iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.42845

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

A Literature Survey on Algorithms for the Optimal Load Balancing in Cloud Computing Environments

Ashutosh Kumar[1], Abhijeet Kumar[2], S. M. Mozammil[3], Ms. C. Vinothini[4]
[1, 2, 3]*Department of Computer Science and Engineering, Dayananda Sagar College of Engineering Bangalore, India*
[4]*Assistant Professor, Department of Computer Science and Engineering, Dayananda Sagar College of Engineering Bangalore, India*

Abstract: ***This paper provides a survey of the existing literature and research carried out in the area of Optimal load balancing in servers. Load balancing in cloud computing is one of the important aspects for efficient delivery of resources and computing . The process of distributing workloads and computing resources in a cloud computing environment is cloud load balancing. It allows enterprises to manage application or workload demands by allocating resources among multiple computers, networks or servers. Hosting the distribution of workload traffic and demands that reside over the internet. It helps enterprises achieve high performance levels for very reasonable costs which is lower than traditional on premises load balancing technology . By taking advantage of clouds' scalability and agility cloud load balancing technology meets rerouted work demands and also improves overall availability. Load balancing can provide health checks for cloud in addition to the workload and traffic distribution .Performance analysis of different existing load balancing algorithms based on different parameters is to be carried out and the algorithm will be optimized for the better performance in cloud. The main purpose of the proposed paper will be to get help in the design of new algorithms in future.***

I. INTRODUCTION

Cloud computing is basically distribution of computing on large scale. The availability of computer storage such as data storage also known as cloud storage, and computing power where direct user management is not required is called cloud computing. It is a growing technology and is becoming more popular everyday as most of the companies want to use this service in order to lower the maintenance and infrastructure cost and increase offerings. One of the most important characteristics of cloud computing is on demand service.

On demand service basically states that a user can access any computing resources without the need for human intervention just by signing up. This is revolutionary since before the introduction of cloud computing, if a user requires extra storage space, it always had to be physically available but with the introduction of cloud computing more than enough cloud space is available to everyone. Not only it solves the need for human intervention but also can be accessed from any location with proper network which is mostly anywhere nowadays as the world is getting more and more digital.

Services provided through cloud computing can be categorized as:

1) *Infrastructure as a Service:* Cloud computing can provide computing hardware infrastructures such as servers or storage. It makes cloud computing very affordable and convenient since users can pay only for infrastructure they need and scale them up or down as needed.
2) *Platform as a Service:* Platform as a service is more like infrastructure management. Hardware and software are hosted by the resource provider on it's own infrastructure. Resource provider then provides the integrated solution to the user. It is very useful for programmers. It provides users with an already built platform linked with a process where users can develop, run and manage their application and not worry about the maintenance of infrastructure.
3) *Software as a Service:* Software as a service is also known as cloud application service and is most like the highest order of service provided by the cloud computing. All the bug fixes, software maintenance and other issues of the application are handled by the resource provider. Applications need not be installed on individual machines as An Api is required for the user to connect through to the application since group access is way more reliable and smoother.

A. *Some pros of Cloud Computing*

1) Low infrastructure cost
2) Only pay for what you use
3) Very easy to grow your applications
4) Everything managed under SLA's
5) Scale up or down at short notice

B. *Some Drawbacks Of Cloud Computing*

1) Too much reliability on service providers
2) Higher ongoing operating costs
3) Too much reliability on a good internet connection
4) Potential privacy and security risk
5) Very limited control of infrastructure

C. *Load Balancing in cloud Computing*

Even distribution of workload and computing resources in a cloud environment in order to achieve greater efficiency and reliability is cloud load balancing.

Proper balance of workload among available nodes is very important. Ensuring efficient resource utilization is a must for effective load balancing technique. The workload among the nodes is equalized by reducing the execution time, communication delay to its minimum and maximizing the resource utilization and throughput. Optimization of all the resources available while ensuring that the delays for app users is minimum is the ultimate goal of cloud load balancing.

This load balancing can be categorized in two types -

1) *Software* -Based load balancer : Software -based load balancers run on standard hardware and standard operating systems.
2) *Hardware* - Based load balancer : These are Application Specific Integrated Circuits (ASICs) adapted for a particular use. This allows high speed promotion of network traffic and is used for transport - level load balancing because hardware-based solutions are faster in comparison to software-based solutions.

