# Using resource allocation for seamless service provisioning in cloud computing

**Mustafa Mahmood Akawee[1], Munef Abdullah Ahmed[2], Raed Abdulkareem Hasan[2]**
[1]Department of Fundamentals of Religion, Imam Aadham University College, Baghdad, Iraq
[2]Faculty of Al-Hweija Technical institute/Noerthern Technical University, Mosel, Iraq

**Article Info**

***Article history:***

Received May 31, 2021
Revised Mar 11, 2022
Accepted Mar 21, 2022

***Keywords:***

Cloud computing
Job scheduling
Load balancing
Resource allocation

**ABSTRACT**

A resources allocation is one of the most important things in cloud computing because the environment in which work depends is heterogeneous environment and there are many who request the same service at the same time. Two criteria are used to be considered. First, the time for completing all the required tasks on the available resources and secondly the energy consumed and is extracted from the number of processes that used to complete those tasks. While uploading an application to the cloud, load balancer allocates the appealed instances for physical computers to balance the computational load of several applications through physical computers. This paper presents some resource allocation problems and issues that can solve with help of load balancing techniques and algorithms. In addition, the performance of the algorithm was measured on seven mathematical functions, and the best results were for the algorithm pyogenes.



***Corresponding Author:***

Raed Abdulkareem Hasan
Faculty of Al-Hweija Technical institute/Noerthern Technical University
Mosel, Iraq
Email: raed.isc.sa@ntu.edu.iq

## 1. INTRODUCTION

Resource allocation is considered important within the cloud environment because of it’s the main role in scheduling tasks received on existing resources and utilizing them appropriately, and therefore in the process of resource allocation there must be a smart strategy that works to make a quick decision in the scheduling process. Two criteria are used to be considered [1]-[4]. First, the time for completing all the required tasks on the available resources and secondly the energy consumed and is extracted from the number of processes that used to complete those task [5]-[9]. We present some resource allocation problems and issues that can solve with help of load balancing techniques and algorithms. We focused on highlighting the importance of resource allocation and its relationship with load balancing, as this research showed the importance of all resource allocation as well as the scope of work in load balancing in order for this research to help in finding areas of research and benefit from them. The following work is contributed: a) Analysis of job Scheduling and its aspects in cloud computing, b) Focus on issues of load balancing by using resource allocation methods, and c) Makespan of resource allocation advantage and challenges.

This research is organized into sections, section 1 is a preface where the problem and the most important goals have been defined. Section 2 went through the studies related to the problem and discussed strengths and weaknesses of it. Section 3 spoke about the load balancing and it relation with resource allocation, and section 4 discusses limitation of resource allocation in cloud computing and the section 5 spoke about the discusion the last part is a conculsion.

## 2. RESOURCE ALLOCATION IN CLOUD COMPUTING

In cloud computing, resource allocation designates the methodology of assigning feasible resources to cloud applications. Cloud resources are arranged on request in a multiplexed and fine-grained manner. In cloud, the assignment of resources is as per the IaaS model. In cloud platforms, resource assignment occurs at two levels: i) While uploading an application to the cloud, load balancer allocates the appealed instances for physical computers to balance the computational load of several applications through physical computers. ii) While an application obtains several incoming appeals, these requests must be allocated to an assigned application instance for balancing the computational load over sequences of instances of a similar application [1], [10]-[14].

### 2.1. Resource allocation methods must fulfill the following yardsticks

Resource allocation methods must fulfill the following yardsticks: i) Resource contention evolves when two applications strive for accessing the same resource at the same time. ii) Resource fragmentation evolves whilst the resources are unconnected. There would be adequate resources, yet, cannot assign it to the required application because of fragmentation. iii) Shortage of resources evolves when there are restrained resources and the necessity of resources is high. iv) Several applications require variant kinds of resources like CPU, I/O devices, memory, and the method must fulfill that plea. The arrangement of resources evolves when the application has more resources than required [15]-[18].

