JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Bio-Inspired Load Balancing Algorithm in Cloud Computing

Kamini Satish Patil[1], Susheel Tiwari[2]

[1]Department of Computer Science Madhyanchal Professional University, Bhopal, India.

[2]Department of Computer Science Madhyanchal Professional University, Bhopal, India.

Abstract

Cloud computing emerges from different technologies and provides online services over the internet. The emerging technology and huge amount of traffic have degraded the performance of cloud computing. The degraded performance of cloud computing has affected the quality of services and market reputation. This paper proposes a load-balancing algorithm for scheduling tasks based on the ant colony optimisation algorithm. The proposed ant colony optimisation algorithm optimised the selection of tasks. The process of task selection reduces the load on virtual machines and improves the capacity of cloud computing. The proposed algorithm is simulated in cloud-sim software for validation. The analysis of the results suggests that the proposed algorithm is better than existing algorithms such as MCT-PSO, MALO, and WOA.

Keywords: - Cloud Computing, Load Balancing, Swam Intelligence, Virtual Machine, Data Centre.

Introduction

The increasing demand for information technology resources over the internet generates huge traffic and declines the service of cloud computing. The management of huge traffic and user requests for data storage and retrieval required a load-balancing approach. The load balancing approach maximize utilization of resources and improves the capacity of cloud networks. Load balancing is adopted in two manners: static load balancing and dynamic load balancing. The dynamic load-balancing approach enhances the performance of cloud computing. The dynamic load balancing approach employed several bio-inspired meta-heuristic functions for load balancing. Recently, several authors proposed swarm intelligence-based dynamic load balancing algorithms such as particle swarm optimization (PSO), genetic algorithm (GA), ant colony optimization (ACO), and many others. Researchers have primarily focused on cloud load balancing and resource scheduling in the majority of research studies. For example, Assi et al. [2] used a novel decomposition approach to address scalable traffic management (STM) in the cloud. STM aids in decreasing the maximum link load, ensuring load balancing among network users. However, multidimensional resource scheduling was not a good fit for this methodology. [3] introduced a scalable workload-driven partitioning scheme to address this problem and increase the throughput and response time for distributed transactions. Another study in [4] offered insight into how honey bee behaviour is used for task scheduling in cloud environments. The behaviour of honey bees was used to achieve the best machine utilisation. In a cloud environment, honey bees' foraging behaviour is used to efficiently distribute the load among virtual machines. Cloud computing has drawn a lot of attention and is acknowledged as the best approach to managing and enhancing resource and data utilisation while providing a range of IT and computing services. [5] looked into the performance and power-constrained load-distributed methods to offer fresh perspectives on data centre optimisation. But resource management is still a problem. To get around this, Batista et al. [6] carried out performance evaluation for resource management by guaranteeing service quality. Numerous bio-inspired algorithms demonstrate their effectiveness in load-balancing systems, like honey bee and ant colony systems. But a majority of them fall short of attaining satisfactory outcomes in every area. As a result, we present hybrid algorithms that leverage the advantages of each individual algorithm. This paper proposes an osmotic hybrid artificial bee and ant colony optimisation algorithm (OH_BAC) for load balancing. Applying the osmosis technique to load balancing will become popular again. It can use a variety of resources to offer services to customers via the Internet [13]. Since popularising the idea of cloud computing, Amazon has created a number of cloud computing platforms, such as Microsoft Azure, EC2, Google Engine, Apache Hardtop, and Apache. The cluster of resource systems is Amazon EC2. The Linux virtual machine resource provided by Amazon Data Centre Amazon Data Centre offers web services [14]. Based on their size, events can be categorised into three groups: very large, large, and small [15]. Cloud computing has a significant impact on the IT sector [16], and there is fierce competition among businesses over how well they can provide services [17]. In order to enable an increasing number of customers to subscribe to cloud services, businesses are also working to further enhance or upgrade their offerings through a variety of resources [18]. As a result, resource allocation and SLA [19], which indicate the degree of user satisfaction, are among the most

significant factors that affect the quality of service. However, setting boundaries and dimensions and overcoming the challenge of reaching upper bounds is crucial [20]. Section II discusses previous work, while Section 3 exhibits the proposed research. Section II discusses previous work; in Section 3, the proposed research is exhibited. Section 4 reveals the experimental results, and the paper is concluded in Section 5.

