A Logical Approach Towards Effective Data Search using Ant Colony Optimization in Cloud Environment

Sudipta Sahana[1*] , Tanmoy Mukherjee[1], Debabrata Sarddar[2]

[1] Department of CSE, JIS College of Engineering, Kalyani, Nadia, India
[2] Department of CSE, University of Kalyani, Kalyani, Nadia, India

Corresponding Author: ss.jisce@gmail.com,   Tel.: +91-9474733974

Abstract— The world has revolutionized over the years with the advent of various technologies and life of mankind has taken a significant turnaround in terms of getting the official problems solved in an effective and efficient manner in no time. One of the most powerful technologies that has come up in recent years is cloud computing. This technology has captured a special place in various Information Technology (IT) sectors and business organizations. Among all the aspects of this technology that are in existence, cloud data search optimization has become a key area of focus for the researchers. Various research works were conducted based on several fundamentals such as Gossip Protocol, Genetic Algorithm, Hybrid Algorithm, Multi-Keyword Synonym Query, Particle Swarm Optimization, Honey Bee Optimization, etc. and all these were put into practical purpose with the primary objective of optimizing the search technique in the cloud. In our paper, we have suggested the use of Ant Colony Optimization Algorithm for an effective data search in database and allocating them to the respective clients through shortest possible network path in no time. We have used the concept of pheromone values to conduct this procedure. Our suggested techniques ensure that our algorithm will achieve a higher degree of performance in terms of increased throughput and increased efficiency as compared to the traditional techniques which were carried out earlier.

Keywords— Database, Client machines, Data Carrier Equipment, Wires, Quadrilateral Obstruction, Pheromone value, Ant Colony Optimization Algorithm

I. INTRODUCTION

The technological world has received a huge boost with the advent of cloud computing. The growing demand of Information Technology (IT) organizations and business firms to store a huge amount of confidential data is well served by this technology. Among a whole lot of talking points encompassing this technology, a significant aspect that has captured the attention of this mechanized world is the data search optimization in cloud computing. The IT sectors continuously strive for searching large amount of data in cloud and making all those data available to clients through shortest possible network route such that the computations and operations can be done in very short interval of time. Several algorithms such as Genetic Algorithm, Particle Swarm Optimization, Bacterial Foraging Optimization, Ant Colony Optimization, Honey Bee Optimization, etc. are available in the market through which the search optimization technique in the cloud can be made more effective and efficient.

The Ant Colony Optimization is one of the most widely used techniques. A large number of computational problems can be solved through the use of this algorithm. It finds the shortest path to data based on the movement of ants. The pheromones spread by ants are served as trail for the other ants to follow, who reach the food through shortest possible path. This logic can be applied in real life IT problems, where the cloud data stored in database can be fetched by Data Carrier Equipment, attached to client systems, through shortest path without any delay. This technique has made the life of the IT organizations and business firms a whole lot easier as large amount of computational problems can be computed in less amount of time.

In our paper, we have suggested the use of Ant Colony Optimization Algorithm for an effective data search in database and allocating them to the respective clients through shortest possible network path in no time. We have also used the concept of pheromone values to conduct this procedure. With all the techniques that we have suggested in our paper, it can be ensured that our algorithm will achieve a high degree of performance in terms of increased throughput and increased efficiency.

The rest of the paper discusses about the following aspects of an effective data search in cloud environment. Section 2 of this paper discusses about all the research works that were carried out pertaining to an effective search optimization in cloud. Section 3 is all about our own algorithm to achieve the same along with a well defined flowchart and network diagram. Section 4 talks about the results that can be obtained by making use of our algorithm. It also talks about some of the parameters such as throughput and efficiency which can be achived with a high degree of precision. In section 5, we present a conclusion on the same with the mention of further scope of research on this field.

II.   RELATED WORK

Elham Azhir et al. [1] proposed some techniques to solve the various factors which affect the execution cost of a query. These techniques were categorized into deterministic and non-deterministic methods. The deterministic and non-deterministic query optimization methods were further divided into three subcategories which were cost-based query plan enumeration, multiple query optimization, and adaptive query optimization methods. The advantages and disadvantages of the algorithms needed to solve the query optimization problems in the cloud environments were presented in this paper. Also, these techniques were compared in terms of optimization, time, cost, efficiency, and scalability. It also offered some key areas to improve the cloud query optimization mechanisms in the future.

