ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 8(9), 814–821, May 2015* DOI : 10.17485/ijst/2015/v8i9/50180

An Analysis on Efficient Resource Allocation Mechanisms in Cloud Computing

K. Shyamala[1*] and T. Sunitha Rani[2]

[1]Department of Computer Science, Dr. Ambedkar Government College, Chennai-39, India; shyamalakannan_2000@yahoo.com
[2]Department of Computer Science, MOP Vaishnav College for Women, Chennai-34, India; sunithathothadri@gmail.com

Abstract

Cloud computing is considered as a striking computing model which allows for the provisioning of resources on-demand. Cloud computing environment enables multiple users to place request for various cloud services simultaneously. Effective and efficient resource allocation is the challenging task in cloud computing. The efficiency of allocation is measured by optimizing appropriate parameters such as execution time, demand, network delay time, capacity of resources and cost. This paper reveals how effectively resource allocation problem can be addressed in the perspective of cloud service provider and also provides a comparative analysis which helps in selecting parameters to meet the objective function for optimizing the demand to maximize the profit.

Keywords: Demand, Capacity, Cloud Computing, Optimization, Resource Allocation, Service provider

1. Introduction

The IT sector initially emerged with mainframe, progressed to minicomputers, personal computers, client-server, IP networks, mobile devices and now witnessing the emergence of cloud computing. Cloud computing has been gaining momentum because it provides infra-structure, platform and software as a service. IT resources required by organizations for a specific period could be hired easily from cloud. Cloud computing technology is a boon for small and medium size business. Research studies indicate that implementation of Cloud Computing has made remarkable changes in business. Cloud computing is considered as a striking computing model as it allows for the provision of resources on-demand.

The concept of Cloud revolves around resource placement, resource provisioning and resource allocation (Figure 1).

Resource Placement is placing of resources across regions which are geographically distributed to allocate when demanded.

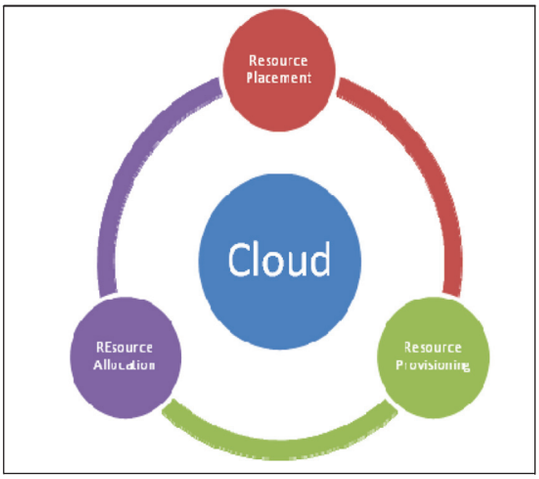

Figure 1. Functioning of cloud service provider.

Resource Provisioning equips and provides users with access to technology resources. It initiates access and alters the state of existing service.

Resource Allocation is a technique used to distribute available resources among competing groups in an

*Author for correspondence

economic way. Resource allocation is a specialized case of automatic scheduling.

Resource Allocation in cloud seems to be a challenging task as it has to satisfy the provider's and user's objectives. The main aim of Resource allocation system is to ensure that the application's requirements are facilitated by provider's infrastructure. A cloud service provider is one who offers IT resources as a service to the consumers either through a private or a public network (Figure 2). The success of cloud computing mainly depends upon the allocation of resources in an efficient and effective way.

Resources in cloud can be provided as measurable utilities by considering parameters like bandwidth, delay, CPU, memory, storage, maximum delay between nodes and topology of the network of all nodes, etc[12,14,17]. Resource optimization refers to resource management of data center in order to improve resource utilization and gain more profit. The objective of a cloud provider is to increase the demand and profit, whereas the objective of the consumer is to maximize the service utilization at a minimized cost. This need has led to the development of efficient resource allocation mechanisms (Figure 3).