II. Related Work

Bio-inspired functions are employed in dynamic load balancing and improve the capacity of resource optimization and allocation. Recently, several authors proposed novel algorithms for load balancing in cloud computing. In [1], the results of the accomplished simulations within the Cloud Analyst platform demonstrated that, depending on the applicable scenario, the proposed technique can improve load balancing in the cloud architecture while reducing response time by up to 82%, processing time by up to 90%, and overall cost by up to 9%. in [2] proposed a load balancing method that minimises make span and maximises resource utilisation by employing modified PSO task scheduling (LBMPSO) to schedule jobs over the available cloud resources. In [3], the proposed approach took into account the Make Span parameters to address the issue with the current modern heuristic approaches. To evenly distribute the load throughout the data centres, the proposed approach relies on the mutation-based Particle Swarm algorithm. The technique in [4] builds a multidimensional fuzzy-based resource scheduling model to achieve cloud-infrastructure resource scheduling efficiency. Comparing the strategy to state-of-the-art works, simulation research reveals that it reduces the reaction time by 35% and increases the resource scheduling efficiency by 7%. In [5], the results of the study show that MMRR has significantly altered cloud services. Out of all the algorithms examined, MMRR did better in terms of cost-effectiveness (89%) and overall response time. According to the report, MMRR should be used to improve cloud service and customer-happiness. in [7] proposed a load balancing method that minimises the make span and maximises resource utilisation by employing modified PSO task scheduling (LBMPSO) to schedule jobs over the available cloud resources. In [8], algorithms take into account the Central Processing Unit's capacity to acquire targets as well as the records of available resources. The Fusion Algorithm has been assessed and contrasted with other algorithms using the Cloud Analyst Simulator. in [9] With regard to load balancing and server consolidation, this work aims to provide a new taxonomy that classifies factors such as dependability, hardware threshold, network traffic, and migration overhead. In [10], the technique offers the best possible transfer of workload-causing virtual machine loads to matching virtual machines in the cloud. The target functions in the proposed optimisation model are the minimization of task execution and transfer time. [11] describes a mobile device load balancing technique in edge-cloud computing systems in this work. The proposed load balancing method has an effective complexity thanks to an implementation based on a genetic algorithm and graph colouring. In [12], three iterations of ALO and GWO were provided in this paper as task schedulers to help shorten the make span. It contrasts the outcomes using FFA and PSO. In [13], one of the primary concerns in cloud environments is the study paper's identification of the necessity of the FT-efficiency parameter in LB algorithms. A new algorithm that incorporates FT in LB is proposed. In [14], in order to guarantee that the workload is distributed equally among the nodes, they present a threshold-based load balancing technique in this study. In [15], in order to provide load balancing and minimise bandwidth costs, this model features a composite goal function. We take into account balancing the demand on the bandwidth of the links and the CPU-processing capabilities of the servers. in [16] proposes a new hybrid approach called PSO-CALBA, a particle swarm optimisation-based content-aware load balancing algorithm. The PSO-CALBA scheduling technique classifies files based on their content type by combining a meta-heuristic algorithm and machine learning. in [17] The various guises, functions, developments, and difficulties associated with AI-based cloud computing models are highlighted in this study. This study examines and summarises research findings and advancements in the field. [18] proposes a hybrid machine-learning (RATS-HM) approach to address those issues by combining resource allocation security with effective task scheduling in cloud computing. In [19], hybrid algorithms would be most appropriate for offering enhanced services to meet the demands of the current situation. This research examined a few, and with further analysis and the addition of other methods and parameters, it is possible to move this forward to the next level. in [20] In order to examine the advantages and disadvantages of different architectures in big data cloud computing networks, we have provided two dynamic models with two variational architectures in this research to address the data duplication issue. In [21], the proposed approach states that end users may receive QoS through improved cache-resource management and energy consumption reduction through the application of stochastic gradient bellman optimality. in [22] To address this issue, we proposed in this study a hybrid scheduling policy called Hybrid Particle Swarm Optimisation Actor Critic (HPSOAC), which is a mix of both the PSO method and the actor-critic algorithm. [23] describes a job scheduling method in a heterogeneous cloud environment based on the Advanced Phasmatodea Population Evolution (APPE) algorithm. Compared to previous algorithms, this approach uses more resources and has a faster convergence time. In [24], a study introduces a novel method for load balancing and task scheduling in cloud computing environments, focusing on how to execute tasks using resources in a way that minimises costs. In [25], the goal is to improve cloud load balancing by integrating load balancer-external factors. In comparison to other well-known LB algorithms, the proposed approach exhibits better load-balancer performance in an experimental test conducted in the Cloud-Sim environment. [26] offers a reinforcement learning (RL) approach that achieves strong scalability and a notable performance improvement in managing load balancing (LB) and dynamic resource allocation (DRA) tasks in a cloud context. Effectiveness can be increased by 30% to 55%, while training time can be shortened by 10% to 45%. The author [28] Swarm intelligence (SI) is proposed in this study as a cloud computing load balancing method. When compared to alternative algorithms, the proposed technique's overall response time is, on average, 12% faster. in [29] tackled the aforementioned scheduling problem, which searches for the best resource for the tasks using a deep Q-learning network model. There were two stages to the process. For simulation, a randomly generated workload is first employed.

III. Proposed Methodology

This section describes a proposed algorithm for load balancing in cloud computing using the ant colony optimisation algorithm. The proposed algorithm, colony optimisation, is modified with the selection of tasks in manners of discontinuity. In the proposed algorithm, the pheromone enables the allocation of a virtual machine (VM) based on the technical capability, cost, and heuristic data of a chosen

server by assessing its adaptability. After every successful virtual machine (VM) assignment to a server, the heuristic component is updated to guarantee optimal results in the following iterations. Some ants begin searching for the ideal configuration that ensures the best load-balancing level and the fastest response time during each iteration. Every ant must visit every node that is accessible. The best solution is found by assessing the capacity of each available node, the desirability of each path, and the load balancing of the network.

Algorithm

Begin: input data

Begin : define parameters($\alpha, \beta, num_{ants}, Max_{iter}$)

Define pheromone

Define ants

While itre< MAX iter do

While i<NUMants do

Ants move on fist location

End while

End while

Terminate iter

Return list of VM

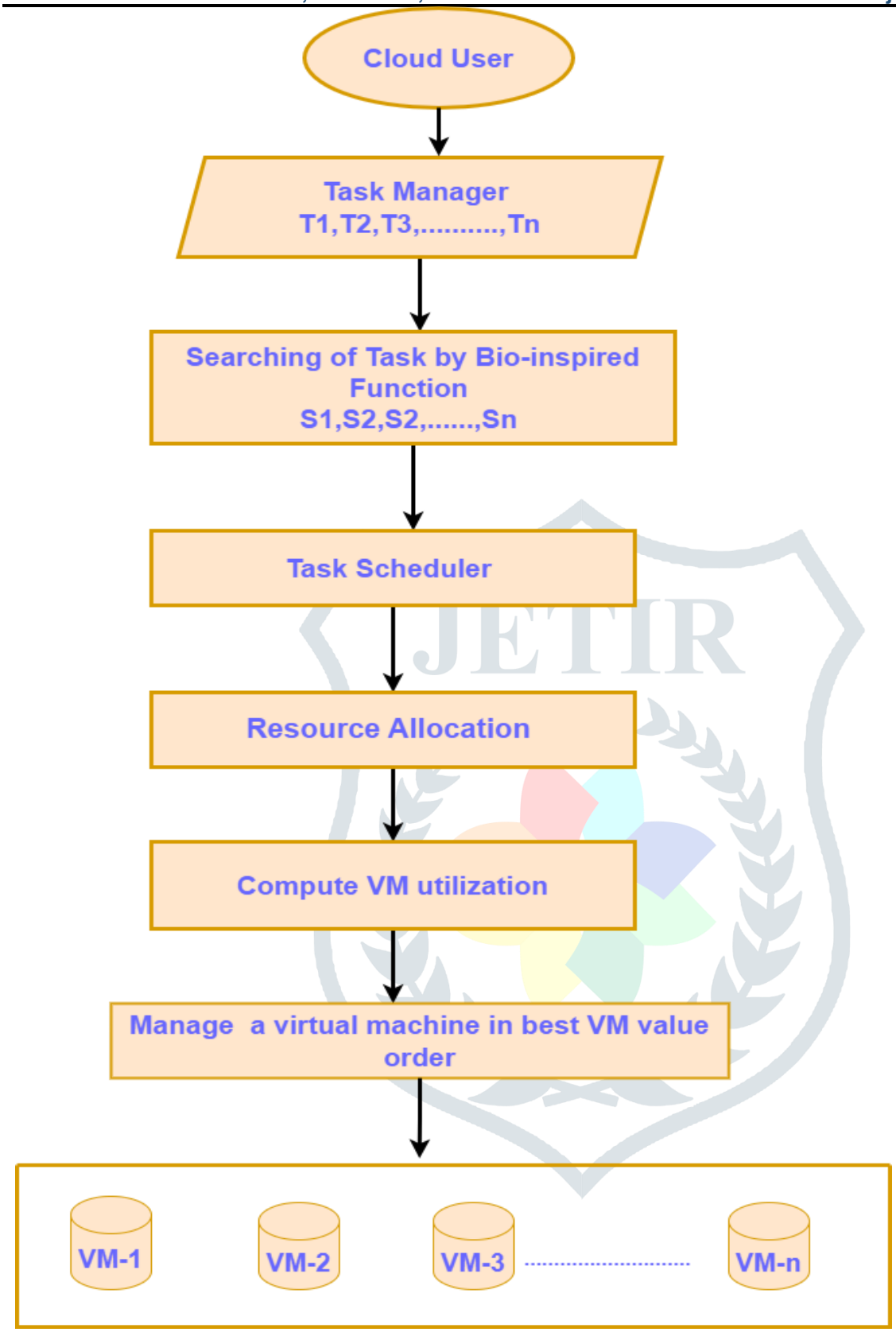

Figure 1 Proposed Model of Load Balancing In Cloud Computing Using Bio-Inspired

IV. Experimental Analysis

To investigate the effectiveness of the proposed algorithm for load balancing in cloud-Sim Simulators. The Cloud-Sim simulator is an open-source toolkit for simulating cloud computing environments. It has been developed in the cloud lab of Melbourne university. This simulator contains a data center, virtual machine, traffic load, users and scheduling algorithm. The complete simulation process deal with the java programing language. The proposed algorithm is implemented in the scheduling algorithm with an existing algorithm for the simulation of the dedicated scenario for different groups of loads and measuring standard parameters of the simulation [28,29,30].

Response Time

The response time of a task refers to the time intervals among tasks to arrive into the system until its completion. Response time Re is expressed as

$$Re = Tc - Ta + Tt$$

where Tc represents the time required to complete a task, Ta represents the arrival time of a task, and Tt represents the transfer time of a task.

Make-span

Make-span is defined as the total time taken to process a set of tasks for its complete execution. Make-span Ma is represented as

$$Ma = \max (CT)$$

where max (CT) is the maximum time required to complete all tasks.

Resource Utilization

Resource utilization denotes the number of resources required during task execution. Resource utilization Ru is expressed as

$$Ru = \frac{T_c}{M_a X\ N}$$

where Tc represents the time taken to complete a task, Ma represents make-span and N represents number of resources.

