A Review on Resource Allocation and VM Scheduling Techniques and a Model for Efficient Resource Management in Cloud Computing Environment.

S.K.Sonkar
, Department of Computer Engineering,
KKWagh IEE&R Nashik,
SP Pune University, India.
e-mail:sonkar83@gmail.com

M.U.Kharat
Department of Computer Engineering,
METIOE Nashik,
SP Pune University, India.
e-mail:mukharat@rediff.com

***Abstract*- The key objective of cloud provider is to deliver the maximum resource utilization and increase the profit by reducing energy consumption and operative cost. While the target of cloud user is minimum response time and enhancing the whole throughput. The exciting task of cloud environment is to allocating the resources using lowest overhead interval along with efficient utilization of existing resources. Using virtual machine scheduling techniques resources are allocated in cloud data center. Many users can simultaneously make the request for service in cloud environment. Therefore to improve the system performance and reducing the cost, an effective resource allocation and vm scheduling technique is needed.**

This paper presents a review on various existing resource allocation and vm scheduling techniques. Based on review proposed an optimal resource management algorithm considering execution time of requests and availability of vm resource. This Scheme which helps to combine various types of workloads nicely and increase the overall utilization of server resources.

Keywords: Cloud Environment, VM Scheduling, Resource utilization, Response time.

I. INTRODUCTION

Today almost all IT organizations want to earn more and more profit with less investment. So they have to take care about the amount invested. For this organizations pay attention towards their IT infrastructure so that the cost will be reduced. So organizations look for solution of this problem and there it is cloud computing act as a dynamic part in this area. Cloud computing provides the on demand services to the organizations that include software, platform and infrastructure as a service to its users [1].

Cloud computing is a computing utility which has infinite capacity, instantaneous scalability and user pays only for the resources they use and only for time duration they are using it [1].

978-1-5090-5515-9/16/$31.00 ©2016 IEEE

According to NIST cloud computing is a model which authorizes on demand, proper entree to the network for a common computing resources (Resources are, CPU, memory, network, storage, etc...) which can be vastly provided with less management efforts. Cloud model encourages availability issue.

Features of cloud computing:

On-demand self-service: A cloud user provided with computing abilities such as server, network, and storage automatically as required by the consumer devoid of human interaction.

Broad Network Access: Available network capabilities are retrieved through standard mechanisms. This mechanism encourages usage of heterogeneous thick or thin client (e.g. mobiles. Laptops etc.).

Rapid Elasticity: Abilities are provided fast and elastically to scale out quickly and released rapidly. These provided capabilities appear to consumer as unrestricted and it can be purchased any time.

A. Cloud Service Models:

Following are the cloud service models

a) Software as a Service (SaaS): In this model software license is provided on subscription basis and it is centrally hosted. User can run their application by using these on demand software on cloud infrastructure. These applications can be accessible from different machines through thin user interface (such as web browsers). The user isn't responsible for managing underlying cloud infrastructure with exception of user specific applications.

b) Platform as a Service (PaaS):In PaaS, user provided with deploying service. User can deploy his own applications coded by using any software development languages. The user does not need to accomplish cloud infrastructure however he has to control over deployed applications.

c) Infrastructure as a Service (Iaas):In IaaS model, user provided with organizational

infrastructure for ex. network storage, and some of other computing resources. By using these resources deploying and running any software is possible to User.

B. Cloud Deployment Models

a) Private Cloud: In this infrastructure is private and worked with a single entity (Organization/User). It is accomplished by the organization or may be user.

b) Community Cloud: In this the cloud infrastructure is shared through multiple entities and provides supports to a specific community that has a shared concerns.

c) Public Cloud: In this general public is provided with cloud infrastructure which is possessed by organization which sells cloud services.

d) Hybrid Cloud: Hybrid cloud contains mixture of private and public cloud.

C. Virtualization

Virtualization is an abstraction of computer resources. It can be possible to access resources in consistent way before and after abstraction through virtualization.

Virtualization supports migration of virtual machines from existing to other physical machines. Migration process transfers the main memory pages and states of virtual machine to a target machine. Migration policy decides when, which and where to migrate a virtual machine. One example is, there is high load on a physical machine and not sufficient resource available for a virtual machine, that virtual machine is migrated to a less utilized physical machine.

Placement policy decides where to put a newly created virtual machine. For example, this may depend on current resource utilization of physical machine or number of already running virtual machines or type of applications that will run on virtual machine. Since in cloud computing, applications experience high variation of demand, virtualization can dynamically help to allocate resources.

D. Resource Management

It is a process which includes resource discovery then allocation of resources and finally monitoring the status of resources as shown in Fig. 1. It manages computing resources for ex. CPU-cores, memory consumption & network bandwidth etc.… [2].These resources as per need it divided and shared between VM's running possibly mixed workloads.

The basic component of resource management is the Resource discovery process. Which determines the suitable types of available resources as per the application requirements [3]. This process is managed by the cloud service provider.

The full information of resources availability is determined by resource discovery procedure. According to [4], resource discovery offer a method for a resource management system (RMS) for determining the status of the resources that are managed by it and other Resource management systems that interrelate with it. The resource discovery works with distribution of resources to provide information about the state of resources to the Server.

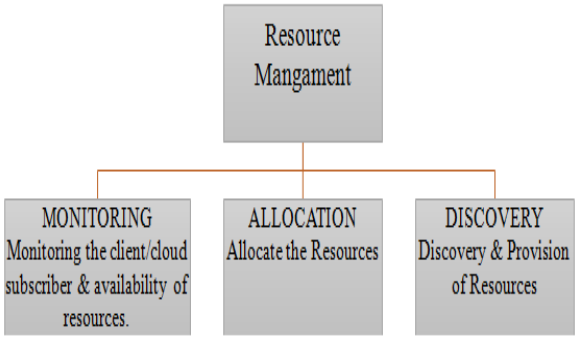

Figure 1. Elements of resource management

The Allocation procedure is the process of assigning an existing resource required for cloud applications over the internet [5]. These resources are allotted based on user demand and pay-per-use method. In this process, scheduling and dispatching method is being useful to assign the resources. Responsibility of scheduler is plan to assign resources to the client. Then, the dispatcher will allocate the assigned resources to the client.

However major problem in cloud computing is optimization of resources which are actuality allocated. Because of resource allocation policy must meet customer demand and their application requirement along with cost to provide should be minimum.

E. Importance of Resource Allocation

Resource Allocation Policy (RAP) which focuses on utilizing and allocating available resources so it need to meet the requests of the cloud application. To complete a user task resource allocation involves the type and some amount of resources. The input for optimal resource allocation policies are order sequence and time of allocation of resources. The optimal resource allocation policy should prevent the conditions as follows:

- Resource contention: when two requests attempt to access the similar resource at the same time then such type of situation arises.
- Scarcity of resources: when there are inadequate resources, then such condition is arises.
- Resource fragmentation: such type of condition occurs when there are sufficient resources but not able to access for required application.
- Over-provisioning: Such type of situation arises when the demand is less and resource availability is more than the demand.

- Under-provisioning: Such type of situation arises when the demand is less and resource availability is more than the demand.

From the cloud user's viewpoint, the major input to the resource allocation policies are the task requirement and SLA. In other hand it require to check resource status and available resources as input which require to manage as well as allocate resources to host applications [6]. The result of any optimal Resource Allocation Policy should fulfil the factors such as throughput, latency and response time. There is a critical problem for dynamically allocate and manages resources across the applications in cloud Even though cloud offers reliable resources.

F. VM Resources Scheduling

At Present in cloud computing environment, virtual machine (VM) resources scheduling most of cases, consider the current state of the system but rarely considers system variation and past data, as a result is always leads to unbalancing the load of the system. In this view of the load balancing problem we can solve the problem of load imbalance as well as high migration cost by applying VM resources scheduling strategy.

The scheduling of VM resources there are many similarities and differences among traditional scheduling algorithms in cloud computing environment. One of the major difference between cloud computing environment and traditional computing techniques is scheduling. Scheduling a task or transferring a small amount data such type of activity are done in in traditional computing environment, whereas transferred data is huge in cloud computing environment and VM resources is the scheduled target so, the granularity is large. Also in cloud computing environment, related with the placement time of VMs, scheduling time of algorithm nearly be neglected.

To maximize utilization of resources, this resources must be appropriately allocated and it should be effectively need to achieve the load balancing. [7]. Therefore important research objective is, scheduling the VM resources in such a way that which realize load balancing and to improve resource utilization in cloud computing.

G. Metrics for Load Balancing

Qualitative metrics consist some parameters which are useful in order to find best algorithm among them [8]. The different qualitative metrics or parameters that are considered important for load balancing in cloud computing are discussed as follows:

- Throughput

Throughput is used for estimation of total number of task, whose execution has been finished successfully. High throughput gives the improved performance of the system.

- Associated Overhead

The execution of the load balancing algorithm creates the quantity of overhead. If an algorithm is successfully implemented then there is least overhead.

- Migration time

Migration time is the time occupied for to transfer a job from one machine to another machine in the system. If migration time is least then there will be the improved system performance.

- Response time

Response time should be minimum, distributed system takings for to responding to executing a particular load balancing Techniques.

- Resource Utilization

Resource utilization is usage to show what amount of resources of the system are utilized. An efficient load balancing algorithm are responsible for maximum resource utilization.

- Scalability

It decides the capability of the system to achieve load balancing algorithm with a constrained number of processors or machines.

- Performance

Performance signifies the effectiveness of the system after execution of load balancing. If all the above factors are fulfilled optimally then system performance is highly improved.

II. LITERATURE SURVEY

Author[9] reviewed the various resource allocation strategies such as Black and Grey box strategies with BG algorithm, Vector dot algorithm, Skewness and load prediction algorithm, resource allocation strategy using feedback control theory, adaptive resource allocation algorithm, Adaptive list scheduling (ALS) and adaptive min-min scheduling (AMMS) algorithms, the main objective of all these algorithms sufficiently satisfy customer demand and their application requirement along with cost to provide should be minimum

Black and Grey box strategies with BG algorithm proposed by T. Wood et al. [10] which uses Xen hypervisor and finds with Nucleus and monitoring engine black box is not possible to make proactive decision making, whereas Grey-box permits proactive decision making. Drawback of Black-box is restricted with reactive decision making and BG algorithm needs higher number of migrations.

Vector dot algorithm introduces by A. Singh et al. [11] is used for load balancing which handle the hierarchical complication of data center. The data center integrated with server and storage virtualization server. is capable and focuses on using Configuration and performance manager. This scheme uses dot product to separate nodes based on the task requirements and eliminates overload on servers, switches and storage nodes.

For Dynamic resource allocation Zhen Xiao [12]

gives Skewness algorithm. Which uses Xen hypervisor Usher controller for resource allocation. The advantages of this technique is no overheads & high performance. It requires very few number of migrations and remaining resource is responsive to virtual machines. Which impacts the scheduling is done effectively. The disadvantage of this scheme is it is not cost effective.

L. Qiang et al. [13] which proposed resource allocation approach using feedback control theory, for appropriate controlling of virtualized resources, which is based on virtual machine (VM).The author explained the advantage of shared space of cloud infrastructure. This technique contains VM-based architecture, in which all hardware resources are collected into common shared space of cloud computing infrastructure so that hosted application can right to use the essential resources as per there requirement in order to meet Service Level Objective of application. The adaptive manager used in this technique is multi-input multi-output (MIMO) resource manager, which contains 3 controllers: CPU controller, memory controller and I/O controller, its objective is control various virtualized resources utilization to achieve SLOs of application by taking controller inputs per-VM CPU, memory and I/O allocation.

For accomplishing self-optimization in distributed autonomic computing systems, Utility functions provide a natural and advantageous framework which is explained by Walsh et al. [14]. Author present a distributed architecture, implemented as precise of model of data center that proves how utility functions can enable a collection of autonomic elements to persistently optimize the utilization of computational resources in a dynamic, heterogeneous environment. The architecture involve two-level structure of autonomic elements that provisions elasticity, modularity, and self-management. Each individual autonomic element finishes application resource usage to optimize local service-level utility functions, and a global arbiter assigns resources among application request environment based on resource-level utility functions obtained from the managers of the applications. The utility function Technique running on a Linux clustering that is suitable to manage realistic, fluctuating Web-based transactional workloads.

The cloud system using preempt able jobs based on the updated actual task executions systems regulate the resource provision efficiently, for this Jiayin Li et al. [15] can proposed an algorithm, adaptive resource allocation. In static resource allocation containing in the static task scheduling Author proposed a two algorithms, Adaptive list scheduling (ALS) and adaptive min-min scheduling (AMMS), generated offline. Online adaptive method is used for re-evaluating the remaining static resource allotment repeatedly with pre-defined frequency, In this case schedulers is re-calculates the final execution time of their respective submitted tasks, for each reevaluation process, no other tasks that are assign to that cloud system. Actually this technique is suitable for static resource allocation.

To reduce the response time along with efficient and balanced resource utilization [16] proposed two dynamic virtual machine scheduling techniques i.e. Best fit and Worst fit. For fully utilization of the resources finding the best fit host for VM which minimizes essential number of physical machine. The binary search method is used to search the best fit host that reduces the allocation time of one VM in order of ($logn$), so complexity of response time will be ($logn$). In Worst Fit VM allocation algorithm, if first Physical machine do not have sufficient resource requirement in all sizes, then power ON the new PM and allocate the VM on the new PM which happened.

Author present[17] proposed Artificial Bee Colony Longest Job First(ABC_LJF) algorithm, in which scheduling is created on the size of jobs, and the Longest Job First (LJF) scheduling algorithm achieved a better performance scheduling strategy in variable environment and corresponding workload which can decrease the data processing time of the system.

Hu et al. [18] proposed the load balancing scheduling technique of VM resources on the Cloud computing environment by implementing a tree structure utilizing Genetic algorithm for scheduling. The system which can resolve the quandary of load imbalance in Cloud computing by considering preceding data and the current state of work in advance to the performance demeanor.

Author [19] has discussed VM Load Balancing Algorithm of Weighted Active Monitoring which can implement in Cloud sim Tools. This Load Balancing algorithm is proposed for the Data center to efficaciously load balance the application requests among the existing virtual machines allocating a certain weight, for to accomplish improved performance parameters such as replication and Data processing time.

The Active Monitoring Load Balancer (AMLB) contains info about every VMs and the total demands presently allotted to the particular VM. When a new request is arrives, and then system recognizes the least burdened VM. If such condition satisfies by number of VM, then the first identified VM is selected. Active Vm Load Balancer returns the vm id number to the Data Center Controller. Then Controller sends the request to the VM identified by that id. Data Center Controller informs the Active VM Load Balancer of the fresh allocation and job request is sent to it.

For parallel processing First come first serve (FCFS) scheduling is very important. In which the resource having the minimum waiting time and it is chosen

for the received job [20]. The default policy accepted by the VM specified is a simple strategy that can distributes a Virtual machine in First Come First Serve scheme. The drawbacks of FCFS are that it is non preventative.

In Round Robin scheduling algorithm uses the loop in queue to store tasks. Every task within the queue has same implementation time. If one of the task cannot be finished during its instant time, then it will be recalled back to the lineup waiting for the succeeding turn [21]. The advantage of Round Robin algorithm is each and every task will be implemented as a result and they do not have to be wait for the earlier task to be completed. The drawback of Round Robin algorithm is that it takes longer time to finish all the tasks.

Generalized Priority Algorithm suggested that the jobs are initially given priority depending on their size such a way that one with the largest size having the largest score .The virtual machine is also similarly prioritized in associated with their MIPS. So highest VM –MIPS is having highest rank. Therefore, the main issue for ranking tasks is their size and MIPS for VM. This scheme is doing better than RR and FCFS scheduling [22].

Author [23] proposed Ena Cloud technique that uses a Virtual Machine to encapsulate the application. This supports applications scheduling and live migration to minimize the number of active machines, so that it save energy.

Especially, the bin packing problem considered for application placement and energy-aware heuristic algorithm is proposed to get a appropriate solution.

Author proposed [24] a balanced system which handles the VM demands in real time and providing assurance for better resource utilization. The total active physical machines and their energy consumption can be reduced as a outcome of high resource utilization.

Author proposed [25] an effective resource management strategy for to constantly combine VMs leveraging live relocation and power off idle nodes to decrease power consumption along with essential Quality of Service.

Author [26] focuses the improvement of dynamic resource provisioning and allocation systems that reflect the interaction among numerous data center infrastructures (such as cooling and software, the hardware, power units,), and comprehensively effort to increase data center energy efficiency and performance.

III. RESEARCH GAPS

- To reduces the resource allocation time (response time) of user demand or request & to deliver maximum use of resources. An efficient dynamic Virtual Machine scheduling method is required to be proposed.
- An efficient migration technique is required for computable improvements for load balancing and server consolidation for the underutilize host and over utilize host.
- An efficient VM placement Technique is need to be proposed so time need to place a Virtual Machine to the host is should be minimize.
- Based on job execution time , need to design effective job priority algorithm to improve the system performance

From the viewpoint of a cloud provider, determine the dynamic behavior of users, user requests, and application demands are impractical. As consider with cloud users, the task must be complete in time with negligible cost. Therefore due to dynamic behavior of resource demand, resource heterogeneity, limited resources, locality restrictions, environmental necessities, we need an efficient resource allocation system that suitable for cloud environments

IV. PROPOSED METHOD

This paper investigates the direction for the optimal resource management algorithm in cloud computing environment. The proposed work implementation will be based on the architecture shown in Fig 2.

A. Scenario Description

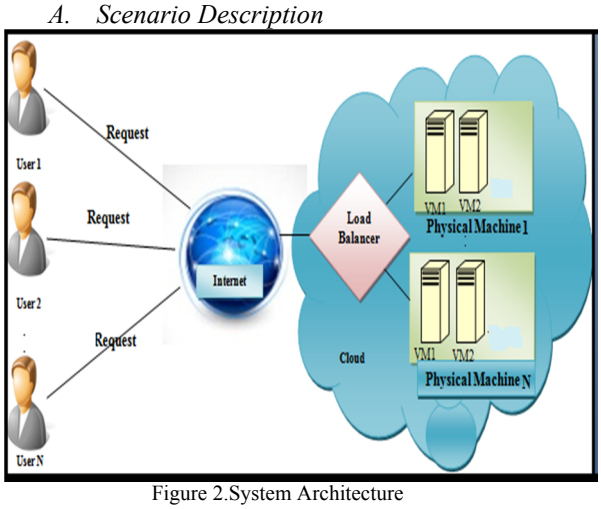

Figure 2.System Architecture

As shown in system arch., proposed system targeting a virtualized cloud that is categorized by an n number of H= (h1; h2; . .) of physical machine given that the hardware infrastructure for creating virtualized resources to assure user's requests. Active host is indicated by Ha with n elements, Ha ⊆ H. For a cloud, given host h_k, it is described by its Performance of CPU which is measured in Million Instructions Per Second, some amount of (RAM) memory, and network capacity which is measured in bandwidth, i.e., $h_k = (c_k; r_k; n_k)$, where c_k, r_k, and n_k stands for CPU capacity, RAM and network bandwidth of the k^{th} host, subsequently. For every host or node h_k, which involves a set of (VM's) virtual machines $V_k = (v_{1k}; v_{2k}; \ldots V|v_k|_k)$ of Virtual Machines. For a given VM v_{jk}, let $c(v_{jk})$, $r(v_{jk})$, and $n(v_{jk})$ which signify the fractions of CPU attainment, amount of RAM, and network capacity

allotted to v_{jk}. Additionally, several VMs with dynamism initiated and stopped on a single host based on the system load. Simultaneously, some Virtual Machines want to migrate across hosts in the interest of to consolidate resources and in addition to cut down energy consumption.

B. Request Model

In proposed system, denoting user requests by R and the set of request is set R = (r1; r2; . . .) of autonomous requests which are arriving dynamically. A request r_i submitted through a user can be measured as a sort of parameters, i.e., r_i =(a_i; l_i; t_i; e_i), where a_i, l_i, t_i and e_i stands for arrival time, request length/size, target time, & end time of request t_i, respectively.

Let r_{tjk} denoted for ready time of VM v_{jk} on host h_k. Similarly, let st_{ijk} stands the start time of task t_i on Virtual Machine vjk. Because there is heterogeneity in nature of CPU processing capabilities of Virtual Machines. Considering e_{tijk} for the execution time of task t_i for VM v_{jk}:

$$e_{tijk} = \frac{li}{c(vjk)} \quad (1)$$

It is expected that f_{tijk} is the finish time of task t_i on Virtual Machine v_{jk}, and which can be simply determined as follows:

$$f_{tijk} = s_{tijk} + e_{tijk} \quad (2)$$

In addition, x_{ijk} is handled to redirect a mapping of tasks to Virtual Machines at diverse set of hosts in a virtualized cloud, where x_{ijk} is "1" if task t_i is allocated to Virtual Machine v_{jk} at host h_k and is "0", otherwise.

The finish time is used to decide whether the task's timing constraint can be guaranteed, i.e.

$$x_{ijk} = \begin{cases} 0, & if\ \text{ftijk} > di \\ 1, & if\ ftijk \leq di \end{cases} \quad (3)$$

The expected output of the proposed model is to make use of available resource efficiently and to finish the task in the minimum cost.

C. Algorithm

Input: User request R= {r_1, r2, r3……..r_n}
Output: Optimal Resource allocation for user request.

1. Every Host H_k in Cloud Contains Set of VM's, V_{jk}={v_{1k},v_{2k},……v_{nk}}
2. Each VM having resources such as CPU measured in MIPS, Memory in MB, and Network bandwidth measured in mbps denoted by$C(V_{jk})$, $R(V_{jk})$, $N(V_{jk})$.
3. Let user request R = {r_1,r_2,r_3……..r_n}
4. Calculate execution time by request r_i for VM, V_{jk}
5. Sequence each request based on their execution time& Schedule the lowest one as first and so on.
6. Check for availability of VM resources { $C(V_{jk})$, $R(V_{jk})$, $N(V_{jk})$ }from system.
7. If request $r_i$is satisfied for VM resources allocate it only if VM resources does not go beyond Hot threshold.
8. Else Move request r_i to next VM
9. End.

V. CONCLUSION

To maximize the server and resource utilization, & also for achieving user satisfaction in cloud computing, an effective resource allocation & programming strategy is incredibly essential. Also Scheduling is very much necessary which also improve the performance of the computer. This paper focused on various resource allocation &scheduling techniques which efficiently allocate the resources & schedules the computational jobs in cloud background. For task scheduling in cloud environment priority remains a significant matter. Based on research gap, proposed the optimal resource allocation scheme based on the execution time of each request on a virtual machines with specific deadline along with load balancer checks the availability of VM resources. To improve overall resource utilization, proposed system checks the availability of resource utilization before assigning the task which prevents load unbalancing issue.

ACKNOWLEDGMENT

FA would like to thanks KKWagh IEE&R, Nashik for providing facilities to make successful work of this paper.

REFERENCES

[1] M. Armbrustet al., "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep., Feb 2009.

[2] Abhinandan S. Prasad, and Shrisha Rao, "A Mechanism Design Approach to Resource Procurement in Cloud Computing", IEEE Transactions On Computers, Vol. 63, No. 1, Pp.17-30, January 2014.

[3] R. Buyya and R. Ranjan, "Special section: Federated resource management in grid and cloud computing systems," FutureGeneration Computer Systems, vol. 26, no. 8, pp. 1189–1191, June, 2010.

[4] K. Krauter, R. Buyya, and M. Maheswaran, "A taxonomy and survey of grid resource management systems for distributed computing," Software: Practice and Experience, vol. 32, no. 2, pp. 135–164. 2002.

[5] V. Vinothina, R. Sridaran, and P. Ganapathi, "A survey on resource allocation strategies in cloud computing," International Journal ofAdvanced Computer Science and Applications, vol. 3, no. 6, pp. 97– 104, 2012.

[6] Patricia Takako Endo et al.: Resource allocation for distributed cloud: Concept and Research challenges (IEEE, 2011), pp.42-46.

[7] Jasmin James and Dr. BhupendraVerma, "Efficient VM Load Balancing Algorithm for a Cloud Computing environment", International Journal on Computer Science and Engineering (IJCSE), ISSN: 0975-3397, Vol. 4 No. 09 Sep 2012

[8] N.K.Mishra,Nishchol Mishra,"Load Balancing Techniques: Need, Objectives and Major Challenges in Cloud Computing- A Systematic Review",

International Journal of Computer Applications (0975 – 8887) Volume 131 – No.18, December 2015

[9] S.K.Sonkar, Dr.M.U.Kharat," A Survey on Resource Management in Cloud Computing Environment", International Journal of Advanced Trends in Computer Science and Engineering, vol4(issue4), Pages: 48 – 5,1 July - August 2015, ISSN: 2278-3091.

[10] T. Wood, P. Shenoy, A. Venkataramani, and M. Yousif, "Black-box and gray-box strategies for virtual machine migration," in Proc. Of the Symposium on Networked Systems Design and Implementation (NSDI'07), Apr. 2007.

[11] A. Singh, M. Korupolu, and D. Mohapatra, "Server-storage virtualization: integration and load balancing in data centers," in Proc. of the ACM/IEEE conference on Supercomputing, 2008.

[12] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen, "Dynamic resource allocation using virtual machine in cloud computing environment," IEEE Transaction on Parallel and Distributed Systems, vol.24, no.6, pp.1107-1117, June 2013.

[13] L. Qiang, Q. Hao, L. Xiao and Z. Li "Adaptive Management of Virtualized Resources in Cloud Computing Using Feedback Control," in First International Conference on Information Science and Engineering, pp. 99-102 ,April 2010.

[14] W. E. Walsh, G. Tesauro, J. O. Kephart, and R. Das, "Utility Functions in Autonomic Systems," inICAC '04: Proceedings of the First International Conference on Autonomic Computing. IEEE Computer Society, pp. 70–77, 2004.

[15] Jiayin Li, MeikangQiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Preemptable Jobs in Cloud Systems," in 10th International Conference on Intelligent System Design and Application, pp. 31-36,Jan. 2011.

[16] Vijaypal Singh Rathor, R. K. Pateriya and Rajeev Kumar Gupta," An Efficient Virtual Machine Scheduling Technique in Cloud Computing Environment", IJCS, Vol.1, No.1 ,pp.1-14, (2014), ISSN: 2287-8491.

[17] B. Kruekaew and W. Kimpan, "Virtual Machine Scheduling Management on Cloud Computing Using Artificial Bee Colony" Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I, IMECS 2014, March 12 - 14, 2014, Hong Kong.

[18] J. Hu, J. Gu, G. Sun, and T. Zhao, "A scheduling strategy on load balancing of virtual machine resources in cloud computing environment," in 3rd IEEE Int. Symp. On Parallel Architectures, Algorithms and Programming (PAAP), 2010, 18-20 Dec. 2010, pp. 89-96.

[19] Panchal et al.,," Dynamic VM Allocation Algorithm using Clustering in Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering 3(9), September - 2013, pp. 143-150

[20] H. Yuan, C. Li, M. Du, "Optimal Virtual Machine Resources Scheduling Based on Improved Particle Swarm Optimization in Cloud Computing", Journal of Software, 2014.

[21] G. Raj, "Effective Cost Mechanism for Cloudlet Retransmission and Prioritized VM Scheduling Mechanism Over Broker Virtual Machine Communication Framework", International Journal on Cloud Computing: Services and Architecture, pp.41-50, 2012

[22] Y. Cao, C. Ro, "Adaptive Scheduling for QoS-based Virtual Machine Management in CloudComputing", International Journal of Contents, pp.7-11, 2012.

[23] Bo Li; Jianxin Li; JinpengHuai; Tianyu Wo; Qin Li; Liang Zhong ,"EnaCloud: An Energy-Saving Application Live Placement Approach for Cloud Computing Environments",IEEE International Conference on Cloud Computing, Pages: 17 – 24, Year: 2009, DOI: 10.1109/CLOUD.2009.72

[24] Xin Li; Zhuzhong Qian; Ruiqing Chi; Bolei Zhang; Sanglu Lu," Balancing Resource Utilization for Continuous Virtual machine Requests in Clouds", Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS),IEEE Conference Publication 2012 ,Pages: 266 - 273, DOI: 10.1109/IMIS.2012.72

[25] Anton Beloglazov; Rajkumar Buyya," Energy Efficient Allocation of Virtual Machines in C loud Data Centers",10th IEEE nternational Conference on Cluster, Cloud and Grid Computing (CCGrid), 2010 , Pages: 577 - 578,Year: 2010, DOI: 10.1109/CCGRID.2010.45

[26] Rajkumar Buyya1,2, Anton Beloglazov1 , and JemalAbawajy," Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges",https://www.researchgate.net/publication/45921163