

oDASuANCO - Ant Colony Optimization based Data Allocation Strategy in Peer-to-Peer Distributed Databases

Dr. D.I. George Amalarethinam¹, C. Balakrishnan²

¹Director-MCA, Associate Professor, Department of Computer Science, Jamal Mohamed College (Autonomous), Tiruchirappalli, India

²Assistant Professor, Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli, India

¹di_geogre@jmc.edu ²balasjc@gmail.com

Abstract: Data Allocation Problem (DAP) incorporates Allocation and Replication methodologies which are involved during the phase of Distributed Database design. Accessibility and availability are the thriving factors for better design of Distributed Databases. High degree of Accessibility and Availability are the outcomes of effective methodologies for Allocation and Replication. This paper proposes a methodology for Reallocation and Replication of fragments in Peer-to-Peer Distributed Database Systems (P2PDDBS) environment using Ant Colony Optimization (ACO) principle, namely, optimized Data Allocation Strategy using Ant Colony Optimization (*oDASuANCO*). 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, Data Allocation Problem, FlexiPeer, Peer-to-Peer Distributed Databases, Reallocation, Replication

1. Introduction

The Distributed Database Systems (DDBS) are more compatible not only with decentralized nature and ability to store growing volume of data, but also to serve the queries in an effective manner by providing a higher degree of parallelism, and improved availability as well as accessibility [1]. Accessibility and availability are the most important aspects to be addressed by any of DDBS environment [2]. 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 leading to accomplish high degree of availability. Particularly, this research 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 information 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 information 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, '*optimized Data Allocation Strategy using Ant Colony Optimization*' (*oDASuANCO*). The Simulation result justifies the effectiveness of *oDASuANCO* based reallocation and replication that the proposed methodology increases the number of successful transactions than initial allocation.

The next section narrates the related works on reallocation and replication management. The section III describes the initial allocation. The section IV describes the criteria for using heuristic algorithm like ACO to solve DAP. Section V explains the methodology of *oDASuANCO*. Section VI elaborates the simulation results of *oDASuANCO*. The section VII concludes the work.

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 DDBS design. Better allocation schemes result in high degree of accessibility. The following section describes the heuristic methodologies used for allocation and reallocation of fragments to the sites in DDBS Rosa Karimi Adl et al., [1] coined a heuristic algorithm for fragment allocation which is based on the ant colony optimization, a meta-heuristic algorithm. The ACO based approach applied for query optimization and integrity enforcement in DDBS. The goal is for efficient data allocation scheme and to minimize the total transaction response time under memory capacity constraints of the sites. Yin-Fu Huang and Jyh-Her Chen [3] proposed a simple and comprehensive model for a fragment allocation problem, also developed two heuristics algorithms to find an optimal allocation of the fragments.

Reza Basseda et al. [4] introduced the fuzzy inference engine to the existing Near Neighborhood Allocation (NNA) algorithm and studied the performance improvement of new Fuzzy Neighborhood Allocation method. Adrian Runceanu [5] drafted an evaluation tool, namely, *EvalTool* to implement a heuristic algorithm for vertical fragmentation.



Rodolfo A. Pazos R, et al.,[6] proposed a vertical fragmentation methodology to optimize the roundtrip response time (query transmission time, query processing time, and response transmission time) by considering the threshold accepting algorithm (a variant of simulated annealing) and ‘tabu’ search algorithm.

Jon Olav Hauglid et al., [7] 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.

Rajinder Singh Virk et al., [8] 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., [9] used the vertical fragmentation technique along with a two-phase allocation 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, Hassan I. Abdalla [10] proposed a methodology for re-allocation of fragments. Umut Tosun et al. [11] compared the heuristic algorithms for fragment allocation by involving Genetic Algorithm principles and proposed a Data Allocation methodology based on Quadratic Assignment Problem. Mehdi Goli and Seyed Mohammad Taghi Rouhani Rankoohi [12] applied the swarm intelligence algorithms with ant collective behavior factors to study the vertical fragmentation problem in Distributed Databases in order to reduce the transaction access cost.

Availability of the data is the thriving requirement in the field of DDBS. Better replication scheme results in improved availability of data in DDBS. During the recent years, enormous amount focuses are given to generate methodologies for replication. The research community are categorized the focusing area on replication as general issues, scalability, Snapshot Isolation (SI) and middleware based replication methodologies. Among these aspects this research focuses on general issues in selective replication. The following section expresses the related works on replication.