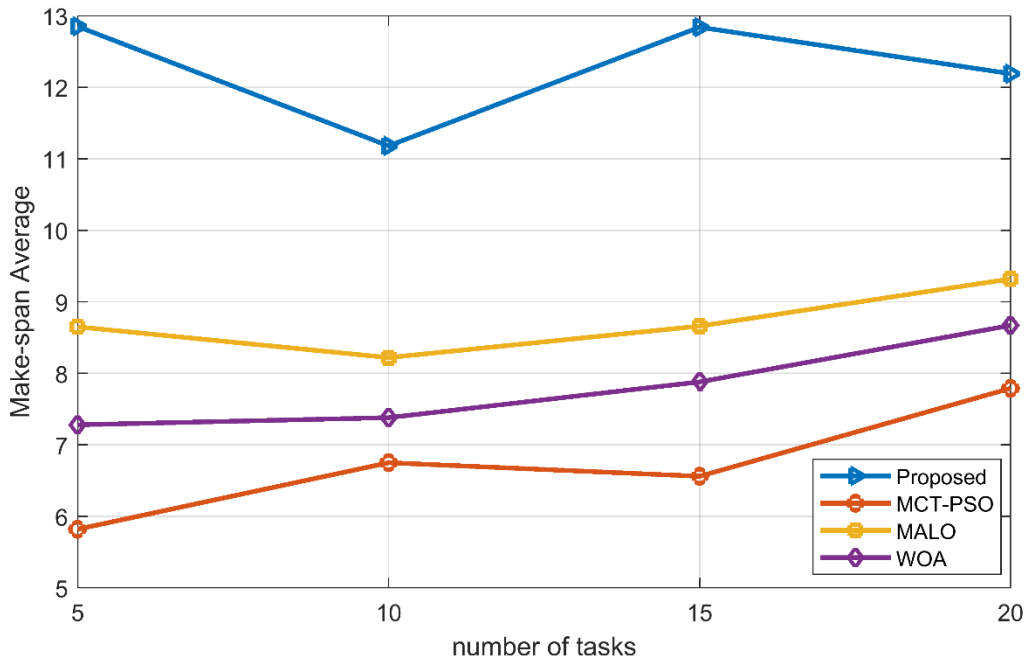

Figure.2 Comparative analysis of make-span average using Proposed, MALO, WOA, and MCT-PSO, techniques with 5, 10, 15, 20, number of tasks.

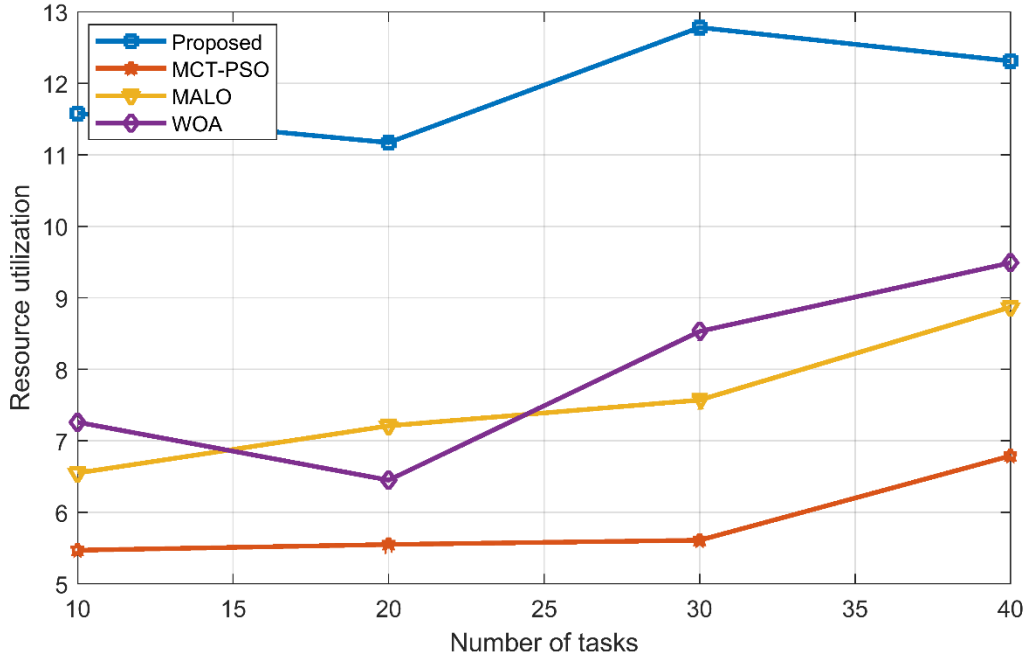

Figure. 3 Comparative analysis of resource utilization using Proposed, MALO, WOA, and MCT-PSO, techniques with 5, 10, 15, 20, number of tasks.

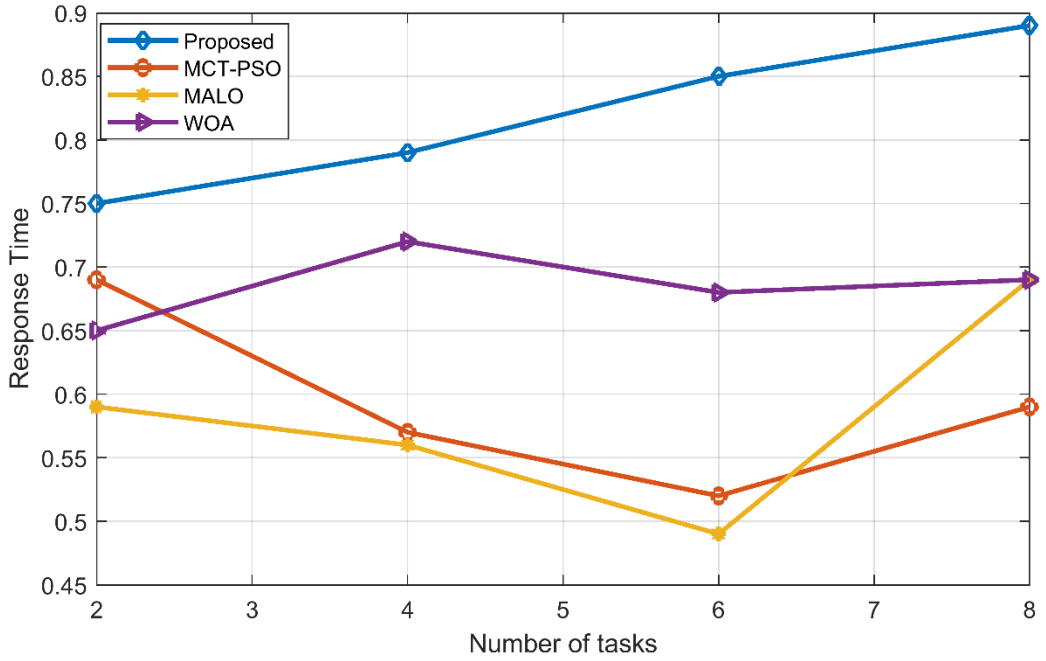

Figure. 4 Comparative analysis of response time using Proposed, MALO, WOA, and MCT-PSO, techniques with 5, 10, 15, 20, number of tasks.

V. Conclusion & Future Work

The objective of this paper is to optimized the selection process of tasks for the allocation of resources in cloud computing. The proposed algorithm of load balancing optimized the phenomenon value of ant colony optimization and reduced the discontinuity of ants for the processing of virtual machines. The selection of virtual machines impacts the management of tasks and improves the capacity of cloud infrastructure. The proposed algorithm of load balancing is simulated in the Cloud Simulator software. The results of the proposed algorithm compare with existing algorithms of load balancing such as MCT-PSO, MALO, and WOA. The analysis of the results suggests that the proposed algorithm is very efficient compared to the existing algorithm for cloud computing. We have verified that each of the four computations is unexpectedly distinct. Documents for the Sim setup. In each scenario, the suggested method provides the best result in the event that different virtual machines arise. Relocation time 19.06, reaction time 224.26, throughput 100%, dependability 89.14, versatility 71.38, accessibility 51.28, and energy consumption 2.55 all come out to be 23.

References

[1]. Ragmani, Awatif, Amina Elomri, Noreddine Abghour, Khalid Moussaid, and Mohammed Rida. "An improved hybrid fuzzy-ant colony algorithm applied to load balancing in cloud computing environment." Procedia Computer Science 151 (2019): 519-526.

[2]. Pradhan, Arabinda, and Sukant Kishoro Bisoy. "A novel load balancing technique for cloud computing platform based on PSO." Journal of King Saud University-Computer and Information Sciences 34, no. 7 (2022): 3988-3995.

[3]. Agarwal, Ronak, Neeraj Baghel, and Mohd Aamir Khan. "Load balancing in cloud computing using mutation based particle swarm optimization." In 2020 International Conference on Contemporary Computing and Applications (IC3A), pp. 191-195. IEEE, 2020.

[4]. Priya, V., C. Sathiya Kumar, and Ramani Kannan. "Resource scheduling algorithm with load balancing for cloud service provisioning." Applied Soft Computing 76 (2019): 416-424.

[5]. Kazeem Moses, Abiodun, Awotunde Joseph Bamidele, Ogundokun Roseline Oluwaseun, Sanjay Misra, and Adeniyi Abidemi Emmanuel. "Applicability of MMRR load balancing algorithm in cloud computing." International Journal of Computer Mathematics: Computer Systems Theory 6, no. 1 (2021): 7-20.

[6]. Gamal, Marwa, Rawya Rizk, Hani Mahdi, and Basem E. Elnaghi. "Osmotic bio-inspired load balancing algorithm in cloud computing." IEEE Access 7 (2019): 42735-42744.

[7]. Pradhan, Arabinda, and Sukant Kishoro Bisoy. "A novel load balancing technique for cloud computing platform based on PSO." Journal of King Saud University-Computer and Information Sciences (2020).

[8]. Tyagi, Neha, Ajay Rana, and Vineet Kansal. "Creating elasticity with enhanced weighted optimization load balancing algorithm in cloud computing." In 2019 Amity International Conference on Artificial Intelligence (AICAI), pp. 600-604. IEEE, 2019.

[9]. Ala'Anzy, Mohammed, and Mohamed Othman. "Load balancing and server consolidation in cloud computing environments: a meta-study." IEEE Access 7 (2019): 141868-141887.

[10]. Alguliyev, Rasim M., Yadigar N. Imamverdiyev, and Fargana J. Abdullayeva. "PSO-based load balancing method in cloud computing." Automatic Control and Computer Sciences 53, no. 1 (2019): 45-55.

[11]. Lim, JongBeom, and DaeWon Lee. "A load balancing algorithm for mobile devices in edge cloud computing environments." Electronics 9, no. 4 (2020): 686.

[12]. Farrag, Aya A. Salah, Safia Abbas Mohamad, and M. El Sayed. "Swarm Intelligent Algorithms for solving load balancing in cloud computing." Egyptian Computer Science Journal 43, no. 1 (2019): 45-57.

[13]. Shahid, Muhammad Asim, Noman Islam, Muhammad Mansoor Alam, Mazliham Mohd Su'ud, and Shahrulniza Musa. "A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach." IEEE Access 8 (2020): 130500-130526.

[14]. Chowdhury, Shusmoy, and Ajay Katangur. "Threshold based load balancing algorithm in cloud computing." In 2022 IEEE International Conference on Joint Cloud Computing (JCC), pp. 23-28. IEEE, 2022.

[15]. Maswood, Mirza Mohd Shahriar, MD Rahinur Rahman, Abdullah G. Alharbi, and Deep Medhi. "A novel strategy to achieve bandwidth cost reduction and load balancing in a cooperative three-layer fog-cloud computing environment." IEEE Access 8 (2020): 113737-113750.

[16]. Adil, Muhammad, Said Nabi, and Summair Raza. "PSO-CALBA: Particle swarm optimization based content-aware load balancing algorithm in cloud computing environment." Computing and Informatics 41, no. 5 (2022): 1157-1185.

[17]. Achar, Sandesh. "Adopting artificial intelligence and deep learning techniques in cloud computing for operational efficiency." International Journal of Information and Communication Engineering 16, no. 12 (2022): 567-572.

[18]. Bal, Prasanta Kumar, Sudhir Kumar Mohapatra, Tapan Kumar Das, Kathiravan Srinivasan, and Yuh-Chung Hu. "A joint resource allocation, security with efficient task scheduling in cloud computing using hybrid machine learning techniques." Sensors 22, no. 3 (2022): 1242.

[19]. Gundu, Srinivasa Rao, Charan Arur Panem, Anuradha Thimmapuram, and R. S. Gad. "Emerging computational challenges in cloud computing and RTEAH algorithm based solution." Journal of Ambient Intelligence and Humanized Computing (2022): 1-15.

[20]. Javadpour, Amir, Ali Majed Hossein Abadi, Samira Rezaei, Mozhdeh Zomorodian, and Ali Shokouhi Rostami. "Improving load balancing for data-duplication in big data cloud computing networks." Cluster Computing 25, no. 4 (2022): 2613-2631.

[21]. Nethaji, S. V., and M. Chidambaram. "Differential Grey Wolf Load-Balanced Stochastic Bellman Deep Reinforced Resource Allocation in Fog Environment." Applied Computational Intelligence and Soft Computing 2022 (2022).

[22]. Pradhan, Arabinda, Sukant Kishoro Bisoy, and Mangal Sain. "Action-Based Load Balancing Technique in Cloud Network Using Actor-Critic-Swarm Optimization." Wireless Communications and Mobile Computing 2022 (2022).

[23]. Zhang, An-Ning, Shu-Chuan Chu, Pei-Cheng Song, Hui Wang, and Jeng-Shyang Pan. "Task scheduling in cloud computing environment using advanced phasmatodea population evolution algorithms." Electronics 11, no. 9 (2022): 1451.

[24]. Sarhadi, Ali, and Javad Akbari2 Abbas Karimi. "Economic Based Scheduling and Load Balancing Algorithms in Cloud Computing Using Learning Automata." (2022).

[25]. Le Ngoc, Hieu, and Hung Tran Cong. "Enhancing Load Balancing in Cloud Computing through Adaptive Task Prioritization." Journal of Computer Science and Technology Studies 5, no. 2 (2023): 01-15.

[26]. Muthusamy, Arvindhan, and Rajesh Kumar Dhanaraj. "Dynamic Q-Learning-Based Optimized Load Balancing Technique in Cloud." Mobile Information Systems 2023 (2023).

[27]. Lahande, Prathamesh Vijay, Parag Ravikant Kaveri, Jatinderkumar R. Saini, Ketan Kotecha, and Sultan Alfarhood. "Reinforcement Learning approach for optimizing Cloud Resource Utilization with Load Balancing." IEEE Access (2023).

[28]. Al Reshan, Mana Saleh, Darakhshan Syed, Noman Islam, Asadullah Shaikh, Mohammed Hamdi, Mohamed A. Elmagzoub, Ghulam Muhammad, and Kashif Hussain Talpur. "A Fast Converging and Globally Optimized Approach for Load Balancing in Cloud Computing." IEEE Access 11 (2023): 11390-11404.

[29]. Mangalampalli, Sudheer, Ganesh Reddy Karri, Mohit Kumar, Osama Ibrahim Khalaf, Carlos Andres Tavera Romero, and GhaidaMuttashar Abdul Sahib. "DRLBTSA: Deep reinforcement learning based task-scheduling algorithm in cloud computing." Multimedia Tools and Applications (2023): 1-29.

[30]. Kakade, Manoj Subhash, Anupama Karuppiah, Mayank Mathur, Priyansh Parikh, Raghav Dhir, and Tejas Gokhale. "Tasks scheduling and load balancing on a cloudlet system using Qualcomm 410c." In Proceedings of the 26th world multi-conference on systemics, cybernetics and informatics (WMSCI 2022). 2022.