Intern. J. Fuzzy Mathematical Archive
Vol. 6, No. 2, 2015, 187-195
ISSN: 2320 –3242 (P), 2320 –3250 (online)
Published on 22 January 2015
www.researchmathsci.org

International Journal of
Fuzzy Mathematical Archive

An Optimized Strategy for Data Allocation in Peer-To-Peer Distributed Databases

D.I.George Amalarethinam [1] **and** ***C.Balakrishnan*** [2]

[1]Department of Computer Science, Jamal Mohamed College (Autonomous)
Tiruchirappalli, India. e-mail: di_george@ymail.com
Corresponding Author

[2]Software DevelopmentSkill Development Centre, Alagappa University, Karaikudi India.
e-mail: balasjc@gmail.com

Received 4 November 2014; accepted 4 December 2014

Abstract. Better accessibility and availability of data are the evergreen aspects to be focussed during the design phase of Distributed Database environments. The accessibility is achieved by means of optimal allocation of fragments into the sites whereas, the high degree of availability attained with better replication strategies. However, the Data Allocation Problem (DAP) incorporates the both Allocation and Replication methodologies. This paper proposes a methodology for Reallocation and Replication of fragments in Peer-to-Peer Distributed Database environment using Ant Colony Optimization (ACO) principle, namely, Ant Colony Optimization based Data Allocation Strategy (ACODAS). ACO algorithm is a meta-heuristic, works with swarm intelligence technique. The experimental results show that ACO based reallocation and replication methodology improves the ratio of successful queries than the initial allocation of fragments.

Keywords: Ant colony optimization algorithm, accessibility, availability, data allocation problem, peer-to-peer distributed databases, reallocation, replication

1. Introduction

Higher degree of parallelism and improved availability as well as accessibility are the promising characteristics of Peer-to-Peer Distributed Database Systems (P2PDDBS). The P2PDDBS are more compatible not only with decentralized nature and ability to store growing volume of data, but also help to serve the queries in an effective manner[14]. Accessibility and availability are the important aspects to be addressed by any of P2PDDBS environment [11]. These aspects improves the QoS factors like fault tolerance, query execution time and possibilities for recovery. Accessibility is achieved by means of better allocation scheme and better replication scheme leads to accomplish high degree of availability. Particularly, this paper considers Reallocation of fragments to appropriate places. Reallocation can be considered as an extended approach of allocation. Reallocation can tune-up the process of allocation and enhances the performance. Reallocation can reduce the drawbacks of allocation and increases the success ratio by allocating the fragments into appropriate sites, for which, reallocation requires the performance analysis data of initial allocation. Hence, the number of frequent transactions tried to access the particular fragment from the particular site is considered as the performance analysis data of initial allocation. The main objective of this paper is to propose a methodology using Ant Colony Optimization (ACO) for deriving Data Allocation Strategy focusing on reallocation and replication, namely, ' Ant Colony Optimization based Data Allocation Strategy' (ACODAS). The simulation result, that isincreased number of successful transactions than initial allocation, justifies the effectiveness of ACODAS based reallocation and replication.

