International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

An Optimal Approach to Load Balancing in Cloud Computing Based On Comparative Study and Analysis

Kushagra Jaiswal[1], Navneet Kaur[2], Mohit Sharma[3]

[1, 2, 3]Department of Computer Science and Engineering, Galgotias College of Engineering and Technology, Greater Noida, India

Abstract: ***Cloud Computing is the latest paradigm in the vast and rapidly expanding computing and Information Technology industry. Cloud computing refers to the practice of using a large number of remote servers that are connected over a network and hosted over the internet to provide various computing facilities as a service. Over the past few years, cloud computing has emerged to be one of the most promising technologies in the IT industry, due to which there is huge competition among the service providers for facilitating the resources at minimal cost without compromising with the performance. This leads to the need of delivering the resources in an optimal way such that the clients may access them seamlessly. There are a number of load balancing techniques that have the capability of distributing the load among the various server systems. In this paper, various load balancing techniques of cloud computing are compared and an effective and highly optimized approach is suggested to minimize the load and hence provide maximum resource utilization.***

Keywords: Cloud Computing, Load Balancing, Ant Colony algorithm, HPAC, metrics

1. Introduction

Cloud computing refers to the delivery of computing as a service rather than a product, by using which the software, shared resources, and information are given to computers and other devices as a utility over a network (mainly the Internet). Cloud computing is a calculation mode that canprovide dynamic flexibility, on-demand services, virtualized resources[1].With the advent and growth of latest technologies and rise in networking infrastructure, the cloud computing has risen to be one of the most widely used services by the Information Technology organizations.

The term 'Cloud' in Cloud Computing basically refers to a distributed system over the different parts of the world that has the capability and the resources, be it software or hardware, to be provided to the large number of clients. Following are the three types of clouds:

Public Cloud

A service provider that hosts the cloud infrastructure makes the public cloud available to the general public. Public clouds are like an in-house enterprise cloud which provides the various cloud computing services to the people on their personal computers. Public cloud service providers like Google, Amazon AWS, Microsoft, etc. own and operate the hardware and software infrastructure which an individual may access on a pay-as-you-go basis from them.

Private Cloud

Private Cloud infrastructure have the computing components that are dedicated to an organization in particular and are not shared with any other organization even if it is an internal subsidiary of the main organization, or any third party. Private cloud can either be hosted internally by the company or can be managed from outside.

Hybrid Cloud

Hybrid Clouds are a combination of two or more clouds that may be private or public in nature. There is a need for implementing more than one type of cloud computing component as per the requirement of the business models. Augmenting a traditional private cloud with the resources of a public cloud can be used to manage any unexpected surges in workload. Hybrid cloud architecture requires both on premise resources and off-site server based cloud infrastructure [2]. Hybrid cloud thus enables a business to spread out its resources in an efficient manner. Following are the fundamental service models of cloud computing:

Software as a Service (SaaS)

The SaaS model provides software services over the internet on demand. The software applications are hosted by the Cloud Computing vendors and are provided to the customers over a network, mainly the internet. Examples of service providers are Microsoft Azure, salesforce.com, Abiquo, Akamai, Oracle on Demand, etc.

Platform as a Service (PaaS)

The PaaS model is a cloud computing model that provides its users a platform to develop, test and deploy various applications that may be operating on a large variety of platforms. It enables the clients to try out a diverse range of software products, their different versions, etc. This model saves the organizations from developing the infrastructure so as to build the software products. A few PaaS vendors are Amazon web services, open stack, Grid Gain, Appistry.

Infrastructure as a Service (IaaS)

The IaaS service model provides all the computing resources in a virtualized form. The resources may be any hardware like storage, computational speed, servers, etc. IaaS providers also provide the facility to host applications and

the environment to deploy the various applications. Backup, maintenance and resilience planning are also a few characteristics of IaaS. Example of vendors are bluelock, cloud scaling, Data pipe, NaviSite, etc.

2. Load Balancing

For specific applications, the cloud service system for the client applications still need pre-allocate the fixed cloud resources. Since the applying for cloud services of client applications is random and the load is also difficult to predict, stationary parts in the cloud scheduling are difficult to solve the problem that some nodes would overload during this progress. Thus, according to the load changing of cloud applications, allocating the dynamic resource self adaptively is meaningful. Such a phenomenon is called load balancing [3].

There are two load balancing approaches namely Static and Dynamic. In a static load balancing technique, the load is evenly distributed among all the nodes on the server side. On the other hand, in dynamic load balancing the load is dependent on the real time status of the nodes. If a node is overloaded, then in a dynamic load balancing the extra load on that node is distributed among the other under loaded nodes. Following figure shows the types of load balancing [4]:

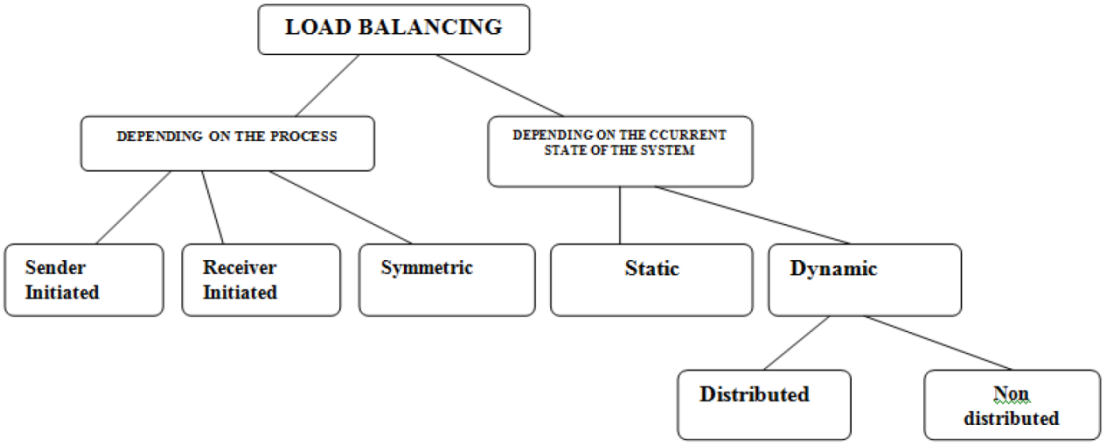

Figure 1: Load Balancing Types

Some dynamic load balancing techniques are discussed as follows:

1) **Honey Bee Foraging Algorithm:** The honey bee foraging algorithm is based on an idea of honey bees looking for food and alarming the other bees to eat the food as soon as they find some. The foraging bees firstly go around and search for food. Once they find the food, they come back to their hives and dance. The strength of their dance signals the other scout bees to follow the forager bees and get the food. This idea is implemented on the cloud servers.
2) **Ant Colony Algorithm:** The Ant colony algorithm works in the following way.Ant originate the from the head node. These ant traverse the width and length of the network in such a way that they know whole location of under loaded node and overloaded node. When these ants traverse the network they update the pheromone table which will store the information of utilization of each node [5]. In the previous work movement of ant two ways:
   a. Forward movement-The ants continuously move in the forward direction in the cloud encountering overloaded node or under loaded node[5].
   b. Backward movement-If an ant encounters an overloaded node in its movement when it has previously encountered an under loaded node then it will go backward to the under loaded node to check if the node is still under loaded or not and if it finds it still under loaded then it will redistribute the work to the under loaded node. The vice-versa is also feasible and possible [5].
3) **Task scheduling algorithm based on Load Balancing:** Y. Fang et al. [6] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources.
4) **Role Based Access Control(RBAC):** In a Role Based Access Control the applications on the cloud servers are assigned to the users based on their role, i.e. an application can be accessed by a user only if he is registered to the company and has a role such that access to the application is granted to that user only.
5) **Throttled Load Balancing Algorithm:** This algorithm makes use of identity of virtual machines. Client requests the ID of virtual machine. Throttled load balancing algorithm returns that ID to the user [7].

There is countless number of load balancing techniques that are developed by the different organizations to optimize the dynamic load of the nodes at their server end. A few of them have been described above. The load balancing techniques are compared on the basis of the following metrics:

a) **Overhead Associated-** It determines the amount of overhead involved while implementing a load-balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently [4].
b) **Throughput-** It refers to the number of tasks performed per unit time. Throughput should be high in an efficient system.

c) **Performance-** It should be high in an efficient system. Good performance of the system means that it is free from failures.
d) **Scalability-** Scalability means the capability of the system to perform optimal load balancing with any finite quantity of nodes.
e) **Resource Utilization-** A good system must utilize each and every available resource to its fullest. There should be minimum occurrences of a process sitting idle or a number of processes in a deadlock state.
f) **Response Time-** It refers to the time taken by the nodes in first response to the request.
g) **Fault Tolerance-** It is the ability of a system to avoid failures due to faults.
h) **Migration Time-** It is the time taken by the algorithm for context switching between the nodes, i.e. the time taken by the system to migrate between jobs.

Based on the metrics so discussed above, following are the merits and demerits of a few load balancing techniques:

Table 1: Comparative study of load balancing techniques

Technique	Merits	Demerits
Honey Bee	High throughput, minimum overhead	Priority load is very low, wastage of resources
Ant Colony	High performance, good scalability, fault tolerant, high resource utilization	Dense network
Task Scheduling	Good performance	Low response to request ratio
RBAC	Low CPU burst time	High response time
Throttled	Load movement factor is very high	Low communication cost effectivity, high network congestion

In the following sections we shall study further on the Ant Colony load balancing technique in cloud computing and further, we shall take a look at the proposed system which is an advanced approach to ant colony optimization.