Ing-Ray Chen et al. [13] used *dynamic voting concept* with *Stochastic Petri nets* as a tool to analyze the reaction characteristics during failures of reconfigurable algorithms for managing replicated data that maximized the data availability. Bettina et al. [14] reviewed the motivation of Postgres-R, a replication model for distributed environment and explained the evolution of replication strategies from lazy replication to cloud based replication.

A book chapter written by Bettina et al. [15] narrated how transactions are executed in a replicated environment and illustrated the design alternatives using a two-step approach. A book written by Alfranio et al. [16] addressed the architectural, integration challenges on the functionality of generally available software components and a set of practical challenges raised between performance assumptions and actual environments in real workloads. FatosXhafa et al. [17] proposed a replication methodology for documents structured as XML files in Peer-to-Peer systems.

3. FlexiPeer Architecture and Initial Allocation Strategy

According to the architecture of *FlexiPeer*, the Accounts relation [18] is taken for analyzing the initial allocation. Table 1 depicts the initial allocation to the sites of four clusters in *FlexiPeer*.

TABLE 1. INITIAL ALLOCATION TO SITES IN CLUSTERS

ACCOUNT N	CATEGOR	CUSTOMER I	DATE	BALANC	REGIO	SIT	CLUSTE
1	A	C1	11/1/1	21000	R1	1	1
10	A	C10	9/4/1	78000	R1	1	1
3	B	C3	2/2/1	18000	R1	3	1
6	C	C6	15/3/1	52000	R1	6	1
8	D	C8	28/3/1	11500	R1	10	1
2	B	C2	21/1/1	13500	R2	2	2
12	B	C12	18/4/1	11800	R2	2	2
7	E	C7	18/3/1	38000	R2	5	2
4	C	C4	8/2/1	22000	R3	4	3
9	A	C9	4/4/1	16800	R3	7	3
5	D	C5	24/2/1	3200	R4	8	4
11	B	C11	11/4/1	23000	R4	9	4

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



TABLE 2. INITIAL ALLOCATION OF FRAGMENTS TO CLUSTER 1

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

Cluster 1 consisting of four sites, hence four fragments are derived [18] and allocated subsequently.

4. Criteria for using Heuristic Algorithms to solve DAP

The performance of any DDBS environment is directly proportional to the efficiency of the way that the data are allocated to the sites. To attain a better data allocation scheme the strategy must be optimal and sustainable for variety of environments. Thus, the data allocation problem can be categorized within resource allocation problems which are defined as 'optimization problems with constraints' [19]. The allocation problem is a NP-complete [1] in nature and requires heuristic methods [12] to solve allocation problem in an efficient manner. Therefore, to achieve a suitable solution for the allocation problem within the acceptable computation time, most of the proposed algorithms [1] [8] [12] are concerned with the usage of heuristic and evolutionary algorithms such as, ACO and Genetic algorithm for solving DAP.

Another important aspect to incorporate heuristic algorithms to solve a problem is that the problem must be belonging to the NP-Hard or NP-Complete. Classes of time complexity are defined to distinguish problems according to their *hardness*. Class P consists of all those problems that can be solved on a deterministic Turing machine in polynomial time from the size of the input. Turing machines are an abstraction that is used to formalize the notion of algorithm and computational complexity. Class NP consists of all those problems whose solution can be found in polynomial time on a non-deterministic Turing machine. Since such a machine does not exist, practically it means that an exponential algorithm can be written for an NP-problem, nothing is asserted whether a polynomial algorithm exists or not. A subclass of NP, class NP-complete includes problems such that a polynomial algorithm for one of them could be transformed to polynomial algorithms for solving all other NP problems. Finally, the class NP-hard can be understood as the class of problems that are NP-complete or harder. NP-hard problems have the same trait as NP-complete problems but they do not necessary belong to class NP, that is class NP-hard includes the problems for which no polynomial algorithms at all can be provided the result. The application of heuristic algorithms in solving NP-complete gives good result than the conventional algorithms. The DAP can be considered as NP-complete [Ros 09] problem. Hence, it is hard to solve the DAP by polynomial algorithms, therefore, heuristics can be used to solve DAP.

After analysing most of the heuristic approaches, the *Swarm intelligence* technique particularly ACO having the characteristics to fit to solve DAP. Due to the high complexity of the DAP. Most of the existing heuristic algorithms for solving DAP are greedy ones, and very recently, stochastic algorithms such as genetic algorithms [8] and swarm intelligence [1] [12] are applied. One of the most significant advantages of stochastic algorithms is their scalability. In particular, one class of stochastic algorithm, which gives better results for solving DAP, is the algorithm based on ant collective behaviour because of close optimal result producing capacity. Scalability and optimal results are being the important properties of DAP, thus, ant behaviour based algorithms [20] are the most desirable candidate to solve DAP than genetic algorithms.

