International Journal of Cloud Computing and Services Science (IJ-CLOSER)
Vol.3, No.2, April 2014, pp. 91~98
ISSN: 2089-3337

A Proposal of Resource Allocation Management for Cloud Computing

Md Imran Alam*, Manjusha Pandey*, Siddharth S Rautaray*
* School of Computer Engineering, Kiit University

Article Info

Article history:

Received Feb 17th, 2014
Revised Mar 25th, 2014
Accepted Apr 05th, 2014

Keyword:

Cloud Computing
Tabu Search
Task Scheduling
CloudSim

ABSTRACT

Cloud computing is an emerging technology that provides computing resources based on demand of the user. This demand or supply of resources can be scheduled based on some policies. The resource allocation management should provide resources in less time and less cost. In this paper, a dynamic optimized resource allocation management algorithm is designed based on three factor that is optimum solution, deadline constraint and cost constraint. The algorithm uses Tabu Search Algorithm followed with prioritization and task grouping.

Corresponding Author:

Md Imran Alam,
School of Computer Engineering,
Kiit University,
Patia, Bhubaneswar, Khurda 751024, India.
Email: imranirfan89@gmail.com

1. INTRODUCTION

Cloud can be defined as parallel and distributed computing system which is a collection of interconnected and virtualized computers helps in computing power, storage, platform and services on the basis of customers demand over internet. Cloud computing can be referred as an application delivery as a service as well as hardware and system software in the datacenters. The characteristics of cloud computing are basically as on-demand-self-service, broad network access, resource pooling, rapid elasticity, measured service, device and location dependency and security. Clouds deploys by four deployment models that are public cloud, private cloud, community cloud and hybrid cloud. Some deployment model provides security, few are for general usages and some are a combination of both. The services provided by cloud computing are software as service, platform as service and infrastructure as service. Infrastructure as a service deals with the infrastructure of the cloud. Platform as a service has the capability to provide an independent platform having deployment capabilities and it may be capable of executing multiple application on a single platform concurrently. Software services are to use applications delivered by the service provider in a particular cloud infrastructure. Scheduling process in cloud can be generalized into three stages namely:

- Resources discovering and filtering
- Resource selection
- Task submission

The simple scheduling steps can be shown by diagram as:

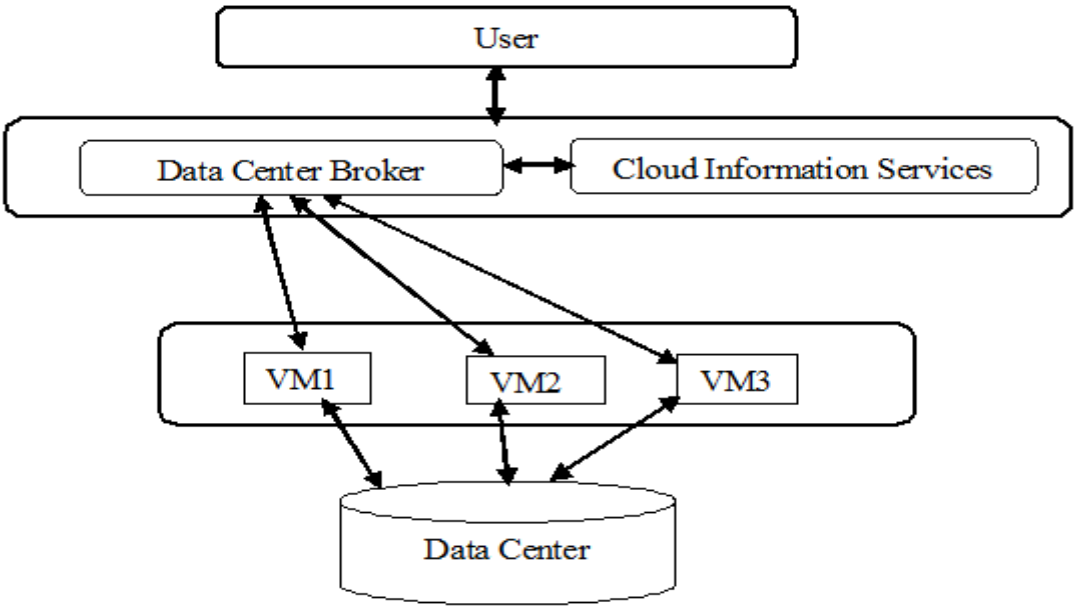

Figure 1. Scheduling in Cloud

Task scheduling is the method by which threads, data flows or processes are given access to system resources. This is normally done to load balance a system effectively or achieve a target quality of service. The demand for a scheduling algorithm arises from the requirement for most modern systems to perform multitasking and multiplexing. Resource allocation is used to assign the available resources in an economic manner. It is part of resource management. In project management, resource allocation is the scheduling of activities and the resources needed by those activities while taking into consideration both the resource availability and the project time.

Tabu Search is a metaheuristic search method employing local search methods used for mathematical optimization. Local searches take a potential solution to the problem and check its immediate neighbors in the hope of finding an improved solution. Tabu search enhances the performance of these techniques by using memory structures that describe the visited solution or user provided set of rules. Metaheuristic is a higher level procedure or heuristic designed to find, generate, and select a lower level procedure that may provide a sufficiently good solution to an optimization problem. Metaheuristics may make few assumptions about the optimization problem being solved and so they may be usable for a variety of problems.

2. RELATED WORK

The dynamic resource allocation is growing need of cloud providers for a number of users and with less response time. Various policies for dynamic resource allocation in cloud computing is shown based on Topology Aware Resource Allocation (TARA), Linear Scheduling Strategy for Resource Allocation and Dynamic Resource Allocation for Parallel Data Processing. Task Scheduling is NP- Complete problem which plays a key role in Cloud Computing. An optimized algorithm based on Fuzzy GA Optimization is proposed which makes a scheduling decision by evaluating the entire group of in the job queue.

The scheduling algorithm where the incoming tasks are grouped on the basis of the task requirement like minimum execution time or minimum cost. Resource selection is done on the basis of task constraints using a greedy approach. Selection of resources can be random, round robin, greedy (resources processing power and waiting time based). Selection of jobs to be scheduled can be based on FCFS, SJF, priority based, coarse grained task grouping [1].

Various algorithms for task scheduling in cloud computing are Min-Min, Max-Min, suffrage, Shortest Cloudlet to Fastest Processor (SCFP), Longest Cloudlet to Fastest Processor (LCFP) and some meta heuristics like Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) and Simulated Annealing (SA). The framework for dynamic optimization algorithm are task grouping, prioritization and greedy allocation. It improves cost and completion time of tasks as compared to

Sequential Assignment. The turnaround [2] time and cost of each job is minimized individually. The cost based task before resource allocation according to resource capacity is not grouped which increases communication overhead.

A genetic algorithm for task scheduling is of two types: Standard Genetic Algorithm (SGA) and Modified Genetic Algorithm (MGA). Users quality of service is not consider. Priority of jobs for multiple users is not considered. Runtime scheduling is not supported. Load Balancing Ant Colony Optimization (LBACO) algorithm inherits the basic ideas from ACO algorithm to decrease the computation time of tasks executing and also considers the loading of each VM. It is not assumed that all tasks are mutually dependent, i.e. precedence constraint between tasks exist. Tasks are not computationally intensive.

Task scheduling can also be done by greedy approach and on priority of jobs. Task scheduling can be based on agent where each node of the resource information will be sealed into a proxy. The high performance in computing applications of the entire cloud system is generally provided by the underlying agent system. Task scheduling algorithm of cloud can be based on PETRI Network. The PETRI network design the whole task flow chart: Data flow, task status, control information and other expressions to the allocation and the using of resource (concurrent style, synchronization and resource contention type) can be presented in this model in a formal comprehensive way [3,4].

Tabu Search Algorithm had been used in Grid Computing for scheduling the task with less time complexity. Even this is used in Cloud Computing for just finding the location of data centres and software components in computing networks.

