iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VI **Month of publication:** June 2017

DOI:

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

An Efficient Multiple Request Resource Allocation in Cloud Environment

Priya Beniwal[1], Anil Arora[2]
[1,2] *Department of Computer Science Gateway Institute of Engineering & Technology (GIET), Deenbandhu Chhotu Ram University of Science & Technology (DCRUST), Sonepat*

Abstract: A cloud environment is the popular shareable computing environments where large number of clients/users are connected to the common cloud computing environment to access the resources and the services. The presented work is focused on the concept of effective resource allocation, de-allocation and reallocation in a cloud environment. To present the concept, we have taken a cloud environment with multiple clouds along with multiple virtual machines. The machines used in our work are all of same types i.e., homogenous machines. 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. When the client stops the task then the service allocated to the client is released & same can be reallocated to another client in the waiting. The client can deallocate all or some specific resources according to requirements. After deallocation the client can further allocate the resources. Hence the work provides efficient allocation, de-allocation and reallocation of cloud services.
Keywords: Resource Allocation, Resource Scheduling, Virtual Machines

I. INTRODUCTION

Cloud computing [1] is a technology that uses the internet and central remote servers to maintain data and applications. It provides facility of centralize memory, storage, applications and processing for powerful computing. Through its centralized computing facilities the cloud computing allows users and organizations to use applications and other services without local storage and local location.

Cloud Computing provides us a means by which we can access the applications as utilities, over the Internet. It allows us to create, configure, and customize applications online. It offers online data storage, infrastructure and application. The term Cloud refers to a Network or Internet.

In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over network, i.e., on public networks or on private networks, i.e., WAN, LAN or VPN. Applications such as e-mail, web conferencing, customer relationship management (CRM), all run in cloud [2].

The cloud makes it possible for users to access information from anywhere anytime. It removes the need for users to be in the same location as the hardware that stores data. Cloud computing comprises of 2 components —the front end and the back end. The front end includes client's devices and applications that are required to access cloud. And the back end refers to the cloud itself. The whole cloud is administered by a central server that is used to monitor client's demands.The presented work is focused on the concept of effective resource allocation, de-allocation and reallocation in a cloud environment.

II. RESOURCE ALLOCATION & SCHEDULING

In cloud computing environments, there are two players: cloud providers and cloud users. On one hand, providers hold massive computing resources in their large datacenters and rent resources out to users on a per-usage basis. On the other hand, there are users who have applications with fluctuating loads and lease resources from providers to run their applications. In most cases, the interaction between providers and users occur as shown in Figure 1 below [3].

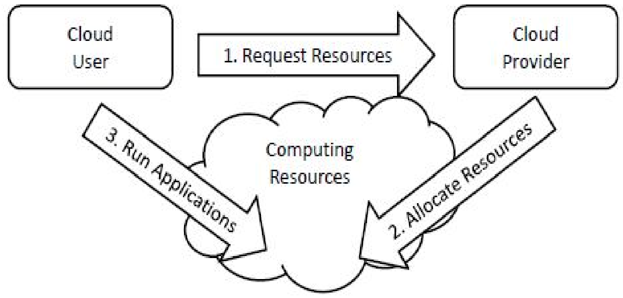


Figure 1: Cloud Usage Scenario

First, a user sends a request for resources to a provider. When the provider receives the request, it looks for resources to satisfy the request and assigns the resources to the requesting user, typically as a form of virtual machines (VMs). Then the user uses the assigned resources to run applications and pays for the resources that are used. When the user is done with the resources, they are returned to the provider [5, 6].

One interesting aspect of the cloud computing environment is that these players are often different parties with their own interests. Typically, the goal of providers is to generate as much revenue as possible with minimum investment. To that end, they might want to squeeze their computing resources; for example, by hosting as many VMs as possible on each machine. In other words, providers want to maximize resource utilization.

Energy efficient Cloud resources allocation consists in identifying and assigning resources to each incoming user request in such a way, that the user requirements are met, that the least possible number of resources is used and that data center energy efficiency is optimized. Figure 2 shows the resource allocation and scheduling scheme for cloud computing.