2. Related Literature

Many authors have approached the issue of Resource Allocation in various ways[15]. They have proposed algorithms which are efficient in allocation[13]. In this paper,

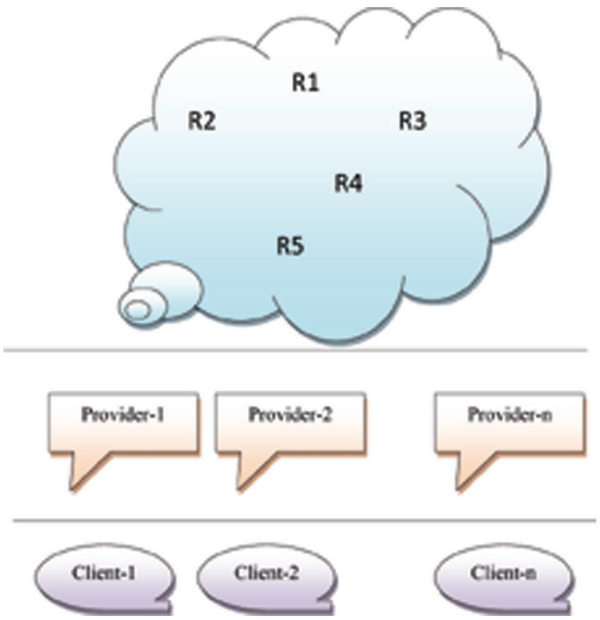

Figure 2. Cloud environment.

Figure 3. Evolution of resource allocation mechanisms.

nine algorithms have been considered to make a comparative study. The algorithm's efficiency is determined with certain parameters. Existing algorithms are analysed in this perspective. For convenience the parameters in these algorithms are denoted as P1, P2…. to P22.

Optimal resource allocation in cloud reveals that all computational needs of an enterprise can be managed by cloud resource reservation. Each resource utilised is associated with a cost. For optimal resource allocation, the number of reserved resources has to be minimized[11,16]. This in turn reduces the operational cost. Resource allocation can be minimized by reusing the resources on tasks that run sequentially[1]. Batch resource allocation has been adopted to study the optimal resource allocation in cloud. Reducing the allocation size reduces the cost. Optimization has been achieved by reducing the number of reserved resources for any request. The algorithm Approx reduces the allocation size by more than 50% when compared to MaxF the base line algorithm. It is also concluded from the evaluation results that 66% of reduction in cost is achieved with Approx.

P1: Resource universe
P2: Batch of number of tasks
P3: Cost

Resource allocation is considered as a great challenge when clients have Service level agreements. The profit depends greatly on meeting these SLAs. SLA based resource allocation problem has been considered and a solution which is very close to optimum has been proposed through a heuristic algorithm[2]. Cluster approach has been adopted to balance the response time as the utility function decreases when the response time increases according to the defined SLA. Local agents are used to parallelize the solution and decrease the response time. Scalability of the solution has been achieved with this approach.

P4: No. of clients with requests for resources.
P5: No. of clusters of resources.

Resource allocation model with processing ability and bandwidth being allocated simultaneously to each

service request on an hourly basis has been developed in [3]. Method I, Method II and Method III have been used to arrive at a conclusion. Method II considers only 'identified resources' in the selection of a center. Best – Fit approach is adopted. Method I uses round robin technique to select the center. If the resources are not enough for the selected center, it selects the next center in a predefined order. Method III has been developed for achieving fair resource allocation among multiple users. As a part of implementation, Multiple Data Centers have been chosen with processing ability and bandwidth. It has been concluded that bandwidth with a minimum delay time should be selected from a group of bandwidths to respond quickly. Method III is used to select a bandwidth with longest network delay time from a group of bandwidths that satisfy the condition on service time. It is for maximizing the possibility to accept requests later, which need a short network delay time. This method can handle more requests by allocating resources in proportion to the expected amount of resources requested by users. Thus Method III can reduce the required amount of resources by up to 20% and also reduces the request loss probability.