D. *Network Layer Load Balancing Algorithms*

1) *Round-robin:* It is one of the most often used algorithms for load balancing. It is very simple. Requests from various clients are distributed to the application server in rotation. Suppose, we have three application servers. Then the first application server shall be requested by the first client, second application server shall be requested by the second client and the third applIcation server shall be requested by the third client.
2) *Weighted Round Robin:* This algorithm is based on a simple Round-robin load balancing algorithm . Servers are rated based on the relative amount of requests each is able to process. Those having higher capacities are sent more requests.
3) *Least Connections:* The server having the least number of active connections receives requests, given that all the connections generate an equal amount of server load.
4) *Weighted Least Connections:* Rating of server is completely based on the processing capacity. Relative capacity of servers and the number of active connections are factors according to which servers are rated.
5) *Source IP Hash:* In this algorithm, the source and destination *IP* address are combined in a request to generate a hash key, which is then designated to a specific server. This lets a dropped connection be returned to the same server originally handling it.

Application Layer Load balancing: The distribution of requests is mostly based on content of the requests being processed which includes its HTTP/S header and message in addition to session cookies. While traveling back from the server then can also track responses, hence, the data on the load each server is processing can be easily obtained.

Least pending requests is the most known application layer algorithm. It monitors pending HTTP/s requests and distributes them to the most available server. Even when it is continuously monitoring the workload of all servers within a server farm it can easily adjust to a sudden influx of new connections.

E. Advantages of LPR

1) *Accurate Load Distribution:* Network layer algorithms distribute requests based on preset rules whereas LPR chooses the best suited server intelligently in real-time.
2) *Request Specific Distribution:* LPR can acknowledge that connection requests take different processing times and distribute load accordingly. As a result, traffic isn't routed to busy servers.

II. BACKGROUND AND MOTIVATION

These days most of the organization and companies are shifting to the cloud based computing technologies because of the ease and efficiency on cloud based server in comparison with the local set-up server. With the increasing data traffic and load on server , it is required at some point of time for the scaling of previously setup servers. Now, scaling can be done in two ways - horizontal scaling and vertical scaling . Horizontal scaling is the preferred way of scaling up the server in the cloud. There are many ways of traffic and requests distribution among the nodes in cloud computing using different techniques and algorithms. There are many factors in load balancing which affect the performance in cloud computing. This comparative study of different load balancing algorithms will help in understanding the deciding factors to be considered when designing or considering new load balancing algorithms.

III. LITERATURE SURVEY

A. A Game Approach to Multi-Servers Load Balancing with Load-Dependent Server Availability Consideration by C. Liu, K. Li and K. Li

Observation - This paper is focused on request migration strategies of multiple set up servers for load balancing in cloud computing . The factors considered in this research is average response time which is defined for each server. The main objective was to minimize the average response time for each server.

Conclusion - Most of the scheduling algorithms in cloud computing ignore the server availability which leads to load imbalance and is also a great waste of computing resources. The experimental results show that proposed IPA algorithm converges to a Nash equilibrium very quickly and significantly decreases the disutilities of all servers by configuring a proper request migration strategy.

B. Load balancing in cloud computing by Mishra SK, Sahoo B, Parida PP

Observation - This paper shows comparative study on load balancing approaches .An abstracted load balancing model is briefly discussed together with activities involved in the load balancing process in cloud computing environments.

Conclusion - It is concluded that there are a lot of issues still open in load balancing process which can be bridged in future by applying an efficient and sophisticated load balancing algorithm.

C. Load balancing based cross - layer elastic resource allocation in mobile cloud by C.Li and L.Li

Observation -In this research paper system status information is used in the hybrid mobile cloud computing system such as the preferences of mobile application , energy , server load in cloud to improve resource utilization and quality of experience and mobile user.

Conclusion -Resource allocation is one of the main operational issues in a MCC environment. After the task is offloaded from the mobile device into the cloud , the cloud provider allocates the task with the desired resources. In the paper, it is discussed the significance and previous resource allocation models and different approaches. To maintain the proper Qos of the MCC system , it is required to adapt the difficulties of the present approaches and overcome the challenges of the resource-allocation strategies.

D. Design of Cloud Computing Load Balance System Based on SDN Technology by S. AL-Mashhadi, M.Anbar , R.A. Jalal, A.Al-Ani

Observation -It was observed that the traditional networking is static and not flexible which is not useful for new business ventures whereas SDN (Software-defined Networking) is programmable during deployment as well as later stage based on the changes in the requirements.

Conclusion - Major issues of cloud computing such as performing customization of system admins can be solved by using Software defined networking .

Client app can be used without much limitations because the system assets will be utilized flexibly with the use of SDN technology.

E. Power aware load balancing for cloud computing by J.M. Galloway , K.L. Smith, S.S. Vrbsky

Observation -It was observed that the major performance between PALB algorithm and the round robin algorithm approach is that compute nodes that are idle using PALB are powered down.

The round robin algorithm is effective in load balancing across the available compute nodes, but such a relatively simple load balancer consumes a large amount of unnecessary power.

Conclusion -This paper concludes with the design of a new load balancing algorithm that balances resources across available compute nodes in a cloud with power savings in nodes.