5. Problem Definition of oDASuANCO

From the facts gained from the earlier sections of this chapter, this research motivated to use ACO for solving DAP. ACO is a meta-heuristic inspired by the behaviour of real ants that cooperate through self-organization [20]. 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 research 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 [21] as in (1) is derived,

$$\sum_{i=1}^n \sum_{j=1}^n P_{ij} = \frac{(S_m + K)^h}{\sum (S_m + K)^h} \quad (1)$$

where $m = 1 \dots n$
 K and h are constants and S indicates the sites



When applying the (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.

TABLE 3. FORMATION OF COST MATRIX BASED ON APF

	site				
		S_1	S_2	...	S_m
	Fragment				
	f_1	P_{11}	P_{12}	...	P_{1m}
	f_2	P_{21}	P_{22}	...	P_{2m}
	f_3	P_{31}	P_{32}	...	P_{3m}

	f_n	P_{n1}	P_{n2}	...	P_{nm}

The cost matrix as shown in Table 3 is attained by an APF update table. APF update table is a table which is formulated for every site of the cluster that measures occurrences the transactions from that site for a particular fragment. The Table 4 shows the model of APF update table for one site and four fragments stored in four different sites of a cluster,

TABLE 4 APF UPDATE TABLE FOR SITE1 OF CLUSTER 1

Fragment	Site S1
f_1	10
f_2	5
f_3	6
f_4	7

The researcher also derives another formula for measuring the ratio of successful queries, as indicated in (2),

$$\text{Ratio of successful queries} = \frac{\sum(CT_s)}{\sum(T)} \tag{2}$$

$$\sum(CT_s) = \text{Sum of cost of successful transactions}$$

$$\sum(T) = \text{Total number of transactions}$$

6. Experimental results and findings of oDASuANCO

This section describes the experimental results and findings on reallocation and replication of fragments by using oDASuANCO. This study takes cluster1 [18] with four fragments and four sites. Here the total transactions are measured as 160. The Table 5 shows the cost matrix of cluster 1,



TABLE 5. COST MATRIX OF CLUSTER 1

Fragment	Sit			
	S ₁	S ₃	S ₆	S ₁₀
f ₁	10	6	18	6
f ₂	5	10	8	17
f ₃	6	7	15	12
f _n	7	8	17	8

In Table 4, the value 15 indicates the APF value of site S6 for fragment f3, which means that there are 15 transactions were tried to access fragment f3 from the site S6 and this can be considered as the cost to form cost matrix.

The initial allocation of fragments is indicated in Table 6 by encircling the respective costs. By applying (2) on the selected costs, the (3) shows the success ratio with initial allocation,

TABLE 6 INITIAL ALLOCATION AND SUCCESSFUL TRANSACTIONS OF CLUSTER 1

site \ Fragment	S1	S3	S6	S10
f1	10	6	18	6
f2	5	10	8	17
f3	6	7	15	12
f4	7	8	17	8

$$\text{Success ratio} = \frac{43}{160} = 27\% \quad (3)$$

(Initial allocation)

TABLE 7 oDASuANCO SUGGESTIONS FOR REALLOCATION OF FRAGMENTS AND RESPECTIVE SUCCESSFUL TRANSACTIONS IN CLUSTER 1

site \ Fragment	S1	S3	S6	S10
f1	10	6	18	6
f2	5	10	8	17
f3	6	7	15	12
f4	7	8	17	8

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

(Based on oDASuANCO)

Applying oDASuANCO into the Table 5, the Table 7 shows the suggestions of oDASuANCO for reallocation of fragments,

The principle of oDASuANCO suggests that to reallocate the fragments into the places which are having high APF value. Applying the (2) on the selected costs of Table 7, the (4) narrates the success ratio with optimized reallocation based on oDASuANCO.

Replication is done to improve the performance by means of strengthening the availability of data. This research utilizes the selective replication mechanism. Selective replication methodology is very much appreciated in non-redundant data allocation methodologies.

A. REPLICATION

Replication process considers the both places, that is the places of initial allocation and suggested places of oDASuANCO. The Table 8 depicts the optimized places for replication of fragments.



TABLE 8. oDASuANCO SUGGESTIONS FOR OPTIMIZED REPLICATION AND RESPECTIVE SUCCESSFUL TRANSACTIONS IN CLUSTER 1

site \ Fragment	S1	S3	S6	S10
f1	10	6	18	6
f2	5	10	8	17
f3	6	7	15	12
f4	7	8	17	8

Applying the (2) on the selected costs of Table 8, the (5) describes the success ratio with optimized replication using oDASuANCO.

$$\text{Success ratio of Replication} = \frac{95}{160} = 59\% \tag{5}$$

(Based on oDASuANCO)

B. RESULTS AND INTERPRETATIONS

Analysing (3), (4) and (5), the following Table 9 shows a comparative analysis of the results produced by oDASuANCO.

