

A SURVEY ON RESOURCE ALLOCATION IN CLOUD COMPUTING

Samah Alnajdi, Maram Dogan, Ebtesam Al-Qahtani

Faculty of Computing and Information Technology, King Abdul-Aziz University
Jeddah, Saudi Arabia

ABSTRACT

Cloud computing is an on-demand service resource which includes applications to data centers on a pay-per-use basis. In order to allocate these resources properly and satisfy users' demands, an efficient and flexible resource allocation mechanism is needed. Due to increasing user demand, the resource allocating process has become more challenging and difficult. One of the main focuses of research scholars is how to develop optimal solutions for this process. In this paper, a literature review on proposed dynamic resource allocation techniques is introduced.

1. INTRODUCTION

Resource management is a major task in cloud computing and in any other computing environments. Cloud computing attempts to provide cheap and easy access to computational resources, which include servers, networks, storage, and, possibly, services. Cloud providers have to efficiently manage, provide, and allocate these resources to provide services to cloud consumers based on service level agreements (SLAs) which both sides agree to prior to the consumer using the services. Therefore, providers must maintain a reliable allocating mechanism in order to satisfy the cloud users' requirements, while stabilizing an appropriate profit margin for themselves. Due to the increasingly high use of the Internet, and thereby cloud services, the typically static allocation and management of resources have become impractical, and the development of dynamic mechanisms have become more appropriate and worth studying. Nonetheless, even these dynamic mechanisms present issues and challenges to be overcome and solutions to be found. Many researchers have tried, and are still trying, to provide the optimal solutions for the resource allocation and management problem in cloud computing environments. In this paper, *we did categorize the existed solutions of resource allocation problem and compare them in each category.*

The contents are organized as follows: section 2 presents a brief background on cloud computing and resource allocation challenges, section 3 discusses the existing dynamic resource allocation techniques and solutions, and, finally, section 4 gives a summary and suggestions for future research directions.

2. BACKGROUND

There are two types of cloud models: the cloud service model and the cloud deployment model.

A. Cloud Computing Services

The cloud offers its services in the form of three types:

1) *Software-as-services (SaaS):*

The cloud provider offers the user SaaS through the Internet to use when he wants without having to install it on his PC. This removes the work of updating and maintenance from the end user, cutting the price of buying the software and removing the need for inconvenient software licenses. The provider and the user will pay-as-used to gain access from any terminal with an operating system and web browser without physical restrictions do this automatically.

2) *Platform-as-a services(PaaS):*

With PaaS, the cloud provides the customer a complex IT environment for the development of applications the consumer uses without having control of the network or server. These services provide the company with a development platform to create their own applications (e.g., Google App Engine).

3) *Infrastructure-as-a-services (IaaS):*

With IaaS, the cloud provides the end user with the storage memory, server, and network on a pay-per-use basis, which reduces the cost and is advantageous to the business. These services provide the company with the ability to add or delete services easily, such as Amazon S3.

B. Cloud Deployment Model

There are various cloud models and classification systems that determine how the services are provided to the end user:

- 1) *Public cloud:* This service is provided to public users and is open to the end user, but the end user cannot control the infrastructure. Examples of public clouds are IBM's Blue Cloud, Sun Cloud, and Google App Engine [1].
- 2) *Private cloud:* This is any service that is provided exclusively to the end user or to a single organization. This offers more security than the public cloud and is more expensive [1].
- 3) *Hybrid cloud:* This cloud is composed of two or more different clouds which may be public or private. More important data can be kept in the private cloud, and the other types of data can be kept in the public cloud. This hybrid arrangement is more secure and less expensive than the purely private cloud [1].
- 4) *Community cloud:* In this cloud, the services are provided exclusively to a group of persons or companies who share the same interests.

C. Resource Allocation In Cloud Computing

In cloud computing, resource allocation (RA) is a field that is taken into account in many computing areas such as datacenter management, operating systems, and grid computing. RA deals with the division of available resources between cloud users and applications in an economic and effective way. It is one of the challenging tasks in cloud computing based on the IaaS. Furthermore, RA for IaaS in cloud computing provides several benefits: it is cost effective because users do not need to install and update hardware or software to access the applications, its flexibility allows access applications and data on any system in the world, and there are no limitations of the medium or usage site.

