Load Balancing in Cloud Computing Using Meta-Heuristic Algorithm: A Review

Ahmad Hamidi
Department of CSE,
School of Engineering and Technology
Sharda University
Greater Noida, India
2020829165.shir@pg.sharda.ac.in

Mayak Kumar Goal
Department of CSE,
School of Engineering and Technology
Sharda University
Greater Noida, India
mayank.goyal2@sharda.ac.in

Rani Astya
Department of CSE
School of Engineering and Technology
Sharda University
Greater Noida, India
rani.astya@sharda.ac.in

Abstract— In cloud computing, load balancing is a method of distributing workload throughout a certain cloud-based area. Enterprises had given permission for the applications to be controlled. These apps, on the other hand, were developed by the network's provider. Aside from that, the cloud computing system is determined by various computers and stakeholders are given assistance in using the cloud computing system for load balancing. The virtual machines had been kept in balance by the system. The load balancing system had logically distributed the load amongst the virtual machines, and the cloud computing system had aided the companies in ignoring the traffic noise as well as the bad workload on the network. It was used to reduce the load balancing gradient between virtual machines. The Bat algorithm was used to encrypt the load balancing vital features and was constructed by ascending from the bottom up. In this paper, we had described the techniques as a meta-heuristic algorithm.

Keywords— Meta-Heuristic algorithm, Cloud computing, load balancing, Encryption, Decryption

I. INTRODUCTION

In this report, a meta-heuristic algorithm has been designed to implement the operation of load balancing in a "cloud computing" environment. The "load balancing" refers to the uniform distribution of workloads and required computing resources in cloud computing. Multiple enterprises can manage their work resources by modifying and tracing the workload demands among various cloud servers and networks [1]. The load balancing operations have been optimized through a higher-level trial and error concept known as meta-heuristic logic. Quality solutions for the real-time load problems and imbalances of cloud computing are obtained and optimized with a meta-heuristic logic. The load balancers required for effective

Cloud computing solutions are the Direct Routing Requesting Dispatching technique and Dispatcher Based Load Balancing Cluster. The cost-effective and scalable servicing of the system can be achieved appropriately with the help of a meta-heuristic method.

The coding language of Python would be utilized to develop the technical meta-heuristic algorithm for the cloud computing environment [2]. Dedicated software, namely PyCharm, would be utilized completely as an integrated programming tool to develop logical coding solutions. PyCharm also provides an optimal Integrated Development Environment (IDE) to develop convenient data science solutions. Furthermore, the meta-heuristic algorithms can solve the conflicting loading issues and imbalances through a set of equations [3]. A diverse range of solutions with an enhanced focus on specific local regions would be generated through a meta-heuristic algorithm. Thus, the main objective of achieving global optimality in programmed algorithms would be achieved effectively. Optimal meta-heuristic algorithms ensure a high-performance and balanced system of cloud servers.

II. RESEARCH RATIONALE

Here, we have discussed the problem of cloud computing, what we have discovered, how to realize, and find the best possible solution for it. Furthermore, using meta-heuristic methods, the research article developed a load balancing model.

A. *What is the issue?*

This research paper includes cloud computing with a Meta-Heuristic Algorithm that is difficult to process in every organization due to the complex implementation of load balancing in cloud computing. Apart from this, the issue is to manage the network security of cloud computing while applying the meta-heuristic algorithm. Further, the below graph shows that the issues regarding cloud computing are improving.

The issue is scalability and managing network security has occurred due to the implementation of load balancing on the cloud computing platform. The above points have been found as an issue as the complexity of cloud computing and the lack of resources available. Moreover, storage efficiency issues are found in cloud computing.

B. *What is the issue now?*

To generate a more effective authentication method that is useful for the safety of important data, several issues have been found. Firstly, the issue is found on the needed time, as well as needed devices for performing the research paper, regarding the several types of approaches analysis.

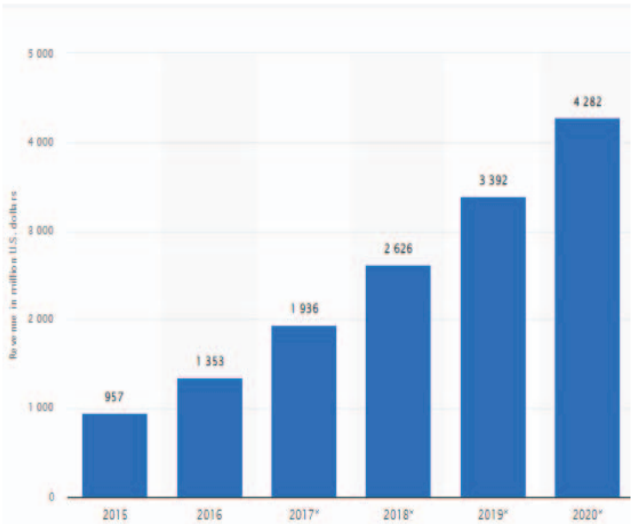

Fig. 1. Issues regarding cloud computing for load balancing

A. What does the research shades light upon?

The research paper has been performed to generate a better approach that has demonstrated to protect the data with several cloud computing methods. Moreover, the research paper has created a load balancing model by utilizing meta-heuristic algorithms.

III. Literature Review

Chaczko Z. et al. [11] concluded that when comparing resource scalability to cloud availability, the ubiquity of information is the most important factor to consider. Load balancing is a mechanism used in clouds to ensure network availability by decreasing the use of computer hardware, software faults, and alleviating resource constraints. This paper addressed cloud load balancing and then shows a case study of system availability using a standard Hospital Database Management System.

Yiqui Fang et al. [12] concluded that a good task scheduling system can suit users' needs while also increasing resource usage and improving the cloud computing environment's overall performance. However, grid computing job scheduling is frequently focused on static task requirements, with a low resource utilization rate. This study addressed a two-level work scheduling technique based on load balancing in cloud computing, in light of the new aspects of cloud computing, such as flexibility, virtualization, and so on. The simulation results in the CloudSim toolkit demonstrated that this work scheduling technique not only meets user expectations but also achieves great resource utilization.

Rajwinder Kaur et. al. [13] concluded that users can share data and access a variety of resources thanks to cloud computing. Users are only charged for the resources they consume. Data and distributed resources are stored in the open using cloud computing. In an open environment, the volume of data storage grows rapidly. As a result, in the cloud environment, load balancing is a major difficulty, Load balancing aids in the distribution of a dynamic workload among numerous nodes, ensuring that no single node is overburdened. It aids inefficient resource utilization. It also helps to boost the system's performance. Load balancing and greater resource usage are provided by many existing algorithms. In cloud computing, many types of loads, such as memory, CPU, and network load, are all feasible. The practice of discovering overloaded nodes and distributing the extra load to other nodes is known as load balancing.

S. K. Mishra et al. [14] examined that load balancing is the method in the recognition of overloaded and underloaded nodes. To maximize distinct performance metrics, researchers developed numerous load balancing methodologies in cloud computing. However, a taxonomy of cloud load balancing algorithms has been presented. This paper provided a quick overview of the performance parameters that have been addressed in the literature and their impact. Results were presented to examine the performance of heuristic-based methods.

Ramezani F. Et al. [15] made a study and found that Particle Swarm Optimization was used to create the optimization model. We enhance the Cloudsim software and utilize PSO as the work scheduling model to assess the suggested strategy. When compared to standard load balancing methodologies, the basic results revealed that the suggested TBSLB-PSO method greatly reduces load balancing time. As a result, the TBSLB-PSO technique reduces the risk of losing a customer's most recent activity, while also improving cloud customers' Quality of Service.

TABLE I. Comparison of Algorithm

Algorithms	Purpose
INS Algorithm	minimize data duplication and data redundancy in cloud computing operations. It integrates both the features of reduplication and optimization of access point selections.
WLC Algorithm	introduces an optimum weight based on the existing capacities of each cloud server. It works on the weight capacities of servers and the number of current clients connected to the servers.
LBM Algorithm	utilizes a three-stage framework of the OLB. It is slower than other algorithms.
Centre-Queue Algorithm	It acts as a central queue where the assigned tasks are classified and arranged in a first-in, first-out (FIFO) order.
Token-Ring Algorithm	No overheads of communication are generated in this method, and thus efficient and quick routing decisions can be taken.
Tabu-Search Algorithm	It is an advanced meta-heuristic algorithm that can be used to solve complex load balancing problems. It avoids forming a loop of optimal solutions.
SA Algorithm	It solves the optimization issues based on hard combinations.
CBD Algorithm	based on the existing Round Robin algorithm with some added modifications.
OLB Algorithm	The nodes are always loaded with numerous requests in the OLB algorithm method. It fails to calculate the current execution time of nodes.

IV. Technical Analysis

Many works and studies have been done in the field of cloud computing, especially in the areas of scheduling (tasks, VMs, and compute), energy management, resource management, and load balancing, to name a few. Load balancing has been an eagle eye among academics, as you may have seen, due to its nature in cloud computing between the stakeholders.

A. Methods

Cloud computing is used to transform the overall computing architecture of an enterprise into several cloud services and applications. However, the usual imbalances of loads of virtual machines in data centers arise a prevailing problem. The load balancing of various computing platforms evolves the technological quality of existing computing services. Cloud computing also helps major organizations to achieve high operational performance at lower costs. The cloud characteristics of scalability and agility are completely utilized by the load balancing algorithms of cloud balancing [4]. These characteristics provide the computing environment to meet rerouted workload demands and thus improve the overall resource availability.

The process of load balancing with the help of a meta-heuristic algorithm can be implemented through two main phases. Firstly, the pre-classification of tasks following requested resources is performed in the initial phase. Later, the process of classification of assigned tasks into odd and even levels occurs as per a meta-heuristic bat algorithm. Various mathematical functions are used to divide the entire functional system into numerous virtual machines. The bat algorithms generate specific levels, and further task allocation is based on the generated levels. The virtual machines obtained from the bat algorithm can deliver equal performances based on almost similar characteristics compared to existing processes [5]. The various met heuristic algorithms used for load balancing of the computing environment can be categorized under static and dynamic types. The dynamic load balancing algorithms are implemented to reduce the number of errors produced by the static balancing algorithms by considering the actual loads of all virtual machines. The following dynamic algorithms can be utilized appropriately in load balancing processes.

B. DARS

The Dynamic adaptive replica strategy (DARS) creates various replicas of nodes to store their values in these replicas to minimize overloads. This method does not compromise with the access delays of cloud servers and reduces the degree of overloaded nodes in a decentralized approach. This method stores data of nodes in their replicas based on their current loading states.

C. Two-phase load balancing

The service nodes, service managers, and request managers platforms utilize this algorithm to perform cumulative functions. However, performance metrics are not provided in this method as it lacks simulation activities.

D. INS algorithm

The main purpose of this method is to minimize data duplication and data redundancy in cloud computing operations. The Index Name Server (INS) algorithm integrates both the features of reduplication and optimization of access point selections. Certain parameters like the Hash code are used for positioning and targeting a block of data to calculate an optimum selection point.

E. WLC

The Weighted least connections (WLC) introduce an optimum weight based on the existing capacities of each cloud server. This load balancer method works on two parameters: the weight capacities of servers and the number of current clients connected to the servers. The node specifications like processing speed, storage capacity, and bandwidth are not considered in this method.

F. LBM

The Load Balancing Min (LBMM) algorithm utilizes a three-stage framework of the Opportunistic Load Balancing (OLB) process. This method is comparatively slower than other load balancing algorithms as it does not consider each node's execution times, and thus, requests keep on waiting for required nodes to become free.

G. Central queue algorithm

In this method, the data center controller acts as a central queue where the assigned tasks are classified and arranged in a first-in, first-out (FIFO) order. If any virtual machine runs into an underload status, the main queue requests new and improved task allocations to the data center controller [9]. Later, the data center controller deletes the old tasks from the queue and sends new task allocations directly to the recognized virtual machines.

H. Token ring algorithm

In this method, various functional tokens move around the entire system so that the system costs can be minimized effectively. A heuristic algorithm is added to this method to remove the existing drawback of token algorithms. No overheads of communication are generated in this method, and thus efficient and quick routing decisions can be taken.

I. Tabu Search

The taboo search is an advanced meta-heuristic algorithm that can be used to solve complex load balancing problems. This search algorithm tracks the regions of cloud spaces that have been previously searched. It relies on the pre-defined concept of a taboo list where short-term memory solutions can be obtained. An additional feature of the short-term memory is that it adds particular smoke attributes of previous solutions and avoids forming a loop of optimal solutions.

J. SA algorithm

The simulated annealing (SA) algorithm is generally utilized to solve the optimization issues based on hard combinations. The controlled randomization process is performed in this method through the simulation of a temperature falling procedure in a thermodynamic system. The current system varies randomly in this technique to better solve the recognized optimization problem [8]. The most beneficial aspect of this method is that it accepts a worse variation as a new and improved solution with optimum probability. As a result, the optimization problem does not become globally trapped at a local minimum; thus, an optimal solution with a probability of 1 can be obtained.

K. Bat-algorithm

The concept of echolocation is utilized in this algorithmic modelling technique where the shortest iteration is effectively identified. The coding of this algorithm starts with the initialization of bat population position, velocity, loudness, pulse rate, and frequency. Furthermore, obtaining an optimized solution is followed to generate a local solution among various selected best solutions.

On the other hand, the static load balancing algorithms assign required tasks to the computing nodes following their respective abilities to process new advanced requests. The static algorithms closely consider the nodes' capabilities and properties, such as the node's processing power, memory and storage capacity, and communication performance. The following static "load balancing" algorithms are mostly utilized.

L. CBD

The Central Load Balancing Decision Model (CLBDM) is based on the existing Round Robin algorithm with some added modifications. The CLBDM can calculate the "connection time" between cloud nodes and clients than the round-robin algorithm. If any issue arises, the current tasks would be urgently terminated, and the current task will be forwarded to other computing nodes following the Round Robin rules.

M. Weighted round-robin algorithm

This method distributes all the task requests among available virtual machines regardless of the characteristics of allocated tasks. Unlike the round-robin algorithm, this method controls the complete workload distribution as it assigns a higher number of requests to nodes with more improved specifications.

N. OLB algorithm

The nodes are always loaded with numerous requests in the "Opportunistic load balancing (OLB) algorithm" method. All the sudden and unexpected tasks are randomly allocated to the nodes that are currently available. The recognized drawback of this method is that it fails to calculate the "current execution time" of nodes.

O. Throttled load balancing algorithm

This method checks whether the current state of a virtual machine is available for new allocations or not. Later, this method sends the state of "acknowledged VM identifier (ID)" to the "data center controllers" to receive a new task allocation for the virtual machines. The data center controls notify the cut-offs required in the algorithms after the virtual machines send results of task processing activities [6].

P. OLB + Min-Min algorithm

This method combines the characteristics of both the "opportunistic load balancing (OLB) algorithm and the Min-min algorithm". In this method, the "execution times" of the tasks are minimized, and thus, nodes can become free much earlier for new task allocations. The entire work allocation efficiency is improved in this method with the help of "better resource utilization". This method is also highly convenient for its massive intolerability towards the operational errors of cloud nodes.

Q. Threshold algorithm

This method involves the immediate assignment of local hosts after their creation to minimize excessive loading conditions. The method copies the load of the current system to determine the loading conditions of the processor, including underloaded, medium loaded, or overloaded. A major drawback of this method is that it allocates all tasks locally, and thus, some processors may be overloaded or under-loaded [7]. This might result in execution issues as the availability of processors cannot be determined beforehand in this algorithm.

R. MapReduce algorithm

The main tasks performed in this static method are "Mapping tasks and Reducing obtained tasks' results". The three different methods executed in this algorithm are "part, comp, and group". The mapping of tasks is primarily initiated in the part method of this algorithm. Moreover, the "request entities" are effectively partitioned into respective parts through the mapping tasks. The comp method compares the parts after these generated parts are saved into a robust "hash key table". The "reduce tasks" is later utilized by the group method to group the effective parts based on similar entities [10]. The "reduce tasks" can process and read the available task entities in a parallel manner, and thus the "reduce tasks" can become overloaded.

V. CONCLUSION

The cloud computing platform requires an advanced solution for load balancing problems in its distributed networks. The dynamic and static load balancing algorithms are provided to achieve a meta-heuristic solution for the general issues. The users and clients can achieve more efficient and high-performance computing solutions with the help of these algorithms. The data handling and optimization techniques can be improved largely with the help of dynamic task scheduling and "dynamic load balancing". The static load balancing algorithms perform satisfactorily in "homogeneous and stable" operating conditions. On the other hand, the dynamic load balancing algorithms are more flexible and thus consider all the attribute types of the system during and before run-times. The clients can use various cloud services by managing the "underlying infrastructure" of the virtual machines.

A "meta-heuristic Bat-algorithm" is found to be most effective in task allocations based on associated levels. Thus, the data center can identify virtual machines with similar performances through advanced mathematical functions. The report has utilized the Python coding platform to produce load balancing algorithms based on a meta-heuristic approach. Therefore, the "PyCharm" software has also been used to model and test solutions to recognized load balancing issues. The "meta-heuristic algorithm" has also helped in improving the "Quality of Service (QoS)" issues of workloads, thus allowing better load balancing practices. The meta-heuristic algorithms also obtained uniform load distributions, enhancing the possible number of task allocations depending upon task levels. Furthermore, the reduction of overloading of servers and cloud nodes can be obtained through a "multiple-server meta-heuristic solution". Thus, the circulation of the traffic and workloads of workflow demands has been optimized effectively through a meta-heuristic approach.

VI. FUTURE SCOPE

First, this work serves as a guide and a comprehensive study for load balancing, which distributes stress among virtual machines in a novel way. The loss of superfluous resource use, as well as not leaving the VMs inactive, is done on purpose. A network-based service in which load balancing is a primary factor is referred to as cloud computing. When a system is overburdened, it performs poorly. As a result, to preserve the QoS position, a creative load-balancing algorithm is necessary; qualification for algorithm techniques has been offered, followed by a comparison of them. Second, in cloud computing, what is the impact of mixing heuristic and meta-heuristic load balancing algorithms? The third step is to compare my method to several existing load balancing algorithms. Finally, I want to identify the best settings for these system characteristics to provide a balanced performance under heavy, light and moderate loads.

REFERENCES

[1] Zatwarnicki, K. Providing Predictable Quality of Service in a Cloud-Based Web System. Appl. Sci. 2021, 11, 2896. https://doi.org/10.3390/app11072896

[2] Li J., Sheng J., Chen Y. "A Web-Based Learning Environment of Remote Sensing Experimental Class with python," The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLIII-B5-2020, pp. 57-61, 2020. DOI: http://dx.doi.org/10.5194/isprs-archives-XLIII-B5-2020-57-2020.

[3] Barrera A., Román P. and Torres-Ruiz F., "T-Growth Stochastic Model: Simulation and Inference via Metaheuristic Algorithms," Mathematics, vol. 9, (9), pp. 959, 2021. DOI: http://dx.doi.org/10.3390/math9090959.

[4] Nguyen, T.A.; Fe, I.; Brito, C.; Kaliappan, V.K.; Choi, E.; Min, D.; Lee, J.W.; Silva, F.A. Performability Evaluation of Load Balancing and Fail-over Strategies for Medical Information Systems with Edge/Fog Computing Using Stochastic Reward Nets. Sensors 2021, 21, 6253. https://doi.org/10.3390/s21186253.

[5] Olivares-Benitez E., "Multi-objective Design of Balanced Sales Territories with Taboo Search: A Practical Case," International Journal of Supply and Operations Management, http://dx.doi.org/10.22034/IJSOM.2021.2.5.

[6] Tabassum N., Ditta A., Alyas T., Abbas S., Khan M. A. "Prediction of Cloud Ranking in a Hyperconverged Cloud Ecosystem Using Machine Learning," Computers, Materials, & Continua, vol. 67, (3), pp. 3129-3141, 2021 DOI: http://dx.doi.org/10.32604/cmc.2021.014729.

[7] Pyszny K., Strzelecki A., "Users' Adoption of Sustainable Cloud Computing Solutions," Sustainability, vol. 12, (23), pp. 9930, 2020. DOI: http://dx.doi.org/10.3390/su12239930.

[8] Ting K. H., Lee L. S. and Seow H. V., "Shared Mobility Problems: A Systematic Review on Types, Variants, Characteristics, and Solution Approaches," Applied Sciences, vol. 11, (17), pp. 7996, 2021. DOI: http://dx.doi.org/10.3390/app11177996.

[9] Wang R. B., Wang W. F., Xu L., Pan J. S., "An Adaptive Parallel Arithmetic Optimization Algorithm for Robot Path Planning," J. Adv. Transport., vol. 2021, DOI: http://dx.doi.org/10.1155/2021/3606895.

[10] Molina-Pérez D., Portilla Flores E. A., Alvarado E. V., "A Novel Multi-Objective Harmony Search Algorithm with Pitch Adjustment by Genotype," Applied Sciences, vol. 11, (19), pp. 8931, 2021. DOI: http://dx.doi.org/10.3390/app11198931.

[11] Chaczko, Z., Mahadevan, V., Aslanzadeh, S., & Mcdermid, C. (2011, September). Availability and load balancing in cloud computing. In International Conference on Computer and Software Modeling, Singapore. 14: 134-140.

[12] Fang Y., Wang F., Ge J. (2010) A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing. In: Wang F.L., Gong Z., Luo X., Lei J. (eds) Web Information Systems and Mining. WISM 2010. Lecture Notes in Computer Science, vol 6318. Springer, Berlin, Heidelberg.

[13] Kaur, R., & Luthra, P. (2012). Load balancing in cloud computing. In Proceedings of the international conference on recent trends in information, telecommunication and computing, ITC.

[14] Mishra. S. K., B. Sahoo, P. P. Parida. 2020. Load balancing in cloud computing: A big picture, Journal of King Saud University - Computer and Information Sciences. 32(2): 149-158.

[15] Ramezani, F., Lu, J. &Hussain, F. K. 2014. Task-Based System Load Balancing in Cloud Computing Using Particle Swarm Optimization. Int J Parallel Prog. 42:739-754.