TABLE 9. COMPARATIVE ANALYSIS OF SUCCESS RATIO USING oDASuANCO

Operation	Success ratio
Initial Allocation	27%
oDASuANCO based Reallocation	42%
oDASuANCO based Replication	59%

The results in Table 9 clearly revealed that oDASuANCO based data allocation methodology gives better success ratio than initial allocation.

The following Table 10 is showing the initial allocation and oDASuANCO based optimized reallocation.

TABLE 10. FRAGMENT PLACEMENTS WITH INITIAL AND oDASuANCO BASED REALLOCATION

SITE ID	FRAGMENT ID Initial allocation	FRAGMENT ID Reallocation with oDASuANCO
S1	f1	--
S3	f2	--
S6	f3	f1, f3, f4
S10	f4	f2



The QoS parameters such as, Success ratio, Resource utilization, Load balancing and Scalability are analyzed based on the findings given in Table 9 and the placements of fragments listed in Table 10.

Success ratio: By analysing the Table 9, the *oDASuANCO* produces better success ratio than initial allocation

Resource utilization and Load Balancing: By analysing the Table 10, the *oDASuANCO* leaves certain sites without assigning the fragments. Hence, the *oDASuANCO* gives unsatisfactory results against these QoS parameters.

Scalability: The scalability of the proposed methodology is studied and the results explained in the following Table 11 by listing the various experimental results for *oDASuANCO*:

TABLE 11. EXPERIMENTAL RESULTS TO TEST THE SCALABILITY OF *oDASuANCO*

Number of Fragments & Site	Total number of Transactions	Success ratio (%)		
		Initial allocation	<i>oDASuANCO</i> Reallocation	<i>oDASuANCO</i> Replication
5	30	25	46	70
10	80	13	29	37
15	120	21	35	52
25	180	30	52	71
30	200	20	47	68

The results displayed for *oDASuANCO* based reallocation and replication in Table 11 clearly explained the scalability nature of *oDASuANCO*. The results show the ability of the *oDASuANCO* to produce consistency upon the various experimentations ranges from 30 transactions to 200 transactions. The following Figure 1 clearly explained the scalability nature of *oDASuANCO*.

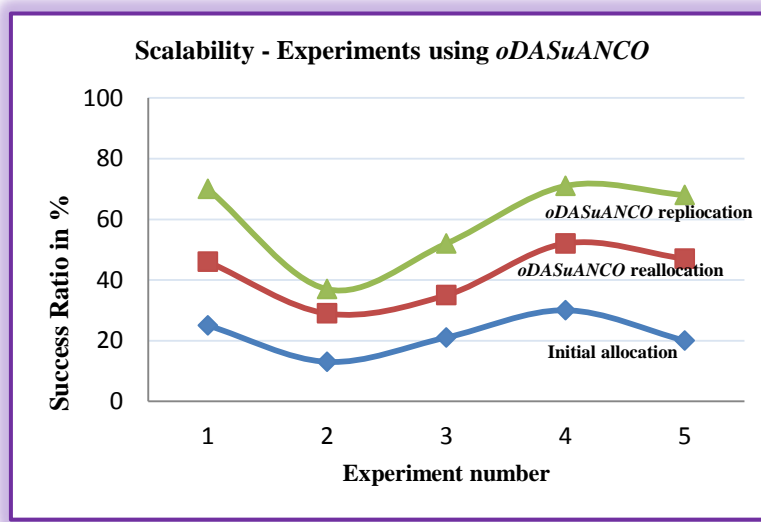


Figure 1 Scalability of *oDASuANCO*

6. Conclusion

This paper concentrated on the issues of data availability and data accessibility in P2PDDBS. This is an optimized model for reallocation and replication of fragments. This methodology is proposed using Ant Colony Optimization algorithm, namely, *oDASuANCO*. 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 *oDASuANCO* are analysed and the results are tabulated and graphically represented.