In addition, there are two major processes of RA via cloud computing

D. Static Allocation

Static Allocation schemes: assign fixed resources to the cloud user or application. In this case, the cloud user should know the number of resource instances needed for the application and what resources are requested and should aim to confirm the application's peak load requests. But the limitation for static allocation is usually affected by the over-utilization or under-utilization of computing resources based on the normal workload of the application. This is not cost-effective and is related to insufficient use of the resource during off-peak periods.

E. Dynamic Allocation

Dynamic Allocation schemes provide cloud resources on the fly when the cloud user or application is requested, specifically to avoid over-utilization and under-utilization of resources. A possible drawback when needed resources are requested on the fly is that they might not be accessible. Thus, the service supplier must allocate resources from different participating cloud data centers [2].

Resource allocation strategy (RAS) is related to combining cloud provider functions for utilizing and assigning scarce resources within the boundaries of the cloud system in order to suit the demand of the cloud application.

As cloud computing has its characteristics, the RAS should avoid the following situations as much as possible:

- 1) *Resource contention*: This situation occurs when multiple users and applications attempt to allocate the same resource simultaneously.
- 2) *Resource fragmentation*: This occurs when applications cannot allocate resources due to isolated resources being small items.
- 3) *Scarcity*: This occurs when multiple applications' requirements for the resources are high and there are limited resources, for example, requests for memory, I/O devices, CPUs, and the techniques that should serve that demand.
- 4) *Over provisioning*: This occurs when the users and applications obtain more resources than those that are requested to fit the quality of service (QoS) requirements.
- 5) *Under provisioning*: This occurs when the users and applications obtain fewer resources than those requested to fit the QoS requirements [2].

From the perspective of cloud users, RA should be achieved at a lower cost and in as little time as possible. However, for the cloud provider, it is impractical to predict the dynamic of user demands, nature of users, and application demands. Therefore, resource diversity, limited resources, locality restrictions, dynamic nature of resource requests, and environmental necessities require an efficient and dynamic RAS that is suitable for cloud environments. Since the dynamic and uncertainty of resource demand and supply are unpredictable, different strategies for dynamic resource allocation are suggested. This research presents different RAS that are utilized in cloud environments.

3. MODELS FOR DYNAMIC RA IN CLOUD COMPUTING

The quality and cost of the services in cloud computing are based on their RA process, and the resource provider should assign the resource to the clients in an optimal way. Yet, there are many RA techniques and proposed models that are used in the area of cloud computing. We are going to present some of the dynamic RA techniques, classifying them based on the main strategy that they use to allocate resources. The result of any optimal RAS must consider certain parameters

such as latency, throughput, and response time. In this paper, we address some of the commonly used strategies: service level agreement-based, utility-based, market-based, and priority-based strategies.

A. SLA-Based Dynamic RA Models

The SLA is an agreement that specifies the QoS between the service provider and the service consumer, and it includes the service price with the level of QoS adjusted by the price of the service [3]. Most of the RA models in cloud computing environments focus on satisfying the agreed specifications of the SLA for the cloud user. Some other models' strategies in RA focus on achieving the objectives of the cloud provider, which could negatively affect some of the users' requirements and the level of QoS provided. One such model is proposed by Popovici et al.[4]. They investigated the QoS parameters such as offered load and price on the SaaS provider's side but did not consider the user's side.

For a multi-cloud environment, Soodeh Farokhi [5] developed a framework for resource allocation in a multi-cloud system from the perspective of the SaaS level, agreed SLA, and service provider conditions. The proposed model utilizes a selection engine, construction engine, and SLA violation detection and monitoring with the use of the service provider's QoS parameters.

There are few models that focus on both the cloud provider and consumer perspectives. One such model was proposed by Wu et al.[6], and it focuses on the QoS parameters of both the SaaS provider and the consumer through proposed RA algorithms aimed at minimizing SLA violations and infrastructure costs, as well as controlling the dynamic change of customers, by specifying customer demands to the infrastructure level aspect and managing dissimilarity of virtual machines (VMs). The two proposed algorithms perform well by decreasing costs by about 50% with fewer VMs and optimizing the means to avoid SLA violations. Also, another work proposed by Lee et al. [7] addresses the issue of profit basis on service request scheduling in cloud computing by taking into account the purposes of both the consumers and service providers.

Zhu et al. [8] proposed architecture to solve virtualized RA problems for multi-tier applications. Their model improved overall performance, reduced the cost, maximized the profit for SaaS providers, and aimed to meet the user's performance requirements.

EARA [9] is an efficient agent-based resource allocation framework designed by Kumar et al. EARA uses agent computing as several agents collect available resource information to allocate it for user requests based on the signed SLA agreement and, therefore, balancing the performance and controlling the cost; however, this model only considers the SaaS level of cloud environments.

Buyya et al. [10] proposed a market-oriented RA scheme by integrating both the customer-driven service management and provider-driven risk management to promote SLA-based RA. Nevertheless, [10] it requires a market maker and a market registry to bring the consumer and providers together and to publish the cloud services and discover their providers.

Pawar et al. [8], proposed a priority-based allocation model by considering various SLA parameters such as bandwidth, memory, and execution time. Using a preemption mechanism associated with the benefit of parallel processing, their model improved utilization, especially in a resource contention situation.

B. Market-Based Dynamic RA Models

Using the market economy to manage RA has been studied extensively in the past, and several researchers have investigated the economic aspects of cloud computing from different points of view. To deal with dynamically fluctuating resource demands, market-driven RA has been proposed, and it has been implemented by many public IaaS providers such as Amazon EC2 [12]. In this environment, the cloud provider could follow the commodity market approach or auction-based mechanisms, with the main goal of achieving maximum revenue while minimizing the cost.

Zaman et al. [13] proposed a combinatorial auction-based mechanism for resource management in clouds. Their algorithm is based on the users' valuation concept, which is that each user desires a specific bundle of VM instances and bids on it. It represents efficient allocation and high profits for the provider, but it still allows users to pay a minimum cost [13].

In addition, Zhang et. al. [14] introduced a mechanism for spot markets, which addressed the problem of allocating resources for different VM types in Amazon spot instances using the model predictive control (MPC) algorithm. The proposed model [14] insures high revenue for the providers over time by changing the price depending on the level of demand, meeting customer expectations, and minimizing energy consumption; however, future or forward markets are not included in the model. From there, Fujiwara et al. [15] improved the market-based allocation by developing a double-sided combinatorial auction-based model, which allows both the users and providers to trade their current and future services in the spot and forward market.

There are other market-based models, such as the one proposed by You et al. [16], the RAS-M model, that define the equilibrium theory and use the GA-based price adjusted algorithm. Their model [16] is efficient and improves utilization and profit; nevertheless, it only considers the physical level of the cloud environment and is limited to CPU resources only.

Lin et al. [17] proposed an RA model for clouds based on a sealed-bid auction, where the users submit their bids to the cloud service providers who collect the bids and determine the price. This mechanism provides efficient allocation of resources, but no profit maximization is ensured due to its truth telling property.

C. Utilization-Based RA Models

In order to overcome the under-utilization of resources that results from allocating fixed resources to applications and services, the main approach of methods that fall into this category is to dynamically manage VMs to maximize utilization of resources and minimize costs. A model that adjusts the VMs according to an application's actual needs has been developed by Lin et al. [18], and it is based on the threshold. The proposed algorithm uses monitoring and predicting the needs of cloud applications, which leads to increased resource utilization and decreased costs.

Yin et al. [19] focus on RA at the application level. The authors proposed a multi-dimensional RA (MDRA) schema using a framework of application allocation to minimize the cost of the data center by assigning small-sized nodes to the processors of users' programs.

Simulated annealing-based RA has been performed by Pandit et al. [20] using a bin packing algorithm with multi-parameters to decrease the unallocated part of resource parameters. The proposed model has improved utilization of cloud resources at the multi-level in the cloud system and has decreased the cost.

The topology aware resource allocation (TARA) schema was introduced by Lee et al [7]. This schema deals with unconcerned of the hosted application's demands for IaaS system. They proposed a prediction engine and genetic algorithm-based search for minimum latency and proper confidence. The authors showed that the TARA experiment could result in a decrease in job compilation time by up to 59%.

Li and Qiu [21] suggested an adaptive RA algorithm for cloud a computing model with preemptable jobs. The authors defined two algorithms, adaptive list scheduling (ALS) and adaptive min-min scheduling (AMMS), which apply to task scheduling [21], and they proved that the proposed algorithm is effective and efficient for use with resources. Younge et al.[22] proposed a resource management model to improve job scheduling using a green cloud framework. Their model maximizes utilization as it reduces performance overload and energy consumption and provides an overall efficiency for data-centers in the cloud computing environment. An online optimization for scheduling preemptable tasks on IaaS cloud system models was proposed by J. Li et al.[23], using a min-min algorithm, and the scheduling process is based on feedback information about actual task execution. Their solution proved to reduce RA execution time and energy consumption; however, the strength of their model is based on the reliability of the feedback information.

Rammohan and Baburaj [24] developed an RA in a cloud computing model based on the interference-aware resource allocation (IARA) technique, providing optimal energy consumption, and it is practical for a resource-constrained environment and supports special hardware.

There are a number of other models. Some of these models mainly focus on improving resource utilization, such as the ones by Minarolli et al. [25], Buyya et al. [26], and Ergu et al. [27].

4. RESEARCH CHALLENGES

The research on RA in cloud systems is still at an early stage. Several existing issues have not been fully addressed while new challenges keep emerging. Some of the challenging research issues are given as follows:

- 1) *Migration of VM:* This migration problem occurs due to the need of the user to switch to another provider in order to get better data storage.
- 2) *Control:* There often is a lack of control mechanism over the resources as they are rented by the users from the remote server.
- 3) *Energy Efficiency:* Due to the emergence of huge data centers that have various computing operations, there is a need for energy efficient allocation. These centers lead to the release of large quantities of carbon emission.
- 4) *The Scheduling of Parallel Jobs:* Parallel jobs in the field of computing increase the job that is serve. There are two types of jobs: dependent and independent. The first type must be done very carefully. These jobs include communication issues. Independent jobs can be performed using several VMs at the same time.
- 5) *Reduction of Cost and Maximizing of Resources:* It is important to handle the constraints that must be met in the allocation of resources in terms of cloud operating costs and to maximize the use of all resources. In other words, the service provider must provide users with low-cost services.
- 6) *Maintaining High Availability:* The availability of resources in the cloud must be guaranteed in case there is a job with long running computations that can take many hours. Thus, there is a need for some techniques to automatically handle any interruption or unavailability in resources and switch the jobs to an available resource. Moreover,

these techniques should support the transparency property by which the user cannot observe the unavailability or any failure problem.

- 7) *Elasticity*: In the cloud, elasticity refers to what extent resource requirements can be handled dynamically. Demand for resources may increase over time, and the cloud should automatically detect the size of these demands to be met and the necessary resources required to meet them.

5. FINDINGS AND FUTURE RESEARCH DIRECTIONS

A great deal of research has been done, and many solutions have been presented in the area of cloud computing in respect to the RA problem; however, there are still some issues and challenges that need further research, and an optimal solution that is practical for most cloud environments has still not been found.\

Some of the findings based on our literature of previous studies are as follows:

- 1) There is a need for reducing the user's SLA violations when maximizing the RA utilization because most of the models affect the QoS in order to reduce the cost and keep high utility algorithms.
- 2) There is a need for an RA framework that is practical for different cloud environments in order to ease the complexity of allocation in heterogeneous clouds.
- 3) There is a need for RA to minimize the cost for cloud consumers and maximize the profit for cloud providers. It is very important for cloud providers to offer efficient utilization and management of the limited amount of resources available.
- 4) There is a need to consider load balance in cloud resources and scheduling the workload in optimal ways in order to satisfy the QoS requirements of users and maximize profit by enhancing the use of resources.

The future direction of research into resource allocation in cloud computing should address each of the above-named challenges and try to implement best practices models.

Table 1 presents the comparative study between various resource allocation techniques in cloud environment and their merits and demerits.

CONCLUSION

Cloud computing technology is increasingly being used in enterprises and business markets. Subsequently, an effective RAS is required for achieving user satisfaction and maximizing the profit for cloud service providers.

This paper provides a survey of some of the models and solutions for the RA problem in the cloud computing environment. These models are classified based on their strategies, and a discussion of their strengths and limitations is supported by a comparison table. Finally, research directions and findings from our literature review are included, and hopefully they will help in motivating future research to determine optimal RA solutions for cloud environments.

ACKNOWLEDGMENT

First of all, we would like to express our sincere thanks to Allah, then we would like to express our gratitude to D. Muhammed Monowar for the continuous support us and for his patience, motivation, and immense knowledge.

TABLE I

Reference	Methodology Used	Strengths	Limitations
[7]	Multi-dimensional resource allocation (MDRA) schema	The proposed algorithm improves resource utilization and reduces cost of data center.	It is inefficient at saving power when user demand increases in a long-run situation.
[20]	Proposed bin packing algorithm use for simulated annealing	It helps to reduce the costs and solves the multi-layer resource allocation problems.	It lacks in handling dynamic resource requests. It relays request availability and restricts bin size.
U4	Topology aware resource allocation (TARA) schema based in IaaS cloud	This minimizes the job completion time of application.	The functions used in this schema are not considered complicated objective functions related to power, infrastructure costs, and reliability.
[23]	Min-min and list scheduling algorithm in pre-emptible tasks	This provides significant improvement in the fierce resource contention, and improves load balancing. This algorithm have a shorter average execution time. It reduces energy consumption.	The algorithm's predictions may not be precise and can then lead to over provisioning or under provisioning. There is a risk whether the feedback information is reliable enough for the scheduler to rely on.
[5]	Proposed framework assists SaaS providers in multi-cloud system	It helps to find proper infrastructure resources which best satisfy the user requirements while monitoring SLA and detection violation.	It does not consider the latency and data traffic included in selected services in a multi-cloud at run time.
[6]	Proposed resource allocation algorithms for SaaS providers	The proposed algorithms reduce the SaaS provider's cost and the number of SLA violations. It manages the dynamic change of customers.	The total profit of the algorithms needs to be enhanced to improve customer satisfaction levels. It should monitor penalty limitations by considering system failures.
[9]	Efficient agent based resource allocation (EARA) framework	This provides appropriate resources for the users. It serves large number of users.	If any agent fails, the system does not function properly.
[11]	Priority and heuristic scheduling algorithms	This provides effective utilization of cloud resources to meet the SLA objective.	The method cannot predict VMs which will be free earlier and based on its capability selecting the task from waiting for the queue for execution on that VM.
[28]	Robust cloud resource provisioning (RCRP) algorithm used for reservation plan in cloud system	The algorithm minimizes the total provisioning costs. It is implemented on the agent side as opposed to the cloud broker side.	It strictly considers a reservation plan for resource provisioning.
[29]	Proposed resource allocation algorithm	The algorithm maximizes the overall profit of the	It's requirements of VMs.

	for infrastructure providers	cloud provider. It minimizes infrastructure costs and SLA violations.	
[24]	Interference-aware resource allocation (IARA) technique based model	It provides optimal energy consumption. It is practical for a resource-constrained environment. It supports special hardware.	Minimum performance and QoS.
[18]	A threshold-based dynamic allocation of system resources using CloudSim toolkit.	This maximizes resource utilization and minimizes costs.	It lacks of considering the overhead of physical resources.
[21]	Adaptive list scheduling (ALS) and adaptive min-min scheduling	The scheme maximize the utilization and benefit for the cloud provider.	Minimizes the control over SLA's violations for users.
[22]	A green cloud framework -based scheme	It maximizes the utilization. It reduces overload and energy consumption. It provides an overall efficiency for data-centers.	The setup phase of the system is high in cost.
[4]	Economics-based approach using utility function	It investigated the QoS parameters such as offered load and price on the SaaS provider's side.	It does not consider SLA's violations toward service customers.
[8]	Agent-based algorithm	It consider allocation for multi-tier applications. It improves the overall performance and reduces the cost. It reduces SLA's violations for both user and provider.	The algorithms is only efficient for SaaS level environment.
[10]	Market-oriented mechanism	It provides both customer-driven and provider-driven resource management.	Many requirements, i.e. market maker and registry maker.
[13]	Combinatorial auction mechanism	It provides efficient allocation and high profit for the provider. Allows users to pay a minimum cost.	Not ensured profit for services users. Users can not trade their services with the providers.
[14]	Model Predictive Control (MPC) algorithm	It provides revenue maximization Allows meeting customer expectations. It provides energy consumption minimization.	Lack future profit prediction because of unconsidering of forward markets.
[15]	Double-sided combinatorial auction-based model	Allows both the users and providers to trade their current and future services in the spot and forward market.	The high requirements of the model. Complex management of resources.
[16]	Equilibrium theory and GA-based price	This provides efficient utilization.	It only considers the physical level of the cloud environment

	adjusted algorithm.	It maximizes the profit for cloud service providers.	and is limited to CPU resources only.
[17]	Sealed-bid auction	It provides efficient allocation of resources.	Not ensured profit maximization due to its truth telling property.

REFERENCES

- [1] Abdulkader, S. J., & Abualkishik, A. M. Cloud Computing And E-Commerce In Small And Medium Enterprises (Sme's): The Benefits, Challenges
- [2] K Delhi Babu, D.Giridhar Kumar "Allocation Strategies Of Virtual Resources In Cloud Computing Networks" Journal Of Engineering Research And Applications,201,Pp.51-55.
- [3] Son, Seokho, Gihun Jung, And Sung Chan Jun. "An Sla-Based Cloud Computing That Facilitates Resource Allocation In The Distributed Data Centers Of A Cloud Provider." The Journal Of Supercomputing 64.2 (2013): 606-637.
- [4] I. Popovici, And J. Wiles, {\Em "Proitable Services In An Uncertain World"} In Proceeding Of The18th Conference On Supercomputing (Sc 2005), Seattle, Wa.
- [5] S. Farokhi, "Towards An Sla-Based Service Allocation In Multi-Cloud Environments," Cluster, Cloud And Grid Computing (Ccgrid), 2014 14th Ieee/Acm International Symposium On, Chicago, Il, 2014, Pp. 591-594.
- [6] Wu, Linlin, Saurabh Kumar Garg, And Rajkumar Buyya.{\Em"Sla-Based Resource Allocation For Software As A Service Provider (Saas) In Cloud Computing Environments." }Cluster, Cloud And Grid Computing (Ccgrid), 2011 11th Ieee/Acm International Symposium On. Ieee, 2011.
- [7] Lee, Gunho, Et Al. "Topology-Aware Resource Allocation For Data-Intensive Workloads." Proceedings Of The First Acm Asia-Pacific Workshop On Workshop On Systems. Acm, 2010.
- [8] Kumar, Ajit, Emmanuel S. Pilli, And R. C. Joshi. "An Efficient Framework For Resource Allocation In Cloud Computing. Computing, Communications And Networking Technologies (Iccnt), 2013 Fourth International Conference On. Ieee, 2013.
- [9] Jiao, Jianxin Roger, Xiao You, And Arun Kumar. "An Agent-Based Framework For Collaborative Negotiation In The Global Manufacturing Supply Chain Network.Robotics And Computer-Integrated Manufacturing 22.3 (2006): 239-255.
- [10] Garg, Saurabh Kumar, And Rajkumar Buyya. "Market-Oriented Resource Management And Scheduling: A Taxonomy And Survey. Cooperative Networking (2011): 277-306.
- [11] Pawar, Chandrashekhar S., And Rajnikant B. Wagh."Priority Based Dynamic Resource Allocation In Cloud Computing With Modified Waiting Queue." Intelligent Systems And Signal Processing (Issp), 2013 International Conference On. Ieee, 2013.
- [12] Amazon, E. C. Amazon Elastic Compute Cloud (Amazon Ec2). Amazon Elastic Compute Cloud (Amazon Ec2), 2010.
- [13] Zaman, Safdar; Grosu, Daniel. "A Combinatorial Auction-Based Mechanism For Dynamic Vm Provisioning And Allocation In Clouds" Cloud Computing, Ieee Transactions On, 2013, 1.2: 129-141.
- [14] Zhang, Qi, Quanyan Zhu, And Raouf Boutaba. "Dynamic Resource Allocation For Spot Markets In Cloud Computing Environments" Utility And Cloud Computing (Ucc), 2011 Fourth Ieee International Conference On. Ieee, 2011.
- [15] Fujiwara, Ikki, Kento Aida, And Isao Ono. "Market-Based Resource Allocation For Distributed Computing" Vol. 2009. Ipsj Sig Technical Report, 2009.
- [16] You, Xindong, Et Al. "Ras-M: Resource Allocation Strategy Based On Market Mechanism In Cloud Computing" In: 2009 Fourth Chinagrid Annual Conference. Ieee, 2009. P. 256-263.
- [17] Lin, Wei-Yu, Guan-Yu Lin, And Hung-Yu Wei. "Dynamic Auction Mechanism For Cloud Resource Allocation." Cluster, Cloud And Grid Computing (Ccgrid), 2010 10th Ieee/Acm International Conference On. Ieee, 2010.
- [18] Lin, Weiwei, James Z. Wang, Chen Liang, And Deyu "A Threshold-Based Dynamic Resource Allocation Scheme For Cloud Computing", Procedia Engineering, 2011.
- [19] Yin, Bo, Et Al. "A Multi-Dimensional Resource Allocation Algorithm In Cloud Computing" Journal Of Information And Computational Science (2012):3021-3028.

- [20] D. Pandit, S. Chattopadhyay, M. Chattopadhyay And N. Chaki, "Resource Allocation In Cloud Using Simulated Annealing," Applications And Innovations In Mobile Computing (Aimoc), 2014, Kolkata, 2014, Pp. 21-27.
- [21] Li, Jiayin, Et Al. "Adaptive Resource Allocation For Preemptable Jobs In Cloud Systems." Intelligent Systems Design And Applications (Isda), 2010 10th International Conference On. Ieee, 2010.
- [22] Younge, Andrew J., Et Al. "Efficient Resource Management For Cloud Computing Environments." Green Computing Conference, 2010 International. Ieee, 2010.
- [23] Li, Jiayin, Et Al. "Online Optimization For Scheduling Preemptable Tasks On IaaS Cloud Systems." Journal Of Parallel And Distributed Computing 72.5 (2012): 666-677.
- [24] Rammohan, N. R., And E. Baburaj. "Resource Allocation Using Interference Aware Technique In Cloud Computing Environment." International Journal Of Digital Content Technology And Its Applications 8.1 (2014): 35.
- [25] Minarolli, Dorian, And Bernd Freisleben. "Utility-Based Resource Allocation For Virtual Machines In Cloud Computing." Computers And Communications (Iscc), 2011 Ieee Symposium On. Ieee, 2011.
- [26] Buyya, Rajkumar, Anton Beloglazov, And Jemal Abawajy. "Energy-Efficient Management Of Data Center Resources For Cloud Computing: A Vision, Architectural Elements, And Open Challenges." Arxiv Preprint Arxiv:1006.0308 (2010).
- [27] Ergu, Daji, Et Al. "The Analytic Hierarchy Process: Task Scheduling And Resource Allocation In Cloud Computing Environment." The Journal Of Supercomputing 64.3 (2013): 835-848.
- [28] Vishnupriya, S., P. Saranya, And P. Suganya. "Effective Management Of Resource Allocation And Provisioning Cost Using Virtualization In Cloud." Advanced Communication Control And Computing Technologies (Icacct), 2014 International Conference On. Ieee, 2014.
- [29] Yuan, Haitao, Et Al. "Sla-Based Virtualized Resource Allocation For Multi-Tier Web Application In Cloud Simulation Environment." Industrial Engineering And Engineering Management (Ieem), 2012 Ieee International Conference On. Ieee, 2012.