3. PROPOSED FRAMEWORK

The framework is mainly done on the basis of task grouping, greedy allocation and a tabu search algorithm.

3.1 Tabu Search Algorithm: Tabu Search Algorithm is extension of search algorithm of local neighborhoods. Tabu Search Algorithm is most successful in group optimum, machine learning, production scheduling, neural networks and function optimum. In this sufferage algorithm is used to get initial value of TSA which develop capacity of TSA and save execution time. It is repeated to get optimum value. The fig 2 shows how the algorithm works.

3.2 Task grouping: The task grouping means that tasks of similar type can be grouped together and then scheduled collectively [5]. It means that some task having similarity can be grouped to form set of tasks. This task is grouped based on deadline constraint or minimum cost. The task grouping can reduce the cost communication ratio. When a task is grouped, it can be determined by priority and then scheduled accordingly.

Prioritization: Priority helps understanding the value of the element of where it is associated. When task scheduling is considered, it determines the order of task scheduling based on the parameters undertaken for its computation [6]. The shorter deadline task is executed first which have higher priority. The task is arranged in ascending order for executing the minimum time constraint task first. The task based on cost can be prioritized in descending order.

3.3 Greedy Allocation: Greedy algorithm is applicable to dynamic heterogeneous resource environments connected to the scheduler through homogeneous communication environment [7]. The greedy approach helps in solving the job scheduling problem.

*Deadline Constrained Task***:** Greedy algorithm is used to minimize the turnaround time of the task which improves overall completion time. The fig 3 shows how the algorithm works.

Turnaround Time= Resource Waiting Time + Task Length/ Processing Power of Resource

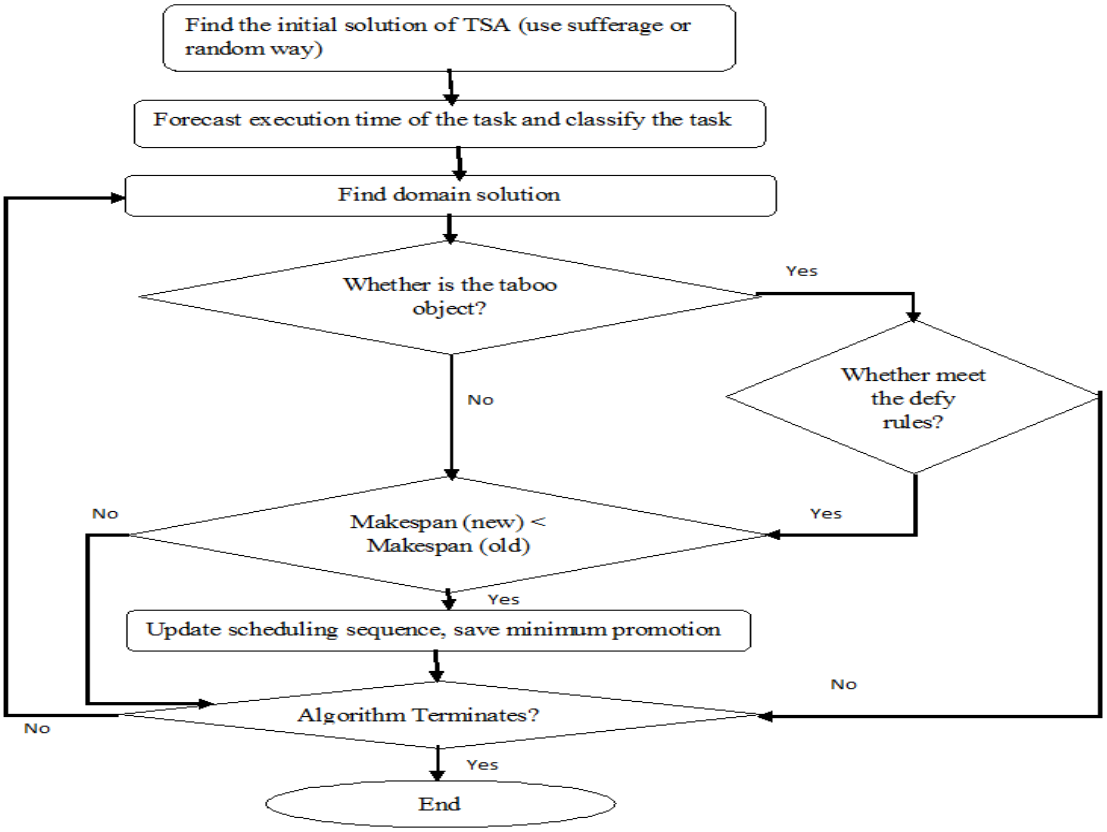

Figure 2. Flowchart of Tabu search algorithm

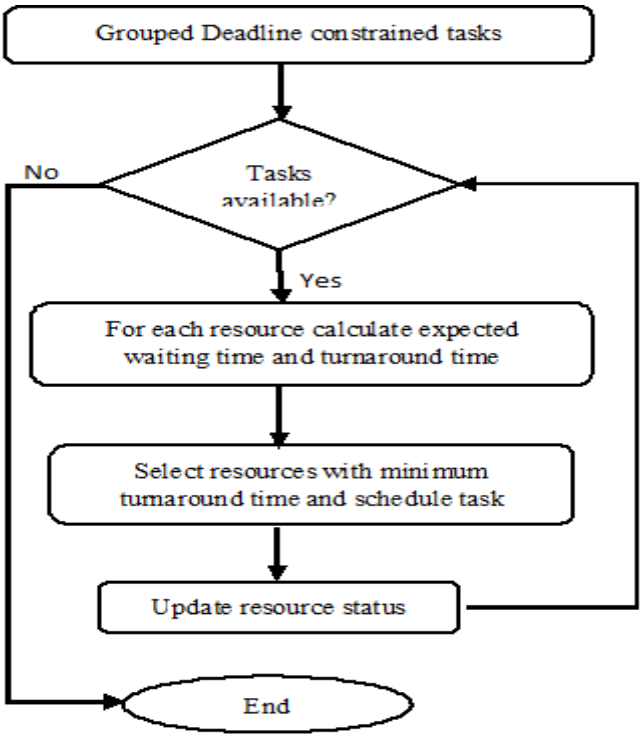

Figure 3. Scheduling of Deadline Constrained Tasks

Minimum Cost Task: The resource having minimum cost is found and scheduled till the capacity is supported. The target task and resource are selected sequentially as per the user's requirement. The fig 4 shows how the algorithm works.

*Cost of Task = (Task length/ Processing Power of Resource) * Resource Cost*

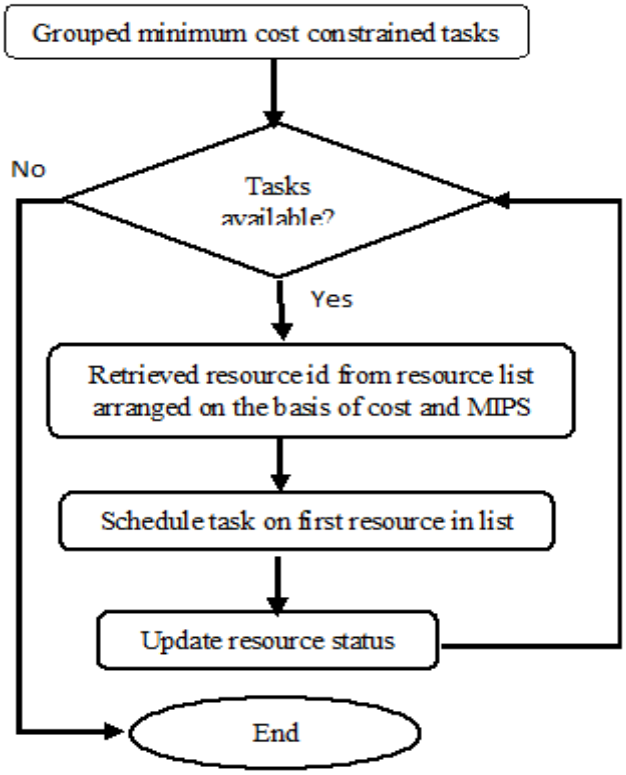

Fig 4. Scheduling of Cost Based Tasks

4. PROPOSED ALGORITHM

The proposed algorithm has tabu search algorithm, deadline constraint and cost constraint.

4.1 Tabu Search Algorithm: First of all find expected execution time of task mode. The base algorithm for this is as follows:

1. Define template set T, make type set C initialization empty.
2. Before beginning to execute each task a:
   a. get type set C_a attaching the a task, according to the features of model in T and the a task;
   b. delete all types of out C in C_a, denoted C_a';
   c. account predicted execution time of each type in C_a';
   d. If C_a' is not empty, then select the average predicted value of all types in C_a' as the predicted result of the a task;

3. After finishing predicting the a task, each type c_i in C_a',
   a. if c_i C, make c_i add to C_a;
   b. if $|c_i|$ is equal to the most numbers of tasks that c_i can deposit, then make the a task replace the most early task among c_i.

Tabu Search Algorithm uses one taboo list to record the arrival local optimum. For getting initial value Sufferage algorithm is used and repeated until to get optimum effect. Let n be number of staying scheduling task and m be number of resources. The resources are encoded and find neighborhood solution by describing the distribution of scheduling task randomly into C_i and C_j. Determine whether it belongs to one same type i.e i=j or else i<>j and select one task between C_i and C_j, exchange to get scheduling sequence. Now find Record Makespan and Promotion of sequence.

Promotion = Makespan of new scheduling sequence – Makespan of old scheduling sequence

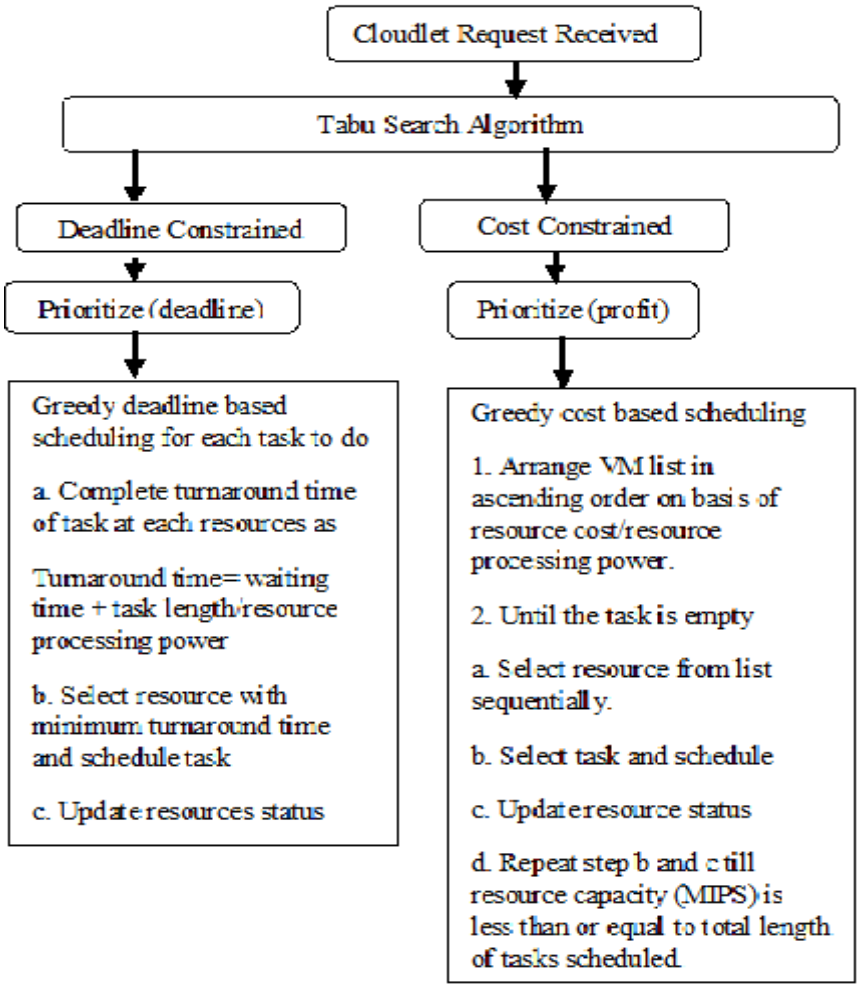

Figure 5. Flowchart of Optimized algorithm of Task Scheduling

Steps are repeated $\sqrt{m}$ times and new scheduling sequence is made where it has least Makespan and least Promotion of neighborhood solution.
Tabu list is m × m 2 dimensional array used for saving the length of banned object which is resource. To find Tabu length dynamic selection is used. The condition here is that if $0.5 * \sqrt{m} < 3$ then length is 3 and even can be selected randomly having an integral number between $[3, 0.5 * \sqrt{m}]$.
Now find an aspiration criterion by first solving involved resource having tabu situation with the current scheduling sequence. Compare the corresponding Promotion of new sequence with promotion in history, if it is less, it is said that the operation is successful. The termination criterion is that the best scheduling sequence appear $\sqrt{m}$ times and execution finishes. Solve Makespan of each scheduling sequence, record it as Makespan $_i$, target f = min {Makespan$_i$}.

4.2 Deadline Constraint: For prioritized deadline constraint task-
1. Turnaround time for every resource are calculated on the basis of following parameters.

- Waiting Time
- Task Length
- Processing Power of Virtual Machine

2. Virtual machine having least turnaround time is selected and scheduled for execution.

3. Updating of waiting time and resource capacity is done accordingly.

4.3 Cost Constraint: For prioritized cost constrained task-
1. Virtual machine is selected based on processing power and cost.
2. Virtual machine cloudlets are scheduled from the group to fill resource capacity.
3. Updating of waiting time and resource capacity is done accordingly.

5. FUTURE WORK

The proposed framework of the algorithm is to be simulated in CloudSim. CloudSim is a toolkit that helps in simulating the cloudlets scenarios. It provides policies for managing the different parts of the system like scheduling, provisioning etc. The different component also helps in evaluating new strategies for utilization of Clouds like policies, mapping, scheduling algorithms as well as load balancing policies.

6. DISCUSSION

The proposed algorithm has been designed for finding quality solution in less time with less resources. So three main factors are taken into consideration that is the optimum solution, deadline constraint and cost constraint. Tabu Search Algorithm helps in finding quality solutions within less time and that solution is not repeated again. This helps in comparing other solution and many solution comes into account. This solution (tasks) is grouped and determined by two factors that is deadline and cost. This factor helps in easy scheduling of the cloudlets. Deadline constraint takes this solution (tasks) and based on priority, it is scheduled. Even the cost constraint works in a similar manner.

7. CONCLUSION

Cloud Computing is continuously expanding for more users, applications and devices therefore the scheduling and resource allocation of tasks should be efficient and in a sustainable manner. In this paper a proposed algorithm is designed based on optimum solution, deadline constraint and cost constraint. For this tabu search algorithm, task grouping and prioritization is used. The simulation can be done by CloudSim.

REFERENCES

[1] Ronak Patel, Sanjay Patel. Survey on Resource Allocation Strategies in Cloud Computing. IJERT Vol. 2 Issue 2 Feb 2013, ISSN: 2278-0181.

[2] Monika Choudhary, Sateesh Kumar Peddoju. IJERA Vol. 2 Issue 3, May- Jun 2012, Page 2564-2568, ISSN: 2248-9622.

[3] Yinfeng Zhang, Yulin Li. Grid Computing Resources Management Scheduling System Based on Evolutionary Algorithm. *Computer Engineering*. 2003: 29(15): 110-175.

[4] Congyi Yuan. Petri net theory. Beijing Publishing House of Electronics Industry, 1998.

[5] P.J. Wild, P. Johnson and H. Johnson, "Understanding task grouping strategies", People and Computers, pp.3-20, 2004.

