iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Review Paper on Dynamic Resource Allocation in Cloud Environment

Rekha[1], Vinod Saroha[2]
[1]Department of Computer Science & I.T Engineering, B.P.S. Mahila vishwavidyalaya Khanpur Kalan, Sonipat
[2] Astt. Prof. (C.S.E &I.T deptt.), B.P.S. Mahila vishwavidyalaya Khanpur Kalan, Sonipat

Abstract*: Cloud computing is an emerging discipline for providing different types of resources to different types of users. It found its application in almost the fields of real life. With cloud computing one can access applications and other resources anytime and from anywhere. Due to scarcity of resources on cloud environment its proper scheduling and allocation is compulsory. Also resources can be allocated statically as well as dynamically therefore proper allocation and de-allocation is an important task for cloud computing. One of the methods statically allocates the resources and other dynamically allocates the resources. With static resource allocation the cloud user has to make prior request for the resources. With dynamic resource allocation the cloud resources are requested by the cloud user on the fly or as and when the application needs. This paper provides review of various dynamic resource allocation techniques. It also provides proposed solution for dynamic resource allocation in cloud.

Keywords: Cloud Computing, Static Resource Allocation, Dynamic Resource Allocation

I. INTRODUCTION

Cloud computing [1] comprises of 2 components —the front end and the back end. The whole cloud is administered by a central server that is used to monitor client's demands (Fig 1).

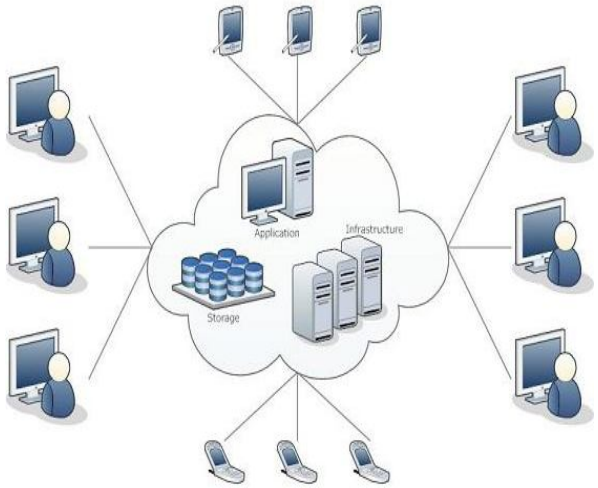


Figure 1: Cloud Computing

The front end includes client's devices and applications that are required to access cloud. The back end refers to the cloud where all the data are stored and resources are available for sharing.

Cloud Computing provides multiple advantages [2] to its users. Some popular advantages are explained below:

Construct & maintain the applications dynamically at any time.

User need not to install any specific software to access the cloud application. Users only need to connect with internet and authenticate them on the specific cloud.

Any type of user can access applications provided on the cloud.

Cloud Computing offers online development and deployment tools, programming runtime environment through

©IJRASET: All Rights are Reserved

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A. Platform as a Service Model

1) Cloud resources are available over the network in a manner that provides platform independent access to any type of clients. Different features provided by cloud computing is shown in figure 2 below:

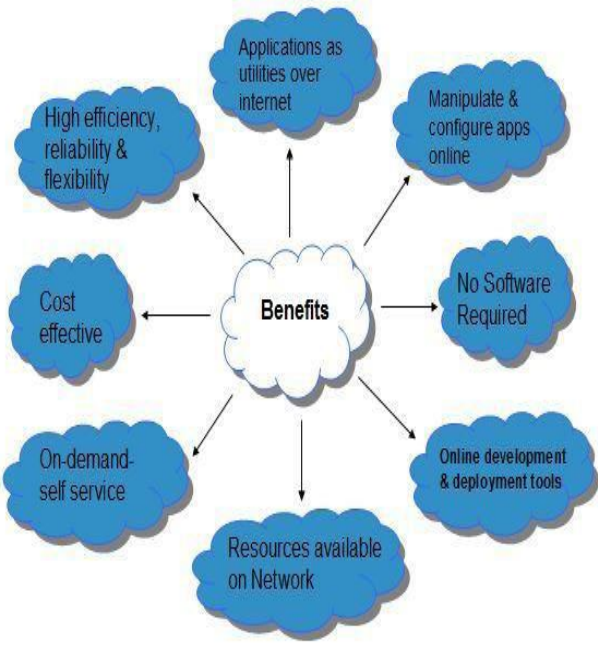


Figure 2: Features of Cloud Computing

One of the main services of cloud environment is resource allocation. The process of assigning the available resources to multiple clients in cost effective and efficient method is called Resource allocation. For efficient resource allocation cloud computing must schedule the available resources to different activities while taking into consideration both the resource availability and the project time.

Resource allocation can be done by two methods. One of the methods statically allocates the resources and other dynamically allocates the resources. With static resource allocation the cloud user has to make prior request for the resources. With dynamic resource allocation the cloud resources are requested by the cloud user on the fly or as and when the application needs. This paper provides review of various dynamic resource allocation techniques. This paper provides review of different dynamic resource allocation strategies in cloud environment.

II. CHARACTERISTICS OF CLOUD COMPUTING

The US National Institute of Science and Technology (NIST) define cloud computing as: "Cloud computing is an efficient and effective model for providing (enabling) ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be quickly and easily scheduled, allocated and de-allocated. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models".

A. Five Essential Characteristics of Cloud Computing as defined by NIST are explained below [2]:

1) On-Demand Self-service: A cloud user can individually provision computing capabilities, such as server time and network storage, thereby removing the need of middleman, because the client can manage automatically and access the resources required as needed without requiring human interaction with each service provider.

2) Broad Network Access: Different clients (users) benefited from the cloud and control them through standard mechanisms without regard of the end-user platform.

3) Resource Pooling: Different Cloud resources (such as memory, processing, secondary storage, and network bandwidth) are

©IJRASET: All Rights are Reserved

pooled to provide for multiple clients using a multi-tenant model, according to the user's demand. Private cloud may only be offsite at a location controlled by the owner or the provider may allow clients to specify general server locations.

4) *Rapid Elasticity:* Within the cloud computing, different resources can be allocated and de-allocated dynamically. Therefore same resource can be shared by multiple clients and hence improve scalability of resources automatically. This is one reason Denial-of-Service (DoS) attacks are decreasing, as companies with adequate cloud accounts are no longer vulnerable.

B. Measured Services

1) The measuring of cloud resource utilization is done automatically using metering capability, according to the type of service storage, processing, bandwidth, and active user accounts. This provides transparency for both the cloud vendor and the clients. The usage charge is send from server to client by showing them report of usage containing date and time, types of usage etc.

III. DYNAMIC RESOURCE ALLOCATION

From the perspective of a cloud provider, predicting the dynamic nature of users, user demands, and application demands are impractical. For the cloud users, the job should be completed on time with minimal cost. Hence due to limited resources, resource heterogeneity, locality restrictions, environmental necessities and dynamic nature of resource demand, we need an efficient resource allocation system that suits cloud environments. Cloud resources consist of physical and virtual resources. The physical resources are shared across multiple compute requests through virtualization and provisioning. The request for virtualized resources is described through a set of parameters detailing the processing, memory and disk needs. Provisioning satisfies the request by mapping virtualized resources to physical ones. The hardware and software resources are allocated to the cloud applications on-demand basis. For scalable computing, Virtual Machines are rented. The complexity of finding an optimum resource allocation is exponential in huge systems like big clusters, data centres or Grids. Since resource demand & supply can be dynamic and uncertain.
Resource allocation is process of assigning the available resources in an economic way and efficient and effective way Resource allocation is the scheduling of the available resources and available activities required by those activities while taking into consideration both the resource availability and the project time. Resource provisioning and allocation solves that problem by allowing the service providers to manage the resources for each individual request of resource. Resource Allocation Strategy (RAS) is all about the number of activities for allocating and utilizing inadequate resources within the limit of cloud environment so as to meet the needs of the cloud application. It requires the type and amount of resources needed by each application in order to complete a user job

IV. LITERATURE REVIEW

Several works related to dynamic resource allocation were done in the past. Some these works are described below.
Valérie Danielle Justafort et. al. [4] discusses about Performance-Aware Virtual Machine Allocation. Cloud Computing has major challenges among which virtual machine (VM) placement is considered as one of the crucial problems. From the cloud provider‟s perspective, this process often consists in choosing the best convenient physical hosts with regards to energy efficiency, resource utilization and revenues maximization. However, to secure the loyalty of their clients, cloud providers should also consider the applications performance. In this paper, VM placement approach is proposed tackling the applications performance concern, in an inter cloud environment. VM placement is stated as a mixed integer programming problem which aims to maximize a global utility function considering VM bandwidth requirements and network latencies.
Bo Yin et. al. [5] discusses about multi-dimensional resource allocation. In this paper, study on the resource allocation at the application level is done, instead to map the physical resources to virtual resources for better resource utilization in cloud computing environment. A multi-dimensional resource allocation (MDRA) scheme for cloud computing is proposed that dynamically allocates the virtual resources among the cloud computing applications to reduce cost by using fewer nodes to process applications. In this model, a two-stage algorithm is adopted to solve this multi-constraint integer programming problem. The algorithm can dynamically reconfigure the virtual machines for cloud applications according to the load changes in cloud applications by assigning new applications on the working node instead of opening a new node. Experiment results show that this algorithm can save resources and increase resource utilization as well as centralize working nodes. MDRA provide resource to deal with requirements more steadily. Moreover, the proposed algorithm in the long run can save power efficiently when the demand of user is decreasing.
Amit Nathani et. al. [6] have proposed policy based resource allocation in IaaS cloud. Most of the Infrastructure as a Service Cloud

©IJRASET: All Rights are Reserved

(IaaS) uses simple resource allocation techniques like immediate and best effort. Immediate Resource Allocation Policy allocates the resource if available otherwise the request is rejected. Best effort policy also resources if currently available otherwise the request is placed in the FIFO queue. So it is not possible for the Service provider to satisfy all the requests as there is finite number of resources at a time. So Haizea is a resource lease manager that addresses these issues by using complex resource allocation strategies. Haizea uses resource leases as resource abstraction and implements these leases by allocating Virtual Machines (VMs). Haizea supports four resource allocation policies: Immediate, best effort, advanced reservation and deadline sensitive. Among the four, deadline sensitive leases by Haizea supports minimal rejection of leases and using this policy, maximum resource utilization is possible. Dynamic Planning based scheduling algorithm is implemented in Haizea which can admit leases and prepare a schedule whenever a new lease can be accommodated.

Sharrukh Zaman et. al. [7] discusses online mechanism for dynamic VM provisioning and allocation. Current cloud provider allocates virtual machine instances via fixed price-based or auction based mechanisms. The limitation in these mechanisms is they are all offline mechanism and they need to collect information and invoked periodically. This limitation is addressed by designing an online mechanism by dynamically provisioning and allocating Virtual machine Instances in Clouds. The algorithm called Mechanism for online VM provisioning and allocation (MOVMPA) is invoked as soon as the user makes a request or some VM instances already become available again. When invoked, the mechanism selects users who would be allocated VM instances they requested for and ensures that those users will continue using for the entire period they requested for.

Mayank Mishra et. al. [8] proposed dynamic resource management using Virtual Machine Migrations. Virtual machine related features such as flexible resource provisioning, and isolation and migration of machine state have improved efficiency of resource usage and dynamic resource provisioning capabilities. Live virtual machine migration transfers “state” of a virtual machine from one physical machine to another, and can mitigate overload conditions and enables uninterrupted maintenance activities. This paper focuses on the details of virtual machine migration techniques and their usage toward dynamic resource management in virtualized environments.

Gihun Jung et. al. [9] discusses about agent-based adaptive resource allocation. Since both cloud users and data centers of a cloud service provider can be distributed geographically, the provider needs to allocate each user request to an appropriate data center among the distributed data centers, so that the users can satisfy with the service in terms of fast allocation time and execution response time. In this paper, an adaptive resource allocation model is proposed that allocates the cloud user‟s job to an appropriate data center. The method to adaptively find a proper data center is based on two evaluations: 1) the geographical distance (network delay) between a cloud user and data centers, and 2) the workload of each data center. The proposed model can successfully allocate cloud users‟ requests to the data center closest to each consumer. Also, the proposed model shows a better response time for allocation than related resource allocation models.

Xiao-Jun Chen et. al. [10] discusses about resource reconstruction algorithms for on-demand allocation. Resource reconstruction algorithms are used to solve the problem of on-demand resource allocation and can be used to improve the efficiency of resource utilization in virtual computing resource pool. The idea of resource virtualization and on the analysis of resource status transition these algorithms are used. These algorithms are designed to determine the resource reconstruction types and they can achieve the goal of on-demand resource allocation through three methodologies: resource combination, resource split, and resource random adjustment. These algorithms can do the resource adjustment on lower cost and form the logical resources to match the demands of resource users easily.

Ilhem Fajjari et. al. [11] discusses about optimized dynamic resource allocation algorithm for Cloud’s Backbone Network. Because of the unpredictable and varying user’s demands, a flexible and intelligent resource allocation scheme is necessary. This paper tackles the fundamental challenge of efficient resource allocation within cloud’s backbone network. The ultimate goal is to satisfy the cloud’s user requirements while maximizing cloud provider’s revenue. The problem consists in embedding virtual networks within substrate infrastructure. A new dynamic adaptive virtual network resource allocation strategy named Backtracking-VNE is investigated to deal with the complexity of resource provisioning within cloud network. The proposal coordinates virtual nodes and virtual links mapping stages to optimize resources usage.

K C Gouda et. al. [12] proposed a new approach that allocates resource with minimum wastage and provides maximum profit. The developed resource allocation algorithm is based on different parameters like time, cost, No of processor request etc. The developed priority algorithm is used for a better resource allocation of jobs in the cloud environment used for the simulation of different models or jobs in an efficient way. After the efficient resource allocation of various jobs, an evaluation is being carried out which

©IJRASET: All Rights are Reserved

illustrates the better performance of cloud computing with profit. A performance study of all the algorithms in various systems and case studies are also presented.

N. Krishnaveni et. al. [13] describe that Cloud computing becomes quite popular among cloud users by offering a variety of resources. This is an on demand service because it offers dynamic flexible resource allocation and guaranteed services in pay as-you-use manner to public. In this paper, we present the several dynamic resource allocation techniques and its performance. This paper provides detailed description of the dynamic resource allocation technique in cloud for cloud users and comparative study provides the clear detail about the different techniques.

Sukhpal Singh et. al. [14] emphasis on the development of energy based resource scheduling framework and present an algorithm that consider the synergy between various data center infrastructures (i.e., software, hardware, etc.), and performance. In specific, this paper proposes (a) architectural principles for energy efficient management of Clouds; (b) energy efficient resource allocation strategies and scheduling algorithm considering Quality of Service (QoS) outlooks. The performance of the proposed algorithm has been evaluated with the existing energy based scheduling algorithms. The experimental results demonstrate that this approach is effective in minimizing the cost and energy consumption of Cloud applications thus moving towards the achievement of Green Clouds.

Riddhi Patel et. al. [15] describe Modified best fit decreasing algorithm (MBFD) is discus for reduce the energy consumption and modified the MBFD algorithm as Energy aware Best Fit Decreasing (EABFD) algorithm. Resource allocation is one of the important challenges in cloud computing environment. It depends on how to allocate the resource to the particular task. Resource allocation can be done by two methods. One of the methods statically allocates the resources and other dynamically allocates the resources. The other challenges of resource allocation are meeting customer demands and application requirements. One of the challenges posed by cloud applications is Quality-of-Service (QoS) management, which is the problem of allocating resources to the application to guarantee a service level along dimensions such as performance, availability and reliability.

V. PROPOSED WORK

The presented work is focused on the concept of effective dynamic resource allocation & de-allocation in a cloud environment. To present the concept, we have taken a cloud environment with multiple clouds along with multiple virtual machines. All the machines are homogenous. These all clouds are assigned by a specific priority. Now as the user request arrive, it performs the request to the priority cloud under its requirements in terms of memory & processor capabilities. As the particular cloud will get the request, it will search for the number of requested processors. If the numbers of processors are available with the current cloud, the resources will be allocated to that particular client. But if the sufficient numbers of processors are not available then the search will be performed for the next particular cloud to perform the resource allocation.

VI. CONCLUSION

With cloud computing one can access applications and other resources anytime and from anywhere. Due to scarcity of resources on cloud environment its proper scheduling and allocation is compulsory. One of the main services of cloud environment is resource allocation. The process of assigning the available resources to multiple clients in cost effective and efficient method is called Resource allocation. For efficient resource allocation cloud computing must schedule the available resources to different activities while taking into consideration both the resource availability and the project time. This paper provided review of various dynamic resource allocation techniques. It also provided proposed solution for dynamic resource allocation in cloud.

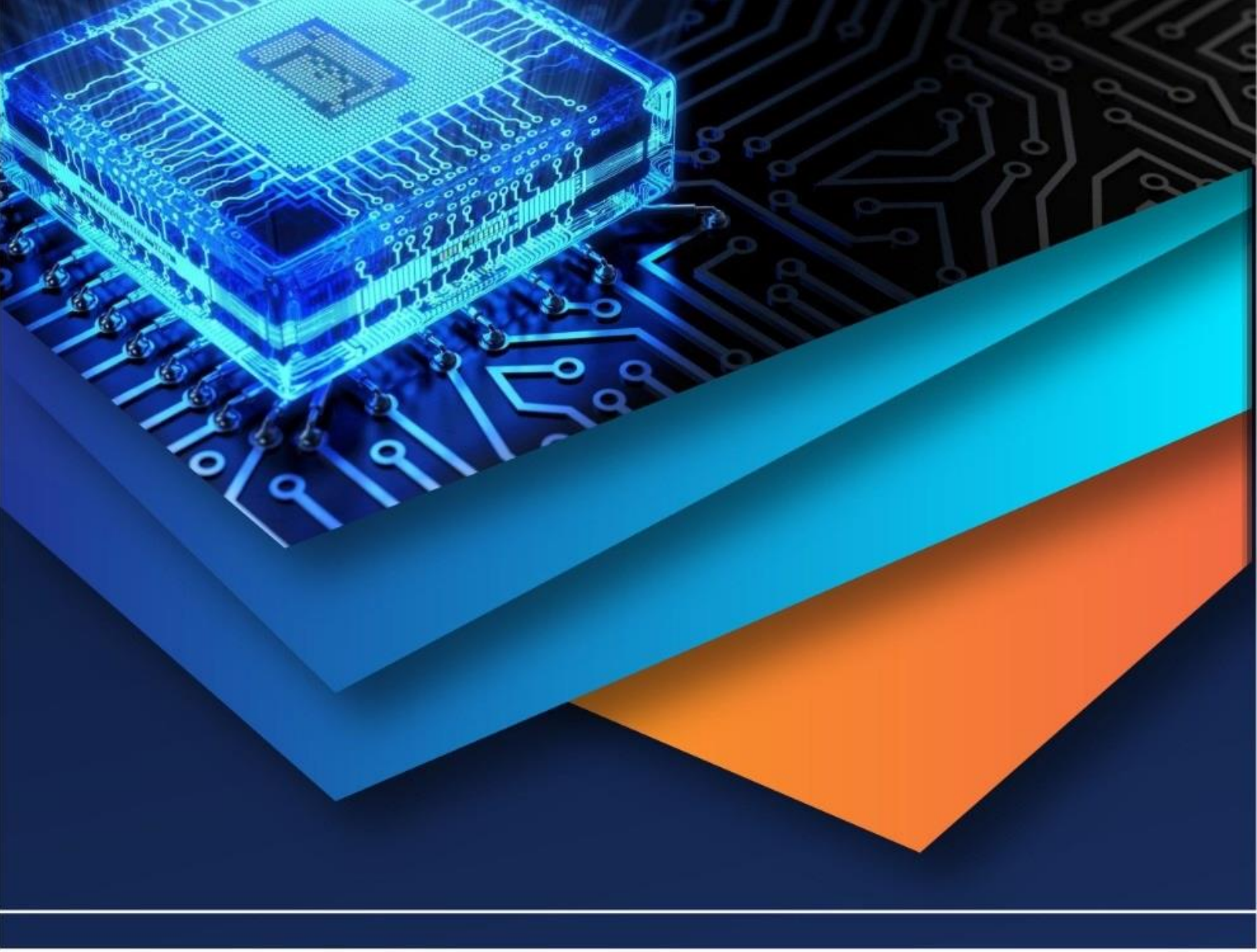
REFERENCES

[1] Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, “Cloud Computing, A Practical approach”
[2] B. Hayes, “Cloud Computing,” Commun. ACM, vol. 51, no. 7, pp. 9–11, Jul. 2008.
[3] V. Vinothina, Dr. R. Sridaran, Dr. Padmavathi Ganapathi, “A Survey on Resource Allocation Strategies in Cloud Computing”, International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012.
[4] Valérie Danielle Justafort, Samuel Pierre, “Performance-Aware Virtual Machine Allocation Approach in an Intercloud Environment”, 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), April 29-May 2012, pp 1-4.
[5] Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu “A Multi-dimensional Resource Allocation Algorithm in Cloud Computing”, Journal of Information & Computational Science, 2012.
[6] Amit Nathani, Sanjay Chaudharya, Gaurav Somani, “Policy based resource allocation in IaaS cloud”, ACM Future generation Computer Systems, Vol. 28, Issue 1, January 2012, pp 94-103.
[7] Sharrukh Zaman, Daniel Grosu, **“**An Online Mechanism for Dynamic VM Provisioning and Allocation in Clouds”, IEEE Fifth International Conference on

©IJRASET: All Rights are Reserved

Cloud Computing, 24-29 June 2012, pp 253-260.
[8] Mayank Mishra, Anwesha Das, Purushottam Kulkarni, Anirudha Sahoo, "Dynamic Resource Management Using Virtual Machine Migrations", IEEE Communications Magazine, Vol 50, Issue 9, September 2012
[9] Gihun Jung, Kwang Mong Sim ,"Agent-based Adaptive Resource Allocation on the Cloud Computing Environment", 2011 International Conference on Parallel Processing Workshops, pp 345-351.
[10]] Xiao-Jun Chen, Jing Zhang, Jun-Huai Li, Xiang Li, "Resource Reconstruction Algorithms for On-demand Allocation in Virtual Computing Resource Pool", Springer, International Journal of Automation and Computing, Vol 9, Issue 2, pp 142-154, DOI: 10.1007/s11633-012-0627-3.
[11] Ilhem Fajjari, Nadjib Aitsaadi, Guy Pujolle, Hubert Zimmermann, "An Optimised Dynamic Resource Allocation Algorithm for Cloud"s Backbone Network", 37th Annual IEEE Conference on Local Computer networks, 22-25 Oct 2012, pp 252-255,
[12] K C Gouda, Radhika T V, Akshatha M, "Priority based resource allocation model for cloud computing", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 1, January 2013
[13] N. Krishnaveni, G. Sivakumar, "Survey on Dynamic Resource Allocation Strategy in Cloud Computing Environment", International Journal of Computer Applications Technology and Research Volume 2– Issue 6, 731 - 737, 2013.
[14] Sukhpal Singh and Inderveer Chana, "Energy based Efficient Resource Scheduling: A Step Towards Green Computing", International Journal of Energy, Information and Communications Vol.5, Issue 2 (2014), pp.35-52
[15] Riddhi Patel, Hitul Patel, Sanjay Patel, "Efficient Resource Allocation in Cloud Computing", International Journal For Technological Research In Engineering Volume 2, Issue 7, March-2015

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)