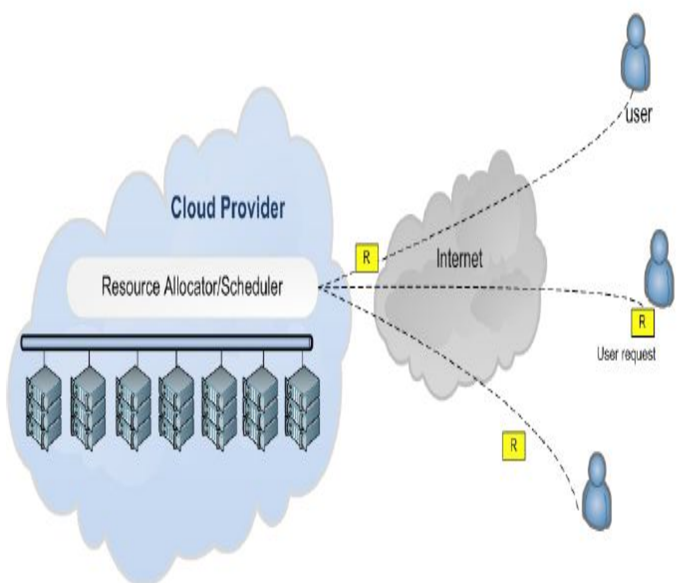


Figure 2: Resource Allocation & scheduling in Cloud Computing

A. Resource Allocation

Resource allocation [8, 9] involves deciding what, how many, where, and when to make the resource available to the user. Typically, users decide the type and amount of the resource containers to request then providers place the requested resource containers onto nodes in their datacenters. To run the application efficiently, the type of resource container need to be well matched to the workload characteristics, and the amount should be sufficient to meet the constraints (e.g., job completion time deadline). In an elastic environment like the Cloud where users can request or return resources dynamically, it is also important to consider when to make such adjustments.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

B. Job Scheduling

Once the resource containers are given to the user, the application makes a scheduling decision. In many cases, the application consists of multiple jobs to which the allocated resources are given. The job scheduler is responsible for assigning preferred resources to a particular job so that the overall computing resources are utilized effectively. The application also has to make sure each job is given adequate amount of resources, or its fair share. Such a scheduling decision becomes more complex if the environment is heterogeneous.

III. PROPOSED MODEL

There are an increasing number of Cloud Services available in the Internet. Cloud services can be a component of a system and different Cloud Servers that would provide different services. In this present work we have defined a multiple cloud environment. Each cloud server is defined with certain limits in terms of memory and the CPU specifications. A middle layer is defined between the cloud servers and the client requests that will perform the allocation of the processes to different clouds in under load and over load conditions. As user requests are performed, some parameters are also defined with each request. These parameters are the number of processor, process time, memory requirement, input output specifications etc. In the general case, the allocation of the processes is performed in a sequential order.

The middle layer will exist between the clouds and the clients. As the request will be performed by the user this request will be accepted by the middle layer and the analysis of the cloud servers is performed by this middle layer. The middle layer is responsible for three main tasks:

A. Scheduling the user request
B. Monitor the cloud servers for its capabilities and to perform the process allocation
C. Deallocate the processes when no services are required by the client.

Figure 3 shows the middle layer along with cloud servers and multiple clients.