In paper [2] a novel cloud task scheduling algorithm based on ant colony optimization was proposed so that tasks of cloud users to virtual machines in cloud computing environments could be achieved in an effective manner. Diversification and reinforcement strategies with slave ants were adopted so that the performance of the task scheduler in cloud computing environments with ant colony optimization could be enhanced. The global optimization problem with slave ants could be solved by the proposed algorithm through the avoidance of long paths whose pheromones were wrongly accumulated by leading ants.

In paper [3] the functioning of the search engines along with the role and importance of search engine optimization were explained.

S. Sahana et al. [4] examined a technique that could reduce the cost of storing data in Cloud using Gossip protocol, which was used to distribute and update information across wireless networks and large-scale database systems. It suggested that using this technique would help the users in reducing time required to search for data. Efficiency was exhibited and a lot more energy was saved through the demonstration of this technique, as suggested in the paper. It also mentioned that using this approach would result in

achieving the objective of a green cloud computing infrastructure. The algorithms for accumulation of new information at regional data center (RDC) and information retrieval were also discussed in this paper.

Li Liu et al. [5] proposed an adaptive penalty function for the strict constraints compared with other genetic algorithms. The use of the coevolution approach helped in adjusting the crossover and mutation probability, which was able to accelerate the convergence and prevent the prematurity. The algorithm was also compared with baselines such as Random, particle swarm optimization, Heterogeneous Earliest Finish Time, and genetic algorithm in a WorkflowSim simulator on 4 representative scientific workflows. This algorithm gave better results than the other state-of-the-art algorithms in the criterion of both the deadline-constraint meeting probability and the total execution cost.

In paper [6] the task scheduling in cloud computing environment was optimized, which was based on a proposed Simulated Annealing (SA) based SOS (SASOS) in order to improve the convergence rate and quality of solution of SOS. It also proposed a fitness function which considered the utilization level of virtual machines (VMs) through which makespan and degree of imbalance among VMs could be reduced. CloudSim toolkit was used to evaluate the efficiency of the proposed method using both synthetic and standard workload. The simulation results proved that hybrid SOS performed better than SOS in terms of convergence speed, response time, degree of imbalance, and makespan.

Mohammed Abdullahi et al. [7] presented a Discrete Symbiotic Organism Search (DSOS) algorithm for optimal scheduling of tasks on cloud resources. Symbiotic Organism Search (SOS) was a metaheuristic optimization technique for solving numerical optimization problems. The symbiotic relationships (mutualism, commensalism, and parasitism) exhibited by organisms in an ecosystem, were mimicked by SOS. The simulation results revealed that DSOS outperformed Particle Swarm Optimization (PSO) which is one of the most popular heuristic optimization techniques used for task scheduling problems. DSOS converged faster when the search got larger which made it suitable for large-scale scheduling problems. Analysis of the proposed method conducted using t-test showed that DSOS performance was significantly better than that of PSO particularly for large search space.

In paper [8] the problems based on synonym-based multi-keyword ranked search over encrypted cloud data were discussed and various techniques were devised to mitigate the same. Some techniques were also proposed to address the challenges of applying an effective searchable system with support of ranked search. The proposed algorithm solved the problem of multi keyword ranked search to achieve more

accurate search result and very important search accuracy as single word search and synonym based search to support synonym queries search data on encrypted cloud. The cloud customers were hugely benefited by the Ranked search as they could find the most relevant information quickly. Ranked search also handled the network traffic problem as server only provided the expected and relevant data. The simulation results showed that encrypted data worked properly for encrypted data on cloud.

George Suciu et al. [9] filled a gap in the cloud computing literature by providing a general overview of the challenges and methodologies for acceleration of search applications for Big Data. This paper analyzed cloud techniques that could be used for faster search of large volumes of data. It also presented and discussed the components and interfaces of the proposed Search Based Applications based on EXALEAD CloudView.

Qiang Xu et al. [10] built a mathematical model of data scheduling between data centers. Generational evolution produced better approximate solution, and got the best approximation of the data placement owing to the global optimization ability of the genetic algorithm. The simulation results showed that genetic algorithm could effectively work out the approximate optimal data placement, and the data scheduling between data centers was minimized.

