

Energy Efficient Cloud Computing Vm Placement Based On Genetic Algorithm

Pooja Daharwal (SOIT RGPV BHOPAL)
Dr. Varsha Sharma (SOIT RGPV BHOPAL)

Abstract- In the age of large data and the large number of users around the world, cloud computing has emerged as a new era of computing. Cloud consists of a datacenter which in turn consists of several physical machines. Each machine is shared by many users and virtual machines are used to use these physical machines. With a large number of datacenters and each datacenter having a large number of physical machines. The VM allocation becomes an NP-Hard problem. Thus, the VM allocation, the VM migration becomes a trivial task. In this article, a survey is carried out on cloud computing in energy cloud, based on scalable algorithms. To solve NP-Hard problems, there are two ways to either give an exact solution or to provide an approximation. The approximate solution is a time-efficient approach for solving NP-hard problems. In this research work, a survey on method for energy efficiency in cloud computing is carried out. The optimization of genetic algorithms has been studied in this research. And the genetic algorithm based VM placement algorithm is implemented for the energy efficiency of cloud operation.

Keywords- Data centre, Energy Consumption, Genetic Algorithm, Virtualization, Virtual Machine (VMs), VM Placement, Cloud Computing.

I. Introduction - Cloud computing is a large scale distributed computing paradigm, driven by an increasing demand for various levels of pay-per-use computing resources [9]. It providing online resources and online storage to the user's. It gives all the data at a lower cost. In cloud computing users can access resources all the time through internet. They need to pay only for those resources as much they use. There are many existing issues in cloud computing. The main problem is load balancing in cloud computing. Load balancing helps to distribute all loads between all the nodes. It also protect that every computing resource is distributed efficiently and fairly. It helps in preventing bottlenecks of the system which may occur due to load imbalance. It provides high satisfaction to the users. Load balancing is a relatively new technique that provides high resource utilization and better response time. [3] Cloud computing provide many advantages to the users.

Increasing cloud computing has resulted in ever-increasing energy consumption and therefore overwhelming electricity bills for data centers. According to Amazon estimates, the energy costs of its data centers account for 42% of total operating costs. In addition, the ever-increasing energy consumption can lead to a dramatic increase in carbon dioxide emissions. So it is desirable to make

every effort to reduce energy consumption in cloud computing. Consolidation of servers using virtualization technology has become an important technology for improving the energy efficiency of data centers [1]. Placement of the virtual machine (VM) is the key to consolidating the server. In recent years, many approaches have been proposed with respect to various VM placement problems.

1.1 Cloud computing consist of several characteristics: [5][6]

- **On demand service-** Cloud computing provide services to users on their demand. Users can access the services as they want.
- **Broad Network Access-** In cloud computing capabilities are available over the network. All the capabilities are accessed through different mechanisms.
- **Resource Pooling-** Different models are used to pooled the resources which provide by the providers to their consumers. All the resources dynamically assigned and reassigned according to consumer demand.
- **Rapid Elasticity-** Quantity of resources is increase at any time according to the customer's requirements.

1.2 Challenges in Cloud Computing

There are many challenges in cloud computing:-

- Security
- Efficient load balancing
- Performance Monitoring
- Consistent and Robust Service abstractions
- Resource Scheduling
- Scale and QoS management
- Requires a fast speed Internet connection.

1.3. Cloud Computing Model

Cloud computing model which consist services of cloud and different deployment models as:

A. Services of Cloud Computing:

Service means different types of applications provided by different servers across the cloud. There are many services are provide to the users over cloud. [7]

1) Software as a Service (SaaS): SaaS provided all the application to the consumer which are provided by the providers. Applications are running on a cloud infrastructure. Interfaces (web browser) are used access the applications. The consumer does not

control the internal function. [8] That Customers who are not able to developed software, but they need high level applications can also be take advantages from SaaS. There are some of applications of software of services:-

- Customer resource management (CRM)
- IT service management
- Accounting
- Web content management

Advantages:

- 1) The main advantage of SaaS is costing less money than buying the whole application.
- 2) It offers reliable and cheaper applications.
- 3) More bandwidth.
- 4) Need less staff.

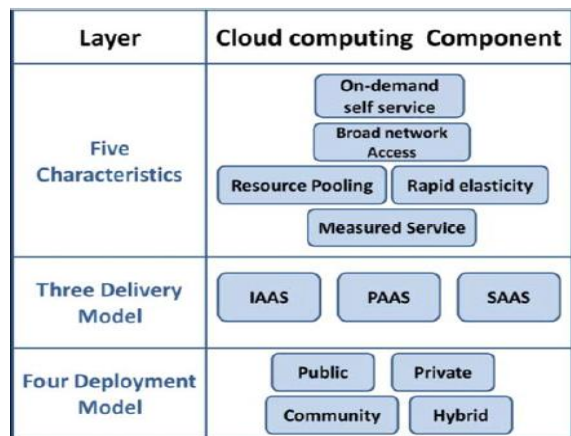


Fig.1 (Model of Cloud Computing)

2) Platform as a Service (PaaS): PaaS offers all the resources to the customers that are required for making applications. It provides all the services on the internet. User not need to download and install the software. Consumers deploy all the application onto the cloud infrastructure .There is different tools and programming languages are gives to the uses to develop the applications. The consumer does not control network, servers, operating systems, or storage. Consumer controls all applications which they deploy. There is very low portability among different providers.

3) Infrastructure as a Service (IaaS): In this service consumer does not manage or control the underlying cloud infrastructure. In infrastructure as a service consumer able to control operating systems, storage, and all applications which they deployed. There is a limited control of customer on the networking components. Infrastructure gives control storing and processing capacity.

B. Service Layers

All services have number of layers which manage by users and suppliers.

Cloud deployment models:

1. Cloud public: The cloud infrastructure is available to the general public or a large industrial group and is owned by an organization. Anyone can use the public cloud as they want without restriction.

2. Private cloud: The cloud infrastructure is used by a single organization. The private cloud is managed only by the organization or a third party. The general public is not able to directly use the private cloud.

3. Community Cloud: The cloud infrastructure is shared by many organizations. The community cloud supports a specific community that has common concerns. Ex: - security requirements, policy, compliance considerations. It can be managed by organizations or a third party.

4. Hybrid Cloud: The hybrid cloud is a combination of two or more clouds (private, community or public). This remains unique entities but is bound by standardized technology that allows portability of data and applications. Ex: - cloud burst for load balancing between clouds.

1.4. Virtualization

Virtualization means that do not exist in true, but it provides everything as real. Virtualization is the software implementation of a machine that will run different programs as a real machine. Through virtualization the user can use the different applications or services of the cloud, so this is the main part of the cloud environment. There are different types of virtualization used in the cloud environment.

Two types of virtualization are:

1. Full Virtualization
2. Para virtualization

1. Full Virtualization: Full virtualization means that a complete machine is installed on another machine. This virtual machine provides all the function that exists on the original machine. It facilitates when real machine not free then the user uses the virtual machine.

2. Para virtualization: Para virtualization, hardware allows multiple operating systems to run only machine. It also allows efficient use of system resources such as memory and processor.

II. LOAD BALANCING

Load balancing is used to distribute greater processing load to smaller processing nodes to improve overall system performance. In the cloud computing environment, load balancing is required to distribute the dynamic local workload consistently across all nodes [10][11][12].

- Load balancing contributes to an equitable distribution of IT resources to achieve high user satisfaction and appropriate use of resources. High

resource utilization and load balancing help minimize resource consumption. It helps to implement failure, scalability and to avoid bottlenecks.

- Load balancing is a technique that has helped networks and resources by providing maximum throughput with minimal response time. Load balancing divides traffic between all servers, so that data can be sent and received without delay with load balancing.

2.1 Goals of Load Balancing

Load balancing has following main goals as define in [2]:

- To increase the performance significantly.
- To have a backup plan in case the system fails even partially.
- To keep the system stable.
- To provide future enhancement in the system.

2.2 Classification of Load Balancing Algorithms

- **Static Load Balancing Algorithms:** Static algorithms are appropriate for systems with low load variations, in the static algorithm the traffic is split evenly among the servers. This algorithm requires prior knowledge of system resources, processor performance is determined at the start of execution, so the decision to shift the load does not depend on the current state of the system. However, static load balancing algorithms have a disadvantage in that tasks are only assigned to the machines after it has been created and the tasks cannot be moved during execution to another machine.

- **Dynamics Algorithms:** Dynamic algorithms [10] are the decision about load balancing based on the current state of the system, i.e. any prior knowledge about the system is not necessary. This will overcome the disadvantages of the static approach. Dynamic algorithms are complex, but they can provide better performance and fault tolerance.

III. EXISTING LOAD BALANCING ALGORITHM

There are many load balancing algorithms that help achieve better throughput and improve response time in a cloud environment. All algorithms have their own advantages. [2][16] [17] [18]

1. LB-based task planning: This algorithm mainly consists of two-level task scheduling mechanisms that are based on load balancing to meet the dynamic requirements of users. It gets a high resource utilization. This algorithm balances the load by first assigning tasks to virtual machines and then to all virtual machines to host resources.

2. Opportunistic Load Balancing: OLB is attempting each node to remain busy, so ignores the

current workload of each computer. The advantage is quite simple and reach the balance of the load, but its gap is not consider each execution time of expectancy of the task, so the complete completion time (Make span) is very poor.

3. Round Robin: In this algorithm all processes are divided among all processors. In this process, each process is assigned to the processor in a round robin order. The workload distributions between processors are equal. Different processes do not have the same processing time. At times, some nodes can be heavily loaded and others remain inactive on web servers where HTTP requests are similar in nature and distributed equally, then the RR algorithm is used.

4. Randomized: This algorithm is static in nature. In this algorithm a process can be handled by a particular node with a probability p . When all the processes are of equal loaded then this algorithm work well. Problem arises when loads are of different computational complexities. This algorithm is not maintaining deterministic approach.

5. Min-Min Algorithm: It starts with a set of all unassigned tasks. In this minimum completion time for all tasks is found. Then the minimum value is selected. Then task with minimum time schedule on machine. After that the execution time for all other tasks is updated on that machine and then the same procedure is followed until all the tasks are assigned on the resources. The main problem of this algorithm is that it has a starvation.

6. Max-Min Algorithm: Max-Min algorithm is almost same as the min-min algorithm. The main difference is the following: In this algorithm first finding out the minimum execution times, then the maximum value is resources. After that maximum time finding, the task is assigned to the particular selected machine. [19] Then the execution time for all tasks is updated on that machine, this is done by adding the execution time of the assigned task to the execution times of other tasks on that machine. Then all assigned task is removed from the list that executed by the system.

7. Honeybee Feed Behavior: This is an algorithm inspired by nature for self-organization. Honeybee performs global load balancing through local server actions. System performance is improved with a wider variety of systems. The main problem is that the flow is not increased with an increase in the size of the system. When the diversity of the population of types of services is required, this algorithm is best suited.

8. Active clustering: In this algorithm, the nodes of the same type of the system are grouped together and work together in groups. It works as a self-

aggregate load balancing technique where a network is reconfigured to balance the system load. Systems optimize the use of similar task assignments by connecting similar services.

9. Comparison and balance: This algorithm is used to reach an equilibrium condition and to manage the load of the asymmetric systems. In this algorithm based on the probability (number of virtual machines running on the current host and the whole cloud system), the current host randomly selects a host and compares its load. If the current host load is greater than the selected host, it transfers an additional load to that particular node. This load balancing algorithm is also designed and implemented to reduce the migration time of virtual machines.

assignment problem (QAP). These algorithms were inspired by the observation of real colonies of ants. The behavior of the ant is more directed towards the survival of the colonies. They do not think for the individual.

Table 3.1 Comparison of existing Load Balancing Algorithms -

Metrics/ Techniques	Performance	Through-put	Overhead	Fault Tolerance	Migration Time	Response Time	Resource Utilization	Scalability	Power Saving
Genetic Algorithm [1]	YES	YES	YES	NO	YES	YES	YES	YES	YES
Based Random Sampling [20]	NO	YES	YES	YES	YES	NO	YES	NO	NO
Task Scheduling Based on LB [18]	NO	NO	NO	NO	NO	NO	YES	NO	NO
Min-Min [16]	YES	YES	YES	NO	NO	YES	YES	NO	NO
Max-Min [16]	YES	YES	YES	NO	NO	YES	YES	NO	NO
Shortest Response Time First [17]	YES	NO	NO	NO	NO	NO	YES	NO	NO
Compare And Balancing [2]	NO	YES	YES	YES	YES	YES	YES	NO	YES
Randomized [16]	YES	NO	NO	YES	YES	YES	YES	YES	NO
Honey Bee Foraging [16]	NO	NO	NO	NO	NO	NO	YES	NO	NO
Active Clustering [18]	NO	NO	YES	NO	YES	NO	YES	NO	NO
Round Robin[2]	YES	NO	YES	NO	NO	NO	YES	NO	NO
Ant Colony Optimization [3]	YES	NO	NO	NO	YES	NO	YES	NO	NO

10. Based Random Sampling: This algorithm is based on the construction of the virtual graph having connectivity between the all nodes of the system where each node of the graph is corresponding to the node computer of the cloud system. Edges between nodes are two types as Incoming edge and outgoing edge that is used to consider the load of particular system and also allotment the resources of the node. [20] It is very good technique to balance the load.

11. Ant Colony Optimization: Ant algorithms are a multivalent approach to difficult combinatorial optimization problems. An example of this approach is the vendor problem (TSP) and the quadratic

IV. VIRTUAL MACHINE PLACEMENT

Virtual machine placement is the process of mapping virtual machines to physical machines. In other words, virtual machine placement is the most appropriate host selection process for the virtual machine. The process involves the categorization of hardware and hardware resources of virtual machines and the anticipated use of resources and the investment objective. The investment objective can be either to maximize the use of available resources or to save energy by allowing some servers to be shut down [3].

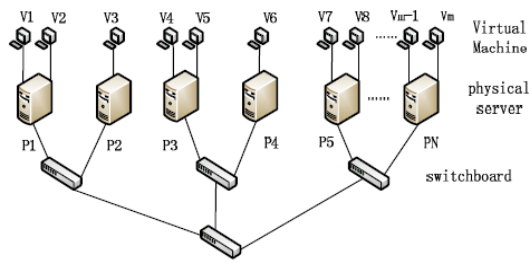


Figure1.Virtual machine placement figure

Classification of Placement Algorithms

The placement algorithms can be broadly classified into two categories on the basis of their placement goal.

- Power Based approach
- Application QOS based approach

TABLE I. VIRTUAL MACHINE PLACEMENT RULES TABLE

Physical Server ID	configuration parameter
Physical Server 1	$CPU(V_1) + CPU(V_2) < CPU(P_1)$ $Mem(V_1) + Mem(V_2) < Mem(P_1)$ $Store(V_1) + Store(V_2) < Store(P_1)$
Physical Server 2	$CPU(V_3) < CPU(P_2)$ $Mem(V_3) < Mem(P_2)$ $Store(V_3) < Store(P_2)$
...	...
Physical Server n	$CPU(V_{m-1}) + CPU(V_m) < CPU(P_n)$ $Mem(V_{m-1}) + Mem(V_m) < Mem(P_n)$ $Store(V_{m-1}) + Store(V_m) < Store(P_n)$

V. GENETIC ALGORITHM

Genetic algorithm is not dependent on the auxiliary information of searching space. It only depends on the fitness function to evaluate individuals, therefore it provides a framework for solving complex problems and is being widely used in various fields at present, for example: function optimization, combination optimization, job scheduling and so on. The advantages of the algorithm are that it begins to search from the population, widely covering and favoring to get the globally optimal solution. Its disadvantages are rare programming realization, weaker local researching, longer searching time and globally optimal solutions are influenced by operator parameters [1].

5.1 Selection Operation

Chromosomes are selected from the population to be parents to crossover. The problem is how to select these chromosomes. There are many methods how to select the best chromosomes, for Example roulette wheel selection, Boltzmann selection, tournament

selection, rank selection, steady state selection and some others.

5.2 Encoding Scheme

Genetic Algorithm consists of various numbers of chromosomes and each chromosomes has various gens that can be represented as V that denotes to the virtual Machine. P denoted the physical Machine on which VMs are to be allocated. P chromosome in this GA consists of /V / genes, each of which stands for a virtual machine. The value of a gene is a positive integer between 1 and /P/, representing the physical machine where the virtual machine is allocated. Fig. 2 shows an example VM placement and its corresponding chromosome.

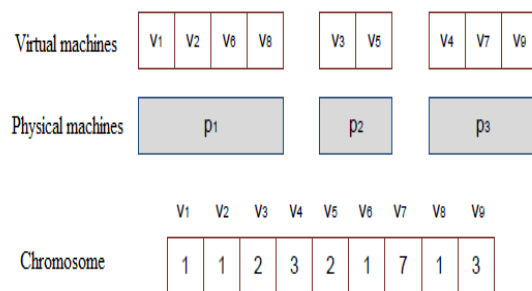


Fig 5.1: Figure representing chromosome for physical machines

5.3 Crossover Operation

The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the test characteristics from each of the parents. In this more than one parent is selected and one or more off-spring produced using the genetic material of the parents. Crossover is usually applied in a GA with a high probability. There are many crossover operator example one point, two point, multi point, arithmetic.

I. One Point Crossover- In this one-point crossover, a random crossover point is selected and the tails of its two parents are swapped to get new off-springs

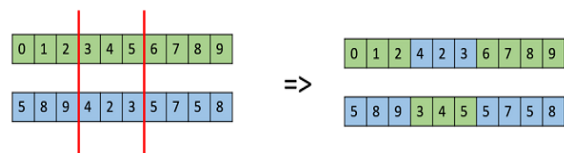


Fig 5.2: Figure showing single point crossover

II. Multi Point Crossover- Multipoint crossover is a generalization of the one-point crossover where in alternating segments are swapped to get new off-springs.

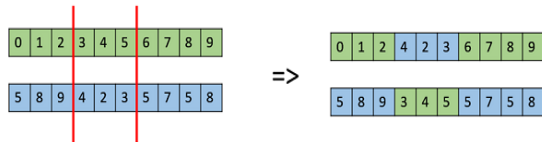


Fig 5.3: Figure showing multi point crossover

5.4 Mutation Operation

Mutation is an important part of the genetic search as it helps to prevent the population from stagnating at any local optima. Mutation occurs during evolution according to a user-definable mutation probability. If it is set to high, the search will turn into a primitive random search. There are many mutation operator types, for example, flip bit, boundary, uniform, non-uniform, Gaussian.

1) **I. Bit Flip Mutation** - In this bit flip mutation, we select one or more random bits and flip them. This is used for binary encoded GAs.

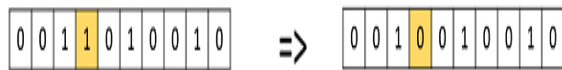


Fig 5.4: Figure showing mutation operation

5.5 Fitness Function-

The fitness function insures that fitness value of infeasible solution is less than of any feasible or optimal solution. The greater fitness value leads to the minimization of energy consumption and VM migrations. The fitness of an individual x in the population of the GA is defined in Eq. Below:

$$\text{Fitness}(x) = \begin{cases} E_{\min}/E(x), & \text{if } x \text{ is feasible;} \\ E_{\min}/(E(x) + E_{\max}), & \text{otherwise} \end{cases}$$

VI. PROPOSED APPROACH

In this work an energy efficient VM placement and migration approach is proposed based on genetic algorithm. The GA was implemented in Java. Since there are no benchmarks available for the new VM placement problem, we must generate random test problems to test the GA. We use a set of experiments to evaluate the proposed GA performance and scalability.

Algorithm 1 is used to categorize physical machine in to appropriate categories. The categories include:

- a) **Over loaded:** When utilization of physical machine is greater than upper threshold it is said to be overloaded.
- b) **Under loaded:** When utilization of physical machine is lower than upper threshold it is said to be under loaded.
- c) **Normal:** When the utilization of physical machine is between lower and upper threshold it is said to be in normal category.

```

Algorithm 1
Categorization PM
{
  Monitor the use of each PM;
  Use = summation of MIPS of current VMs /
  total MIPS of PM;
  If (use > upper threshold)
    Add a PM to an overloaded list
  If (use < lower threshold)
    Add PM to the Under-Loaded List
  If (lower threshold < use < upper threshold)
    Add PM to the normal category
}
    
```

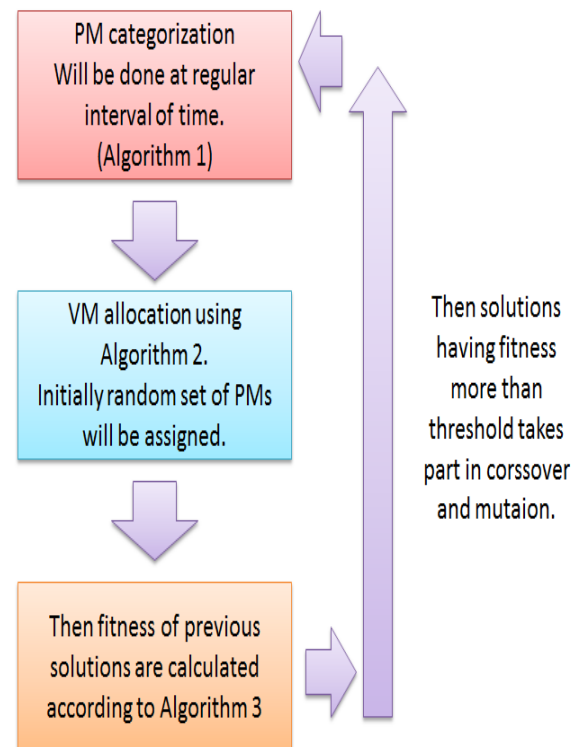


Fig 6.1: Flow chart of proposed approach

Algorithm 2 Describe VM allocation will be done using Genetic approach. But initially VMs will be assigned to random set of physical machines. Then according to the performance and situation fitness of solution will be calculated and then this solution if having fitness more than threshold will take part in crossover and mutation.

```

Algorithm 2
VM_Allocation
{
Initialize the cloud data center;
Initialize physical machines;
Assign a set of clouds to run on the cloud;
Next, the respective cloud will be mapped to the
virtual machine of the corresponding MIPS.
These virtual machines will be attributed to the
random play of physical machines initially and to
the solution obtained by genetics after a few
iterations.
}
    
```

Algorithm 3 Describes fitness calculation of the solutions assigned in previous algorithm. Fitness is calculated based on energy consumption. If energy consumption is high fitness will be low.

```

Algorithm 3
Genetic_VM
{
Take random set of assigned PMs from
population for the given number of VMs.
Calculate energy consumption for the assigned
solution.
Calculate fitness value of the solution and for
all the solutions obtained in every iterations
If ( fitness value > threshold)
{
Move solution to the crossover and mutation set
}
}
    
```

```

Algorithm 4
Crossover
{
Take two solutions from crossover set.
These solutions will be having fitness value above
the threshold.
Apply crossover to obtain the new solution.
Send the solution to last step of VM_Allocation
algorithm,
}
    
```

VII. RESULTS

All the implementations done in this research work are done using Cloud Sim configured in Eclipse. In this paper, we propose a genetic algorithm (GA) [2] for a new VM placement problem that considers energy consumption in both the physical servers (PMs) and the communication network in the data center. The experimental results show that the genetic algorithm works well with various test problems and balances well when the size of the problem increases. The proposed algorithm is compared with the previous approaches in terms of two major factors which include:

- a) Energy consumption (KWH).
- b) Number of VM Migrations.

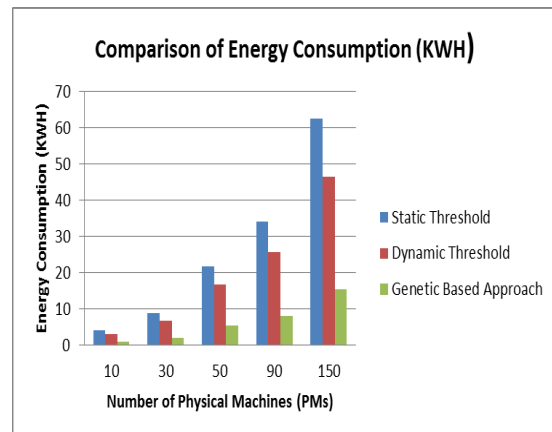
7.1 Energy Consumption – To calculate energy consumption in this research work, Power datacenter class of CloudSim is used which considers three factors to calculate energy consumption.

- a) Number of active physical machines.
- b) Number of VM Migrations.
- c) MIPS capacity of VMs running.

Table 7.1
Table showing comparison of energy consumption

Comparison Of Energy Consumption In Kwh				
No. Of Pms	No. Of Vms	Energy Consumption For Static Threshold (Kwh)	Energy Consumption For Dynamic Threshold (Kwh)	Energy Consumption For Genetic (Kwh)
10	15	4.07	3.05	0.98
30	30	8.93	6.71	2.12
50	70	21.62	16.79	5.53
90	110	34.21	25.64	8.1
150	200	62.49	46.45	15.43

Table 1 holds the comparison of energy consumption of three approaches for cloud resource allocation (VM allocation and VM migration). The results are analyzed for different number of physical machines and virtual machines. The comparison is done between static double threshold method for cloud, dynamic double threshold method and proposed genetic based method. It is clear that proposed approach has lowest energy consumption.



Graph 7.1: Graph illustrating comparison of energy consumption

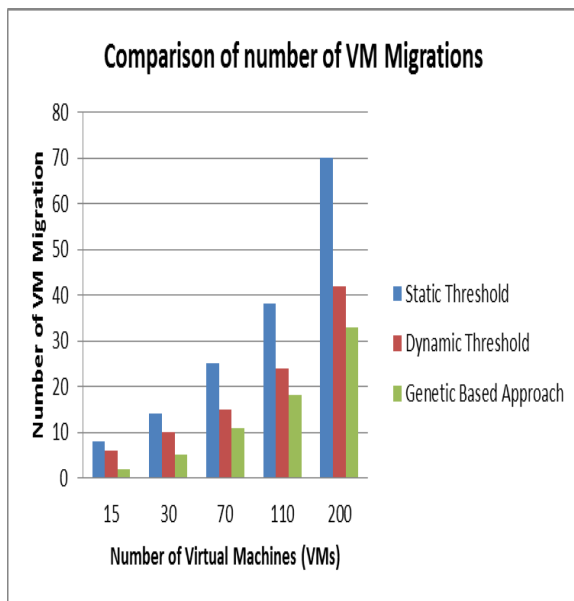
7.2 Number of VM Migrations – A VM running on any PM will be migrated only in two cases either the

machine is overloaded or machine is under loaded. When the physical machine is over loaded and VM is migrated to some other physical machine such condition is called VM migration. When VM is migrated because machine is under loaded such condition is termed as VM consolidation.

Table 7.2
Table showing comparison of number of VM migrations

Comparison Of Number Of VM Migration				
No. Of Pms	No. Of Vms	No. Of Vm Migrations For Static Threshold (Kwh)	No. Of Vm Migrations For Dynamic Threshold (Kwh)	No. Of Vm Migrations For Genetic (Kwh)
10	15	8	6	2
30	30	14	10	5
50	70	25	15	11
90	110	38	24	18
150	200	70	42	33

It is clear from Table 2 that number of VM migrations are least for proposed approach as compared to dynamic double threshold and static double threshold approaches.



Graph 7.2: Graph illustrating comparison of number of VM Migrations

VIII. CONCLUSION

Cloud computing has been a trending technology these days. It allows one to access data from any geographical part of the world. In this paper a survey is carried out on the types of cloud, architecture of cloud and different resource allotment and

scheduling algorithms for cloud. Cloud consists of several physical machines and on top of them many virtual machines are executed. So cloud resource scheduling becomes a NP-Hard problem. Therefore in this paper some of the naturally inspired meta-heuristic and evolutionary algorithms are studied which includes Genetic algorithm, Ant colony optimization and migrating Bird optimization.

In future a resource scheduling algorithm by modifying the Genetic algorithm fitness function can be proposed which may give energy efficient scheduling for the cloud in the real time.

References

- [1] Meikang Qiu, Senior Member, "Phase- Change Memory Optimization for Green Cloud with Genetic Algorithm" In IEEE dec 2015.
- [2] Rajwinder Kaur and Pawan Luthra "Load Balancing in Cloud Computing," Proc. of Int. Conf. on Recent Trends in Information, Telecommunication and Computing, ITC 2014.
- [3] Xiao-Fang Liu, Zhi-Hui Zhan, Jeremiah D. Deng, Yun Li, Tianlong Gu and Jun Zhang "An Energy Efficient Ant Colony System for Virtual Machine Placement in Cloud Computing" In IEEE 2016.
- [4] Dan Liu, Xin Sui, Li Li, "An Energy-efficient Virtual Machine Placement Algorithm in Cloud Data Center" 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) 2016.
- [5] <http://blog.nextright.com/?cat=6> "Cloud Computing « Nextright Blog".
- [6] A. Sidhu, S. Kinger, "Analysis of load balancing techniques in cloud computing", INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY 4 (2) (2013) pages737–741.
- [7] <http://www.qualitytesting.info/group/cloudcomputing/forum/topics/software-as-a-s>, "Software as a Service (SAAS) - Quality Testing".
- [8] <http://letslearncloud.wordpress.com/> "Cloud Computing | Learn Cloud and its tips."
- [9] Sahu, Yatendra and Pateriya, RK, "Cloud Computing Overview with Load Balancing Techniques", International Journal of Computer Applications, 2013, vol. 65, Sahu2013
- [10] Chaudhari, Anand and Kapadia, Anushka, "Load Balancing Algorithm for Azure Virtualization with Specialized VM", 2013, algorithms, vol 1, pages 2, Chaudhari
- [11] Nayandeep Sran, Navdeep Kaur, "Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing", vol 2, jan 2013
- [12] Bala, Anju and Chana, Inderveer, "A survey of various workflow scheduling algorithms in cloud environment", 2nd National Conference on Information and Communication Technology (NCICT), 2011.
- [13] Chaczko, Zenon and Mahadevan, Venkatesh and Aslanzadeh, Shahrzad and Mcdermid, Christopher, "Availability and load balancing in cloud computing", International Conference on Computer and Software Modeling, Singapore, chaczko2011availability.

- [14] S.-C.Wang, K.-Q. Yan, S.-S.Wang, C.-W. Chen, “A three-phases scheduling in a hierarchical cloud computing network”, in: Communications and Mobile Computing (CMC), 2011 Third International Conference on,IEEE, 2011, pp. 114–117.
- [15] O. Elzeki, M. Reshad, M. Elsoud, “Improved max-min algorithm in cloud computing, International Journal of Computer Applications”vol 50 (12) (2012) pages 22–27.
- [16] U. Bhoi, P. N. Ramanuj, “ Enhanced max-min task scheduling algorithm in cloud computing”, International Journal of Application or Innovation in Engineering and Management (IJAIEM), ISSN,2013,pages 2319—4847.
- [17] Y. Hu, R. Blake, D. Emerson, “An optimal migration algorithm for dynamic load balancing”, Concurrency: Practice and Experience 10 (6) (1998) pages 467–483.
- [18] Nidhi Jain Kansal, Inderveer Chana, "Cloud LoadBalancing Techniques: A Step towards Green Computing", IJCSI, Vol. 9, Issue 1, January 2012.
- [19] Nusrat Pasha, Dr. Amit Agarwal and Dr. Ravi Rastogi, “Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment” International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 5, May 2014.
- [20] Sreenivas Velagapudi, M.Prathap and Kemal Mohammed, “Load Balancing Techniques: Major Challenge in Cloud Computing – A Systematic Review,”IEEE, International Conference on Electronics and Communication Systems (ICECS) - Coimbatore, India (2014).
- [21] S. Wang, K. Van, W. Liao, and S. Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICC SIT), Chengdu, China, pp.108-113, September 2010.
- [22] Shahrzad Aslanzadeh, Venkatesh Mahadevan, Christopher Mcdermid, “Availability and Load Balancing in Cloud Computing,” International Conference on Computer and Software Modeling IPCSIT vol.14 IACSIT Press, Singapore 2011.
- [23] T. Kokilavani and Dr. D.I. George Amalarethinam, “Load Balanced Min-Min Algorithm for Static Meta-Task Scheduling in Grid Computing,” International Journal of Computer Applications Volume 20– No.2, pp.0975-8887, April 2011.