The *oDASuANCO* gives better 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, *oDASuANCO* 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]. Rosa KarimiAdl and Seyed Mohammad TaghiRouhaniRankoohi, "A new ant colony optimization based algorithm for data allocation problem in distributed database", Knowledge and Information Systems, Volume 20, Number 3, 349--373, DOI: 10.1007/s10115-008-0182-y. 2009.
- [2]. M. T. Ozsu and P. Valduriez, Principles of Distributed Database Systems, 3rd ed., Springer, 2011.
- [3]. Yin-Fu Huang and Jyh-Her Chen, "Fragment Allocation in Distributed Database Design", Journal Of Information Science And Engineering 17, pp. 491-506, 2001.
- [4]. Reza Basseda, MaseudRahgozar, and CaroLucas, "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, Volume 6, Part 2, 834--837, DOI: 10.1007/978-3-540-89985-3_114. 2009.
- [5]. Adrian Runceanu, MarianPopescu, "EvalTool for Evaluate the Partitioning Scheme of Distributed Databases", Technological Developments in Networking, Education and Automation, 627--632, DOI: 10.1007/978-90-481-9151-2_109, 2010.
- [6]. Rodolfo A. Pazos R., Graciela Vázquez A., Joaquín Pérez O., José and A. MartínezF., "Vertical Fragmentation Design of Distributed Databases Considering the Nonlinear Nature of Roundtrip Response Time. Knowledge-Based and Intelligent Information and Engineering Systems", LNCS, Springer, Volume 6277/2010, 173--182, DOI: 10.1007/978-3-642-15390-7_18. 2010.
- [7]. Jon Olav Hauglid, NorvaldH.Ryeng and KjetilNørvåg, "DYFRAM: dynamic fragmentation and replica Management in distributed database systems. Distributed and Parallel Databases", 157--185, DOI:10.1007/s10619-010-7068-1, Springer, 2010.
- [8]. Rajinder Singh Virk and GurdinderSingh, "Optimizing Access Strategies for a Distributed Database Design using Genetic Fragmentation", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.6. 2011.
- [9]. Ali A. Amer, Hassan I. Abdalla, "A heuristic approach to re-allocate data fragments in DDBSs", In. International Conference on Information Technology and e-Services (ICITeS), 2012.
- [10]. Hassan I. Abdalla, "A New Data Re-Allocation Model for Distributed Database Systems", International Journal of Database Theory and Application, Volume 5, Number 2. 2012.
- [11]. UmutTosun, TanselDokeroğlu and AhmetCosar, "Heuristic Algorithms for Fragment Allocation in a Distributed Database System", In. 27th International Symposium on Computer and Information Sciences, Paris, 2012.
- [12]. Mehdi Goli and Seyed Mohammad TaghiRouhaniRankoohi, "A new vertical fragmentation algorithm based on ant collective behavior in distributed database systems", Knowledge and Information Systems, Springer. Volume 30, Number 2. 435--455 DOI: 10.1007/s10115-011-0384-6. 2012.
- [13]. Ing-Ray Chen, Ding-Chau Wang and Chih-Ping Chu, "Response time behavior of distributed voting algorithms for managing replicated data", Information Processing Letters, Volume 75, Issue 6, pp. 247-25, Nov. 2000.
- [14]. Bettina Kemme, Ricardo Jiménez-Peris, Marta Patiño-Martínez and Gustavo Alonso, "Database replication: a tutorial", Replication, LNCS, vol. 5959/2010, pp.219-252, DOI: 10.1007/978-3-642-11294-2_12, Springer-Verlag, Jan. 2010.
- [15]. Bettina Kemme, Gustavo Alonso, "Database replication: a tale of research across communities", in Proc. VLDB Endowment, Volume 3 Issue 1-2, pp.5-12, Sep. 2010.
- [16]. Alfrânio Correia, José Pereira, Luís Rodrigues, Nuno Carvalho and Rui 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.
- [17]. Fatos Xhafa, Alina-Diana Potlog, Evjola Spaho, Florin Pop, Valentin Cristea and Leonard 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.
- [18]. D.I. George Amalarethinam and C. Balakrishnan, "A Study on Performance Evaluation of Peer-to-Peer Distributed Databases", IOSR Journal of Engineering, Vol. 2(5) pp. 1168-1176, May. 2012.
- [19]. Lee Z, Su S and Lee C, "A heuristic genetic algorithm for solving resource allocation problems", Knowledge Information Systems, Vol. 5, No. 4, pp. 503--511, 2003.
- [20]. Dorigo M., Birattari M. and Stützle T, "Ant colony optimization", Technical Report 2006-023, Université Libre de Bruxelles, Belgium, 2006.
- [21]. J.-L. Deneubourg, S. Aron, S. Goss, and J. M. Pasteels, "The Self-Organizing Exploratory Pattern of the Argentine Ant", Journal of Insect Behavior, Vol. 3, No. 2, 1990.