P6: Computation time as processing ability size
P7: Network delay time as size of Bandwidth

Optimization of Resource utilization can be first attained at the application level where applications arrive[4]. Choosing of nodes seems to be the primary technique that optimizes the utilization of resources. Resource utilization can be improved by assigning the arriving applications at fewer nodes and at the second level by detecting the nodes that are less utilised. The goal of this paper is to dynamically allocate virtual resources[11,18] among the applications based on the load. When the workload of the application increases, new set of VMs is given for this application. Changing applications are processed at the smallest size of nodes. Less utilized nodes are consolidated on smaller servers by virtual machine migration. The node selection problem is modelled as a binary integer programming problem. The main focus of the paper is to reduce the cost of cloud providers. For this purpose, a frame work consisting of virtual machines has been proposed to encapsulate applications and hence minimize the number of work nodes. Resource allocation problem has been modelled as a binary integer programming problem and an improved MDRA algorithm was proposed to solve this problem. This method provides better time complexity and leads to fine economic benefits. MDRA ensures better utility with lower cost. It also saves resources, increases resource utilization as well as centralized working nodes.

P8: Capacity of resources
P9: nodes

A bidding based approach[5] for resource allocation has been proposed to identify the malicious participants whose goal is to damage the system or increase the own utility. The algorithm helps in resisting damage caused to system by malicious participants[19]. Game theoretical framework has been adopted to solve the problem. Resource Allocation Game (RAG) with two participants and then n participants has been used. It is assumed that each participant is provided with unique registered identity verified by the central authority. Optimal bidding function is generated for each rational participant and a mechanism called DMF is adopted to detect the cheating behaviour.

P10: No. of Participants
P11: Deadline for current stage

Method 3E is an enhancement of joint multiple resource allocation method to handle multiple heterogeneous resource-attributes[6]. Resources are allocated based on the 'key resource attribute'. This attribute is decided by the system and not by the user. This key resource attribute can differ from request to request. In this technique, the heterogeneous resource attributes are allocated simultaneously for specified periods which are fetched for a common pool. This facilitates an efficient allocation when there is more demand for resources. It also helps in load balancing. As compared to Method 3 which considers only one attribute Method 3E is effective on hybrid cloud and can reduce the total resources by up to 30%. Variable P6 and P7 are used in this algorithm, as it also revolves around the same type of parameters.

P6: Computation time for processing ability
P7: Network delay time for bandwidth

Resource allocation algorithm has been modelled based on priority in [7]. An algorithm has been proposed to allocate the demanded resources with minimum wastage and maximum profit. Various parameters like cost, profit, users, time, processor requests, etc., have been considered. The devised technique is based on the threshold of all parameters for a dynamic cloud environment. The resource allocation sequence is decided by the priority algorithm.

P12: Threshold value
P13: Time
P14: Priority
P15: Price
P16: Node
P17: Server name

Turnaround time scheduling[8] has been used to determine the gain function and the loss function of a task using priority. The performance of the algorithm was tested with the pre-emptive and non pre-emptive approach. With this approach, the overall resource utilization can be improved and the processing cost can be reduced. It also increases the efficiency of scheduling algorithm.

P18: No. of tasks
P19: Execution time
P20: Utility density threshold

The problem of resource allocation has been approached with the aim to reduce the power consumption at the data centers. A technique called Bacterial Foraging[9,10] has been used to improve the energy efficiency at the data centers.

P21: Number of VMs & physical machines
P22: Min & Max resource requirement

Advantages of the resource allocation algorithms provided by different authors are summarized in Table 1.

3. Perform ance Analysis of Resource Allocation Algorithms - Cloud Service Provider's Perspective

Table 2 reveals that parameters P2, P4, P11 and P19 can be linked as a network diagram (Figure 4).

In this section, an effort has been taken to make a comparative study of various algorithms to maximize

Table 1. Positive aspects of resource allocation algorithms

Author & Year	Resource Allocation Algorithms	Advantage
Fangzhe Chang, Jennifer Ren, Ramesh Viswanathan 2010.	Optimal Resource Allocation in clouds	Reduces allocation size and cost
Hadi Goudarzi, Massoud Pedram 2011	Maximizing profit in cloud computing via resource allocation	Decreases response time Increases profit
Shin-ichi Kuribayashi 2011	Optimal Joint Multiple Resource Allocation Method for Cloud Computing Environments	Handles more requests Reduces the request loss probability
Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu 2012	A Multi Dimensional Resource Allocation Algorithm in Cloud Computing	Improves resource utilization by assigning applications at fewer nodes Consolidates underutilized nodes Provides better time complexity
Haiyang HU, Zhongjin LI, Hua HU 2012	An Anti-cheating Bidding Approach for Resource Allocation in Cloud Environment	Identifies the malicious participant Detects the cheating behaviour
Yuuki Awano, Shin-ichi Kuribayashi 2012	Proposed Joint Multiple Resource Allocation Method for Cloud Computing Services with Heterogeneous QoS.	Very effective on hybrid cloud Can easily accommodate requests with stringent requirements
K C Gouda, Radhika T V, Akshatha M 2013	Priority Based Resource Allocation Model for Cloud Computing	Allocate the requested resources with minimum wastage Maximum profit
Suhas Yuvaraj Badgujar, Anand Bone 2014	Cloud Resource Allocation as Preemptive Sch edul ing Approach	Overall resource utilization is improved Processing cost is reduced.
Akshat Dhingra, Sanchita Paul 2014	Green Cloud Computing Towards Optimizing Data Centre Resource Allocation	Energy efficiency at data centers is improved Saves Cost even under dynamic work load

Table 2. An analysis of demand based efficient resource allocation algorithms

Author & Year	Resource Allocation Algorithm	P1	P2	P3	P4	P5	P10	P11	P18	P19	P20
Fangzhe Chang, Jennifer Ren, Ramesh Viswanathan 2010.	Optimal Resource Allocation in clouds	✓	✓	✓	X	X	X	X	X	X	X
Hadi Goudarzi, Massoud Pedram 2011	Maximizing profit in cloud computing via resource allocation	X	X	X	✓	✓	X	X	X	X	X
Haiyang HU, Zhongjin LI, Hua HU 2012	An Anti-cheating Bidding Approach for Resource Allocation in Cloud Environment	X	X	X	X	X	✓	✓	X	X	X
Suhas Yuvaraj Badgujar, Anand Bone 2014	Cloud Resource Allocation as Preemptive Scheduling Approach	X	X	X	X	X	X	X	✓	✓	✓

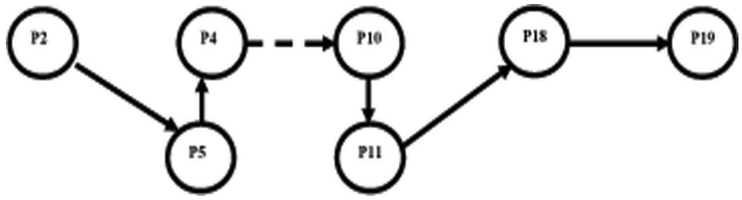

Figure 4. Network of parameters.

the profit, minimize the malicious attempts and also determine the nodes that can be used to allocate resources efficiently for any given request. Efficient Resource Allocation Mechanisms help in improving the Quality of Service (QoS). Increase in QoS, increases demand and profit. Increase in demand also increases the power consumption as data centers consume enormous amount of energy. Cloud environment is prone to malicious attacks as demand increases. Malicious participants not only damage the system but also increase their own utility. Various authors have addressed these issues with Resource Allocation Mechanisms. These algorithms help in overcoming all the problems faced during resource allocation. A performance study on the parameters of several algorithms has been presented to attain the **objective function** which optimizes the demand to maximize the profit in the view of cloud provider's perspective.

Therefore it is concluded from the link that P2→P19 as P2→P5, P5→P4, P4→P10, P10→P11, P11→P19. The parameter batch of tasks can determine the resources required by different clients, the deadlines for the number of tasks and the execution time required for the completion of the tasks. The focus can be shifted more towards the parameters P2 and P19 rather than other parameters. This process helps in reducing the parameters to a minimum number and choosing the appropriate ones to meet the objective function.