A new algorithm i.e. Hybrid improved particle swarm optimization with mutation crossover (HIPSOMC) was proposed in paper [11]. The main objective was to get maximum benefit from resources, optimized utilization of resources and scheduling played an important part to achieve all these. The proposed algorithm turned out to be much more efficient than the previously designed algorithms i.e. particle swarm optimization, simulated annealing and improved particle swarm optimization and consumed less execution time.

Medhat Tawfeek et al. [12] presented a cloud task scheduling policy based on Ant Colony Optimization (ACO) algorithm and compared with different scheduling algorithms First Come First Served (FCFS) and Round-Robin (RR). The makespan of a given task set could be minimized with the help of this algorithm. The Ant Colony Optimization (ACO) algorithm was used for allocating the incoming jobs to the virtual machines. The cloudsim toolkit package was used in the simulation of the algorithm. The simulation results showed that cloud task scheduling based on ACO outperformed FCFS and RR algorithms.

An overview of the framework for keyword searching with summaries, was given in paper [13]. It also described a ranking algorithm for ranked keyword search and their results. Users could actively utilize clouds to query a collection with the help of keyword search. It discussed a study design on user presenting a query in a cloud.

In paper [14] a new methodology was implied based on Search Engine Optimization (SEO) without getting sandboxed by search engines like Google, Bing and other. This process was involved in implementing safe link building techniques with link velocity as its key without compromising the on page optimization. The latest algorithmic updates were considered and the strategy was developed to rank for a keyword. Organizations could take advantage of the traffic from the search engines and had a good online presence. It was done based on the basic guidelines recommended by all the search engines for proper indexing without sandboxing. This method assured that the online progress of any business would not be hindered.

Lizheng Guo et al. [15] formulated a model for task scheduling and proposed a particle swarm optimization (PSO) algorithm which is based on small position value rule, for minimizing the cost of the processing. Comparisons were drawn between PSO algorithm and the PSO algorithm embedded in crossover and mutation and in the local research. As a result, the simulation results showed that the PSO algorithm not only converged faster but also ran faster than the other two algorithms in a large scale. It also proved that the PSO algorithm was more suitable to cloud computing.

III. METHODOLOGY

We need to have some of the prerequisites to carry out this algorithm. They are given below:

　i.　Database containing a large number of data sets.

　ii.　Client machines (Desktops, Laptops, Tablet, etc.).

　iii.　Data carrier equipment.

　iv.　Wires.

Now, the algorithm (*Figure 1*) for an effective data search in cloud platform using the concept of Ant Colony Optimization can be detailed in the following steps:

Case 1 (When there is no network hurdle between the client machines and database):

　a.　At first, there is a database containing m number of data sets. Let us consider, that each data set is meant for performing a particular operation. Hence, there is m number of operations that are to be performed.

　b.　Next, we need to install certain number of client machines on the opposite side of database. Each

client machine is capable of performing a particular operation. Let us consider that there is n number of client machines. Then the number of client machines should be at least equal to the number of data sets in the database. Mathematically, m<=n.

c. Next, if the number of client machines are falling short than the number of data sets in the database, then we need to install more number of client machines to fulfill the criteria given in step b. Mathematically, if m>n, then install more number of client machines to make it m<=n.

d. Next, there should be a data carrier equipment attached to each of the client machines. Each data carrier equipment should be able to travel through the network to the database and bring each data set, through the network, to each of the client machines, where each operation needs to be performed. If there is m number of data sets, then there should be m number of data carrier equipment actively taking part in searching of database through shortest routes in the network and bringing the data sets through the network and allocating the same to equal number of client machines.

e. Next, there should be a well-defined wired connection between the client machines and the database. The connection should be such that the outgoing wires from the client machines should merge into a single wire and that single wire should be connected to the database.

f. Next, each of the data carrier equipment should be allocated a pheromone value. The pheromone values should be the same for each and every data carrier equipment. When each of the data carrier equipment travels through the network, the pheromone values should be represented in the form of glowing lights over the network.

g. The pheromone values, represented in the form of glowing lights over the network, should serve as indicators for other data carrier equipment to follow.

h. Hence, all the data carrier equipment would be following the same path to reach the database and bring each of the data sets to the respective clients. Once, all the data sets in the database are collected and brought by equal number of data carrier equipment to the respective clients, the database becomes empty and hence an effective search optimization through the network can be achieved.

Case 2 (When there is presence of network hurdle between the client machines and database):

i. Let us consider the same set up (steps a. to f.) as in case 1.

j. Now, there is an obstruction which has come in the path of the single wire, merged from outgoing wires from different client machines, and the database. In our case, let us assume the obstruction to be a network of the shape of quadrilateral.

k. The quadrilateral obstruction within the path is such that it is shorter on one side of the network and longer on the other side of network. Let us assume that data carrier equipment cannot move on the straight path of network because of this obstruction.

l. For our convenience, let us take the sides of the quadrilateral obstruction as a, b, c and d. The shorter sides are a and c and the longer sides are b and d. Let a be the shorter side, located above the single wire of network, facing the client machines and c be the shorter side, located above the single wire of network, facing the database. Similarly, b is the longer side, located below the single wire of network, facing the client machines and d is the longer side, located below the single wire of network, facing the database.

m. As there is no pheromone value in the quadrilateral obstruction path, so half of the data carrier equipment will go towards the side a of the quadrilateral obstruction and the other half will go towards the side b of the quadrilateral obstruction. Let us suppose, there are m number of data carrier equipment originating from m number of client machines out of n number of client machines, then m/2 number of data carrier equipment will go towards the side a of the quadrilateral obstruction and the other m/2 will go towards the side b of the quadrilateral obstruction.

n. Based on the logic given in the above step (step m.), it can be inferred that half of the data carrier equipment, carrying data sets from the database to the client, will move towards the side c of the quadrilateral obstruction and the other half will move towards the side d of the quadrilateral obstruction. Let us suppose, there is m number of data carrier equipment carrying m number of data sets from database to m number of clients, then m/2 number of data carrier equipment will go towards the side c of the quadrilateral obstruction and the other m/2 will go towards the side d of the quadrilateral obstruction.

o. Let us suppose that the length of the shorter sides (a and c) is x each and the length of the longer sides (b and d) is 2x each.

p. Now, let us assume that each data carrier equipment takes 1 second to cover 2x distance. Also, pheromone values, represented in the form of glowing lights, remain as it is for 1 second, become ½ in next second and then $1/4^{th}$ in next second and so on and so forth. Hence, after 1 second, side a of the quadrilateral obstruction has pheromone values for twice the data carrier equipment of side b. Also, side c of the quadrilateral obstruction has pheromone values for twice the data carrier equipment of side d.

q. Depending on the above logic (given in step p.), $2/3^{rd}$ of the data carrier equipment will move towards side a of the quadrilateral obstruction and the remaining $1/3^{rd}$ will move towards side b of the quadrilateral obstruction, based on the pheromone values, represented in the form of glowing lights. Hence, more and more data carrier equipment starts to move towards side a than side b, based on the concentration of pheromone values. Also, the concentration of pheromone values and accordingly the number of data carrier equipment will start increasing towards side a and the same will keep on decreasing towards side b of the quadrilateral obstruction. The same logic is applicable for sides c and d of the quadrilateral obstruction.

r. After a certain amount of time, it can be observed that all data carrier equipment will be following the sides a and c of the quadrilateral obstruction which is the shortest path from the client machines to the database and vice-versa. Hence, an effective and efficient path to data sets in the database can be achieved using the concept of Ant Colony Optimization Algorithm.
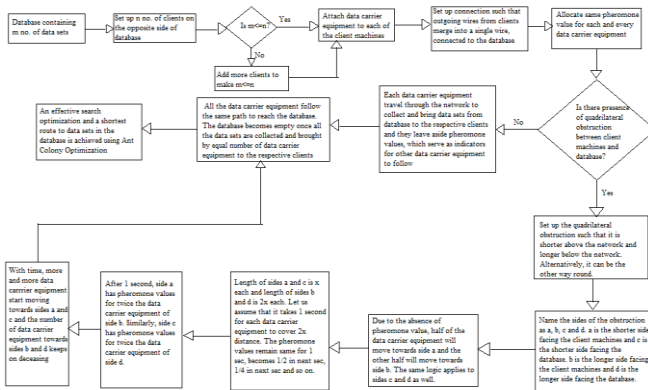

