
BJ. Hubert Shanthan, A.Stanislas, Dr L.Arockiam
M.Phil Scholar, PhD Scholar, Associate Professor
Dept. of Computer Science, St. Joseph College Trichy, Tamilnadu, India

Abstract
Cloud Computing is a novice technology and it is very popular among IT users by providing its large variety of resources among its consumers. Resource Allocation is one of the pretentious concerns in the cloud. It allocates the resources to its consumer on demand. But in cloud environments, it has only a finite number of resources so it is very difficult for cloud providers to provide all the demanded services to its consumer. In this paper, we present several resource allocation techniques and models which focus on key issues related to the existing resource allocation in the cloud. This paper gives a detailed description about resource allocation strategies and mechanism in the cloud for cloud users and it would benefit for the researchers to explore more about the resource allocation in a cloud environment.

Keywords
Cloud Computing, Resource Allocation, cloud users, cloud service providers

I. Introduction
In an IT Industry, there are many technologies that are developed and are used by the more number of end users. Main Objective of all types of technologies is to satisfy the needs of the customers. In computer we are having technologies such as cluster, grid, etc. In order to overcome this difficulties of advent technologies, the new technology was discovered and it is boon for the IT industry and it is called cloud computing. Cloud is one of highly creditable and most valuable technology in computer science. Cloud computing, has aimed at allowing access to large amounts of computing power in a fully virtualized manner.

Vaquero et al. [1] have stated that clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized Service Level Agreements (SLA).

As Shown in Figure 1, the Cloud service offerings from these vendors can broadly be classified into three major Streams, the Infrastructure as a Service (IaaS), the Platform as a service (PaaS), and the Software as a Service (SaaS).

A. Infrastructure as a Service [IaaS]
In this service Model, it provides machine, storage and network resources that developers can manage their own operating system, application and support resources. It provides a virtual data centre within the cloud. Company system administrators must manage both hardware and software resources. Amazon Elastic Compute Cloud [EC2 Cloud] is a good example for IaaS service model.

B. Platform as a Service [PaaS]
In this service model, it creates the platform for the cloud users to develop new cloud environment and developers can deploy their applications PaaS mainly includes hardware, operating system, development tools and administrative tools. Windows Azure and Google App Engine are the well knows PaaS service providers in the cloud.

C. Software as a Service [SaaS]
In this Service Model, the users can access the service by using the Web portal. The Web browser is used as a user interface tool. The advantage of SaaS is simplicity of integration, and it is mainly used and developed for the end users, Salesforce.com is an example for SaaS service Model.

Cloud computing is mainly emerged for public utilities it has large deployment models with variations in physical location and distribution of resources. Service class of cloud is mainly classified into public cloud, private cloud, hybrid cloud or community cloud.

D. Public Cloud
It is a cloud which is made available to the users in a pay-as-you-go manner to the general public.

E. Private Cloud
It has an internal data centre within organization and it is not available to the public usage.

F. Hybrid Cloud
A cloud is said to be hybrid whenever a private cloud is supplemented with computing capacity of public cloud, it is a combination of the public and private cloud.
II. Scalability

Scalability is a desirable property of a system, which indicates its ability to either handle growing amounts of work in a graceful manner or its ability to improve throughput when additional resources (typically hardware) are added. A system, whose performance improves after adding hardware, proportionally to the capacity added, is said to be a scalable system. There are two types of scalability [8]:

- Horizontal Scalability (Scale out)
- Vertical Scalability (Scale up)

Horizontal Scalability

It refers to the cloud’s ability to connect multiple hardware or software entities. So they make single unit [8].

Vertical Scalability

It refers to the cloud’s ability to extend the capability of existing hardware or software by adding the resources [8]. The key terms which are linked up to scalability are listed below [9].

- Load balancing
- Workload
- Scheduling
- Resource allocation
- Quality of service
- Service Level Agreement

III. Resource Allocation

According to Vinothina et al [10], they have purposed in their survey on resource allocation strategies in cloud describes that Resource allocation is the process of allocation of capable of being used resources that are needed for the applications in the internet.

IV. Guidelines for Resource Allocation

1. Resource Contention

It mainly arises when the two or more cloud enabled applications are requesting resources to the service provider.

2. Scarcity of Resources

It mainly arises when they are limited number of resources that are available in the cloud service provider.

3. Resource Fragmentation

This situation mainly arises whenever the cloud application requests resources to the server and the resources are isolated which means there will be enough resources but not able to allocate to the needed applications.

4. Over provisioning

Cloud service consumers get more resources from the service providers than the demanded resources.

5. Under provisioning

It mainly occurs when the application is assigned to less number of resources than the demand.

V. Merits of Resource Allocation in Cloud

- Service Providers can easily rent their resources, by having a mutual agreement with consumers in terms of SLA (Service Level Agreement)
- They can get their applications and data anywhere in the world or any system Cloud providers can share their resources to overcome the problem of scarcity of resources.
- In cloud effective resource allocation helps to reduce the cost of the consumers by cutting down capital expenses and infrastructure expenses.

Demerits of Resource Allocation In Cloud [11]:

- In Cloud there is no proper security mechanism is involved in the data security and as well in the resources.
- Migration problem occurs while transferring the resources to traditional system to cloud based system
- In public cloud environment there is a chance of external and internal threats in virtual resources allocation.

VI. Dynamic Resource Allocation Architecture in Cloud

They have suggested a new framework for efficient dynamic resource allocation it is called Nephele [12]. It is the first information processing framework to exploit dynamic resource allocation for parallel data processing in cloud and its resulted are interpreted results in Map Reduce framework and it is compared with popular data processing framework called hadoop. Nephele framework is implemented and running in an IaaS [Infrastructure as-a-Service] cloud

VII. Resource Strategies, Mechanisms and Algorithms in Cloud

In a cloud environment there are several resource allocation policies, strategies, and mechanisms are used to meet the guidelines of effective resource allocation. They are as follows

IaaS Resource Provisioning Policy [13]

It mainly focuses on two important aspects such as pricing and profit. They had suggested the policy for increasing resource utilization, by outsourcing unoccupied resources from other service providers. There are two types of VMs, they are spot VM and on demand VM. On demand VM allows customers to pay for compute capacity on hourly basis not for long term commitment. In Spot VM , allows customers to reduce the cost by accepting risk of VM being cancelled for the customers willing pay more for the same resources .The main aim of this paper is to increase resource utilization and increase profit[ ].
Adaptive Resource Allocation Algorithm (ARA) Mechanism

In order to overcome the problem of sudden increase of on-demand requests for Virtual Machines, as it automatically degrades the performance of the application. In order to satisfy peak user demands and meet constraints of Service Level Agreements (SLA) an efficient adaptive resource allocation mechanism schemes are highly demanded in the cloud. So, to overcome, this problem they proposed new burstiness-aware algorithms to balance overflow of amount of the work done across in all computing sites, and thus to enhance overall system performance. They had designed two algorithms such as static and online for greedy and random prediction of bursty workloads.

Skewness Algorithm [15]

It mainly supports green computing, by reducing the usage of physical servers in data centres in the cloud by optimizing the number of servers in usage. They have suggested by introducing the concept of skewness algorithm to measure the mismatched resources in the multi-dimensional resource utilization of a server and they developed set of heuristics that prevent the workload of the systems. By minimizing the skewness we can improve overall application of server resources and prevent the overload of the system and these algorithm is proved to be efficient by trace driven simulation and experimental results. Skewness algorithm achieves overload avoidance and green computing for systems with multi-resource constraints.

Topology Aware Resource Allocation Model (TARA) [16]

This architecture is used for effective utilization of resource allocation in infrastructure as a Service and it mainly uses prediction engine with a light weight stimulator and also a genetic algorithm to find a best solution in a large space Topology Aware Resource Allocation. (TARA) model is tested in server clusters in Map Reduced Benchmarks and the job completion time is up to 50% when compared with application-independent allocation policies. Prediction Engine is mainly composed of Objective Function, Application Description and Available Resources on IaaS.

Automated Negotiation Strategy for resource Allocation [17]

They had developed a mechanism for automated negotiation between consumers and service providers. It automatically negotiates resource contracts between consumers and providers agents. It mainly solves the problem of GENI [16] prototype it is initially operates as a nonprofit platform, it allocates the resources dynamically from multiple providers donate or withdraw the available resources In this Model, there is a negotiation protocol, between seller and buyer. It mainly consists of four types they are Buyer reasoning, Seller Reasoning, Fixed Contract and Tentative contract. Buyer negotiation strategy contains four important factors they are Deadline Pressure, Seller cost and Single Provider.

Service Level Agreement (SLA) based Resource Allocation in SaaS (Software as a Service) [18]

They had proposed resource allocation algorithms for SaaS providers who want to minimize infrastructure cost and SLA violations. The algorithms are designed in a way to ensure that SaaS providers are able to manage the dynamic change of customers, mapping customer requests to infrastructure level parameters and handling heterogeneity of Virtual Machines. Algorithms are ProfitMinVmMinAvailSpace it maximizes profit by reusing VM which as maximum available space. ProfitMinVmMinAvailSpace it maximize the profit by reusing VMs, which have minimum available space.

Dynamic Resource Allocation Model:

Chao-Tung Yang et al [19] had purposed, a dynamic resource allocation model for virtual machine management on cloud, it presents a system which puts into effect of optimization with dynamic resource allocation dealing with virtualization machines on physical machines and results comprises that whenever the virtual machine is loaded heavily automatically the resource is allocated or migrated to another virtual machine without service interruption. The main purpose of this model is to reach best load balance between each physical machine. Open Nebula is a tool used to implement dynamic resource allocation in the cloud.

Priority Based Scheduling Algorithm for Dynamic Resource Allocation

Chandrasekhar S. et al [20] had proposed, an algorithm for preemptive job scheduling and multiple Sla parameters such as memory, network bandwidth, and required CPU time. Resource contention fierce the algorithm for better utilization of the resources. They had also devised load balancing algorithm, to know how to balance the load among the Virtual machines Priority Based Scheduling Algorithm (PBSA) and Cloud min-min scheduling (CMMs) algorithm is compared, and from the simulation results the authors proved that PBSA is more better than CMMs algorithm.

Efficient Framework for Resource Allocation in Cloud

Amman Kumar et al [21] proposed a framework called Efficient Agent based Resource Allocation (EARA) model for resource allocation based on agent computing on SaaS level in Cloud Computing. The powerful self adaptive resource allocation mechanism, is directly implies the capability of the system to serve a large number of users simultaneously. EARA contains five different agents, each equipped with functionality to collect information regarding all resources available in actual cloud deployment. Based on signed SLA agreement, and then replies to the user with appropriate allocation or response code. Resource Computation and Scheduling is done based on Vector Space Model of information Retrieval.
**Deferred Acceptance Algorithm [23]**

Hong Xu et al had proposed architecture, [22] for general resource management in the cloud. It uses the secure matching framework to decouple policies from mechanisms when mapping virtual machines to physical servers. The authors had utilized, both online and offline algorithms for stable match making many to one by with matching between, VM with heterogeneous resource needs to the servers. The authors had suggested this anchor Model as a unifying fabric for resource management in the cloud, where policies are decoupled from the management mechanisms by the stable matching framework. It mainly overcomes the failure in DA (Deferred Acceptance Algorithm). Efficiency and scalability of this new anchor framework are implemented as a prototype and large-scale trace driven simulations.

**Tenant Based Resource Allocation Model**

Javier Espadas et al [24] had suggested a tenant based resource allocation for SaaS application. It mainly solves a problem, of under used and over usage of cloud resources. The Major aim of this work attempts to establish formal measurements for under and over provisioning of virtualized resources in cloud infrastructures, specifically for SaaS platform deployments and proposes a resource allocation model to deploy SaaS applications over cloud computing platforms by taking into account their multitenancy, as one factor and also by creating a cost-effective scalable environment. This model contains three complementary they are Isolation, Load Balancing and Allocation of VM instances. By using these approaches, they suggest this model can solve under utilization and over utilization of resources

**Efficient Framework for Resource Management in the Cloud**

Patrizia Scandurra et al [25] had proposed, a framework that co ordinates layers in the cloud for optimizing resource allocation in cloud by using pareto optimal solution Pareto[20] solution is to minimize multiple objectives by required level of minimum reliability. It has four steps they are, Software as a service (SaaS) framework must use Pareto solution to reduce the cost of software adoption. Platform as a Service (PaaS) framework must the satisfy the demands of SaaS framework by utilizing Pareto solution. Infrastructure as a Service (IaaS) framework must satisfy the (PaaS) demands and it must also minimize the objectives by using Pareto solution. Finally, it must renegotiate the contract between these three layers.

**Rule Based Resource allocation model (RBAM):**

T.R Gopalakrishnan et al [26] had proposed, a Rule Based Resource allocation model (RBRA) along with a Supply-Demand analysis of the resources in a time marching paradigm. In this framework, the analysis shows an improved performance of the system, achieved through efficiently allocating the resources to the jobs submitted to the cloud. The extension or the future work can be a sequence of research on other parameters of Quality of Service (QoS) of a Cloud System.

**VIII. Comparison of Resource Allocation Models and Techniques**

<table>
<thead>
<tr>
<th>Title</th>
<th>Tool</th>
<th>Technique/Algorithm</th>
<th>Future Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated Negotiation with Decommitment for Dynamic Resource Allocation in Cloud computing</td>
<td>Cloud Simulator</td>
<td>Mathematical Model is used to Increase Resource Provisioning In IaaS</td>
<td>Increase Resource Availability</td>
</tr>
<tr>
<td>A Versatile and Efficient Framework for Resource Management in the Cloud</td>
<td>Not Mentioned</td>
<td>Deferred Acceptance Algorithm</td>
<td>Live Virtual Machine is needed</td>
</tr>
<tr>
<td>Topology-Aware Resource Allocation for Data-Intensive Workloads</td>
<td>Map Reduce Light Weight simulator</td>
<td>Genetic Algorithm</td>
<td>Multiple Virtual Machines needed to test this algorithm</td>
</tr>
<tr>
<td>Resource Provisioning Policies to Increase IaaS Provider’s Profit in a Federated Cloud Environment</td>
<td>Cloud Simulator</td>
<td>Mathematical Model is evaluated for Resource provisioning in IaaS</td>
<td>Prediction of availability of resources in IaaS</td>
</tr>
<tr>
<td>Dynamic Resource Allocation Model for Virtual Machine Management on Cloud</td>
<td>Open Nebula</td>
<td>Mathematical Model</td>
<td>It must be tuned for efficient Resource Allocation</td>
</tr>
</tbody>
</table>

Fig. 4: Implementation of EARA Model [21]
<table>
<thead>
<tr>
<th>Resource Allocation Method</th>
<th>Tool/Algorithm</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient resource arbitration and allocation strategies in cloud computing through virtualization</td>
<td>Not Used Rule Based Resource Allocation Model</td>
<td>It can be deployed for different Platforms in cloud</td>
</tr>
<tr>
<td>Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud,</td>
<td>Nephele Simulator Framework Model for dynamic Resource Allocation in Cloud</td>
<td>Underutilization and Job execution can be done automatically</td>
</tr>
<tr>
<td>SLA-based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments,</td>
<td>Not Mentioned ProfminVmMinAvaiSpace, ProfminVmMinAvaiSpace</td>
<td>SLA terms must be changed to increase the efficiency of the algorithm</td>
</tr>
<tr>
<td>A Layered Coordination Framework for Optimizing Resource Allocation in Adapting Cloud-based Applications Layer</td>
<td>Not Used An Coordination Layered Architecture Model</td>
<td>Validation and application is needed to this model</td>
</tr>
<tr>
<td>Priority Based Dynamic resource allocation in Cloud Computing, Not mentioned Load Balancer, Priority Based Scheduling Algorithm (PBSA), Cloud min min Scheduling (CMMS) Algorithm</td>
<td>Prediction VM which will be free earlier and according its capability selecting the task from waiting queue for Execution on that VM.</td>
<td></td>
</tr>
<tr>
<td>An Efficient Framework for Resource Allocation in Cloud Computing, Eucalyptus, Java Agent Development Environment (JADE)[31]</td>
<td>Vector Space Model</td>
<td>Capability of computing pricing policy for energy aware resource management</td>
</tr>
</tbody>
</table>

**IX. Conclusion**

Cloud computing is a boon for this modern technological world. It is one of the most essential technology needed to share and allocate resources in online. Cloud is permanent all time storage environments, it is portable and flexible to save and share the resources. In this paper, we can analyse various resource allocation mechanisms, techniques used in a cloud environment. It would benefit both researchers, academicians to learn more about cloud. It also gives a brief review about resource allocation in the cloud with future enhancements. Many authors have proposed algorithms and methods for dynamic resource allocation in cloud computing.

In summary, an efficient Resource Allocation Technique should meet following criteria’s. Quality of Service (QoS) aware utilization of resources, cost reduction and power reduction / energy reduction. Some of the authors have focused on IaaS based resource allocation with VM scheduling. The ultimate goal of resource allocation in cloud computing is to maximize the profit for cloud providers and to minimize the cost for cloud consumers.

**X. Acknowledgement**

The authors are extremely thankful to all the PhD Scholars, for their invaluable suggestions towards the improvement of the paper. We also thank our beloved parents and family members for their encouragement and support.

**References**


[8] Divyakant Agrawal, Amr El Abbadi, Sudipto Das, and Aaron
J. Elmore, Database Scalability, Elasticity, and Autonomy in the Cloud, Anonymous.


Authors

BJ. Hubert Shanthan is doing his M.Phil in Computer Science in St. Joseph’s College (Autonomous), Thiruchirapalli, Tamil Nadu and India. He received his M.Sc at Loyola college, Chennai. He worked as Teacher in Acharya Siksha Mandir, Pondicherry. He has attended one International Conference on World Congress on Communication and Information Technology in St Joseph’s College Trichy. His research area is Cloud Computing.

A. Stanislas is doing his Ph.D. in Computer Science in St. Joseph’s College (Autonomous), Thiruchirapalli, Tamil Nadu and India. Prior he received his M.CA at Loyola College, Chennai. He worked as Lecturer in St. Xavier’s College, Dumka, and Jharkhand. He has attended many national and international conferences and workshops, presented papers and published a few papers. He has also delivered a few guest lectures on Green and Cloud Computing in Seminars. His Research area includes Networking, Green Computing and Cloud Computing.

Dr. L. Arockiam is working as Associate Professor in the Department of Computer Science, St. Joseph’s College (Autonomous), Thiruchirapalli, Tamil Nadu, and India. He has 25 years of experience in teaching and 16 years of experience in research... He received “Best Academic Researcher of 2012” award from ASDF Techno Forum Group, Pondicherry. He has chaired many technical sessions and delivered invited talks in National and International Conferences. His research interests are: Software Measurement, Cognitive Aspects in Programming, Data Mining, Mobile Networks, and Cloud Computing.