### 2.2. Advantages of using resource allocation

Advantages of using resource allocation: i) No hardware or software overhead is available. ii) Minimized location overhead. The data and applications are used from anywhere. iii) Resource distribution amidst cloud providers may be completed during a shortage of resource supply.

## 3. LOAD BALANCING IN CLOUD COMPUTING

Load balancing (LB) is measured as one of the salient concerns and challenges in cloud environs; it has been the procedure of transmission and reassigning of load amidst possible resources for optimizing output, while mitigating the expensiveness and response time, enhancing performance, resource utilization and energy conservation. Service level agreement (SLA) and user fulfillment might be proffered with brilliant load balancing methodologies. Thence, presenting the effective load-balancing algorithms and mechanisms has been a solution for the triumph of cloud computing environs. Numerous researches were accomplished in the load balancing and task scheduling field in cloud environs [19]-[22]. For considering, some of the developed load balancing algorithms include: i) First agent-based dynamic load balancing of cloud computing algorithm (ABDLB); ii) First come first serve algorithm (FCFS); iii) Round Robin algorithm (RR); iv) Load balancing ant colony optimization algorithm (LBACO); v) Longest job first algorithm (LJF); vi) Shortest job first algorithm (SJF); vii) Random allocation algorithm (R); ix) Modified particle swarm optimization (modified PSO). These algorithms are not enough, hence, meta-heuristics such as evolutionary algorithms and swarm intelligent algorithms (SI) are examined. The presented algorithm of this analysis is a potential load balancing methodology assumed to be apt for application in a cloud computing background [23]-[25].

Cloud computing comprises web services, software, visualization, distributed computing and networking. Cloud might have numerous facets such as multiple distributed servers and client data center. The operational conception of LB algorithm can resolve the load essentially in CPU convention, a process in the network, or request delay, and memory capacity. LB is the method that can allocate the load amongst the several nodes of variable distributed system for improving the resource usage in addition to job response time. In all processors, every node or system on the network does the equal amount of work at any time. They focus mainly on using operational LB algorithm to decrease the latency and exploit the throughput on the cloud environment [26]-[28]. The following are certain aims of LB: i) To increase the service availability. ii) To maximize the usage of resources. iii) To enhance the performance. iv) To maintain system stability, v) To make a system fault tolerant. vi) To achieve user satisfaction. vii) To minimize the waiting time and execution time of every task from the diverse locality.

While balancing the load, it is compulsory to effectuate some factors which can be utilized to range the enactment of an LB algorithm. Those factors are: i) associated overhead, ii) fault-tolerant, iii) migration time, iv) migration overhead and other associated overheads, v) response time, vi) resources utilization, and vii) scalability. The growing claim for computational power has directed the enlargement of data centers' overwhelming enormous energy; therefore, it is compulsory to balance the load in a manner that reduces energy usage.

## 4. LIMITATION OF RESOURCE ALLOCATION IN CLOUD COMPUTING

Limitation of resource allocation in cloud computing: i) As operators do not have ownership, they lease resources from isolated servers to their intention; they fix no constraints on their resources. The users or clients are commonly regarded as tenants and not the owners. ii) Migration issue occurs whilst the users demand switching to several other providers to have improved data storage. It is not simple to transfer large data from one provider to the others. iii) More and deeper knowledge is required to allocate and handle resources in cloud, as all knowledge regarding the active role of the cloud considerably relies on the constraint satisfaction problem (CSP) [29]. Table 1 shows the framework in mobile cloud computing-based optimization objectives and types of connections to make comparison between them in one place. The purpose of scheduling is to allocate resources and increase the productivity of shared resources. So that, this research will face challenges, including cost, time, security, reliability, inefficiency, and a lack of control [30]-[35].