Mathematical Perspective

The links between the parameters can be represented with Composition of functions. One parameter can be represented as a function of an another appropriate parameter

P2 = f[P4], P4 = g[P11] and P11 = h[P19]
Let P2 = z, P4 = y, P11 = x and P19 = s
Then z = f(y), y = g(x) and x = h(s)
z(s) = f(g(h(s)))
z(s) = f o g o h(s)
z = f o g o h

The parameters P2, P4, P10 and P18 in Table 3 represent the demand of resources in different forms such as batch of tasks, number of clients, number of participants and number of tasks respectively. P4 and P10 are one and the same. From Table 3, it can be inferred that, high demand for resource provides maximum profit to the provider as shown in Figure 5. The demand can be optimized to maximize the profit.

Table 3. P2→(P4 = P10) → P18

Author & Year	Resource Allocation Algorithm (RAA)	P2	P4	P10	P18
Fangzhe Chang, Jennifer Ren, Ramesh Viswanathan 2010.	Optimal Resource Allocation in clouds	✓	X	X	X
Hadi Goudarzi, Massoud Pedram 2011	Maximizing profit in cloud computing via resource allocation	X	✓	X	X
Haiyang HU, Zhongjin LI, Hua HU 2012	An Anti-cheating Bidding Approach for Resource Allocation in Cloud Environment	X	X	✓	X
Suhas Yuvaraj Badgujar, Anand Bone 2014	Cloud Resource Allocation as Preemptive Scheduling Approach	X	X	X	✓

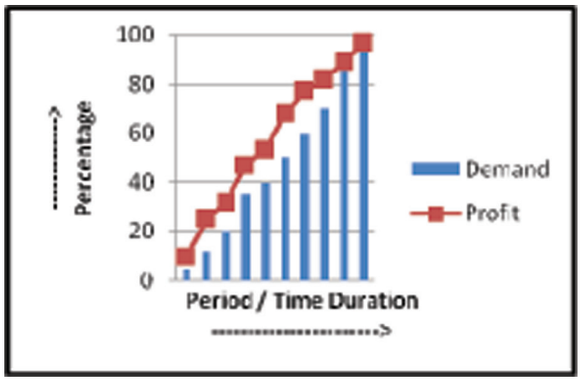

Figure 5. Impact of demand on profit.

Mathematical Perspective of Resource Allocation Algorithms with Reference to Table 3

Matrix representation

$$I_4 = \begin{array}{c} \begin{matrix} P2 & P4 & P10 & P18 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Table 4 reveals that the parameters P6, P7 have been discussed by two different authors in [3] and [6]. P6, P7, P8, P12, P22 represent the capacity of resources in various forms at different nodes that are available for allocation. Making a choice of node for allocation of resources is an important task in the cloud environment. Identifying the node's capacity for resources can determine whether a particular node can be used for allocation of resources for a particular request from the user.

Mathematical Perspective

The Matrix representation of the Table 4 for parameters which represent capacity helps in solving the problem.

$$A = \begin{array}{c} \begin{matrix} P6 & P7 & P8 & P12 & P22 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

As a particular case if the demand and the supply are the same then this can be solved by transportation problem method.

4. Conclusion

The growing need for resources by multiple users has fostered the cloud computing technology. Effective resource allocation strategy facilitates the cloud providers to provide resources in an efficient manner to more number of users. A summary of nine resource allocation methods have been discussed to provide an overall picture of resource allocation. From the above discussion, it can be concluded that efficient resource allocation can optimize cost, time and power consumption. It can also minimize the underutilization of resources, balance load,

Table 4. Analysis of capacity based resources allocation algorithms

Author & Year	Resource Allocation Algorithms	P6	P7	P8	P9	P 12	P 13	P 14	P 15	P 16	P 17	P 21	P 22
Shin-ichi Kuribayashi 2011	Optimal Joint Multiple Resource Allocation Method for Cloud Computing Environments	✓	✓	X	X	X	X	X	X	X	X	X	X
Bo Yin, Ying Wang, Luoming Meng, Xuesong Qiu 2012	A Multi Dimensional Resource Allocation Algorithm in Cloud Computing	X	X	✓	✓	X	X	X	X	X	X	X	X
Yuuki Awano, Shin-ichi Kuribayashi 2012	Proposed Joint Multiple Resource Allocation Method for Cloud Computing Services with Heterogeneous QoS.	✓	✓	X	X	X	X	X	X	X	X	X	X
K C Gouda, Radhika T V, Akshatha M 2013	Priority Based Resource Allocation Model for Cloud Computing	X	X	X	X	✓	✓	✓	✓	✓	✓	X	X
Akshat Dhingra, Sanchita Paul 2014	Green Cloud Computing Towards Optimizing Data Centre Resource Allocation	X	X	X	X	X	X	X	X	X	X	✓	✓

request loss and leasing cost. An efficient resource allocation not only amortizes the administration overhead but also monitors the malicious users causing damage to the system and curbs them from increasing their own utility. This comparative analysis helps in selecting parameters that need much concentration to meet the objective function. It also shows how demand and capacity affect efficient resource allocation. Hence Resource allocation algorithms in future can be designed by considering these two most important parameters.

5. References

1. Chang F, Ren J, Viswanathan R. Optimal resource allocation in clouds. 3rd International Conference on Cloud Computing; 2010.
2. Goudarzi H, Pedram M. Maximizing profit in cloud computing system via resource allocation. ICDCS Workshops; IEEE Computer Society; 2011. p. 1–6.
3. Kuribayashi S-I. Optimal joint multiple resource allocation method for cloud computing environments. Int J Res Rev Comput Sci. 2011; 2(1).
4. Yin B, Wang Y, Meng L, Qiu X. A multi-dimensional resource allocation algorithm in cloud computing. J Inform Comput Sci. 2012; 9(11):3021–8.
5. Hu H, Li Z, Hu H. An anti-cheating bidding approach for resource allocation in cloud computing environments. J Comput Inform Syst. 2012; 8(4):1641–54.
6. Awano Y, Kuribayashi S-I. Proposed joint multiple resource allocation method for cloud computing services with heterogeneous QoS. The third International Conference on Cloud Computing, GRIDs and Virtualization. IARIA; 2012.
7. Gouda KC, Radhika TV, Akshatha M. Priority based resource allocation model for cloud computing. IJSETR. 2013 Jan; 2(1):215–9.
8. Badgujar SY, Bone A. Cloud resource allocation as preemptive scheduling approach. Int J Comput Appl Tech. 2014 Feb; 88(18):14–8.
9. Dhingra A, Paul S. Green cloud computing: to optimizing data centre resource allocation. IJERT. 2014 Feb; 3(2):1037–42.
10. Dhingra A, Paul S. Green cloud: heuristic based BFO technique to optimize resource allocation. Indian Journal of Science and Technology. 2014 May; 7(5):685–91.
11. Sireesha G, Bharathi L. Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. IJERT. 2012 Aug; 1(6).
12. Rajakumari SB, Nalini C. An efficient cost model for data storage with horizontal layout in the cloud. Indian Journal of Science and Technology. 2014 Mar; 7(3S):45–6.

13. Niu D, Li B, An efficient distributed algorithm for resource allocation in large-scale coupled systems. Proceedings of IEEE INFOCOM 2013 Main Conference; 2013 Apr 14-19; Turin, Italy: IEEE; 2013. p. 1501–9.
14. Rajathi A, Saravanan N. A survey on secure storage in cloud computing. Indian Journal of Science and Technology. 2013 Apr; 6(4):4396–401.
15. Patel P, Singh AKr. A survey on resource allocation algorithms in cloud computing environment. Golden Research Thoughts. 2012 Oct; 2(4):6–11.
16. Shelke R, Ranjani R, Dynamic resource allocation in cloud computing. IJERT. 2013 Oct; 2(10).
17. Neela TJ, Saravanan N. Privacy preserving approaches in cloud: a survey. Indian Journal of Science and Technology. 2013 May; 6(5):4531–5.
18. Zhang Q, Gürses E, Boutaba R, xiao J. Dynamic resource allocation for spot markets in clouds, Utility and Cloud Computing (UCC). 2011 Fourth IEEE International Conference.
19. Sharifi AM, Amirgholipour SK, Alirezanejad M, Aski BS, Ghiami M. Availability challenge of cloud system under DDOS attack. Indian Journal of Science and Technology. 2012 Jun; 5(6):2933–7.