F. Suboptimal mechanism for load balancing in cloud environment by S.Pandey , A.K. Upadhaya, C.K. Jha

Observation -It was observed that during repetition of execution of proposed mechanism shows that based on the input data, cloud analyst outputs the response and processing time of requests , service time and other related metrics.

Conclusion -The proposed mechanism helps in building a framework for embedding a security since there is involvement of aggregated nodes during load balancing.

G. Load balancing in Cloud Computing :Issues and Challenges by Balaji, K.Turkish Journal of Computer and Mathematics Education

Observation - Load balancing is the most driving technology in cloud computing but with the diverse techniques and limitations it comes with it's own set of issues and challenges. The issues and challenges faced here in load balancing are discussed.

Conclusion -The paper concludes with categorization of various load balancing techniques that helps in differentiating them.The merits, demerits and issues regarding each technology is presented in individual section

H. A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications by Shafiq, Dalia Abdulkareem

Observation - In this paper, the main parameter considered for comparison in this research is Makespan time . The proposed algorithm in paper reduces the makespan time and tries to provide efficient utilization of resources . As compared to existing Dynamic load balancing algorithm this proposed algorithm provides efficient resource utilization of 78%. The algorithm is able to handle the large size requests compared to the existing approaches.

Conclusion -It is proven that considering Qos parameters such as the Deadline can significantly improve the utilization of resources reducing the makespan time and providing efficient allocation techniques in virtual machines.

IV. CONCLUSION

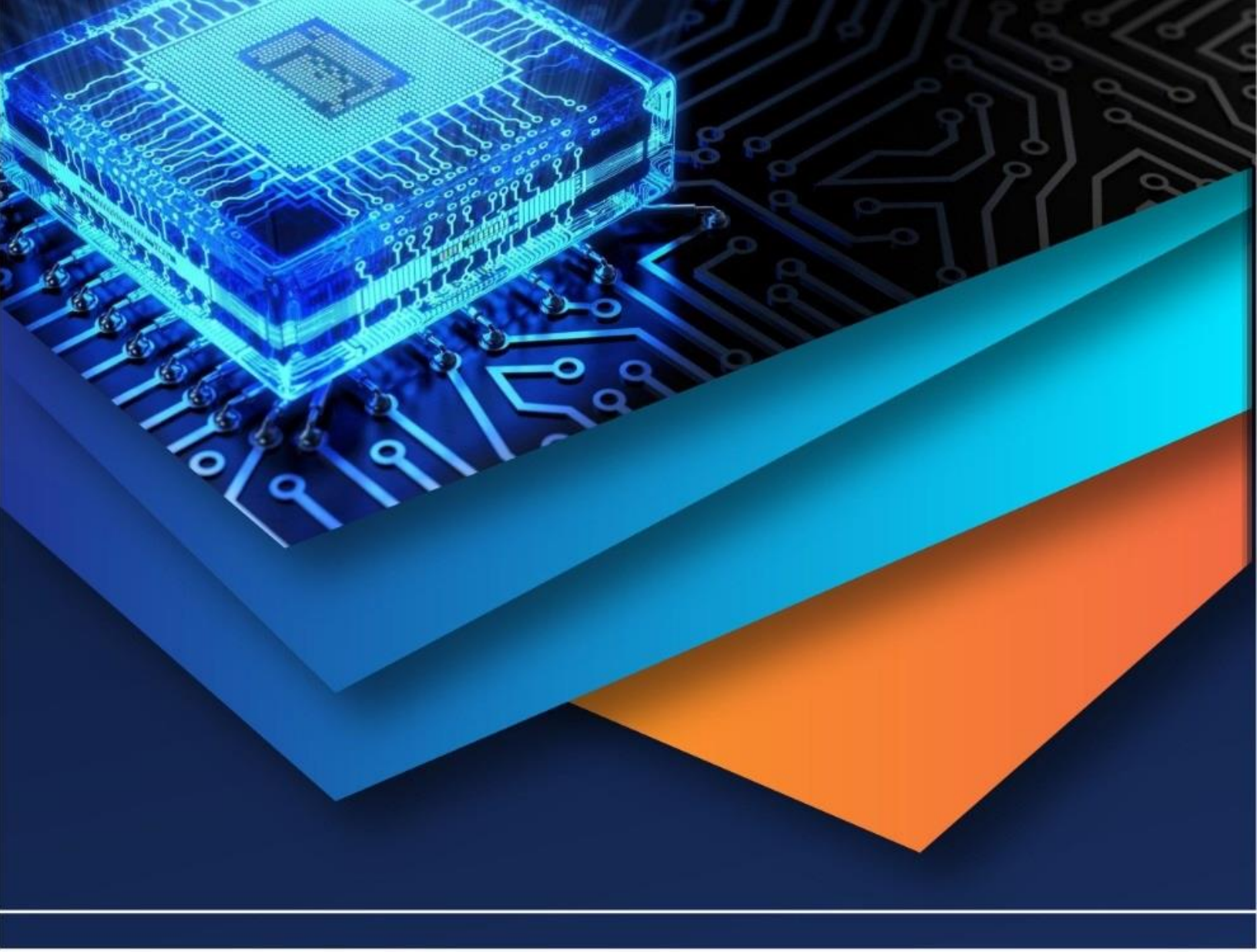
Throughout the study and research of the papers , it was observed there are many factors affecting the resource utilization in load balancing like average response time, makespan time , availability of servers and many others. Our comparative study of Load balancing algorithms will help to choose the deciding factors to be considered when designing algorithms. This will help to understand the effect of various load balancing algorithms on resource utilization and efficiency based on different factors.