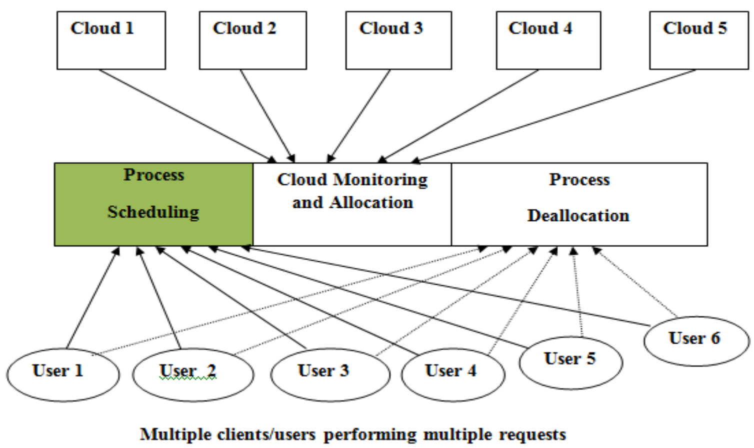


Fig 3: Process Scheduling, Allocation & Deallocation

Whenever user sends application demands requesting for datacenter resources than each and every request is encapsulated in virtual machine (vm) requests with the help of virtualization to the middle layer. For each vm request in domain V, server from the resource pool of server nodes is allocated with the help of vm scheduler. The vm scheduler calculates the utilization of each running server and finds the most suitable server for placing the vm request according the algorithm proposed and updates the utilization after each placement of vm request. It also ensures that upper threshold (maximum number of services) must not be crossed. The queue manager at vm scheduler maintains a queue of allocated vm requests for each server. No other request can placed on the server until these allocated queues of requests keep executing. On deallocation vm scheduler releases the resources occupied by the client [12, 13].

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

IV. PROPOSED ALGORITHM

In this paper, we consider the problem of efficient allocation & deallocation of cloud services. The algorithm for allocation of cloud resources using multiple requests performed by clients is listed below:

```
// allocation process
1. For i ← 1 to req.length do
2. Satisfied Flag ←false;
3. For t ← 1 to start do
4. if(serv[t].ReamainingCap >= req[i] and serv[t].startedStatus = true)
5. Place the request on the server serv[t]
6. satisfiedFlag ← true
7. Update the utilization and remaining capacity of server serv[t]
8. cl[clientId][i].servID ← serv[t].servID
9. cl[clientId][i].numCPU ← noOfCPU
10. cl[clientId][i].allocatedstatus ← true
11. break
 // end of inner for loop
// end of outer for loop
12. If(satisfiedFlag = false)
13. For t ← 1 to MAX
14. if(serv[t].startedstatus = true)
15. continue;
16. else
17. serv[t].startedstatus ← true;
18. serv[t].util ← 0
19. start ← start + 1
20. call allocation process
21. break
// end of if
// end of for
22. if(t = MAX)
23.  Print "Request cannot be satisfied because data center is out of servers"
 // end of allocation
```

A. The main goals of our thesis are

1) Minimize the number of servers used
2) Maximize the resource utilization until upper threshold of utilization is reache
3) Optimize the resource allocation strategy by allocating the released resources to other requests in the waiting queue.

For achieving the above goals upper threshold have been defined. When utilization of all the running reaches upper threshold i.e., all the services/resources of a server are consumed then we start the new server for mapping the requests.

V. RESULTS & DISCUSSIONS

We compare result of proposed algorithm with the work of other previous researchers. The metric used for comparison is the number of servers started with respect to number of virtual machines requested. We compare our work with previous work of load balancing using migration algorithm.

Figure 4 below shows the comparative chart of existing techniques and proposed algorithm for efficient service allocation and deallocation on cloud computing.

Volume 5 Issue VI, June 2017
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

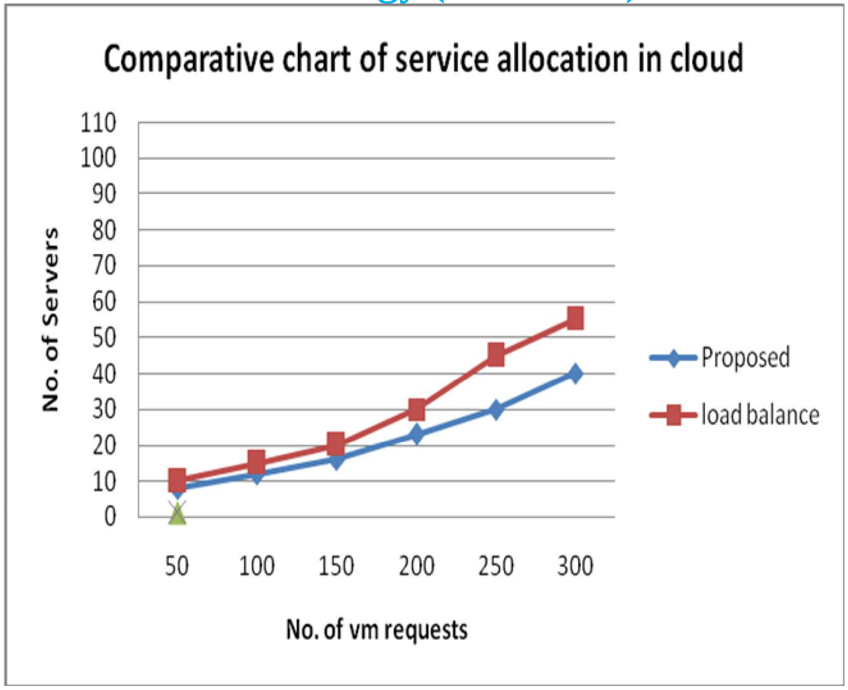


Figure 4: Simulation Results

VI. CONCLUSION

Cloud computing is a fastest growing technology that maintains data and applications using the concept of internet. It does not require installing costly hardware and software resources to store, process and access the data, information and applications by the organization and customers. This technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth. The present work provided efficient allocation, de-allocation and reallocation of cloud services.

REFERENCES

[1] Hayes, "Cloud Computing," Commun. ACM, vol. 51, no. 7, pp. 9–11, Jul. 2008.

[2] P. Mell and T. Grance, "The NIST definition of cloud computing (draft)," NIST special publication, vol. 800, no. 145, p. 7, 2011.

[3] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, "Cloud Computing, A Practical approach"

[4] D. Warneke, O. Kao, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud", IEEE Transactions on Parallel and Distributed Systems, Vol. 22, No. 6, pp 985 - 997, June 2011, DOI: http://doi.ieee computersociety.org/10.1109/TPDS.2011.65.

[5] V. Vinothina, Dr. R. Sridaran, Dr. Padmavathi Ganapathi, "Resource Allocation Strategies in Cloud Computing", International Journal of Advanced Computer Science and Applications [IJACSA], Vol. 3, No.6, 2012. ISSN: 2158-107X (Print), DOI: 10.14569/issn.2156-5570.

[6] Sowmya Koneru, V N Rajesh Uddandi, Satheesh Kavuri, "Resource Allocation Method using Scheduling methods for Parallel Data Processing in Cloud", International Journal of Computer Science and Information Technologies[IJCSIT], Vol. 3(4), 2012, pp 4625 - 4628 4625, ISSN: 0975-9646.

[7] Gihun Jungand, Kwang Mong Sim, "Location-Aware Dynamic Resource Allocation Model for Cloud Computing Environment", International Conference on Information and Computer applications (ICICA 2012), pp 37 - 41, IPCSIT, Vol. 24, IACSIT Press, Singapore.

[8] Thangaraj P, Soundarrajan S, Mythili A, "Resource allocation policy for IaaS in Cloud computing", International Journal of Computer Science and Management Research, Vol 2, Issue 2, pp 1645 - 1649, February 2013, ISSN 2278-733X.

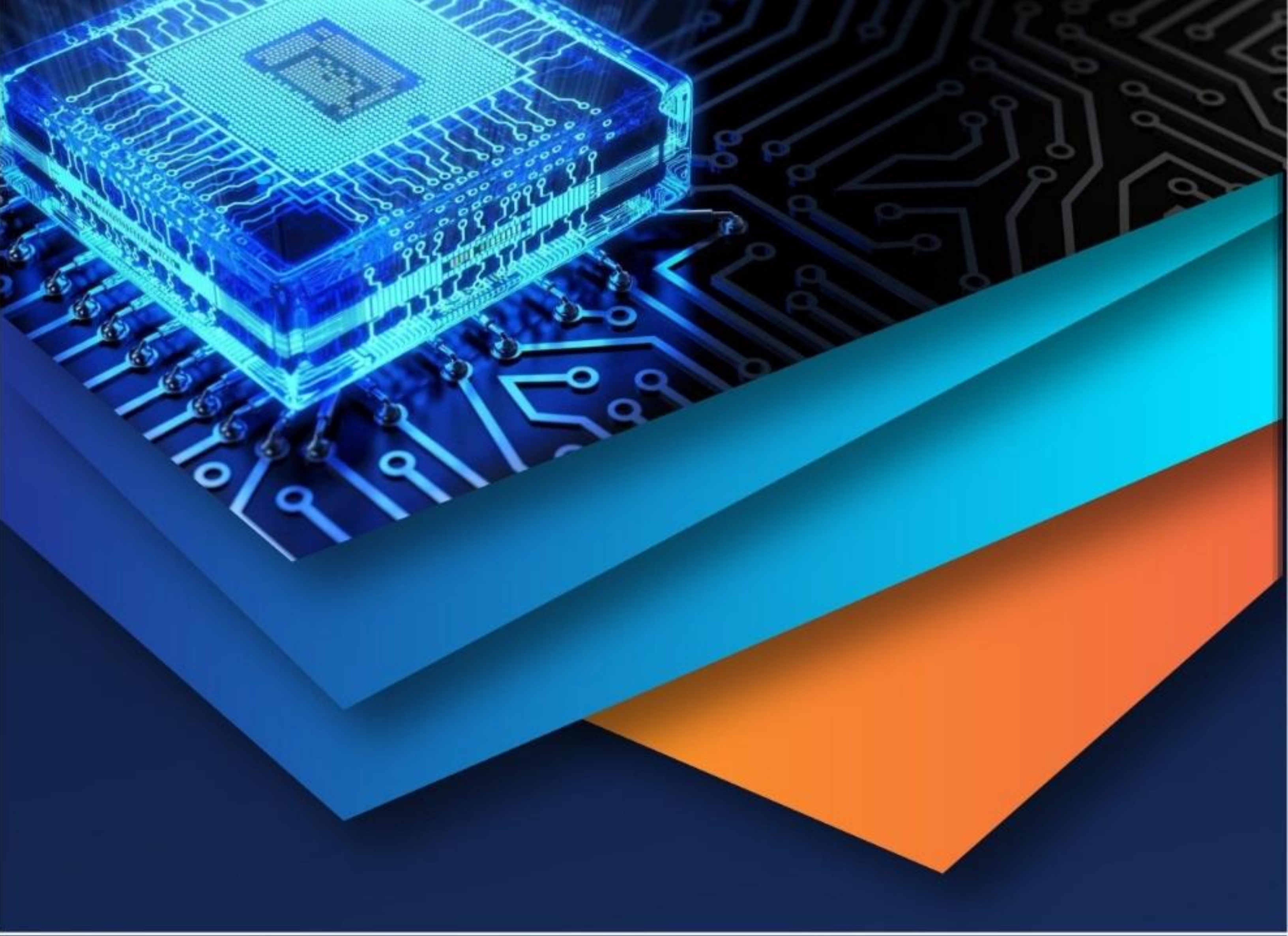
[9] Mohd Hairy Mohamaddiah, Azizol Abdullah, Shamala Subramaniam, and Masnida Hussin, "A Survey on Resource Allocation and Monitoring in Cloud Computing", International Journal of Machine Learning and Computing, Vol. 4, No. 1, February 2014.

[10] Pratik P. Pandya, Hitesh A. Bheda, "Dynamic Resource Allocation Techniques in Cloud Computing", International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 1, January 2014.

[11] Nguyen Trung Hieu, Mario Di Francesco, and Antti YlaJaaski, "A Virtual Machine Placement Algorithm for Balanced Resource Utilization in Cloud Data Centers", 2014 IEEE International Conference on Cloud Computing.

[12] B. Rajasekar, S. K. Manigandan, "An Efficient Resource Allocation Strategies in Cloud Computing", International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 2, February 2015.

[13] Sumita Bose, Jitender Kumar, "An Energy Aware Cloud Load Balancing Technique using Dynamic Placement of Virtualized Resources", Advances in Computer Science and Information Technology (ACSIT) Volume 2, Number 7; April – June, 2015 pp 81 – 86.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY