Efficient Load Balancing Algorithm using Improved Particle Swarm Optimization for Cloud Computing Environment

Sanjay Gupta

sanjgupta24@gmail.com

Department of computer Science & Engg.

Lecturer, Polytechnic College, Satna (M.P)

Abstract

Cloud computing is an emerging technology in the current scenario of data storage and computation. The umbrella of cloud computing offers various services over the internet, such as infrastructure, software, and application platforms. The process of delivering services required multiple resources in cloud computing. The virtual machine is a utility component of cloud computing systems. The performance of a cloud computing system depends on the management of virtual machines and the allocation of resources. The dynamic load balancing approach deals with swarm intelligence algorithms. In this paper, we propose a meta-heuristic algorithm called MPSO based on particle swarm optimization for virtual machine (VM) scheduling and balancing the load in cloud computing. The particle swarm optimization set the diverse properties of the virtual machine and requested the job. The define fitness constraints function partially allocates jobs for dedicated machines and distributes them according to the process job scheduler. For the evaluation of performance, we used cloud simulator software, which is called cloud analyst. The cloud analysis software is a bag of composition of cloud environment and load balancing policy. In the scenario of policy design, there are two types of policies: one is a genetic algorithm policy and the other is a PSO-based policy. The PSO-based policy reduces the load effect by approximately 10–12% in the compression of the genetic algorithm.

Keywords: - Cloud Computing, Load Balancing, Virtual Machine, CloudSim, PSO

Introduction

Cloud computing is an open platform for on-demand access to services. The stack of cloud computing offers three types of services: platform as a service (PAAS), infrastructure as a service (IAAS) and software as a service (SAAS). Cloud computing services make use of emerging computing technologies such as fog computing and edge computing. Cloud computing's centralized management is a significant impediment to resource allocation and task scheduling [1,2,3]. The management of task scheduling and resource allocation handles load balancing. The unbalanced scenario of resources and allocation management suffers from an unsupported capacity of load and a decline in the system capacity of cloud computing. Load balancing approaches deal with the conflict between resource allocation and task scheduling. The load balancing approach is divided into two sections: static load balancing and dynamic load balancing. The static load balancing approach uses conventional scheduling algorithms such as FCFS, SJF, and round robin (RR). For resource and task allocation, the dynamic load balancing approach employs a partition-based scheduling algorithm and a heuristic-based function. Because of the exponential growth in the number of customers using the service, an effective load balancing scheme is essential in cloud computing [4,5]. It is important in digital devices, servers, and networks because it allows for maximum resource utilization, increased scalability, and the elimination of disruptions and oversupply. In a cloud, there are various types of loads such as memory, network, and so on. The load balance mechanism must identify the load type and redistribute the extra loads to the undercoated VMs to maximize resource utilization and minimize server response time. Load balancing is a term in parameters such as data processing time and VM response time that has no effect on data centres with the request of time changing in the cost computation of the algorithm proven in the efficient terms of load balancing. Load balancing is the most commonly used simplicity algorithm, and it is sufficient to be efficient in either case[6,7]. Load balancing reduces execution time, waiting time, response time, and throughput on cloud computing. The load balance has been used in optimization for simple methods such as loading processors in tasks for transferring the loaded. In a cloud environment, it is required to schedule computational resources in such a way that service providers get most out of their resources and also users manage their functional programs with lowest cost. This paper proposed an improved particle swarm optimization algorithm for load balancing. The proposed algorithm enhances the balancing factor of load in response to resource. The rest of the paper describes in section II related work, in section III proposed algorithm, in section IV experimental analysis, in section conclusion & future work.

II. Related work

Wei-Tao Wen and others [1,] in this research, we provide a brand-new distributed VM migration method based on the Ant Colony Optimization met heuristic algorithm. The local migration agent in our ACO-VMM strategy independently monitors resource usage and initiates the migration. In this research, we present a highly scalable and reliable distributed VM migration approach dubbed ACO-VMM. The ACO-VMM technique uses Ant Colony Optimization to discover a nearly optimal solution to the overload balance problem in data centres because finding the ideal mapping relationship is strictly NP hard. Kai Pan and others [2] this technique creates a suitable resource-task allocation model by taking into account the peculiarities of complex networks. The simulated studies demonstrated that this methodology can enhance cloud resource consumption and load balancing. These traits are fully combined in this paper to set up a cloud computing load balancing model based on complex networks. Prior references discussed optimization strategies, but they haven't completely taken on sound effects.Akash Dave and others [3] This paper presents a comprehensive overview of the load balancing optimization methods based on evolutionary and swarm-based algorithms that will aid in resolving optimization issues or inefficient resource use. In this work, various load balancing approaches are examined and described. This looked at a few evolutionary and swarm-based optimization strategies and examined how well they applied to various environments. Jigna Acharya and coworkers [4] A technology called cloud computing uses dynamic virtual machine (VM) allocation to simplify activities. Users bill resources depending on their requests as they utilise them. The difficulties faced by cloud providers are numerous. One of his greatest challenges is load balancing, which involves using an optimization technique that is inspired by nature. With this in mind, we compared our algorithm to another FCFS algorithm. The PSO algorithm produces superior results to FCFS. Based on minimal MAKESPAN time, a comparison is made. Ronak R. Patel and others [5] Numerous soft computing techniques, including ant colony, honey bee, stochastic hill climbing, genetic, honey bee, and throttled algorithms, are used to optimise the scheduling of resources and jobs on the cloud. There are many different soft computing approaches available. We applied them all and discovered that they all either use random selection or natural selection, so in that case, those resources are also taken into account if they are unable to handle the tasks. The cloud environment is now overburdened as a result. Kanakala, V. Ravi Teja, and colleagues [6] In this paper, we cover load balancing, various proposed algorithms for spreading the load among the nodes, as well as the factors that are taken into consideration when determining the optimum algorithm to balance the load. The classification of load balancing algorithms was covered in depth in this work, and the current load balancing methods for cloud computing are summarised together with the performance enhancements that have been suggested. In order to balance the load in cloud computing, the performance of the implemented algorithms fell short of cloud criteria. Priyadarshini, A.P., et al. [7] This work conducts an experiment employing certain well-known algorithms, such as the Particle Swarm Optimization Load Balancing Algorithm (PSO), the Cat Swarm Optimization Load Balancing Algorithm (CSO), and the Genetic Load Balancing Algorithm (GA). The Genetic Load Balance Algorithm is primarily concerned with balancing, whereas the Particle Swarm Optimization Load Balancing Algorithm and Cat Swarm Optimization Load Balancing Algorithm are distributed balancing algorithms. These three methods all belong to a dynamic environment. The most important factor in the cloud computing environment is cost details. Prabhjot Kaur and coworkers [8] In this, we suggest a technique for resource allocation based on the load of virtual machines (VMs). A load predictor is used to forecast the load on each node and assign resources in accordance with that prediction in order to solve the challenge of scheduling VM resources in a cloud computing environment. In this research, a VM load balancing technique was suggested, which optimises the number of servers in use while simultaneously supporting green computing by dynamically allocating data centre resources based on application demands. R. Rajeshkannan and colleagues [9] This research aids in locating the most efficient load balancing algorithm for maximising throughput, minimising response times, and preventing overload. The load balancer is recommended in every circumstance to provide service continuity and handle increased traffic. This study investigated the methodological analysis of several load balancing methods for cloud network services. Huangning Chen and colleagues [10] in this paper, we suggest a particle swarm optimization-based soft real-time job scheduling technique for cloud computing. Along with cost and makespan, the optimum objectives also take deadline missing ratio and load balancing degree into account. The proposed approach is efficient, according to the simulation studies. To increase the dependability of cloud systems, we will concentrate on offering fault-tolerant mechanisms in the future. A.I. Awad and others [11] Task scheduling, which is crucial to the effectiveness of the entire cloud computing infrastructure, is the most crucial requirement in a cloud computing environment. In cloud computing, task scheduling refers to allocating the optimal resources for a task's execution while taking into account a variety of factors, including time, cost, scalability, make-up, dependability, availability, throughput, resource consumption, and so forth. In order to demonstrate how much LBMPSO can save in terms of make span, execution time, round trip time, and transmission cost, we compared it to regular PSO, randomness, and the Longest Cloudlet to Fastest Processor (LCFP) algorithm. Ashalatha, R and others [12] Applications' dynamic allocation causes sessions to move from one location to another as needed. The primary focus of this study is on various dynamic load balancing techniques utilised for effective cloud resource management. The two main elements in the cloud computing paradigm are load balancing and resource management challenges. The primary area of research in the modern world is the dynamic load balancing strategies that are covered in this paper. These problems can be more thoroughly explained and used in a real-time open cloud setting. Shubhakankshi Goutam [13] is one example. The survey of load balancing methods now in use is presented in this study, along with a list of their drawbacks and potential research areas. In this essay, we'll talk about current technologies in use as well as their benefits and drawbacks. Researchers continue to try to increase the performance aspects and overcome the drawbacks of different strategies. The subject of distributed load balancing is equally significant and open for research into new algorithms and the enhancement of existing ones. In addition to Emanuel Ferreira Coutinho [14], this study suggests an elastic cloud computing architecture based on autonomous computing ideas. We created two studies using micro benchmarks and applied them to both private and hybrid clouds to validate them. According to experiments for verifying the architecture, it is feasible to employ autonomous computing and cloud computing along with various technologies and different providers. Different criteria should be utilised for rules design in future work for the evolution of the proposed architecture, such as average request response time. citing Nidhi Bansal et al. [15] This algorithm works well to balance the load, but it is inefficient when it comes to cost performance. In this work, Clouds Simulator has also conducted a comparative analysis of several scheduling strategies. According to its rising popularity day by day, cloud computing is highly crucial for proper resource usage and optimal solutions with work scheduling algorithms. It was demonstrated that the virtual machine tree optimised task scheduling algorithm's cost parameter is not very effective when load balancing and other factors are taken into account. Others include Zhu, Yongfei, and others [16] The Particle Swarm Optimization (PSO) algorithm is neoteric and is highly commended in the field of cloud computing load balancing, but recently a more neoteric approach that deploys the classifier into load balancing has been presented. Additionally, the absence of training samples has reduced classification accuracy, and the cloud work allocation approach is inadequate, resulting in an unfavourable load balancing degree. The red-black tree algorithm has good running speed, but there is a significant issue with its cloud job management and degree of balance. Finally, the excellence in balance degree and the flow time of cloud work using the PSO algorithm are combined for the upgraded PSO algorithm. Aslanzadeh, Shahrzad, et al. [17] in this research, we offer a novel load balancing method based on the Endocrine algorithm, a human hormone system regulatory behaviour model. Our suggested approach applies the self-organizing mechanism between overloaded VMs to achieve system load balancing. We modified the open source cloud simulator Cloudsim in order to examine our suggested approach. According to the testing findings in a simulated environment, Endocrine-PSO provides the best opportunity for moving tasks from a VM that is overcrowded to a VM that is more powerful. Elhossiny Ibrahim and colleagues [18] an improved task scheduling method for the cloud computing environment has been developed in this study. An improved work scheduling technique for the cloud computing environment has been proposed in this paper. The improvement algorithm's fundamental concept is to assign the available VMs to the desired tasks according to their processing power. Pooja Mangla and colleagues [19] Task scheduling techniques are a significant issue that has emerged with the introduction of cloud computing. The effectiveness and efficiency of the services provided by the clouds are significantly influenced by task scheduling. Task scheduling's primary goal is to cut down on the amount of time needed to complete all of the running jobs, or to shorten the make span. Rakshanda and others [20] the popularity of the Genetic Algorithm (GA) and all of its variations can be attributed mostly to their simplicity and intuitiveness. The mechanics of PSO, a popular heuristic search technique, were influenced by the swarming or typical behaviour of natural populations. In order to handle the job scheduling problem in cloud computing and reduce response time, this research provides a unique solution based on a hybrid PSO-GELS algorithm. Every atom stands for a workable answer. Based on rounding off actual values, the position vector is changed from continuous values to discrete values.

III. Proposed Methodology

In this section discuss the proposed algorithm for load balancing in cloud computing. The proposed algorithm based on the concept of dynamic load balancing technique. The dynamic load balancing process accrued the process of scheduling based on task and resource allocation. For the allocation of task and resource used two different technique one is searching of task according to the dedicated job and other is execution of task incorporation of process. For the execution of task used particle of swarm optimization. The particle swarm optimization the load and perform the task[11,12]. The particle of swarm optimization selects the input of virtual machine load in terms of M^R theM show the value of domain machine load and R shows that Job component of R real load. The R Job content {r1, r2, r3, r4,…….rn} describe the artificial particle as population. The unique job relation of input load set the velocity of particle. If the job selection value is change then next iteration moves the update of velocity. These terms describe as, the particle's job value, Rid and its near value of particle; Rgd is a velocity value of optimization job space. The random values for job are fet1 and fet2 are used for the local and global value selection of particle, that is, to make the optimal solution. The values of c1 and c2 manage the value of velocity of Rid and Rgd in deciding the particle's next movement velocity. At that each iteration changes the velocity of swarm and creates new job subset for selection of job. The derivation of equation in (c) and (d)

$$v_{id=W\times V_{id}+C1\times fet1\times\left(R_{id-x_{id}}\right)+C_2\times rand2\times(R_{gd-X_{id}})}c \ldots\ldots (c)$$

$$X_{id=X_{id}+V_{id}} \ldots\ldots\ldots\ldots (d)$$

where w denotes the value of job matrix; Rid is the position of particle value, Rgd is the position of global value best fitness value, c1 and c2 are constants and are known as acceleration coefficients; d denotes the dimension of the problem space; fet1, fet2 are random values in the range of (0, 1). The finally optimal job is executed to data centre. Here first define the fitness constraints function for the allocation of job corresponding to virtual machine. here define the Mi is the set of virtual machine and Ri is the set of job and Lk is total load over the cloud. The fitness constraints function define as

$$L_k\left(M_{i,R_j}\right) = \frac{\left[\tau\left(M_i,R_j\right)\right]\alpha[\rho(M_i,R_i)]\beta}{\sum_{h\in lk}(Mi,Rj)^{[\tau(Mi,h)]\alpha[\rho(Ri,h)]\beta}}, Ri\epsilon Lk(Mi,Ri) \ldots\ldots\ldots\ldots (1)$$

Here $\tau(Mi,Ri)$ is the partial load and available virtual machine.

Here discuss the step

1. Input R1,R2…………………………Rn in M1,M2,………..Mn for total population
2. Map(Mi,Ri)
3. Define particle as job and distribute all in Ri
4. Do
5. For each particle do
6. For each Mi do
7. Select next job
8. End for
9. Calculate individual job cost based on formula 4.6.1
10. If estimated value Vi equal to job then
11. Load map (Mi, Ri)
12. Update the velocity of job using formula(c)
13. End if
14. End for
15. Population is empty
16. Terminate the process.

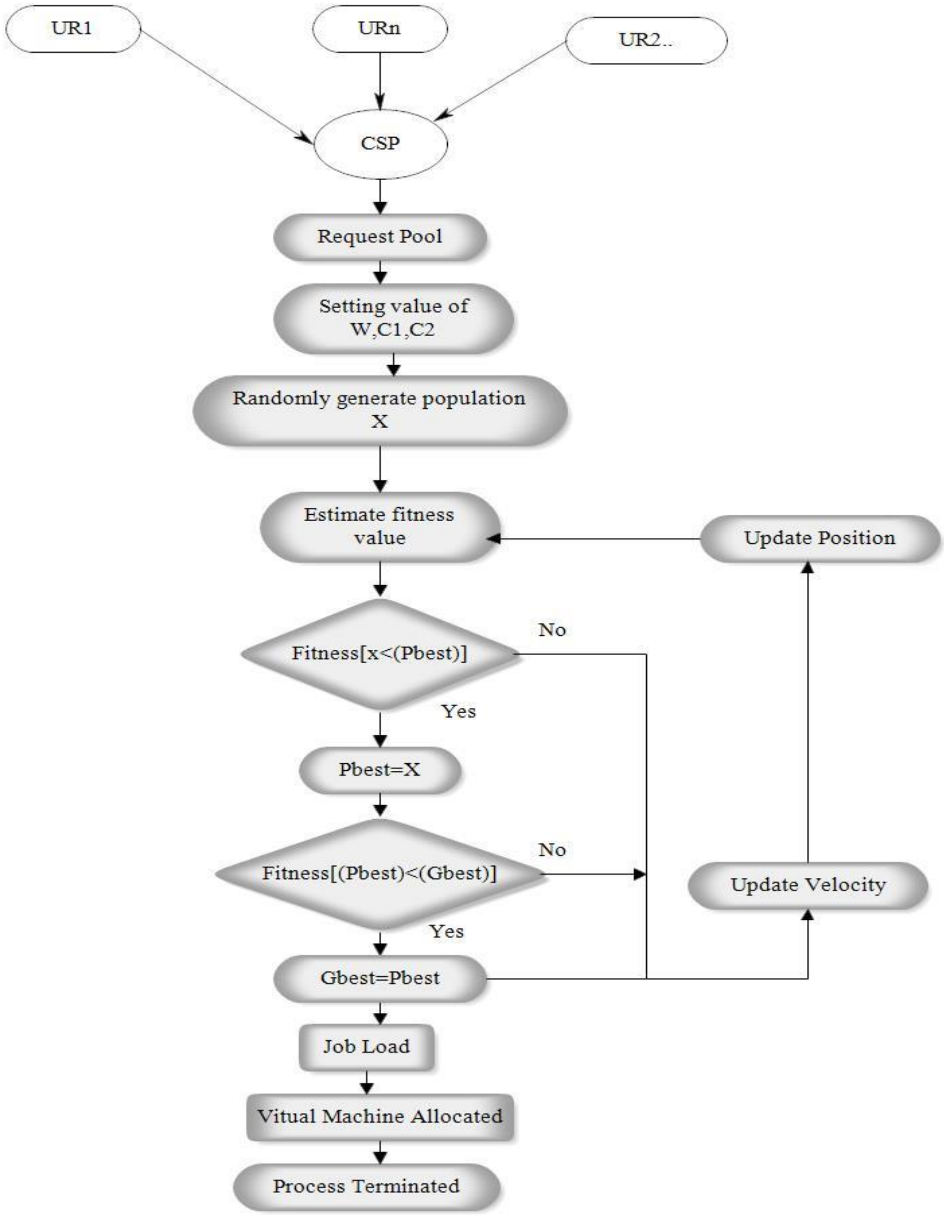

Figure 1 process block diagram of load balancing process based on PSO

IV Experimental analysis

To validate the proposed algorithm of load balancing in cloud computing environment simulates in CloudSim 3.0 Simulator. The CloudSim 3.0 Simulators is java based program and provides the data centers, virtual machine, traffic data and all variables and parameters of cloud environments [13,14,15].

Table 1: Show the Response Time and Processing Time analysis for PSO, ACO, and Method

Data Set	Techniques	ORT		DCPT	
UB5 DC4	PSO	Average	318.771	Average	0.376
		Minimum	255.635	Minimum	0.035
		Maximum	388.632	Maximum	0.656
	IPSO	Average	325.795	Average	0.386
		Minimum	258.638	Minimum	0.08
		Maximum	389.665	Maximum	0.658

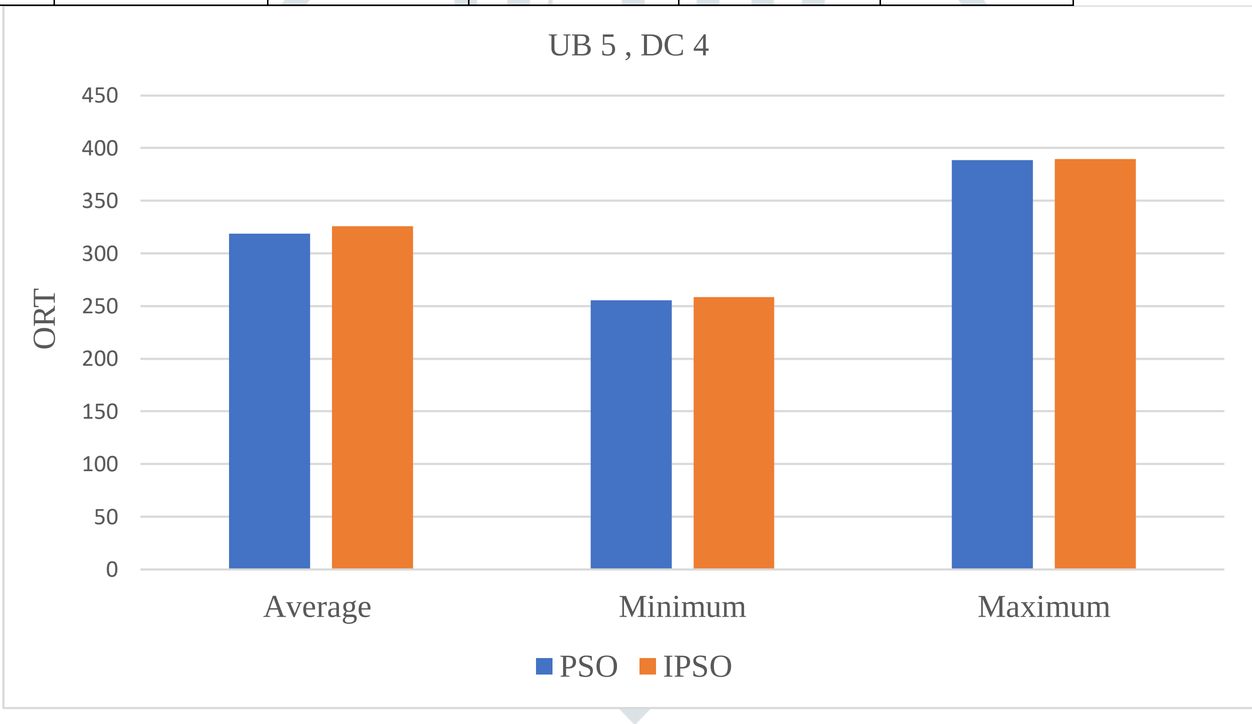

Figure 1: shows the comparative performance of overall response time (ORT) for UB 5 and DC 4 using PSO, IPSO, in terms of average, minimum, and maximum values in milliseconds.

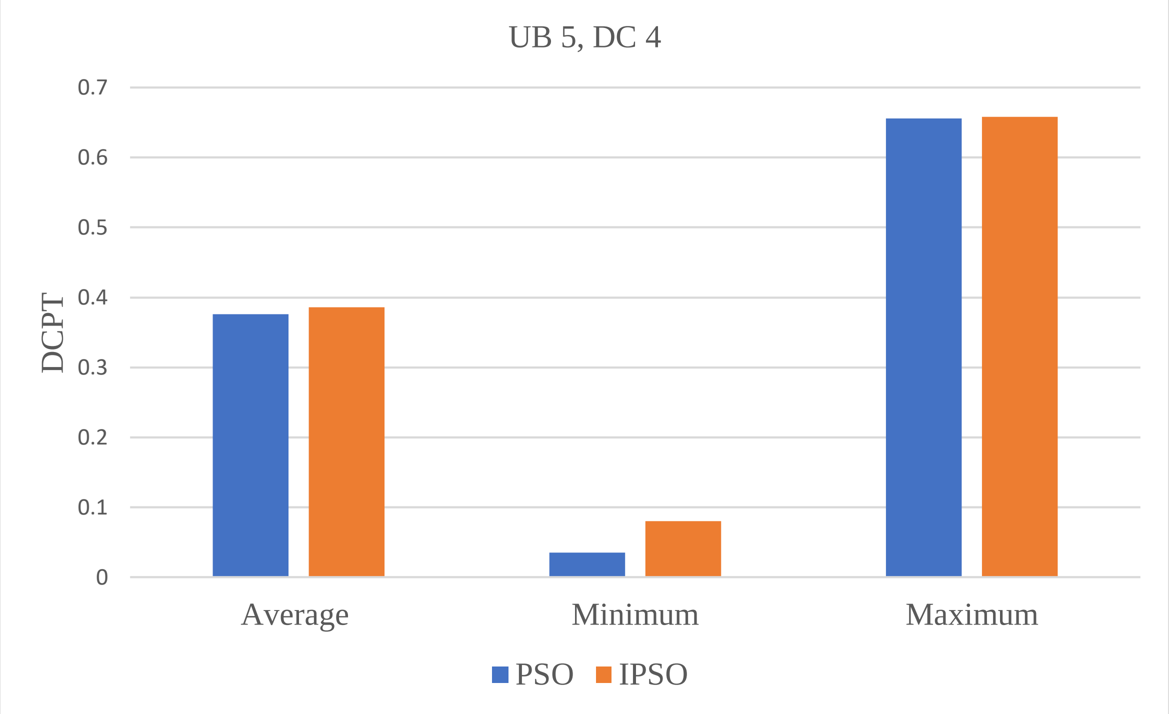

Figure 2: shows the comparative performance of overall response time (DCPT) for UB 5 and DC 4 using PSO, IPSO, in terms of average, minimum, and maximum values in milliseconds.

V Conclusion & Future Work

The improved particle swarm optimization is proposed in this paper to form swarm computing and find load balancing for VM placement. Proposed method for providing an efficient cloud computing environment. PSO and COMODL collaborate in the proposed algorithm to select the best VM to migrate to the best PM. Furthermore, PSO activates the most appropriate osmotic host among all PMs in the system to reduce power consumption. In two experiments with fixed and variable loads, the proposed algorithm was simulated to calculate performance at various metrics. The particle swarm optimization-based load balancing policy is very efficient for assigning jobs to dedicated virtual machines. The minimum time span was a problem for the partial allocation of job allocation policy. The efficiency factor of particle swarm optimization policy is affected by the minimum time span factor.

References

[1]. Wen, Wei-Tao, Chang-Dong Wang, De-Shen Wu, and Ying-Yan Xie. "An ACO-based scheduling strategy on load balancing in cloud computing environment." In *2015 Ninth international conference on frontier of computer science and technology*, pp. 364-369. IEEE, 2015.

[2]. Pan, Kai, and Jiaqi Chen. "Load balancing in cloud computing environment based on an improved particle swarm optimization." In *2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 595-598. IEEE, 2015.

[3]. Dave, Akash, Bhargesh Patel, and Gopi Bhatt. "Load balancing in cloud computing using optimization techniques: A study." In *2016 International Conference on Communication and Electronics Systems (ICCES)*, pp. 1-6. IEEE, 2016.

[4]. Acharya, Jigna, Manisha Mehta, and Baljit Saini. "Particle swarm optimization based load balancing in cloud computing." In *2016 International Conference on Communication and Electronics Systems (ICCES)*, pp. 1-4. IEEE, 2016.

[5]. Patel, Ronak R., Swachil J. Patel, Dhaval S. Patel, and Tushar T. Desai. "Improved GA using population reduction for load balancing in cloud computing." In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2372-2374. IEEE, 2016.

[6]. Kanakala, V. RaviTeja, V. Krishna Reddy, and K. Karthik. "Performance analysis of load balancing techniques in cloud computing environment." In *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1-6. IEEE, 2015.

[7]. Priyadarshini, A. P., R. Sharmistha, and K. P. Prasant. "Load balancing adaption of some evolutionary algorithms in cloud computing environment." *African Journal of Computing & ICT* 8, no. 2 (2015).

[8]. Kaur, Prabhjot, and Pankaj Deep Kaur. "Efficient and enhanced load balancing algorithms in cloud computing." *International Journal of Grid Distribution Computing* 8, no. 2 (2015): 9-14.

[9]. Rajeshkannan, R., and M. Aramudhan. "Comparative study of load balancing algorithms in cloud computing environment." *Indian Journal of Science and Technology* 9, no. 20 (2016): 1-7.

[10]. Chen, Huangning, and Wenzhong Guo. "Real-time task scheduling algorithm for cloud computing based on particle swarm optimization." In *Second International Conference on Cloud Computing and Big Data in Asia*, pp. 141-152. Springer, Cham, 2015.

[11]. Awad, A. I., N. A. El-Hefnawy, and H. M. Abdel_kader. "Enhanced particle swarm optimization for task scheduling in cloud computing environments." *Procedia Computer Science* 65 (2015): 920-929.

[12]. Ashalatha, R., and Jayashree Agarkhed. "Dynamic load balancing methods for resource optimization in cloud computing environment." In *2015 Annual IEEE India Conference (INDICON)*, pp. 1-6. IEEE, 2015.

[13]. Megharaj, Geetha, and K. G. Mohan. "A survey on load balancing techniques in cloud computing." *IOSR Journal of Computer Engineering (IOSR-JCE)* 18, no. 2 (2016): 55-61.

[14]. Coutinho, Emanuel Ferreira, Danielo Gonçalves Gomes, and José Neuman de Souza. "An autonomic computing-based architecture for cloud computing elasticity." In *2015 Latin American Network Operations and Management Symposium (LANOMS)*, pp. 111-112. IEEE, 2015.

[15]. Bansal, Nidhi, Amit Awasthi, and Shruti Bansal. "Task scheduling algorithms with multiple factor in cloud computing environment." In *Information systems design and intelligent applications*, pp. 619-627. Springer, New Delhi, 2016.

[16]. Zhu, Yongfei, Di Zhao, Wei Wang, and Haiwu He. "A novel load balancing algorithm based on improved particle swarm optimization in cloud computing environment." In *International conference on human centered computing*, pp. 634-645. Springer, Cham, 2016.

[17]. Aslanzadeh, Shahrzad, and Zenon Chaczko. "Load balancing optimization in cloud computing: Applying Endocrine-particale swarm optimization." In *2015 IEEE International Conference On Electro/Information Technology (Eit)*, pp. 165-169. IEEE, 2015.

[18]. Ibrahim, Elhossiny, Nirmeen A. El-Bahnasawy, and Fatma A. Omara. "Task scheduling algorithm in cloud computing environment based on cloud pricing models." In *2016 World Symposium on Computer Applications & Research (WSCAR)*, pp. 65-71. IEEE, 2016.

[19]. Mangla, Pooja, and Sandip Kr Goyal. "A study of meta-heuristic based techniques of load balancing in cloud computing environment." *International Journal in Management and Social Science* 4, no. 4 (2016): 194-7.

[20]. Vinod, Rakshanda1 Dr Kanwal2 Garg. "Particle Swarm Optimization and Gravitational Emulation Based Hybrid Load Balancing Strategy in Cloud Computing."