[6] Q. Cao, B. Wei and W. M. Gong, "An optimized algorithm for task scheduling based on activity based costing in cloud computing", In International Conference on eSciences 2009, pp. 1-3.

[7] S. Singh and K. Kant, "Greedy grid scheduling algorithm in dynamic job submission environment", in International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), 2011, pp. 933-936.

[8] Federico Larumbe, Brunilde Sanso. A Tabu Search Algorithm for the Location of Data Centres and Software Components in Green Cloud Computing Networks. IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013: 22-35.

[9] Harun Pirim, Engin Bayraktar and Burak Eksioglu (2008). Tabu Search: A Comparative Study, Tabu Search, Wassim Jaziri (Ed.), ISBN: 978-3-902613-34-9, InTech, Available from: http://www.intechopen.com/books/tabu_search/tabu_search__a_comparative_study

[10] Fred Glover and Rafael Martí. Tabu Search. Available from: http://www.uv.es/rmarti/paper/docs/ts2.pdf

[11] XU Hai yan. The Grid Task Scheduling based on Tabu Search Algorithm. World Automation Congress (WAC), 2012. ISSN: 2154-4824, ISBN: 978-1-4673-4497-5

[12] Chandrashekhar S. Pawar and R. B. Wagh. A review of resource allocation policies in cloud computing. World Journal of Science and Technology 2012 Vol. 2 Issue 3, ISSN: 2231-2587, Page 165-167.

[13] Sandeep Tayal. Tasks Scheduling optimization for the Cloud Computing Systems. IJAEST Vol. 5 Issue 2, Page 111-115, 2011.

[14] Mastering Cloud Computing: Cloud Foundation and Applications Programming. Rajkumar Buyya, Christian Vecchiola, S. Thamarai Selvi. MK Publisher Edition 2013.

[15] V. Vinothina, Dr. R. Shridaran, and Dr. Padmavathi Ganpathi, A survey on resource allocation strategies in cloud computing, International Journal of Advanced Computer Science and Applications, 3(6):97--104,2012

[16] Atsuo Inomata, Taiki Morikawa, Minoru Ikebe and Md. Mizanur Rahman, Proposal and Evaluation of Dynamic Resource Allocation Method Based on the Load Of VMs on IaaS, IEEE, 2010.

[17] Bahman Rashidi, Mohsen Sharifi, Talieh Jafari. "A Survey on Interoperability in the Cloud Computing Environments", IJMECS, vol.5, no.6, pp.17-23, 2013.DOI: 10.5815/ijmecs.2013.06.03
[18] Dr. Kumar Saurabh, Rishi Ranjan. Cloud Management Simulation and Design. Computer Modelling and Simulation (UKSim), 2012 UKSim 14th International Conference on Modelling and Simulation. ISBN: 978-1-4673-1366-7

BIOGRAPHY OF AUTHORS

Md. Imran Alam received his B.E degree in the year 2011 from GTEC, Vellore affiliated to Anna University, Chennai, and currently pursuing M.Tech in Computer Science Engineering from KIIT University, Bhubaneswar. His research area includes Distributed Database systems, Task Scheduling, Cloud Computing and big data.

Manjusha Pandey is presently working as an Assistant Professor in the School of Computer Engineering, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha. She is pursuing her PhD from Indian Institute of Information Technology, Allahabad. She has more than 10 research publications to her credit in journals and conferences of repute. Her research interest areas include Wireless Sensor Network, Security and Privacy in Wireless Sensor Network, Human Computer Interaction and Computer Networks.

Siddharth Swarup Rautaray is presently working as an Assistant Professor in School of Computer Engineering, KIIT University, Bhubaneswar, Odisha. He is pursuing his Ph.D. work in the Department of Information Technology at Indian Institute of Information Technology, Allahabad, India. His current research interests include Human–Computer Interactions, User Interface Design and Computer Vision Image Processing. He has published around 15 research papers in different international and national journals and conferences. He is a member of the IEEE Computer Society, ACM SIGCHI and CSI.