Figure 1. Flowchart for our algorithm

No Network Hurdle Between the Client Machines and Database

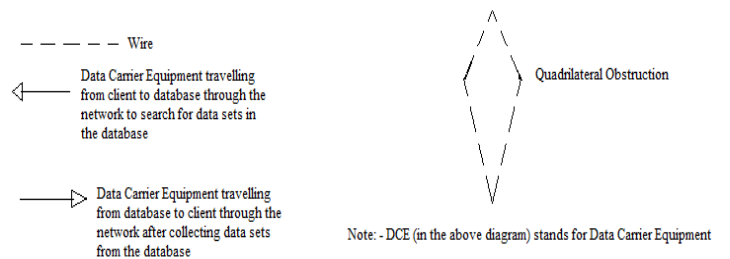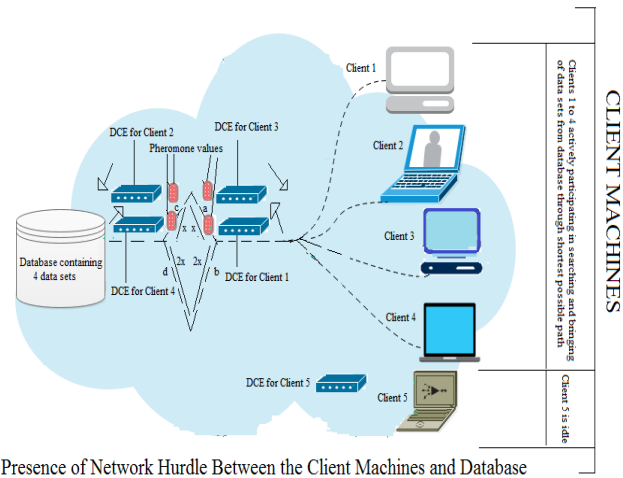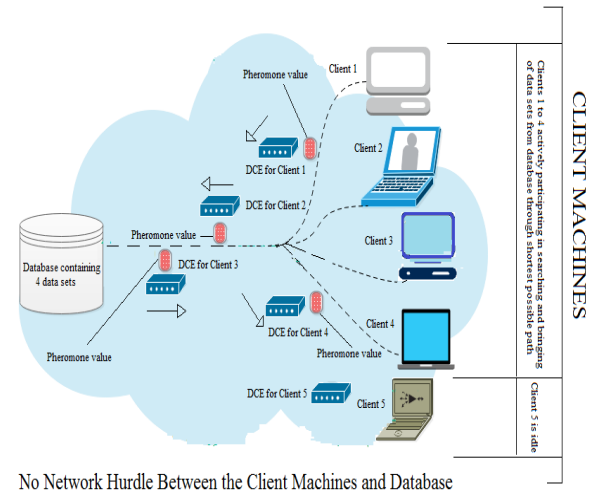Presence of Network Hurdle Between the Client Machines and Database

Figure 2. Diagrammatic Representation of an Effective Data Search through Shortest Possible Path using Ant Colony Optimization in Cloud Environment

When there is no network hurdle between the client machines and database, Data Carrier Equipment of clients 1 to 4 travel through straight path in the network for searching the data sets and bringing them from the database and allocating them to the respective clients. However, when

there is network hurdle (quadrilateral obstruction) between the client machines and database, Data Carrier Equipment of clients 1 to 4 travel through paths a and c of the quadrilateral obstruction for searching the data sets and bringing them from the database and allocating them to the respective clients. Hence, an effective data search through shortest possible path can be achieved using Ant Colony Optimization in Cloud Environment (*Figure 2*).

IV. RESULTS AND DISCUSSION

In this paper, we have used the Ant Colony Optimization Algorithm for an effective data search through shortest possible path in Cloud Environment. When our algorithm is compared with other optimization algorithms, it can be observed that our algorithm performs a lot better in terms of throughput and efficiency. In our model, even if the database contains a large number of data sets, it can be guaranteed that the data sets will be allocated to the respective clients a lot quicker. Irrespective of the distance between the client machines and the database, our algorithm will go a long way in allocating large number of data sets to the respective clients through shortest possible path and in less time.

We have suggested the use of pheromone values in our paper, which serve as indicator for other Data Carrier Equipment to follow. More number of data sets in the database suggests the active participation of more number of client machines and the active participation of more number of client machines means the active participation of more number of Data Carrier Equipment and more the number of actively participating Data Carrier Equipment, more the concentration of pheromone values in the network. Hence, through the use of our algorithm it can be inferred that the process of searching and allocating the data sets to the respective clients will be a lot quicker.