2. Literature review

This section of the paper states the related works that are stimulated to do research on reallocation and replication methodology. Allocation is the next phase that follows fragmentation in Distributed Data Base Systems (DDBS) design. Better allocation schemes results in high degree of accessibility. The following section describes the heuristic methodologies used for allocation and reallocation of fragments to the sites in DDBS.Adl et al. [14] have coined a heuristic algorithm for fragment allocation which is based on the ant colony optimization scheme to minimize the total transaction response time under memory capacity constraints of the sites. Huang and Chen [16] proposed a simple and comprehensive model for a fragment allocation problem, also developed two heuristics algorithms to find an optimal allocation of the fragments.Basseda et al. [13] introduced the fuzzy inference engine to the existing Near Neighborhood Allocation (NNA) algorithm and studied the performance improvement of new Fuzzy Neighborhood Allocation method. Hauglid et al.[8] proposed a de-centralized approach for dynamic table fragmentation and allocation in distributed database systems based on observation of the access patterns of sites to tables. The fragmentation, replication, and re-allocation based on recent access history, aimed at maximizing the number of local accesses compared to accesses from remote sites.Virk et al.[12] proposed a Genetic Algorithm based probabilistic approach to find near optimal fragmentation plan for selecting the various nodes or sites for placing recursively the vertically fragmented data attributes.Ali et al.[2] used the vertical fragmentation technique along with a two-phase allocaton method for fragmentation and allocation in relational Distributed Databases in order to minimize the communication cost and query response time. With the help of communication and update cost values for each individual fragments, Abdalla [7] proposed a methodology for re-allocation of fragments. Tosun et al. [15] compared the heuristic algorithms for fragment allocation by involved Genetic Algorithm principles and proposed a Data Allocation methodology based on Quadratic Assignment Problem.Availability of the resources is the thriving factor in the field of distributed computing. During the recent years, enormous amount focuses were given to generate methodologies for replication in Distributed Database Systems. The research community were categorized the focusing area as general issues, scalability, Snapshot isolation (SI) and middleware based replication methodologies. Among these aspects this paper focuses on general issues in selective replication. The following section expresses the views of the related works on replication. A book chapter written by Bettine et al. [3] narrates how transactions are executed in a replicated environment and illustrated the design alternatives using a two-step approach. A book narrated by Alfranio et al. [1] toaddresses the architectural and integration challenges on how required functionality is provided by generally available software components and a set of practical challenges raised between performance assumptions and actual environments in real workloads.Xhafa et al. [6] proposed a replication methodology for documents structured as XML files in Peer-to-Peer systems.

3. Initial allocation strategy

The following 'Accounts' relation [4] is taken for the initial allocation of fragments to the sites of the Clusters [4]. Table 1 depicts the initial allocation to sites of four clusters [4].

ACCOUNT NO	CATEGORY	CUSTOMER ID	DATE	BALANCE	REGION	SITE	CLUSTER
1	A	C1	11/1/14	21000	R1	1	1
10	A	C10	9/4/14	78000	R1	1	1
3	B	C3	2/2/14	18000	R1	3	1
6	C	C6	15/3/14	52000	R1	6	1
8	D	C8	28/3/14	11500	R1	10	1
2	B	C2	21/1/14	13500	R2	2	2
12	B	C12	18/4/14	11800	R2	2	2
7	E	C7	18/3/14	38000	R2	5	2
4	C	C4	8/2/14	22000	R3	4	3
9	A	C9	4/4/14	16800	R3	7	3
5	D	C5	24/2/14	3200	R4	8	4
11	B	C11	11/4/14	23000	R4	9	4

Table 1: Initial allocation to sites in clusters

The Table 2 explains the allocation of fragments in the initial stage in cluster1, this allocation is done based on assumption.

SITE ID	FRAGMENT ID
S1	f1
S3	f2
S6	f3
S10	f4

Table 2: Initial allocation of fragments to cluster 1

Cluster 1 consisting of four sites [4] hence four fragments are derived and allocated subsequently.

3.1. Criteria for using heuristic algorithms to solve dap

The performance of any DDBS environment is directly proportional to the efficiency of the way the data are allocated to the sites. Hence, the data allocation problem can be categorized within resource allocation problems which are defined as 'optimization problems with constraints' [10]. Therefore, in order to achieve a suitable solution for the allocation problem within an acceptable computation time, the most of the related proposed algorithms [12,14] are concerned with the usage of heuristic algorithm and evolutionary algorithm such as, ACO and Genetic algorithm for solving DAP.

Most of the existing heuristic algorithms are greedy ones, and only recently, the stochastic algorithms such as genetic algorithms [12] and swarm intelligence [14] have been widely usedfor solving DAP. One of the most significant advantages of stochastic algorithms is their scalability. After analysing most of the heuristic approaches, we find that, the Swarm intelligence technique is possessing the characteristics to solve DAP. In particular, due to the scalability nature and ability to produce near optimal solution, the ant behaviour based algorithm [5] isthe most desirable candidate to solve DAP than genetic algorithms.