Table 1. Some service and also resource allocation structures in mobile cloud computing

| Framework Name | Optimization Objective | Framework Type | Connection Type |
|---|---|---|---|
| Analytical framework (Wen *et al.* 2012) | Energy saving and execution time | Convex optimization | Not provided |
| Analytical framework (Wang and Dey, 2012) | Energy saving | Game theory | WIFI |
| Clone cloud (Kumar *et al.* 2013) | Execution time and energy saving | Virtualization | WIFI, 3G |
| Clone clone (Kosta *et al.* 2012) | Execution time and energy saving | Virtualization | WIFI, 3G |
| Cloudlet (Keke *et al.* 2016) | Execution time | Virtualization | WIFI |
| Analytical framework (Kosta *et al.* 2012) | Energy saving and price | Virtualization | WIFI, 3G |
| DG algorithm (Kaliappan et al. 2016) | Energy saving and execution time | Client/server | WIFI |
| Chroma (Lewis *et al.* 2014) | Lessening execution time | Client/server | WIFI |
| Cuckoo (Yang *et al.* 2013) | Execution time and energy saving. | Client/server | WIFI, 3G, Bluetooth |
| Calling the cloud (Dammert *et al.* 2014) | Execution time, proxy cost and code size | Client/server | WIFI, Bluetooth |

## 5. MATETICAL CALCULATION

If we assume n number of tasks/requests and m number the virtual mechins, and number of solutions is sth. So that we need to assign the tasks to its resources according to the requirement of each request, and the cost assign task i to resource *j* is $C(Xij)$ and accumulated calculation will be as: F(solutions)=$\sum_{i=1}^{n} C(Xij); j\{1,2 \ldots \ldots, m\}$. Here are the conditions that restrict the mathematical model: i) all missions must be completed, ii) each task performed by one supernumerary machine (resource), iii) the assignment is made for one time, vi) each task has a volume with a lot of data to be processed, v) each F-fiction machine has different computing characteristics from the other, and iv) the missions are independent of each other.

There are many problems that cannot be solved within an acceptable time, for example, there is no clear algorithm and similarly there is no clear solution to the traveling salesman problem (TSP) to solve the constraint satisfaction problem (CSP). Its main objective is to classify NP-completeness problems because of this phenomenon a theory has been developed inherently intractable based on the NP-completeness point of view based on the difficulty of solving it problems arithmetic. NP-This is a problem that is categorized by the bin packing problem. We have a problem called It is a set of boxes and a set of items. Completeness is requested boxes at the lowest possible cost. It is similar to our problem in that we have a number of tasks (anatomists) and a group from resources (boxes) and asks to assign tasks (elements) to resources (boxes).

## 6. DISCUSION

It is based on an algorithm simulation not falling into the problem of local examples (Exploitation) and exploiting and developing good solutions (exploration). Thus, increasing the radius of access to the general optimal solution. Based on the strengths of some (local optima) algorithms to overcome vulnerabilities $j$, where ABDLB located. We have expanded the space of solutions dynamically, and dynamic-opposite learning (DOL) performance was increased to overcome the problem of local examples using. In addition, particle swarm optimization (PSO) was compared with all the algorithms presented by this will be verified by implementing an optimal (LBACO) stopping rule. Optimizing the sample space construct using process outperformed all algorithms. The revised algorithm within the cloud environment does not allocate resources, in addition, the performance of the algorithm was measured on seven mathematical functions and the best results were for the algorithmpyogenes. In this study, mentioning its benefits and place of use to begin with, we will mention all the strategies that were used inin aas long as it is. The Figure 1 shows the difference between algorithms and its performance.

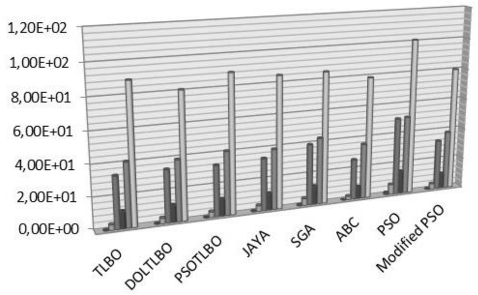


Figure 1. The difference between algorithms and its performance

## 7. CONCLUSION

Resource allocation is considered important within the cloud environment, the process of resource allocation there must be a smart strategy that works to make a quick decision in the scheduling process. This paper focused on highlighting the importance of resource allocation and its relationship with load balancing it also discussed limitation of resource allocation in cloud computing. In addition, the performance of the algorithm was measured on sevenmathematical functions, and the best results were for the algorithmpyogenes.

## REFERENCES

[1] R. Achary, V. Vityanathan, P. Raj, and S. Nagarajan, "Dynamic job scheduling using ant colony optimization for mobile cloud computing," in *Advances in Intelligent Systems and Computing*, vol. 321, 2015, pp. 71–82, doi: 10.1007/978-3-319-11227-5_7.

[2] B. An, V. Lesser, D. Irwin, and M. Zink, "Automated negotiation with decommitment for dynamic resource allocation in cloud computing," *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, vol. 2, 2010, pp. 981–988, doi: 10.1145/1838206.1838338.

[3] N. Desai and W. Cirne, "Job scheduling strategies for parallel processing," in *19th and 20th International Workshops, JSSPP 2015*, Hyderabad, India, May 26, 2015, Dec. 2015, doi: 10.1007/978-3-319-61756-5.

[4] D. Ergu, G. Kou, Y. Peng, Y. Shi, and Y. Shi, "The analytic hierarchy process: Task scheduling and resource allocation in cloud computing environment," *Journal of Supercomputing*, vol. 64, no. 3, pp. 835–848, Jun. 2013, doi: 10.1007/s11227-011-0625-1.

[5] A. G. García, I. B. Espert, and V. H. García, "SLA-driven dynamic cloud resource management," *Future Generation Computer Systems*, vol. 31, no. 1, pp. 1–11, Feb. 2014, doi: 10.1016/j.future.2013.10.005.

[6] H. Ghouma and M. Jaseemuddin, "Context aware resource allocation and scheduling for mobile cloud," *in 2015 IEEE 4th International Conference on Cloud Networking*, CloudNet 2015, Oct. 2015, pp. 67–70, doi: 10.1109/CloudNet.2015.7335282.

[7] S. Guo, J. Liu, Y. Yang, B. Xiao, and Z. Li, "Energy-efficient dynamic computation offloading and cooperative task scheduling in mobile cloud computing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 2, pp. 319–333, 2019, doi: 10.1109/TMC.2018.2831230.

[8] S. Javanmardi, M. Shojafar, D. Amendola, N. Cordeschi, H. Liu, and A. Abraham, "Hybrid job scheduling algorithm for cloud computing environment," in *Advances in Intelligent Systems and Computing*, vol. 303, 2014, pp. 43–52, doi: 10.1007/978-3-319-08156-4_5.

[9] T. Kleinjung *et al.*, "A heterogeneous computing environment to solve the 768-bit RSA challenge," *Cluster Computing*, vol. 15, no. 1, pp. 53–68, Mar. 2012, doi: 10.1007/s10586-010-0149-0.

[10] Y. Li, M. Chen, W. Dai, and M. Qiu, "Energy optimization with dynamic task scheduling mobile cloud computing," *IEEE Systems Journal*, vol. 11, no. 1, pp. 96–105, Mar. 2017, doi: 10.1109/JSYST.2015.2442994.

[11] M. Liaqat *et al.*, "Federated cloud resource management: Review and discussion," *Journal of Network and Computer Applications*, vol. 77, pp. 87–105, Jan. 2017, doi: 10.1016/j.jnca.2016.10.008.

[12] X. Lin, Y. Wang, Q. Xie, and M. Pedram, "Task scheduling with dynamic voltage and frequency scaling for energy minimization in the mobile cloud computing environment," *IEEE Transactions on Services Computing*, vol. 8, no. 2, pp. 175–186, Mar. 2015, doi: 10.1109/TSC.2014.2381227.

[13] J. Liu, X. Luo, X. Zhang, and F. Zhang, "Job scheduling algorithm for cloud computing based on particle swarm optimization," *Advanced Materials Research*, vol. 662, pp. 957–960, Feb. 2013, doi: 10.4028/www.scientific.net/AMR.662.957.

[14] S. T. Maina, "Job scheduling in grid computng using simulated annealing," Thesis, University of Nairob, 2012.

[15] S. Mousavi, A. Mosavi, A. R. Várkonyi-Kóczy, and G. Fazekas, "Dynamic resource allocation in cloud computing," *Acta Polytechnica Hungarica*, vol. 14, no. 4, pp. 83–104, 2017, doi: 10.12700/APH.14.4.2017.4.5.

[16] D. Nagaraju and V. Saritha, "An evolutionary multi-objective approach for resource scheduling in mobile cloud computing," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 1, pp. 12–21, Feb. 2017, doi: 10.22266/ijies2017.0228.02.

[17] M. Nir, A. Matrawy, and M. St-Hilaire, "An energy optimizing scheduler for mobile cloud computing environments," in *Proceedings - IEEE INFOCOM*, Apr. 2014, pp. 404–409, doi: 10.1109/INFCOMW.2014.6849266.

[18] C. S. Pawar and R. B. Wagh, "Priority based dynamic resource allocation in cloud computing with modified waiting queue," in *2013 International Conference on Intelligent Systems and Signal Processing*, ISSP 2013, Mar. 2013, pp. 311–316, doi: 10.1109/ISSP.2013.6526925.

[19] H. Peng, W. S. Wen, M. L. Tseng, and L. L. Li, "Joint optimization method for task scheduling time and energy consumption in mobile cloud computing environment," *Applied Soft Computing Journal*, vol. 80, pp. 534–545, Jul. 2019, doi: 10.1016/j.asoc.2019.04.027.

[20] P. S. Pillai and S. Rao, "Resource allocation in cloud computing using the uncertainty principle of game theory," *IEEE Systems Journal*, vol. 10, no. 2, pp. 637–648, Jun. 2016, doi: 10.1109/JSYST.2014.2314861.

[21] V. Ramharuk, "Survivable cloud multi-robotics framework for heterogeneous environments," Thesis, School of Computing, College of Science, Engineering and Technology (CSET), University of South Africa, 2015.

[22] M. J. Singh, "A survey on QoS based task scheduling approach in grid computing," *International Journal of Engineering Trends and Technology*, vol. 8, no. 7, pp. 359–366, Feb. 2014, doi: 10.14445/22315381/ijett-v8p265.

[23] C. Tang, S. Xiao, X. Wei, M. Hao, and W. Chen, "Energy efficient and deadline satisfied task scheduling in mobile cloud computing," in *Proceedings - 2018 IEEE International Conference on Big Data and Smart Computing, BigComp 2018*, Jan. 2018, pp. 198–205, doi: 10.1109/BigComp.2018.00037.

[24] M. A. Mohammed, I. A. Mohammed, R. A. Hasan, N. Tapus, A. H. Ali, and O. A. Hammood, "Green energy sources: issues and challenges," in *Proceedings-RoEduNet IEEE International Conference*, Oct. 2019, vol. 2019-October, pp. 1–8, doi: 10.1109/ROEDUNET.2019.8909595.

[25] O. A. Hammood *et al.*, "An effective transmit packet coding with trust-based relay nodes in VANETs," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 2, pp. 685–697, Apr. 2020, doi: 10.11591/eei.v9i2.1653.

[26] M. A. Mohammed, A. A. Kamil, R. A. Hasan, and N. Tapus, "An effective context sensitive offloading system for mobile cloud environments using support value-based classification," *Scalable Computing*, vol. 20, no. 4, pp. 687–698, Dec. 2019, doi: 10.12694/scpe.v20i4.1570.

[27] R. A. Hasan, M. N. Mohammed, M. A. Bin Ameedeen, and E. T. Khalaf, "Dynamic load balancing model based on server status (DLBS) for green computing," *Advanced Science Letters*, vol. 24, no. 10, pp. 7777–7782, Oct. 2018, doi: 10.1166/asl.2018.13016.

[28] R. A. Hasan, S. S. Najim, and M. A. Ahmed, "Correlation with the fundamental PSO and PSO modifications to be hybrid swarm optimization," *Iraqi Journal for Computer Science and Mathematics*, pp. 25–32, Jul. 2021, doi: 10.52866/ijcsm.2021.02.02.004.

[29] S. I. Jasim, M. M. Akawee, and R. A. Hasan, "A spectrum sensing approaches in cognitive radio network by using cloud computing environment," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 2, pp. 750–757, 2022, doi: 10.11591/eei.v11i2.3162.

[30] R. A. Hasan, Hadeel W. Abdulwahid, and Arwa Sahib Abdalzahra, "Using ideal time horizon for energy cost determination," *Iraqi Journal for Computer Science and Mathematics*, pp. 9–13, Jan. 2021, doi: 10.52866/ijcsm.2021.02.01.002.

[31] W. A. Hammood, R. A. Arshah, S. Mohamad Asmara, and O. A. Hammood, "User authentication model based on mobile phone IMEI number: A proposed method application for online banking system," in *Proceedings - 2021 International Conference on Software Engineering and Computer Systems and 4th International Conference on Computational Science and Information Management, ICSECS-ICOCSIM 2021*, Aug. 2021, pp. 411–416, doi: 10.1109/ICSECS52883.2021.00081.

[32] W. A. Hammood, S. M. @Asmara, R. A. Arshah, O. A. Hammood, H. Al Halbusi, and M. A. Al-Sharafi, "Factors influencing the success of information systems in flood early warning and response systems context," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 18, no. 6, pp. 2956–2961, Dec. 2020, doi: 10.12928/TELKOMNIKA.v18i6.14666.

[33] A. A. Ayoob, G. Su, D. Wang, M. N. Mohammed, and O. A. Hammood, "Hybrid LTE-VANETs based optimal radio access selection," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 5, 2018, pp. 189–200, doi: 10.1007/978-3-319-59427-9_21.

[34] R. Daoud, W. Ali, Z. A. Husain, Z. A. Majeed, and A. Hasan, "Neural network based assessment the performance of the triangular in-tegrated collector," *NTU Journal of Renewable Energy*, vol. 1, no. 1, pp. 43–49, 2021.

[35] A. H. Ahmed, R. Daoud, and Z. H. Ali, "Theoretical and experimental study of the effect of concentration ratio on CTPTC performance," *NTU Journal of Renewable Energy*, vol. 1, no. 1, pp. 30–37, 2021, doi: 10.1234/ntujre.v1i1.5.


## BIOGRAPHIES OF AUTHORS

**Mustafa Mahmood Akawee** currently works at the al-imam Adhm. Mostafa does research in Information Systems (Business Informatics), Computer Communications (Networks) and Communication Engineering. Their current project is 'offloading in mobile cloud computing. He has finish his master from computer science department at Kharkiv university. He can be contacted at email: It.diyala2@gmail.com.

**Munef Abdullah Ahmed** currently works at the Al-Hawija Technical Institute, Northern Technical University. Munef does research in Natural languge processing. Their current project is Data Analysis. He has finish his master from computer science department at Technology university. Baghdad and he finish Ph.D in Computer science and IT at University polytechnic of Bucharest, Romania. He can be contacted at email: Munif.hwj@ntu.edu.iq.

**Raed Abdulkareem Hasan** currently works at the Al-Hawija Technical Institute, Northern Technical University. Raed does research in Information Systems (Business Informatics), Computer Communications (Networks) and Communication Engineering. Their current project is 'offloading in mobile cloud computing'. He has finish his master from computer science department at BAMU university India. He can be contacted at email: raed.isc.sa@ntu.edu.iq.