HABC: Hybrid Optimization Method Based Resource Allocation for Improving the Energy Efficiency in Green Cloud

Mrs.J.K.Jeevitha¹, Dr.G.Athisha²
1Assistant ProfessorDepartment of Information Technology,
PSNA College of Engineering and Technology,India
2Professor & Head,Department of Electronics and Communication Engineering
PSNA College of Engineering and Technology,India

Abstract: Cloud computing is one of the most important and plays a significant role in computing world. It makes all the cloud users happy in providing various kinds of services under SAAS, PAAS, and IAAS. Due to the increased number of customer and customer demands on services the number of services deployed in the cloud is also getting increased. Services accessed by the customers by Pay-N-Use. The resources allocated related to the user's request virtually by considering the numbers of cloud users and resources. But actual resource allocation in heterogeneous cloud environments is complicated, and the accuracy is less, so that, these problems are leads to the high level of energy consumptions. Various methods, potential experiments, and algorithms are proposed in the earlier research works to increase the efficiency of the resource allocation with QoS. This paper discourse the problem of optimum resource allocation to a user with improved Quality of Service (QoS) using HABC (Hybrid Artificial Bee Colony) method and identify the better influencing parameters using Design of Experiment (DoE) tool. Initially, the user requests and the resources are validated, then map the scheduled resources and finally choose the best and allocate the right resources using HABC algorithm. The simulation is conducted using CloudSim tool. The obtained results verified and the performance is evaluated by comparing the results to different evolutionary algorithms. The parameters such as energy, time, cost, valid request, optimum number resources, accurate mapping, and allocation considered as the QoS parameters which are calculated from the simulation and presented here.

Keywords: Cloud Computing, Resource Allocation, Resource Mapping, Request Validation, Scheduling, Quality of Service.

Nomenclature
QoS: Quality of Service
DoE: Design of Experiment
SaaS: Software as a Service
PaaS: Platform as a Service

I. INTRODUCTION

In distributed computing, several computing methods were proposed and utilized from desktop computing to cloud computing. Different things are given as a service to different people in the cloud. All the resources can be shared with any person from anywhere at any time based on pay-n-use method [1]. The virtualization concept and the scalability problem of the cloud are represented by different definitions and experiments conducted by the authors in [2] and [3] respectively. Various applications utilize cloud-based services in multiple forms like software hardware and data.
Three different users interconnected in the cloud such as service provider, customer and end users. Cloud computing provides a un-countable service to cloud customers regarding software, hardware and data centers. Major IT companies such as Google, Amazon, IBM and, Microsoft have their cloud data centers around the world to support cloud services. Due to the un-countable services and the customers, maintaining the response and allocating the resources become a mesh. The authors in [4, 5] discussed the job scheduling and the authors in [6, 7] presented the resource allocation policies for effective resource utilization and efficient service provision to the customers. The cloud subscribers determine the Resources needed by the user through the Service Level Agreement (SLA) which are provided by the cloud data centers and tend to satisfy the Quality of Service (QoS). SLA provides a two-sided contract between the cloud provider and the user; it also determines the level of performance, price, the content of the service supplied and the penalties for not providing the services to the user. Any failure to observe of the QoS leads to SLA violation, and consequently, and hence penalty must be paid by service providers [8]. The speedy growth of cloud services and their interrelated technologies, cloud infrastructures have become more problematic and more convoluted. Due to this Resource management becomes a significant issue in a current cloud environment and precisely affects the development of cloud services. Modern data centers provide a high level of performance and optimization; yet, a new concern in energy consumption has occurred.

A model of the cloud architecture with the virtual machine and physical machine sharing process illustrated in Figure-1. A scheduler and a monitor used for scheduling and monitoring the resources (VM, PM) in the cloud. There is \( N \) number of resources inter-connected with the \( K \) physical machine used for computing cloud data in host computing system. These are the resource going to allocate to run applications are various time on "on-demand." In this paper, it assumed that the cloud is centralized and hosted in the datacenter where it comprises of a more number of heterogeneous servers. Different jobs assign all the servers. The virtualization procedure permits the cloud can create any number of a virtual machine for any available physical machine. Hence, assigning a task is more flexible under any server. All the servers can save their energy in an idle state where it performs the maintenance and has all the other sub-systems are ready while it waits for the task to attain. When a task comes inside, the allocated VM executes the task, and the host has to spend some more additional energy related to the number of resources required by the task, and it called as resource utilization in workload model. Cloud is distributed across in more geographical locations and mainly connected to a specific physical location. It assumed that the all the resources are homogeneously associated with their capacity and their computing capability. It can conclude by applying virtualization methods [9]. A message is transmitted from one resource to another resource when a task executed on the destination resource, which is possible in many systems. The lower and upper limit of the energy consumption of the server in cloud computing is represented as a pick-load-state and idle-state.

II. RELATED WORKS

According to the Environmental Protection Agency (EPA), the energy consumption of data center would rise from 2006 (61 billion kWh) to 2011 [12]. In 2009, data centers accounted for 2% of worldwide electricity consumption with an economic impact of US $30 billion [13]. Gartner Group forecasted data center hardware expenditure for 2012 to be as the US $106.4 billion, a 12.7% increase from 2011, whereas cloud computing revenue is predicted to jump from the US $163 billion in 2011 to the US $240 billion in 2016 [14]. IT sector contributes indirectly to carbon dioxide emission as a significant amount of carbon is used in the electricity production process, and they emit dioxide and causes greenhouse gases [15]. The IT sector was responsible for 2% of carbon dioxide worldwide in 2005, a figure that is estimated to grow by 6% per year [12]. Aggressive energy-efficiency measures for all devices inside the data center can reduce 80% of energy costs and 47 million metric tons of carbon dioxide emissions [12]. Power usage effectiveness (PUE) is the ratio
of total data center energy usage to IT equipment, energy usage, which helps in measuring the data center efficiency [12]. The average PUE value of data centers in a 2005 survey was found to be 2 [16]. It estimated that the average PUE value would fall to 1.9 by 2011, although aggressive energy efficiency measures can result in typical PUE value of 1.2 [12]. The higher PUE value indicates that most of data center energy consumed in cooling measures instead of computing. An energy-efficient data center with a lower PUE index can lead to several benefits such as:

a) Reduced energy consumption, hence, lesser operational expenses;
b) Lesser greenhouse gas emissions; and
c) Lesser device costs.

Data center operators are also interested in the total cost of ownership: Which is the sum of CApitalEXpenses (CAPEX) required in setting up the data center and the OPerational EXpenses (OPEX) needed in running the data center [12].

While considering task scheduling and resource allocation, various earlier works [17 -22] discussed the same. Some of them focused on reducing the task completion time and cost of the resource utilization. The author in [23] stated that scheduling is a systematic task in built-in supercomputers. Currently, resource allocation becomes emerging research. The author in [24] focused on Service Level Agreement for service providing for SaaS applications. In the SLA agreement, it mainly concentrated on reducing the time and cost of resource allocation. In [25-26], it said that the service providers have to allocate dedicated VMs for all the users, then it reduce the time and cost for all the responses. Similarly, various scheduling algorithms proposed for cloud computing [27-29]. The algorithms discussed in [23, 17] suggested for cloud scheduling based on their suitability. This paper also utilizes the scheduling method to reduce the time and cost and to optimize the resource allocation. Due to the speedy growth of cloud users and resources, recent research works are focused on optimization methods for resource scheduling and resource allocation. An optimal task scheduling and resource allocation is obtained by Particle Swarm Optimization (PSO) method which utilizes a fitness function [30]. The makespan reduced by increasing the fitness function and it also increases the execution capacity. In [32-33] it is proved from the experimental results that PSO can reduce the time and cost of the resource allocation. The author in [34] used Genetic Algorithm for scheduling and resource allocation and obtained better results.

From the above discussion, various earlier approaches focused on reducing the energy consumption using multiple strategies like resource allocation, task completion, proper resource utilization, SLA based resource allocation and optimized resource allocation. Hence it is identified that it is necessary for improving the energy efficiency by resource allocation.

III. ARTIFICIAL BEE COLONY

Artificial Bee Colony is one of the optimization algorithm referred from [35-37], used to select an optimal value from a mass of values. Foraging behavior of the honey bees used for optimization process. All the bees are involved in searching their foods. There are three different bees are considered here such as scout bees, onlooker bees, and employed bees. Among these bees, scout bees always search for bees in a random direction within the solution space. The employed bees found the location of the food and distribute the information to other onlooker bees. Finally, the fitness function calculated by the onlooker bees and searches the food. All the three types are bees seek their food in long distance and all directions. One they got the food the scout bees come back to original places and do "Waggle Dance". While dancing, the information about the food-source and an amount nectar to be spent for food source. At that nectar quality and the wastage of energy is evaluated. Then the onlooker bees select the best food source from the scout bees' information as the best food source.

ABC algorithm is best in searching the optimal solution, the searching time and the searching ability is little poor in ABC comparing with other optimization algorithms. Also, SA is an analogous method; it can obtain the best solution. Hence, the searching function of SA is integrated
into ABC to increase the efficiency of the ABC to achieve the optimal solution efficiently and speedily in the search space. The best food source information is searched by SA randomly. By integrating two different optimization algorithms, the efficiency of the best resource selection is increased.

The entire process of ABC algorithm is given in the form of pseudo code, and it is presented in Figure-3 and Figure-4. There is N number of VMs considered in the cloud environment for assigning to various users. Different servers (s) are interconnected, and the incoming requests maintained by CMP (Cloud Management Policies) where it maintained the load in each S based on the present time. The entire scheduling process focused on the VM and load. According to the load and available VM, the scheduling approach works. This function reduces the time and makespan. Let us assume, the N, VMs are

\[ VM = \{VM_1, VM_2, \ldots, VM_N\} \]
\[ i = \{1, 2, \ldots, N\} \forall i \in N \]
\[ Task = \{task_1, task_2, \ldots, task_K\} \]

It is known that VMs are allocated parallel to the incoming requests without any collision. The complete process of VM scheduling and load balancing using ABC is illustrated in Figure-5.

**Numerical Illustration of HABC**

1. **Initializing the Population**
   - N number of Selected Bees (SB) = SB\(_1\), SB\(_2\), SB\(_3\), \ldots, SB\(_N\)

2. **Fitness Evaluation**
   - Using the following formula the fitness value of each bee is calculated as:
Where,

\[ \text{fit}_{ij} = \frac{\sum_{i=1}^{n} t_{\text{length}_{ij}}}{\text{cap} - \text{of} - \text{VM}_j \, \text{(the cap is j)}} \]

and the capacity is calculated according to:

\[ \text{capacity}_j = p_{\text{num}}_{j} \times p_{\text{mips}}_{j} + \text{vm}_{\text{bw}}_{j} \]

Where,

\( p_{\text{num}}_{j} \): denotes the number of processors in \( \text{VM}_j \)

\( p_{\text{mips}}_{j} \): denotes the millions of instruction per second over the processor in \( \text{VM}_j \)

\( \text{vm}_{\text{bw}}_{j} \): denotes the network bandwidth

3. for searching the neighbors \( m \) sites are selected

The highest fitness value based neighbor is selected as SB among \( m \) VMs.

4. Recruit Bees for Selected Sites

The high amount of bees is deployed in the search space to increase the best solution. Among these bees, the fitness is computed as

\[ \text{fit}_{ij} = \frac{\sum_{i=1}^{n} t_{\text{length}_{ij}} + \text{InputFile}\text{length}}{\text{capa} - \text{of} - \text{VM}_j \, \text{(the capa is j)}} \]

Where,

\( \text{InputFile}\text{length} \): Length calculated before execution of a task.

5. Select the best fitness bees from each set of population and assign a task to each \( \text{VM}_j \).

Based on the absolute fitness values the best bees are selected in all the round of operation and allocate the VM using first in first out priority (best fitness, time).

IV. EXPERIMENTAL RESULT AND DISCUSSION

The proposed HABC algorithm implemented in CloudSim software and the obtained results verified. The algorithm implemented and executed so many times, with various numbers of users, data centers, tasks, and VMs. It is a simulation tool can do simulations for cloud applications based experiments. For example, a set of sample parameters used in the simulation is given in Table-1. The parameters used for optimization in HABC are given in Table-2. To execute the HABC, a set of VM, DC, PC, Laptops, and Servers assigned in the cloud, and they are all interconnected by wire or wireless mediums.

In this simulation, any user can send a request to cloud regarding any resources. This request is passed immediately to the closest server presented in the cloud. Whenever a request comes, it always directed to the server with the help of http-address. During the request and resource allocation, the HABC maps the request and the available resources using the optimization ability.
After resource allocation, the state of the resource is changed into "busy," after completion of the request tasks; the state of the resource is changed into "available/free/idle." While searching and mapping the request and the resources, the HABC looks only for the resource's whose state is "free," the distance is decidedly less comparing with the other resources and time taken for the entire resource allocation, and task completion process is very less.

Figure-5: HABC Based Scheduling and Load Balancing
Table-1: Simulation Parameters Used in CloudSim

<table>
<thead>
<tr>
<th>Device Type</th>
<th>Parameter</th>
<th>Values Assumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Center</td>
<td>Number of Data Centers</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Number of Hosts</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Manager Type</td>
<td>Space Shared/Time Shared</td>
</tr>
<tr>
<td></td>
<td>Number of PEs per Host</td>
<td>2 to 4</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>Host Memory</td>
<td>2048 to 10240</td>
</tr>
<tr>
<td></td>
<td>Cost of the Datacenter</td>
<td>10</td>
</tr>
<tr>
<td>Virtual Machine (VM)</td>
<td>Number of VMs</td>
<td>30 to 210</td>
</tr>
<tr>
<td></td>
<td>MIPS of PE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1000 to 2000</td>
</tr>
<tr>
<td></td>
<td>Number of PE/VM</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>VM-Memory size (MB)</td>
<td>512 to 2048</td>
</tr>
<tr>
<td></td>
<td>Bandwidth</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>Manager Type</td>
<td>Time Shared</td>
</tr>
<tr>
<td>Task</td>
<td>Number of Tasks</td>
<td>100 to 700</td>
</tr>
<tr>
<td></td>
<td>Length of Task (M)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5000 to 20000</td>
</tr>
<tr>
<td></td>
<td>Number of PE per requirement</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Manager Type</td>
<td>Space shared</td>
</tr>
</tbody>
</table>

Where,

MB denotes Megabyte

PE denotes Processing Element

MIPS denotes (Million Instructions per Second) is a measure of the processing speed of the computer

MI denotes Million Instruction.

Table-2: Parameters Used in ABC Algorithm

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>Total number of SB</td>
<td>1000</td>
</tr>
<tr>
<td>( m )</td>
<td>Selected number of sites from ( n ) sites visited</td>
<td>5</td>
</tr>
<tr>
<td>( e )</td>
<td>Number of best sites taken from ( m ) sites</td>
<td>1</td>
</tr>
<tr>
<td>( nep )</td>
<td>Number of bees recruited from ( e ) sites</td>
<td>800</td>
</tr>
<tr>
<td>( nsp )</td>
<td>Number of best bees elected for other sites (( m ) – ( e ))</td>
<td>200</td>
</tr>
</tbody>
</table>

Initially, the number of optimized VMs selected from the total number of VM initialized is computed the results verified. The calculated results regarding optimal resource allocation are illustrated in Figure-6. The efficiency of the HABC is verified by changing the number of requests and VMs allocated. From the simulation, the obtained results regarding optimal resource allocation for the valid number of requests is shown in Figure-7. From the results, it is identified that when the deployed number of VM and requests are increased then the optimal allocation becomes faster and smooth. Obtaining available optimized resource is easy.

Resources are the VMs, will be allocated only for the valid requests sent by the authorized users. It also increases the cloud security. It is validated by verifying the user information previously stored during user registration. After the requests and the VMs are validated all the requests, and the VMs are analyzed, mapped and select the best VM for appropriate requests. This process is experimented by changing the number of VMs and requests and executed many times. In each round of operation, it is verified that how many optimized VMs are available for requests. The number of optimized VM selected is given in Figure-6. When the number of VM is increased then the optimal
number of VM is also increased, and it is obtained by HABC speedily. Based on the increased number of VMs the number of optimal VMs is also increased, and it proved in the experiment. Unless otherwise, the user request of VM is valid or available they cannot match with one another.

From the various rounds of the experiment, it is noticed that the HABC is efficient regarding optimization based resource allocation. Hence it is necessary to evaluate the performance of the proposed HABC. To do that, obtained results of the proposed approach are compared with the other traditional existing methods such as FCFS (First Come First Served), SJF (Shortest Job First) and Longest Job First (LJF). In FCFS, the requests processing is done based on the queue model, in SJF the requests processing is done based on the size and distance, whereas in LJF the requests processing is done based on the big size of requests are processed first. Here all the three models are experimented along with the HABC and the obtained results are compared with one another.
According to the changed number of requests and the VMs, the average makespan is also computed and compared. The compared results among the various traditional scheduling algorithms with the HABC are given in Figure-8. From the comparison results, it is noticed that the performance of HABC is better than the other scheduling approaches.

Also, the average makespan of the scheduling approaches is calculated based on the number of VMs. Here the ABC is integrated with the existing FCFS, LJF and the average makespan is computed. Comparing with the obtained results the lesser average makespan is obtained using HABC-SJF than the other scheduling approaches. Also, it is noticed that when the number of VMs increased then the average makespan is decreased gradually and it is shown in Figure-9. From the entire performance evaluation, the ABC integrated scheduling approaches provides a better result than the other scheduling approaches. But, combining HABC with different scheduling approaches HABC based scheduling outperforms regarding average makespan.

It is experimented and verified under various parameters and compared with the earlier research works whose focus is also same to evaluate the performance of HABC. The entire performance of HABC is computed regarding total energy, energy consumed in server, mean value of the response time, number of completed tasks and number of uncompleted tasks. The obtained results are given in Figure-10. Comparing with the other existing approaches, HABC is better than the different methods.

From the results and discussion, it is identified and concluded that the proposed HABC is better and suitable for resource scheduling and resource allocation efficiently.

![Performance Evaluation in Terms of Makespan](image)

Figure-8 Comparison Of various Methods regarding Makespan versus Tasks
The primary objective of this paper is to do optimum resource allocation using optimized scheduling, request validation, and VM validation. Artificial Bee Colony algorithm is used to do optimization here. The algorithm is experimented using CloudSim tool, and the results are verified. From the obtained results the incoming requests are validated, the appropriate VMs (resources) are optimized using ABC and makespan are obtained. The results illustrated that the proposed ABC algorithm is more efficient and suitable for cloud computing environment regarding reduced makespan for increased resources in a cloud environment. The performance is also evaluated by comparing among various scheduling approach, and it is concluded that ABC_LJF is proved itself is more efficient for maintaining system stability and efficiently scheduling. In future, the ABC_LJF is compared with the other optimization algorithms, and the performance is evaluated.
REFERENCES


XII. Octavio Gutierrez-Garcia.


XXX. Yang Z., Qin X., Li W., Yang Y., Optimized task scheduling and resource allocation in cloud computing using PSO based fitness function, Information Technology Journal 12(23) (2013).

XXXI. Xu M., Cui L., Wang H., Bi Y., A multiple QoS constrained scheduling strategy of multiple workflows for cloud computing, IEEE International Symposium on Parallel and Distributed Processing with Applications (2009), 629-634.