3.2.Problem definition of acodas

The facts gained from the earlier sections of this paper have given motivation to use ACO for solving DAP. ACO is a meta-heuristic inspired by the behaviour of real ants that cooperate through self-organization [5]. For using ACO, a very important aspect called Artificial Pheromone Factor (APF) to be formulated. This Pheromone factor is actually inherited from the behaviour of real ants. A substance called Pheromone is deposited on the ground while ants are foraging. Pheromone trails are formed on the ground by this way which also reduces stochastic fluctuations at the initial phases of search. The shorter trails will be used more frequently by ants and they gain more pheromone.A two-dimensional cost matrix is to be formulated for implementing the ACO to find optimal places for fragment reallocation. For which, this paper considers, the number of frequent transactions tried to access the particular fragment from the particular site as a Pheromone substance.

To determine the APF, the following probabilistic function [9] as in equation 1 is used, $\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij} = \frac{(S_m + K)^h}{\sum (S_m + K)^h}$ (1)where m = 1 ... n and K, h are constants, S indicates the sites When applying the equation 1, it is possible to measure the number of transactions trying to access a fragment from a site, the resultant value can be considered as the cost for constructing the cost matrix. For example, P_{21} indicates the APF (i.e., total number of frequent transactions tried to access the particular fragment from the site) of site 1 on fragment 2. The following Table 3 shows the structure of a cost matrix.

Fragment \ site	S_1	S_2	…	S_m
f_1	P_{11}	P_{12}	…	P_{1m}
f_2	P_{21}	P_{22}	…	P_{2m}
…	…	…	…	…
f_n	P_{n1}	P_{n2}	…	P_{nm}

Table 3: Formation of cost matrix based on APF

The cost matrix as shown in Table 3 is attained by an APF update table, which is formulated for every site of the cluster that measures occurrences the transactions from that site for a particular fragment. The table 3.1 shows the model of APF update table for one site and four fragments stored in four different sites of a collection,

Site / Fragment	S1
f1	**8**
f2	**18**
f3	**6**
f4	**5**

Table 3.1: APF update table for Site1 of cluster 1

This paper also derives another formula for measuring the ratio of successful queries, as indicated in Equation 2,

Ratio of successful queries = $\frac{\sum(CT_s)}{\sum(T)}$ (2)where $\sum(CT_s)$ = Sum of cost of successful transactions, $\sum(T)$ = Total number of transactions

3.3. Experimental results and findings of acodas

This section describes the experimental results and findings on reallocation and replication of fragments by using ACODAS. This paper takes Cluster1 [4] with four fragments and four sites. Here the total transactions measured are 160. The table 4 shows the cost matrix of cluster 1.

site / Fragment	S_1	S_3	S_6	S_{10}
f_1	**8**	**12**	**10**	**10**
f_2	**18**	**10**	**7**	**5**
f_3	**6**	**20**	**6**	**8**
f_4	**5**	**17**	**8**	**10**

Table 4: Cost matrix of Cluster 1

In Table 4, the value 17 indicates the APF value of site S_3 for fragment f_4, which means that there were 17 transactions tried to access fragment f_4 from the site S_3 and this can be considered as the cost to form the cost matrix.

The initial allocation of fragments is indicated in Table 5 by encircling the respective costs. By applying Equation 2 on the selected costs, the Equation 3 shows the success ratio with initial allocation,

Fragment \ site	S_1	S_3	S_6	S_{10}
f_1	**8**	**12**	**10**	**10**
f_2	**18**	**10**	**7**	**5**
f_3	**6**	**20**	**6**	**8**
f_4	**5**	**17**	**8**	**10**

Table 5: Initial allocation and successful transactions ofCluster 1

$$\text{Success ratio} = \frac{34}{160} = 21.25\ \% = 21\% \qquad (3)$$

(Initial allocation)

Applying ACODAS into the Table 4, the Table 6 shows the suggestions of ACODAS for reallocation of fragments.

Fragment \ site	S_1	S_3	S_6	S_{10}
f_1	**8**	**12**	**10**	**10**
f_2	**18**	**10**	**7**	**5**
f_3	**6**	**20**	**6**	**8**
f_4	**5**	**17**	**8**	**10**

Table 6: ACODAS suggestions for reallocation of fragments and respective successful transactions incluster 1

The principle of ACODAS suggests that to reallocate the fragments into the places which are having high APF value. Applying the Equation 2 on the selected costs of Table 6, the Equation 4 narrates the success ratio with optimized reallocation based on ACODAS.

$$\text{Success ratio of Reallocation} = \frac{67}{160} = 41.87\ \% = 42\% \qquad (4)$$

(Based on ACODAS)

Replication is done to improve the performance by means of strengthening the availability of data. This paper utilizes the selective replication mechanism, which is highly appreciated in non-redundant data allocation methodologies. Replication process considers the both places, that is the places of initial allocation and suggested places of ACODAS. The Table 7 depicts the optimized places for replication of fragments.

site / Fragment	S_1	S_3	S_6	S_{10}
f_1	8	12	10	10
f_2	18	10	7	5
f_3	6	20	6	8
f_4	5	17	8	10

Table 7: ACODAS suggestions for optimized replication and respective successful transactions incluster 1

Applying the Equation 2 on the selected costs of Table 7, the Equation 5 describes the success ratio with optimized replication using ACODAS.

$$\text{Success ratio of Replication} = \frac{101}{160} = 63.12\% = 63\% \quad (5)$$

(Based on ACODAS)

3.3.1. Interpretations

In the following section, the QoS parameters such as, Success ratio, Resource utilization, Load balancing and Scalability are analyzed based on the placements of fragments listed in Table 8 and the findings as given in Equations (3) (4) and (5). The following Table 8 is showing the placements of fragments in both initial allocation and ACODAS based optimized reallocation.

SITE ID	FRAGMENT ID Initial allocation	FRAGMENT ID Reallocation with ACODAS
S_1	f_1	--
S_3	f_2	--
S_6	f_3	f_1, f_3, f_4
S_{10}	f_4	f_2

Table 8: Fragment placements with initial and ACODASbased reallocation

By analysing the Equations (3) (4) and (5), the ACODAS produces better success ratio than initial allocation. From Table 8, it is found that, the ACODAS leaves certain sites without assigning the fragments. Hence, the ACODAS gives unsatisfactory results against the QoS parameters, such as, Resource utilization and Load Balancing.The scalability of the proposed methodology is studied by conducting various experiments and the results are explained in the following Table 9.

Number of Fragments & Sites	**Total number of Transactions**	**Success ratio (%)**		
		Initial allocation	ACODAS Reallocation	ACODAS Replication
5	30	18	39	65
10	80	20	45	78
15	120	21	35	68
25	180	30	62	88
30	200	40	59	88

Table 9: Experimental results to test the Scalability of ACODAS

The results displayed for ACODAS based reallocation and replication in Table 9 clearly explained the scalability nature of ACODAS. The results show the ability of the ACODAS to produce consistency upon the various experimentations ranges from 30 transactions to 200 transactions. The Figure 2 clearly explained the scalability nature of ACODAS.

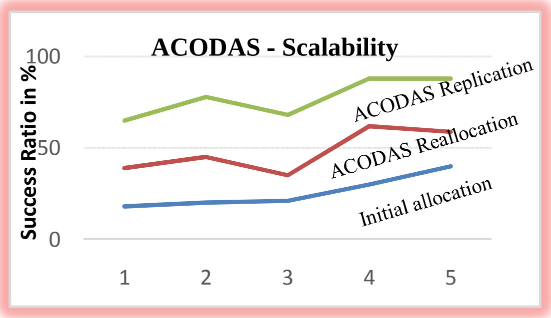

Figure 2: Scalability of ACODAS

4. Conclusion

This paper concentrated on the issues of data availability and data accessibility in P2P Distributed databases. This is an optimized model for reallocation and replication of fragments. This methodology is proposed using Ant Colony Optimization algorithm, namely, *ACODAS*.The process of finding APF to formulate cost matrix using APF update table is explained. Based on the principles of ACO, the reallocation and replication procedure explained. The investigation of QoS parameters against *ACODAS* are analysed and the results are tabulated and graphically represented. The *ACODAS* gives good results in success ratio and scalability, but, the results for resource utilization and load balancing are not satisfactory. Due to the meta-heuristic nature of ACO, *ACODAS* is able to produce only near optimal solution.

The future work may consider this issue and resolve by proposing another Data Allocation Strategy using proven optimization techniques.

REFERENCES

1. A.Correia, J.Luís Rodrigues, N.Carvalho and R.Oliveira, Practical database replication, Replication, *LNCS*, vol. 5959/2010, pp.253-285, DOI: 10.1007/978-3-642-11294-2_13 Springer-Verlag, Jan. 2010.

2. A.A.Amer, H.I.Abdalla, A heuristic approach to re-allocate data fragments in DDBSs, In. International Conference on Information Technology and e-Services (ICITeS), 2012.
3. B.Kemme, Gustavo alonso, database replication: a tale of research across communities, in *Proc. VLDB Endowment,* 3(1-2) (2010) 5-12.
4. D.I.George Amalarethinam and C.Balakrishnan, A study on performance evaluation of peer-to-peer distributed databases, *IOSR Journal of Engineering,* 2(5) (2012) 1168-1176.
5. M.Dorigo, M.Birattari and T.Stützle, Ant colony optimization, Technical Report 2006-023, UniversitéLibre de Bruxelles, Belgium, 2006
6. F.Xhafa, A.Diana Potlog, E.Spaho, F.Pop, V.Cristea and L.Barolli, Evaluation of intra-group optimistic data replication in P2P groupware systems, in Concurrency and Computation: Practice and Experience, Ed. Geoffrey C. Fox and Luc Moreau: John Wiley & Sons, Ltd., 2012
7. H.I.Abdalla, A new data re-allocation model for distributed database systems, *International Journal of Database Theory and Application*, 5(2)(2012) 45-60.
8. J.O.Hauglid, N.H.Ryeng and K.Nørvåg, DYFRAM: dynamic fragmentation and replica Management in distributed database systems,*Distributed and Parallel Databases*, 28 (2-3) (2010) 157-185.
9. J.-L.Deneubourg, S.Aron, S.Goss and J.M.Pasteels, The self-organizing exploratory pattern of the argentine, Ant, *Journal of lnsectBehavior*, 3(2) (1990) 159-168.
10. Z.Lee, S.Su and C.Lee, A heuristic genetic algorithm for solving resource allocation problems, *Knowledge Information Systems*, 5(4) (2003) 503–511.
11. M.T.Ozsu and P.Valduriez, Principles of Distributed Database Systems, 3rd ed., Springer, 2011.
12. R.S.Virk and G.Singh, Optimizing access strategies for a distributed database design using genetic fragmentation, *International Journal of Computer Science and Network Security*, 11(6) (2011) 180-183.
13. R.Basseda, M.Rahgozar and C.Lucas, Fuzzy neighborhood allocation (FNA): a fuzzy approach to improve near neighborhood allocation in DDB,*Advances in Computer Science and Engineering Communications in Computer and Information Science*, 6(2) (2009) 834-837.
14. R.K.Adl and S. Mohammad, T.R.Rankoohi, A new ant colony optimization based algorithm for data allocation problem in distributed database, *Knowledge and Information Systems*, 20(3) (2009) 349—373.
15. U.Tosun, T.Dokeroglu and A.Cosar, Heuristic algorithms for fragment allocation in a distributed database system, In. 27th International Symposium on Computer and Information Sciences, Paris, 2012.
16. Y.-F.Huang and J.-H.Chen, Fragment allocation in distributed database design, *Journal of Information Science and Engineering,* 17 (2001) 491-506.