Considering all the factors, we can safely tell that the throughput of our system will increase with more number of data sets in the database. The graphical representation for both the traditional approach and our proposed approach is shown below (*Figure 3*).
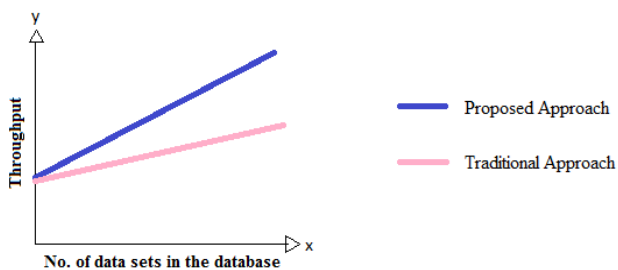

Figure 3. Comparing the proposed approach with the traditional approach showing the throughput will increase with increasing number of data sets in the database

Next, the efficiency of our system will increase with more number of data sets in the database. The graphical representation for both the traditional approach and our proposed approach is shown below (*Figure 4*).
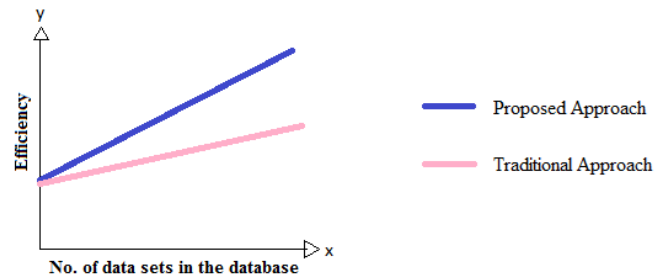

Figure 4. Comparing the proposed approach with the traditional approach showing the efficiency will increase with increasing number of data sets in the database

V. CONCLUSION AND FUTURE SCOPE

An effective search optimization in cloud has been one of the primary objectives for IT industries and business firms who are looking to operate with a large amount of data in a limited amount of time with reduced manpower and cost effectiveness. The other parameters which are taken into account are throughput and efficiency. The profitable organizations are the ones which apart from serving the clients' requests effectively and efficiently, achieves a high degree of performance in terms of having all these parameters to be met. In our paper, we have suggested an effective search optimization technique through the use of Ant Colony Optimization Algorithm and the pheromone values. Our suggested technique ensures that a large number of data sets present in database can be fetched by actively participating Data Carrier Equipment and can be allocated to large number of clients through shortest possible path in the network in a short interval of time. Apart from that, our technique will also go a long way in achieving increased throughput and increased efficiency with the increase in number of data sets in the database. In the future, we are planning to put our suggested technique into practical purpose, where we will look to optimize the data search in cloud and getting a large number of clients' requests allocated effectively and efficiently.

REFERENCES

[1] Elham Azhir, Nima Jafari Navimipour, Mehdi Hosseinzadeh, Arash Sharifi, Aso Darwesh, "*Deterministic and non-deterministic*

query optimization techniques in the cloud computing", Concurrency and Computation Practice and Experience, 5th March 2019, DOI: 10.1002/cpe.5240.

[2] YoungJu Moon, HeonChang Yu, Joon-Min Gil and JongBeom Lim, "*A slave ants based ant colony optimization algorithm for task scheduling in cloud computing environments*", Human-centric Computing and Information Sciences, 9th October 2017, DOI: 10.1186/s13673-017-0109-2.

[3] Gurneet Kaur, "*Role And Importance Of Search Engine Optimization*", International Journal Of Research-Granthaalayah, Volume **5**, Issue **6**, June 2017, DOI: 10.5281/zenodo.818213.

[4] Sudipta Sahana, Rajesh Bose, Debabrata Sarddar, "*An Enhanced Search Optimization Protocol Based on Gossip Protocol for the Cloud*", International Journal of Applied Engineering Research, Volume **12**, Number **19**, pp. **8436-8442**, ISSN **0973-4562**, **2017**.

[5] Li Liu, Miao Zhang, Rajkumar Buyya, Qi Fan, "*Deadline-constrained coevolutionary genetic algorithm for scientific workflow scheduling in cloud computing*", Concurrency and Computation Practice and Experience, WILEY, 22nd July 2016, DOI: 10.1002/cpe.3942.

[6] Mohammed Abdullahi, Md Asri Ngadi, "*Hybrid Symbiotic Organisms Search Optimization Algorithm for Scheduling of Tasks on Cloud Computing Environment*", PLOS One, 27th June 2016, DOI: 10.1371/journal.pone.0158229.

[7] Mohammed Abdullahi, Md Asri Ngadi, Shafi'i Muhammad Abdulhamid, "*Symbiotic Organism Search optimization based task scheduling in cloud computing environment*", Future Generation Computer Systems, 24th August 2015, DOI: 10.1016/j.future.2015.08.006.

[8] Manish M. Pardeshi, R. L. Paikrao, "*Enhanced and Efficient Search Multimedia Data by Using Multi-Keyword Synonym Query*", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume **4**, Issue **6**, June 2015.

[9] George Suciu, Ana Maria Sticlan, Cristina Butca, Alexandru Vulpe, Alexandru Stancu and Simona Halunga, "*Cloud Search Based Applications for Big Data - Challenges and Methodologies for Acceleration*", ARMS-CC 2015, LNCS **9438**, Springer International Publishing Switzerland 2015, pp. **177–185**, **2015**, DOI: 10.1007/978-3-319-28448-4_13.

[10] Qiang Xu, Zhengquan Xu, Tao Wang, "*A Data-Placement Strategy Based on Genetic Algorithm in Cloud Computing*", International Journal of Intelligence Science, 2015, **5**, **145-157**, April 2015, DOI: 10.4236/ijis.2015.53013.

[11] Gunvir Kaur, IEr. Sugandha Sharma, "*Research Paper on Optimized Utilization of Resources Using PSO and Improved Particle Swarm Optimization (IPSO) Algorithms in Cloud Computing*", International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014), Vol. **2**, Issue **2**, Ver. **3**, April-June 2014.

[12] Medhat Tawfeek, Ashraf El-Sisi, Arabi Keshk and Fawzy Torkey, "*Cloud Task Scheduling Based on Ant Colony Optimization*", The International Arab Journal of Information Technology, Vol. **12**, No. **2**, 23rd April 2014.

[13] Vimmi Makkar, Sandeep Dalal, "*Ranked Keyword Search in Cloud Computing: An Innovative Approach*", International Journal of Computational Engineering Research, Volume **3**, Issue **6**, June 2013.

[14] Dr S. Saravanakumar, K Ramnath, R Ranjitha and V.G.Gokul, "*A New Methodology for Search Engine Optimization without getting Sandboxed*", International Journal of Advanced Research in Computer and Communication Engineering, Volume **1**, Issue **7**, ISSN: **2278-1021**, September 2012.

[15] Lizheng Guo, Shuguang Zhao, Shigen Shen, Changyuan Jiang, "*Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm*", Journal of Networks, Volume **7**, No. **3**, March 2012, DOI: 10.4304/jnw.7.3.547-553.

Authors Profile

Mr. Sudipta Sahana is an assistant professor of a renowned engineering college of West Bengal. More than 4 years he has worked in this region. He has passed his M.Tech. degree in Software Engineering and B.Tech. Degree in Information Technology from West Bengal University of Technology with a great CGPA/DGPA on 2010 and 2012 respectively. He is recently working in Ph.D. on the domain of "cloud computing". He is a member of the Computer Science Teachers Association (CSTA), and also a member of International Association of Computer Science and Information Technology (IACSIT).

Mr. Tanmoy Mukherjee is a M.Tech final year student of CSE in JIS College of Engineering, West Bengal. He received his B.Tech degree from Bengal College of Engineering and Technology, West Bengal with a good CGPA/DGPA in 2013. Currently, he is working in the domain of "cloud computing". He has written a conference paper on Virtualization Security which is about to be published. He worked for 2 years in ISAC, Bangalore, where he played a significant role in the testing of onboard satellites.

Mr. Debabrata Sarddar is an Assistant Professor in the Department of Computer Science and Engineering, University of Kalyani, Kalyani, Nadia, West Bengal, INDIA. He has done PhD at Jadavpur University. He completed his M. Tech in Computer Science & Engineering from DAVV, Indore in 2006, and his B.E. in Computer Science and Engineering from NIT, Durgapur in 2001. He has published more than 75 research papers in different journals and conferences. His research interest includes wireless and mobile system and WSN, cloud computing.