Moving forward , new efficient load balancing algorithms can be designed and implemented based on the comparative study.

REFERENCES

[1] C. Liu, K. Li and K. Li, "A Game Approach to Multi-Servers Load Balancing with Load-Dependent Server Availability Consideration," in IEEE Transactions on Cloud Computing, vol. 9, no. 1, pp. 1-13, 1 Jan.-March 2021, doi: 10.1109/TCC.2018.2790404.
[2] Mishra SK, Sahoo B, Parida PP (2018) Load balancing in cloud computing: a big picture. J King Saud Univ Comp Infor Sci:1–32
[3] K. Balaji, P. Sai Kiran, M. Sunil Kumar, An energy efficient load balancing on cloud computing using adaptive cat swarm optimization, Materials Today: Proceedings, 2021, ISSN 2214-7853,
[4] C. Li, L. Li, Load-balancing based cross-layer elastic resource allocation in mobile cloud, Wireless Pers. Commun. 97 (2) (2017) 2399–2437
[5] S. Al-Mashhadi, M. Anbar, R.A. Jalal, A. Al-Ani. 2020. Design of Cloud Computing Load Balance System Based on SDN Technology. In Computational Science and Technology (pp. 123-133). Springer, Singapore.
[6] J.M. Galloway, K.L. Smith, S.S. Vrbsky. 2011, October. Power aware load balancing for cloud computing. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 19-21).
[7] S. Pandey, A.K. Upadhaya, C.K. Jha, Suboptimal mechanism for load balancing in cloud environment, 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), 2017.
[8] R. Aggarwal, L. Gupta, Load balancing in cloud computing, Int. J. Comput. Sci. Mobile Comput. 6 (6) (2017) 180–186.
[9] Balaji, K. "Load balancing in Cloud Computing: Issues and Challenges." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.2 (2021): 3077-3084.

[10] Shafiq, Dalia Abdulkareem, N. Z. Jhanjhi, and Azween Abdullah. "Load balancing techniques in cloud computing environment: A review." Journal of King Saud University-Computer and Information Sciences (2021).
[11] Shafiq, Dalia Abdulkareem, et al. "A Load Balancing Algorithm for the Data Centres to Optimize Cloud Computing Applications." IEEE Access 9 (2021): 41731-41744.
[12] Annie Poornima Princess, G., Radhamani, A.S. A Hybrid Meta-Heuristic for Optimal Load Balancing in Cloud Computing. J Grid Computing **19,** 21 (2021).
[13] Afzal, S., Kavitha, G. Load balancing in cloud computing – A hierarchical taxonomical classification. J Cloud Comp **8,** 22 (2019). https://doi.org/10.1186/s13677-019-0146-7
[14] Gupta H, Sahu K (2014) Honey bee behavior based load balancing of tasks in cloud computing. Int J Sci Res 3(6)
[15] Mishra SK, Puthal D, Sahoo B, Jena SK, Obaidat MS (2017) An adaptive task allocation technique for green cloud computing. J Supercomp 405:1–16
[16] Ibrahim AH, Faheem HEDM, Mahdy YB, Hedar AR (2016) Resource allocation algorithm for GPUs in a private cloud. Int J Cloud Comp 5(1–2):45–56
[17] Jebalia M, Ben Letafa A, Hamdi M, Tabbane S (2015) An overview on coalitional game-theoretic approaches for resource allocation in cloud computing architectures. Int J Cloud Comp 4(1):63–77
[18] Noshy M, Ibrahim A, Ali HA (2018) Optimization of live virtual machine migration in cloud computing: a survey and future directions. J Netw Comput Appl:1–10
[19] Gkatzikis L, Koutsopoulos I (2013) Migrate or not? Exploiting dynamic task migration in mobile cloud computing systems. IEEE Wirel Commun 20(3):24–32
[20] Jamshidi P, Ahmad A, Pahl C (2013) Cloud migration research:



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY