Load Balancing In Cloud Computing Using Optimization Algorithm

Deepika Arora
Assistant Professor
deepikacse@geeta.edu.in
Geeta Engineering College, Panipat

MUNEER AHMAD SOFI
Research Scholar
munir.sofi7@gmail.com
Geeta Engineering College, Panipat

Abstract - Cloud computing may be viewed as a broad spectrum of networks to connect many private or public networks. Delivering a scalable framework for application, data and file storage is the core of this technology. Cloud systems may malfunction when virtual machines are distributed with excessive load. In this work, a flexible task scheduling strategy based on ACO algorithm has been applied. This strategy can effortlessly assign the task to the most reliable VM (virtual machine). The weight counted on a VM is the measure of its consistency. This work uses MATLAB tool to implement the recommended and old strategies. The simulations are carried out, showing that ACO based task scheduling strategy significantly decreases the total execution time with no performance loss than the existing solutions.

Keywords - Virtual Machine Migration, Load Balancing, ACO, Weight Based Algorithm

I. INTRODUCTION

The cloud computing can be described as an on-demand service pool which connects various servers to each other for providing services to aiming clients. The cloud providers may contain direct access to these services. Therefore the resources can be used according to the requirement[1]. The user can extracts and modifies the data stored in the clouds. The different services to the user are provided on demand using a feature called "cloud service provider". This trait makes certain that the amount of services being utilized for any number of times can be employed for calculating the expense of the user to access that service. The cloud computing system provides extremely complicated applications in different environments. In addition, some skilled concentrated services are provided in each environment. In cloud computing, common group of resources is provided to the users. Using cloud computing[2], the users can utilize these resources according to their need everywhere. The main objective of this technology is to maintain the minimum cost to access the services. It is analyzed that the software and hardware assets obtained using internet remain present in the virtual system and supports to provide the services. The user accesses a common group of resources using cloud computing on the basis of demand[3]. The virtualization allows user to subscribe and use the services for a certain time period by getting access of the common group of resources using cloud computing. The cloud computing reduces application cost reduced and allows the user to access more hardware parts [4][5]. The load balancing is a technique using which the whole load of the network is divided and allocated to the several resources or devices. This approach improves the competence of resource employment and enhances the response time of any task as well. Moreover, this approach avoids the condition in which some centers are under loaded while others are overloaded. The trustworthiness and accessibility of the data can be improved with the help of redundancy by using various components for load balancing. The load is measured within the system in terms of the network load, memory usage, and CPU load [6][7]. For classifying the load balancing, following two methods are applied:

A. Static: The current information of the system is not needed by this technique. This technique works on the earlier information. In order to establish this cloud service provider, the homogeneous resources are utilized in this method. In addition, the overall resources do not show flexibility within the static environment. The memory, performance, capability and processing power are needed on the basis of the need. The changes are rejected in such situations in runtime[8]. This technology is not very competent in the heterogeneous environments. This technology can be adopted easily in the static environments. The resources are distributed in this environment on the basis of first-come-first-server strategy. They may be classified as:

- ➔ **Sender initiated**: In this situation, the sender starts the procedure of load balancing.
- ➔ **Receiver initiated:** In this situation, the receiver starts the load balancing procedure.

B. Dynamic: This approach uses the current state of the system and does not need earlier information. In this

environment, Individual resources are established in the clouds. The resources are dynamic in nature in this environment. This approach does not need previous information and just uses the execution time information.

➔ **Symmetric**: In this case, both sender and receiver begin the load balancing process.

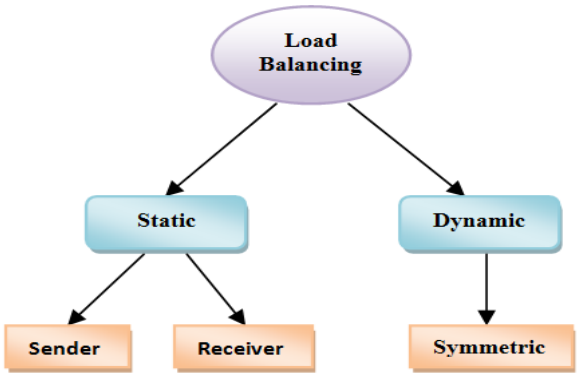

Figure 1: Types of load balancing

The process in which a request is shifted from one virtual machine to other virtual machine without informing the end user is called virtual machine migration. The systematic upgrades, disaster recovery and load balancing are some of the reasons behind the virtual machine migration [9].

Virtual Machine Migration Types: The two types of migrations are described below:

i. **Hot (live) Migration:** In this migration, the virtual machines run continuously. This status is not lost even after the switching from one virtual machine to other virtual machine.

ii. **Cold (non-live) Migration:** The status of the virtual machine is not lost because of the suspension of virtual machine. However, the client notices the interruption in the service. Towards the end, the machine performs continuously at the target node [18].

The genetic algorithm is an approach that is recognized as a branch of computer science. This algorithm is applied to find answers of the optimization related issues. The genetic algorithms are also termed as evolutionary algorithms. In order to perform genetic algorithm, the developmental science involves various techniques[10]. The fitness potential is described by representing the inherited computations. The inherited computations install system randomly. This algorithm is modified by the monotonous applications. This algorithm needs various factors such as crossover operators, mutation and selection. Various researchers have enhanced the inherited computations as a solution to modernize different fields. The inherited computations provide solutions for resolving a problem. These algorithms are becoming very popular in recent times as the modification handles a lot of problems [11]. The natural parallelism encourages the usage of distributed machines likewise the distribution network planning. The genetic algorithm faces Scheduling and State Assignment issues. The issues having all earmarks of being fitted can be arranged. In order to remove the color issue, the genetic algorithm is implemented a number of times. This phenomenon depicts the effectiveness of this algorithm[12]. The experts uses genetic algorithm to handle the problems related to the booking likewise the shop planning issue. The genetic algorithm can be combined with other algorithms. The half-breed GA technique can be used to handle the higher quality arrangements in a lesser amount of time. One of the instances is Traveling Salesman Problem (TSP). The scientists can inspect the force of the genetic algorithm in different test domains [13]. The ambiguous qualities can be identified from the expression information using hereditary study.

II. LITERATURE SURVEY

WANG Bei et.al (2016) proposed a new method in which the task scheduling problems arising in load balancing were avoided. This method was named as Multi-Population Genetic Algorithm (MPGA) [14]. For avoiding the previous concurrence in cloud computing, the proposed algorithm was provided as a substitute to the GA such that the task scheduling related problems could be resolved. To initialize the population such that the exploration effectiveness could be expanded, the min-min and max-min algorithm were utilized. Further, for exhibiting the offspring such that a sure possibility could be adopted by users, the metropolis standard was applied. The preservation of variety of population and local optimal were presented with the help of this idea. It was seen that by applying MPGA-based task scheduling algorithm, improved task scheduling results could be achieved as per the simulation outcomes. An efficient task scheduling was provided by the proposed algorithm. Thus, in comparison to the adaptive genetic algorithm, the efficiency of proposed algorithm was seen to be better.

Sukhpreet Kaur, et.al (2017) proposed a novel algorithm which aimed to assign the tasks of users to VMs and it was named as IGA (Improved Genetic Algorithm) [15]. Increasing the usage of resources along with ensuring that least power expenditure was spent and the task implementation cost was reduced were the important objectives of proposed algorithm.

For balancing the load, all the tasks were assigned to the available VMs. Thus, a pair of VMs was used to allocate the dynamic workload. Either over-utilization or under-utilization of VM was avoided here. With respect to power effectiveness and cost, the proposed algorithm provided good results. For load balancing, the tasks were distributed among all the VMs. It was seen that in comparison to the previous GAs, the proposed improved GA provided better results which were shown by the graphs presented towards the end of this research. The power competence and cost were the two important parameters calculated here.

Sheetal Karki, et.al (2018) studied that in a cloud, the data is saved in a central virtual machine. Help can be assigned to clients by the cloud supplier companies. Mostly based on the requirements, assistance was provided by clients. The services accessed by clients were only the ones for which they had to pay [16]. The need for load balancing was increased with the increase in amount of requests. For increasing the valuable resource exploitation and power usage, it was important to perform load balancing. To supporting the task migration such that the virtual machines at the time of cloudlet implementation, the virtual machines were overloaded. From one virtual machine to another the tasks were migrated. For reducing the power, resource usage and processing time, the tasks were waiting in line as presented by check point algorithm and threshold.

T. Deepa, et.al (2017) studied that the major reasons behind the popularity cloud computing were the services provided by them. For cloud computing network, load balancing was considered as an imperative concern [17]. Because of the increase in number of users and their necessities for different kinds of services, the effective usage of assets was important in cloud structure. Based on the demands of clients a competent load balancing algorithm must provide the resources such that the effective usage of resources can be ensured. For assigning priority to clients, the scheduling was used by load balancing mechanism. In terms of response time and waiting time, the performance displayers of load balancing algorithms were recognized. Providing an organized comparative analysis of accessible load balancing algorithms was the major objective of this research.

Pramod Kumar, et.al (2018) studied that an important method using which the parallel computing was performed in conjunction with distributed computing was called cloud computing [18]. Depending upon the global demands of clients, various attributes were provided like the software packages, storage, exchanged resources and so on. A complete answer related to computing and storage was provided through this approach. The handling of service requests was difficult for running the feedback thread such that the data privacy could be protected. Therefore, few performance issues were faced by fast usage of technology among which few were data redundancy, fault tolerance and data loss. One of the major challenges of performance analysis was cloud technology. Here, in certain time period, the information stored in clouds could not be accessed by the user due to the load balancing related problems. Therefore, for performance analysis, load balancing was considered as an important dynamic. The request time and response time were considered as important factors for load balancing. By improving the load balancing dynamic, the performance scrutiny of system was improved.

P. Geetha, et.al (2016) presented that as an improved version of cloud computing a new technology was proposed named as green cloud computing [19]. A network that interlinks computing assets all over the world such that various resources can be shared is known as cloud computing. In an eco-friendly manner, the computing resources of green computing were used due to which the ecological advantages were provided. As an integration to the green computing and cloud computing, green cloud computing was designed in which the effectiveness and efficiency of this approach were provided. Further, mobile computing and cloud computing was integrated as a fusion for mobile cloud computing. To generate intensive data, the computation science was transformed. Thus, a method called load balancing was used to distribute the load diagonally which was known as green cloud network. A detailed study related to load balancing algorithms was provided by the proposed structure. For reassigning the complete load to single nodes in a known system load balancing in cloud was considered. Further, in terms of quality parameters, the comparison of various load balancing algorithms was studied.

Snehal A. Narale, et.al (2018) presented that in the cloud computing technology, the data is stored within the cloud data hub which was then accessed by the users through the cloud service providers [20]. Access to data worldwide could be assessed by the users. For balancing the load among two or more cloud servers, load balancing technology is applied which distributes the load to different resources equally. Advancing the resource utilization, reducing the data center's cost and VM, increasing the throughput, minimizing the response time and preventing overload are some of the major aims of load balancing. For reducing the data center transmit cost, data hub processing time and virtual machine outlay, this research was proposed. The throttled load balancing mechanism was applied along with advanced response time service based mechanism such that the response time could be minimized. This research used cloud forecaster simulator and data modeling to perform simulation. The throttled load balancing and their relevant advancements were assessed by

this research. The data center processing time and outlay were reduced by this method.

Guilin Shao, et.al (2016) proposed a new load balancing related strategy in which the data correlation in cloud scenarios was used as base [21]. Few issues that were related to the data correlation in relevance to the virtual machines were also included in this research. Individual virtual machine was migrated in the cloud to reduce the resource utilization as well. Based on the association among virtual machines and data, the migration module was discovered by the proposed method. Similar kind of data was handled by the virtual machines. To perform an absolute migration, the load-intensive data suite was generated by the virtual machines. It was seen that by implementing the load balancing strategy, the communication cost was reduced by the proposed method. Similar kind of data was handled using the virtual machines. To perform an absolute migration, load-intensive data suite was built by the virtual machines. It was seen that this research helped in reducing the contact cost. Also, to a certain level, the resource utilization was also improved here.

Pradeep Kumar Tiwari, et.al (2016) studied that any cloud system includes several resources like network, memory, and CPU [22]. The load balancing approach uses these resources in an effective manner. Cloud computing was considered to be an appropriate choice such that the CAPEX and OPEX of the enterprises could be reduced. For achieving higher QoS and less SLA breach, load imbalance needs to be handled in a very effectual manner. On appropriate usage of physical and reasonable assets, the effectiveness of cloud service could be calculated. On the effectual load balancing mechanism huge research has been done and different techniques have been proposed. The VM migration structures play an important role in load management. This research mainly emphasizes on using the CPU capability of VMs to the maximum extent. To manage the load inequity based problem, a new method was proposed named as Dynamic Weighted Live Migration (DWLM). With respect to few parameters, the performance of proposed algorithm was evaluated by comparing it with other algorithms like Push Pull algorithm. Throughput, scalability and migration time were some of the evaluation parameters calculated here. This research was applied to analyze the policies of various load balancing algorithms.

Venkateshwarlu Velde, et.al (2017) studied that there is huge growth in the cloud computing technology [23]. There are several web services, distributed computing, virtualization and software programs provided in this technology. Many applications have adopted cloud computing to avail its benefits. The services to main source of computing power are provided such that certain services can be provided to specific packages. This research proposed a new load balancing algorithm in which the fuzzy method was used. It was a complex issue to balance load in cloud computing based applications. To steady the load in cloud environments, the fuzzy logic was used in this research along with the speed of computer and distributed load of VM were used.

III. RESEARCH METHODOLOGY

The endowed algorithm is an immediate variant of the fuzzy logic-based Cloudlet implementation strategy. The existing algorithm generates fuzzy rules for distributing tasks to different virtual machines for execution. The creation of fuzzy rules relies upon the number of cloudlets and the resources that virtual machines contain. The resources of VMs (Virtual Machines) are calculated with regard to the failure rate and execution time. The presented algorithm follows different steps for redistributing tasks:

1. Cloudlet Assignment: - The first task is cloudlet assignment. This task involves finding the most suitable virtual machine for implementing the Cloudlet. The discovery of a virtual machine depends on the execution time and failure rate. Virtual machines start executing cloudlets. In addition, virtual machines start maintaining check points on the server so that cloudlets can be executed successfully.

2. Overloaded virtual detection and virtual machine migration: - A virtual machine will be treated overloaded if it does not respond back. To migrate task, the load on each VM (Virtual Machine) is counted. The virtual machine with the maximum load is selected for cloudlets migration.

3. Apply ACO algorithm for the optimization: - The performance result of the weight-based algorithm will be given as input to the ACO algorithm to select the optimized virtual machine for executing the task execution. ACO is a search heuristic based on the swarm intelligence of ant colonies that uses pheromones as chemical messengers. Ants first randomly search the area around their colony in search of food. When the ants move, traces of pheromone are left on the ground, this pheromone vanishes over time. Next, the ants select their routes with a probability based on the thickness of pheromones on the discovered routes. By virtue of this indirect communication, the ants discover a direct route between the colony and the food source.ACO algorithms define the transition probabilities in the following way:

$$p(c_i|\mathrm{s}) = \frac{[\tau_i]^{\alpha}[\eta(c_i)]^{\beta}}{\sum_{c_j \in N(s)}[\tau_j]^{\alpha}.[\eta(c_j)]^{\beta}}, \forall c_i \in N(s) \ldots . (4)$$

In this equation, ηdenotes the optimal weighting function. A heuristic value $\eta(c_j)$ is allocated by it to each probable solution component $c_j \in N(s)$. Also, τ_i is the pheromone

concentration on the corresponding route; α and β are positive parameters and establish the connection between heuristic and pheromone information. There are various ACO algorithms that depend on the strategy adopted to update pheromone. Evaporation of the pheromone equally reduces each pheromone value. A commonly used formula for increasing the values of the pheromone pathway on the solution elements as a part of these strategies is as follows:

$$\tau_i \leftarrow (1-\rho).\tau_i + \rho. \sum_{s \in S_{upd} | c_i \in s} \omega_s \,.F(s) \ldots\ldots (5)$$

Here, S_{upd} denotes a solution set employed for update, $\rho \in (0,1]$ refers toanevaporation rate parameter, $F: S \rightarrow R^+$ is a quality function, and $\omega_s \in R^+$ is the weight of a solution s

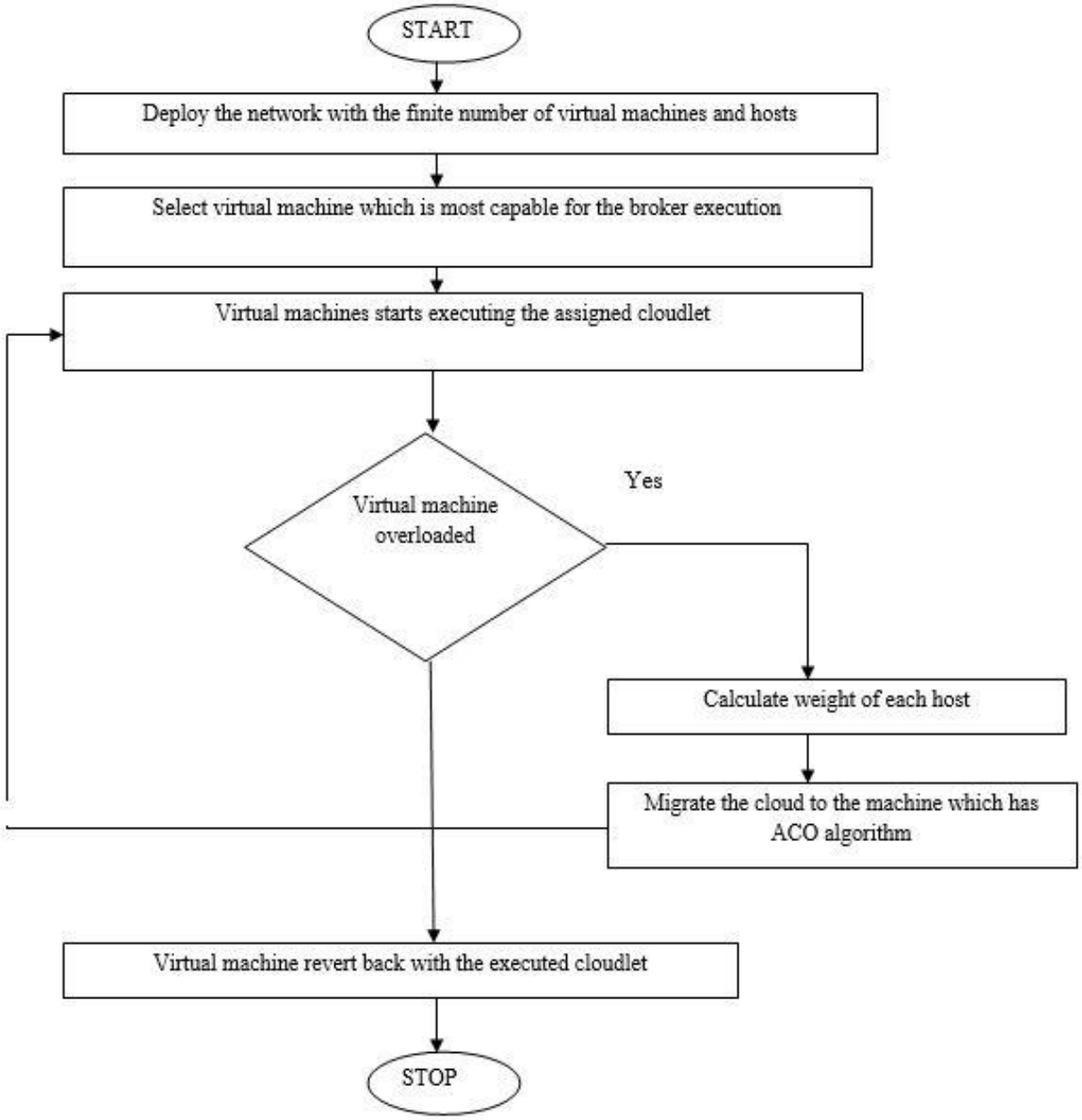

Figure 1:Proposed Model

IV. RESULT AND DISCUSSION

The MATLAB is the abbreviation for Matrix Laboratory. The LINPACK (Linear System Package) and EISPACK (Eigen System Package) projects have developed matrix software that MATLAB can accessed with ease. It is a high-performance language executed in technical computing. The core of MATLAB software is incorporating computation, visualization and programming environments. Apart from this, it is a new programming language that includes sophisticated data structures, built-in editing and debugging tools and supports object-oriented programming

Table 1: Parametric Analysis

Parameter	**Weight Based Algorithm**	**ACO Algorithm with weight-based algorithm**
Response Time	2.8 seconds	2.5 seconds
Finish Time	2.9 seconds	2.2 seconds
Energy Consumption	0.57 joules	0.5 joules
Cost	0.56 buffers	0.45 buffers
No of migration	10	7

.

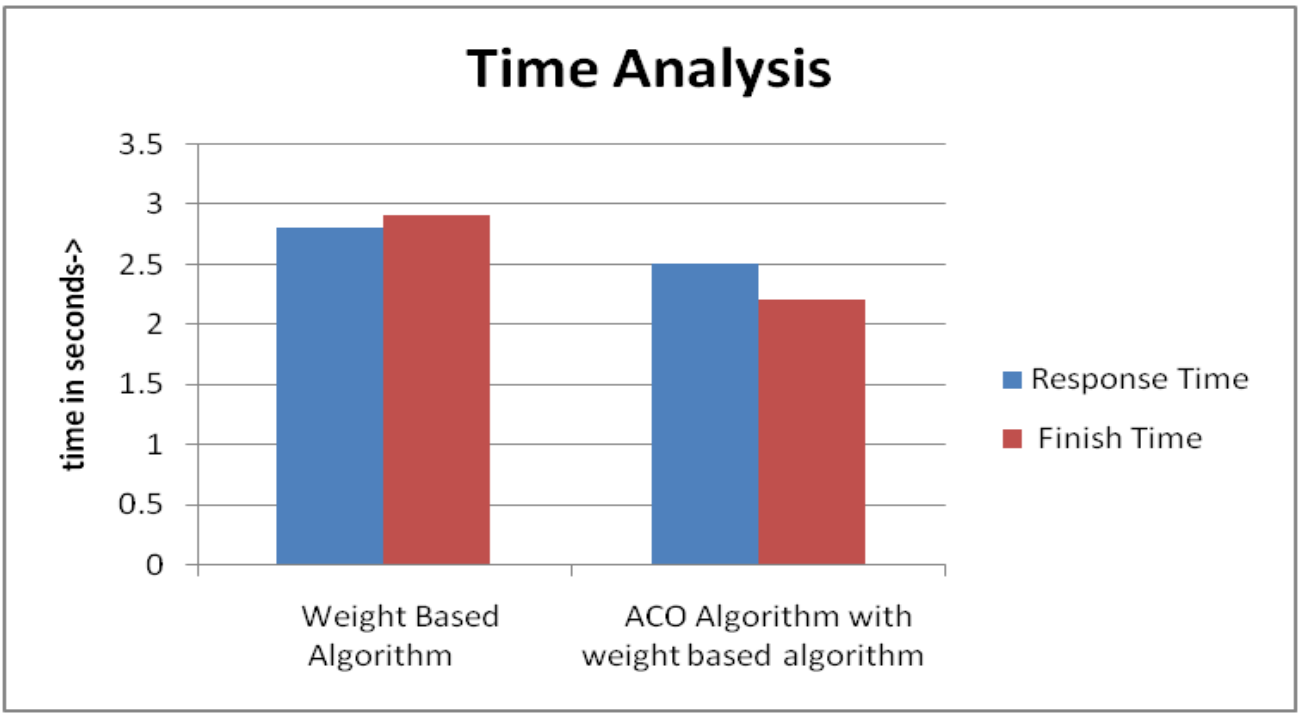

Fig 2: Time Analysis

Figure 2 depicts the response time and finish time of the old weight based algorithm. This algorithm is compared against the ACO + weighting-based algorithm. The performance-based analysis of both algorithms reveal that the proposed algorithm outperforms the weight-based algorithm.

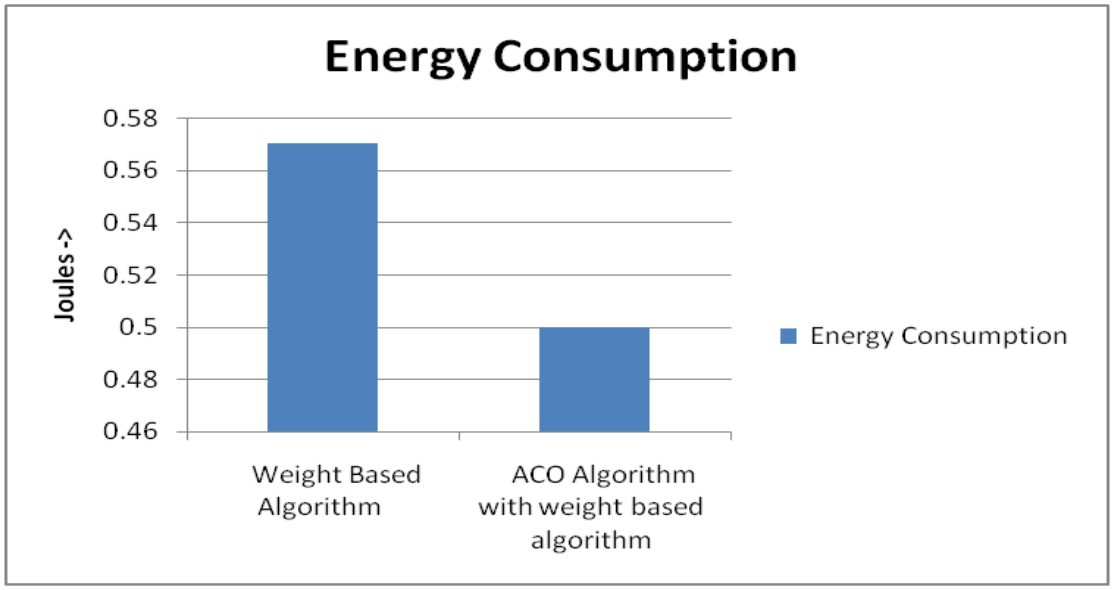

Fig 3: Energy consumption Analysis

As shown in figure 3, the energy consumption of existing algorithm which is weight based algorithm is compared with ACO +weight based algorithm. It is analyzed that proposed algorithm is more efficient as compared to weight based algorithm in terms of energy consumption

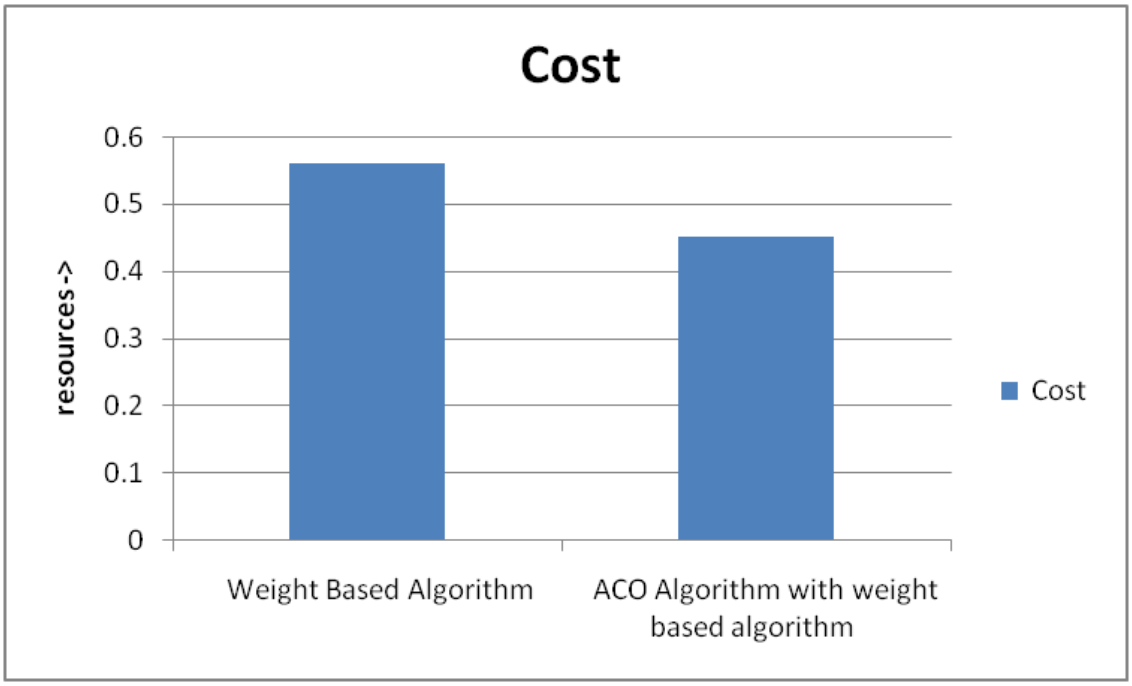

Fig 4: Cost Analysis

Figure 4 depicts the comparison between the old weight-based algorithm and the new ACO +weight-based algorithm in terms of cost. The performance-based analysis of both algorithms reveal that the proposed algorithm outperforms the weight-based algorithm. The total number of consumed resources are the measure of the resource consumption.

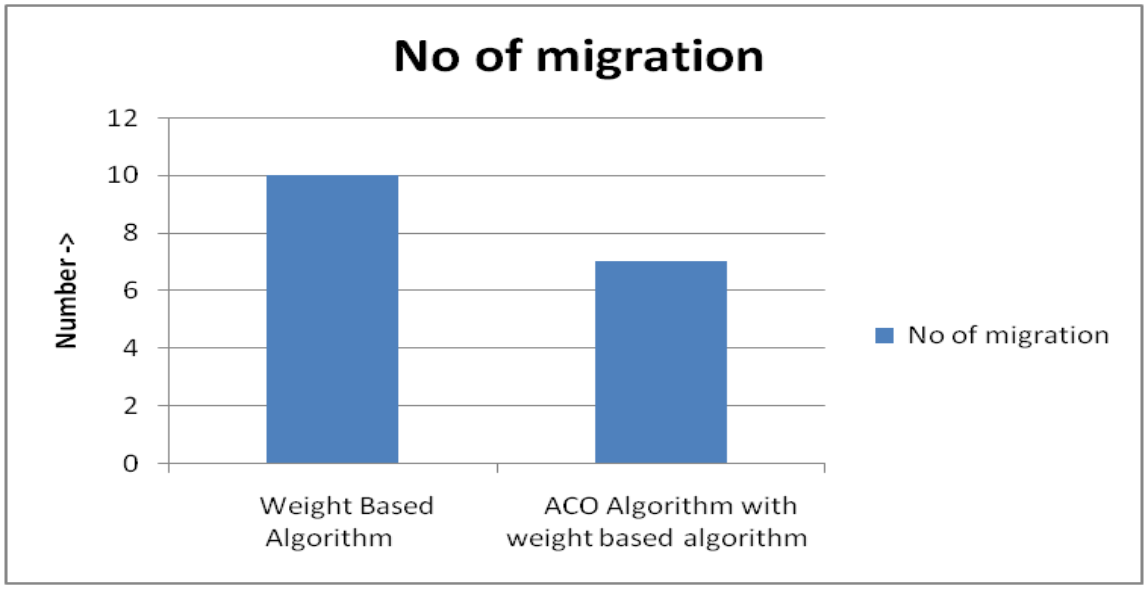

Fig 5: No of migration Analysis

Figure 5 depicts the comparison between the old weight-based algorithm and the new ACO +weight-based algorithm in terms of no of task migration. The performance-based analysis of both algorithms reveal that the proposed algorithm outperforms the weight-based algorithm in terms of no. of migrations.

V. CONCLUSION

Cloud computing services are available directly and on demand through data storage and processing to software such as CRM (Customer Relationship Management) systems. Cloud load balancing involves the distribution of workloads and the measurement of resources in a cloud system Companies can control the demand for application or workload by allocating resources amongst a large number of computers, networks or servers facilitated by load balancing. Distributing workload traffic and demands present on the web is involved in cloud load balancing. It helps that the resource count is well assigned and a high level of user satisfaction as well as proper utilization of resources can be achieved. The weight-based algorithm for load balancing in cloud computing is implemented with the ACO algorithm. The recommended algorithm is implemented in MATLAB and the results are examined by considering some metrics. The proposed algorithm for load balancing in cloud computing has improved by 10 percent.

VI. REFERENCES

[1] Randles, M.; Lamb, D.; Taleb-Bendiab, A., "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA), April 2010, pp.551-556.

[2] Fang, Yiqiu, Fei Wang, and Junwei Ge, "A task scheduling algorithm based on load balancing in cloud computing", Web Information Systems and Mining. Springer Berlin Heidelberg, 23 October 2010, Pages 271-277.

[3] Chaczko, Zenon, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, "Availability and load balancing in cloud computing", International Conference on Computer and Software Modeling, Singapore, Vol. 14, 2011.

[4] Samrat Kumar Dey, Md. Raihan Uddin, Kh. Mohaimenul Kabir and Md. Mahbubur Rahman, "Enhancing the Security of Cloud Computing: Genetic Algorithm and QR Code Approach", Proceedings of the 2017 4th International Conference on Advances in Electrical Engineering, 2017.

[5] Shui Han, Jianchuan Xing, "Ensuring Data Storage Through A Novel Third Party Auditor Scheme in Cloud

Computing", IEEE computer science & Technology, 2011 pp 264-268.

[6] Sean Carlin, Kevin Curran, "Cloud Computing Security", International Journal of Ambient Computing and Intelligence, January 2011, pp 14-19.

[7] Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing", Recommendations of the National Institute of Standards and Technology, September 2011, Special Publication 800-145.

[8] Dr Nashaat el-Khameesy,Hossam Abdel Rahman, "A Proposed Model for Enhancing Data Storage Security in Cloud Computing Systems", 2012, vol-3.

[9] Arora Pankaj, Wadhawan C.Rubal, Er.Ahuja P.Satinder, "Cloud Computing Security Issue in Infrastructure as a Service", International Journal of Advance Research in Computer Science and Software Engineering, 2012.

[10] Habib, S. M., Hauke, S., & Ries, S., "Trust as a facilitator in cloud computing: a survey", Journal of Cloud Computing , 2012, page 01-18.

[11] WANG Bei, LI Jun, "Load Balancing Task Scheduling based on Multi-Population Genetic Algorithm in Cloud Computing", Proceedings of the 35th Chinese Control Conference, July 2016, pp 27-29.

[12] Srinivas.J, K. Venkata Subba Reddy, Dr. A. Moiz Qyser, "Cloud Computing Basics", International journal of advanced research in computer and communication engineering , 2012, pp. 343-347.

[13] Tushar Desai, Jignesh Prajapati, "A Survey of Various Load Balancing Techniques and Challenges in Cloud Computing", International Journal of Scientific & Technology Research, November 2013, Volume 2, Issue 11.

[14] Cloud Computing Bible, Wiley Publishing, Inc., Indianapolis, Indiana, p25.

[15] Sukhpreet Kaur, Dr. Jyotsna Sengupta, "Load Balancing using Improved Genetic Algorithm(IGA) in Cloud Computing", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 6, Issue 8, August 2017, pp 2278-1323.

[16] Sheetal Karki, Anshika Goyal, "Performance Evaluation of Check Pointing and Threshold Algorithm for Load Balancing in Cloud Computing", International Journal of Computer Sciences and Engineering, Vol.-6, Issue-5, May 2018, pp 2347-2693.

[17] T. Deepa, Dhanaraj Cheelu, "A comparative study of static and dynamic load balancing algorithms in cloud computing",2017, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)

[18] Pramod Kumar, Dr. Mahesh Bundele, Mr. Devendra Somwansi, "An Adaptive Approach for Load Balancing in Cloud Computing Using MTB Load Balancing",2018, 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)

[19] P. Geetha, C.R. Rene Robin, "A comparative-study of load-cloud balancing algorithms in cloud environments", 2017, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)

[20] Snehal A. Narale, P.K. Butey, "Throttled Load Balancing Scheduling Policy Assist to Reduce Grand Total Cost and Data Center Processing Time in Cloud Environment Using Cloud Analyst",2018, Second International Conference on Inventive Communication and Computational Technologies (ICICCT)

[21] Guilin Shao, Jiming Chen, "A Load Balancing Strategy Based on Data Correlation in Cloud Computing", 2016, IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC)

[22] Pradeep Kumar Tiwari, Sandeep Joshi, "Dynamic weighted virtual machine live migration mechanism to manages load balancing in cloud computing", 2016, IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)

[23] Venkateshwarlu Velde, B. Rama, "An advanced algorithm for load balancing in cloud computing using fuzzy technique",2017, International Conference on Intelligent Computing and Control Systems (ICICCS)