :- The data is devided into files F^p , F^q , F^r each file is subdivided into f^p (No. of fragments of file Fp) Allocation of data file F^p on network of N (n ∊ N) nodes is represented by (f^p X n) allocation matrix $AM^p = [a^p_{ij}]$ $a^p_{ij} = 0,1$ depending upon the storage of ith fragment of F_p at node j or (not) otherwise .

F_p is non redundant

F_p is redundant

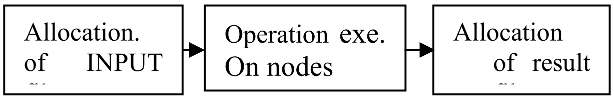
F_p is completely duplicated on nodes.A transmission of F^p is transformation of its allocation matrix

$$AM^p \rightarrow AM^p$$

An N X N transmission cost matrix c is unit transmission cost between any two nodes in network $C \equiv C_{jk}$

C_{jk} = cost of transmission of one byte from j to k node.costs are symmetric $c_{jk} = c_{kj}$

2. Global transaction's operator tree(preparation of operations list) : A global transaction can be represented by an operator tree with several operations .This operator tree shows strategy of executing this transaction with the help of selection ,projection and join operations.1.Initial allocation of input files -→ 2.



operation execution on nodes -→3. final allocation of result file formed during above process 2

Some file fragments are to be transmitted from one node to another and operation allocation determines which file fragments are processed at each node so that the transmission costs are minimized.

3. Distributed operations *(* DDB operation distribution):- Operations in operator tree of global transactions prepared in second step can be described in terms of selection ,projection and join type of operations. A selection operation consists of taking out those records from file which have some common property.Projection operation has two steps Ist step removes some items from the records of file .Second step removes duplicate records which may result from the Ist step .A join operation finds these records of two different files F^p , F^q which have some matching property so need comparison of each record of one file with all the records of other one. $E \subseteq N$ one of two files F^p or F^q is completely distributed and other is completely duplicated. Then following correctness condition holds $\{ cdi(F^p,E) \wedge cdu(F^q,E)] \vee [cdi(f^q,E) \wedge cdu(F^p,E)]$

This condition reflects either some fragments of F^q are compared with whole file F^q or vice versa..

4. Result File :-In all above 3 operations a result file Fr is formed which is itself distributed.In case of selection and projection F^r is part of input file F^p while in case of join operation the result file F^r fragmentation depends on which of two files F^p F^q is completely distributed on on execution nodes. Let F^p is completely distributed then $f^r = f^p$ and $AM^r = AM^p$ .solution of operation allocation problem lies in assignment of allocation matrix of input and output files to the set E of execution nodes at execution time.

Allocation matrix AM^P of same file are different before and after the operation execution which is due to transmission of AM^P required by solution of operation allocation problem.This transmission is defimed by the transformations of allocation matrix.

$$AM^p \rightarrow \overline{AM^p}, \quad AM^q \rightarrow \overline{AM^q}, \quad \overline{AM^r} \rightarrow Am^r$$

Tools used in DCS & DDBA may be different which is due to difference in shared resources in two cases .Tools used inDCS task allocation include global table(GT),Task precedence relationship,task clustering,and task concurrency.In DDBS the distributed data (redundant &duplicate),different query schedules &distributed operations and transformed result file can be taken as tools.As these tools are nothing but characteristics of a DDBS ,they will surely lead to a better DDBS.

7. Conclusion

This paper presents task allocation problem of DDBS design research area.The DCS & DDB systems are taken where task allocation appears to be same but basically is quite different which is due to their basic difference which lies in the type of shared resources in each case.Task allocation is found to be more complex in DDBS as compared to DCS .Mojor aspects in DDB system design are as follows:-

1. Communication Network details (Location of nodes, allocation of computer resources network topology, selection of link capabilities)It is same as processor details in DCS.
2. Data allocation includes unit of data allocation ie data fragmentation type and amount of data replication(at which nodes and how much)
3. operating strategies related to query execution and concurrency control includes which copy of data to access and where to process the data how to route the data and locking the data.

These 2 and 3 are much more complex & interdependent than tasks details in DCS.

This complexity is because of two factors of consideration ie. Data & task in the former and only task in later but the final goal and objective of task allocation is same. That is to Increase the throughput of the system by minimizing the network's processor turnaround time. More research work on task allocation in DDBS needs to be done to have more practicaliy working & efficient DDB systems. In DDBS lot of work has been done on data allocation and query processing but not on task allocation. This field of research needs consideration.A better task allocation in DDBS is only possible using **efficient tools** like proper fragmented data definition and distribution ,transformation of input cost matrices into result matrix and **optimized operations allocation** on nodes. A better **task allocation algorithm** will only finally produce efficient DDB systems.

Reference

[1] Perng-Yi Richard Ma,Edward Y.S. Lee,"A Task Allocation Model for Distributed Computing Systems",IEEE Transactions on Computers,Vol c-31 No. 1 ,January 1982,pp 41 –47.

[2] Chen-Chung Shen & Wen-Hsiang Tsai ,"A graph Matching Approach to optimal Task Assignment in Distributed Computing Systems using a Minimax Criteria",IEEE Transactions on Computers ,Vol. C-34 ,No. 3 ,March 1985. pp 197-203.

[3] Deo Prakash Vidyarthi,Anil Kumar Tripathi& Biplab Kumar Sarkar,"Multiple Task Management in Distributed Computing System",Journal of The CSI,,Vol.31, No. 1,March 2001 .

[4] Muhmmad Kafil and Ishfaq Ahmad ,"Optimal Task Assignment in Heterogenous Distributed Computing Systems,"IEEE Concurrency,July- September 1998,Pp 42-49.

[5] Sol M.Shartz,Jia-Ping Wang,"Models & Algorithms for Reliability –oriented Task – Allocation in Rendundant Distributed Computer Systems",IEEE Transactions on Reliability ,Vol. 38 No. 1 April ,1989,pp16-26.

[6] Wesley W.chu,and Lance M-T. Lan ,"Task Allocation and Precedence Relations for Distributed Real Time Systems", IEEE Transactions on Computer, Vol. C-36 ,No. 6 ,June 1987 ,Pp 667-679.

[7] Muhammad Kafil and Ishfaq Ahmad ,"Optimal Task Assignment in Heterogenous Computing Systems",IEEE Concurrency, September ,Pp, 42-51,1998.

[8] Imtiaz Ahmad ,Muhammad K. Dhodhi and Anil Ghafoor ,"Task Assignment in Distributed Computing Systems",IEEE 1995 Pp 49-53.

[9] Stephen S. Yau and Venkeepuram R.Satish,"A Task Allocation Algorithm for Distributed Computing Systems",IEEE 1993,Pp 336-342.

[10] Sub Ramakrishnan ,II Hyung Cho and Larry A.Dunning ,"A close look at Task Assignment in Distributed Systems",IEEE 1991,Pp806 –812.

[11] Verginia Mary Lo. ,"Heuristic Algorithm for Task Assignment in Distributed Systems",IEEE Transactions on Computers, Vol 37,No. 11,November 1988,Pp1384-1397.

[12] Deo Prakash Vidyarthi,Anil Kumar Tripathi ,Biplab Kumar Sarkar,"Cluster based task Allocation in distributed systems", Proceedings of 18th Intl. Parallel and Distributed Processing Symposium,IEEE 2004

[13] Ravi Mukkamala ,"Storage efficient and Secure Replicated Distributed databases",IEEE Transactions on Knowledge and Data Endineering, Vol 6, No 2,apr,1994, pp337-341

[14] Ajit Kimar Verma,Mangesh Trimbak Tamhankar,"Reliability –Based optimal Task –Allocation in Distributed Database Management Systems"IEEE Transactions on reliability Vol 46 , No. 4 ,1997, Dec. pp 453-459

[15] Yucai Feng Dong Li ,Hengshan Wu YiZhang,"A Dynamic load balancing Algorithm Based on Distributed Database system", IEEE ,2000,pp 949- 952

[16] Stefano ceri and Giuseppe Pelagatti, "Allocation of operations in Distributed Database Access",IEEE Transactions on Computer Vol. C-31,No. 2 ,Feb,1982.

[17] Douglas W. Cornell and Philip S. Yu,"On Optimal Site Assignment for Relations in the Distributed Database Environment",IEEE Transactions on Software Engineering,Vol. 15,No. 8 ,August,1989.

[18] Salvatore T.march and Sangkyu Rho,"Allocating Data and Operations to nodes in DDB Design", IEEE Transactions on Knowledge and Data Engineering,Vol,7,No 2,April 1995.

[19] Syam Menon ,"Allocating Fragments in Distributed Databases",IEEE Transactions on Parralel and Distributed systems ,Vol 16 ,No 7,July 2005.

[20] LinWujuan and Bharadwaj Veeravalli,"An Adaptive Object Allocation and Replication Algorithm in Distributed Databases",Proceeding of the 23rd International Conference on Distributed Computing Systems Workshop",2003.

[21] Anna Brunstrom ,Scott T. Leutenegger,Rahul Simha,"Experimental Evaluation of Dynamic Data Allocation Stretegies in a Distributed Database With Changing Workloads", Acm , 1995 ,pp395-402.

[22] Seong-Jin Park ,Doo-Kwon Baik ,"A Data Allocation Considering Data Availability in Distributed Database Systems",IEEE 1997,pp708-713.

[23] Peter M.G. Apers, "Data Allocation in Distributed Database Systems",ACM Transactions on database Systems ,Vol. 13, No. 3 ,September 1998, ,pp 263-304.

[24] Ravi Mukkamala,"Measuring the effects of Data Distribution Models on Performance Evaluation of Distributed Database",IEEE Transactions on Knowledge and Data Engineering,Vol 1,No 4,December 1989, pp 494-505.

[25] L.W. Dowdy & D.V. Foster ,"Comparative model of the file assignment problem,"ACM computing Survey ,vol. No 2 PP 287-314 june 1982.