3. Previous Work

Here we discuss the traditional Ant Colony algorithm for load balancing in cloud computing environments. Dorigo M. introduced the ant algorithm based on the behavior of real ants in 1996[4][8], the algorithm is based on heuristics to solve the combinatorial optimization problems. As per the observations the ants have the capability to find the shortest and optimal path from food to their nest. While moving, the ants lay some pheromone on the ground; while an isolated ant encounter a previously laid trail, this ant can detect it and decide with high probability to follow it. Hence, the trail is reinforced with its own pheromone [4]. The probability of an ant choosing a path is dependent on the amount of pheromone laid on that path. Through this positive feedback mechanism, ant can find an optimal way finally [8]. ACO is inspired from the ant colonies that work together in foraging behavior. Many researchers, in fact have taken inspiration from the behavior of the ants to further understand the concept of feedback mechanisms and combinatorial optimization problems. The ants work together in search of new sources of food and simultaneously use the existing food sources to shift the food back to the nest [4].

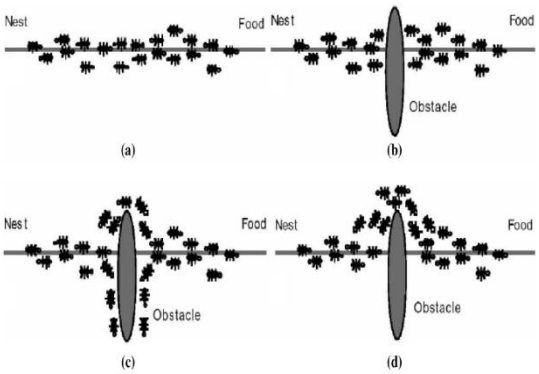

Figure 2: Modification in ants' path upon encountering an obstacle [9]

The flowchart of the traditional ant colony algorithm is as follows [9]

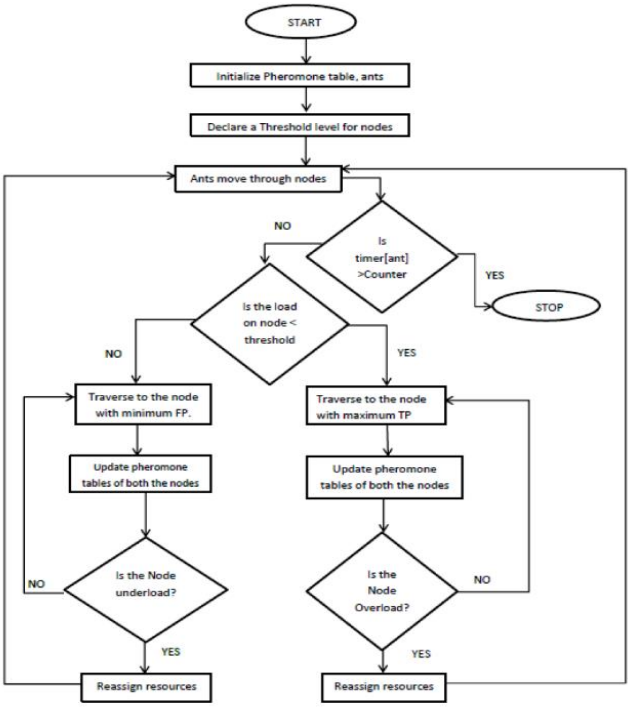

Figure 3: Flowchart of traditional ant colony algorithm [9]

4. Proposed System

Now we shall discuss our proposed system. The traditional model had a shortcoming of Migration Time which we try to reduce in our proposed model entitled High Pace Ant Colony (HPAC) load balancing technique. Following diagram provides an insight to the proposed system:

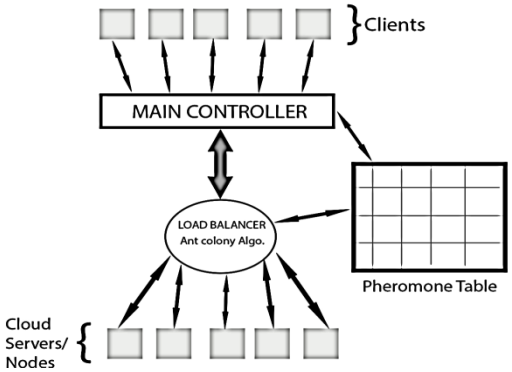

Figure 4: HPAC architecture

The HPAC technique works in the following ways:

1) The Main controller waits for a request from the client side. Once it gets a request it reads the load that the client is demanding from the system and stores it in a temporary variable.
2) Once the input is taken, it looks for the existing load value with respect to that node in the pheromone table and looks for the least value in the table which is large enough to be allocated to the client and virtually allocates that memory to the client.
3) In the next step, the main controller sets the threshold value for the nodes so that their load be balanced.
4) After setting the threshold, it looks for the node with the load value crossing the threshold value and immediately signals the load balancer that the values on the pheromone have been changed. It also sends the node no. of the node whose value has crossed the threshold value.
5) The final step involves the functionality of the load balancer which operates on the Ant Colony Load Balancing algorithm as discussed earlier

Following flowchart gives a better understanding of the model:

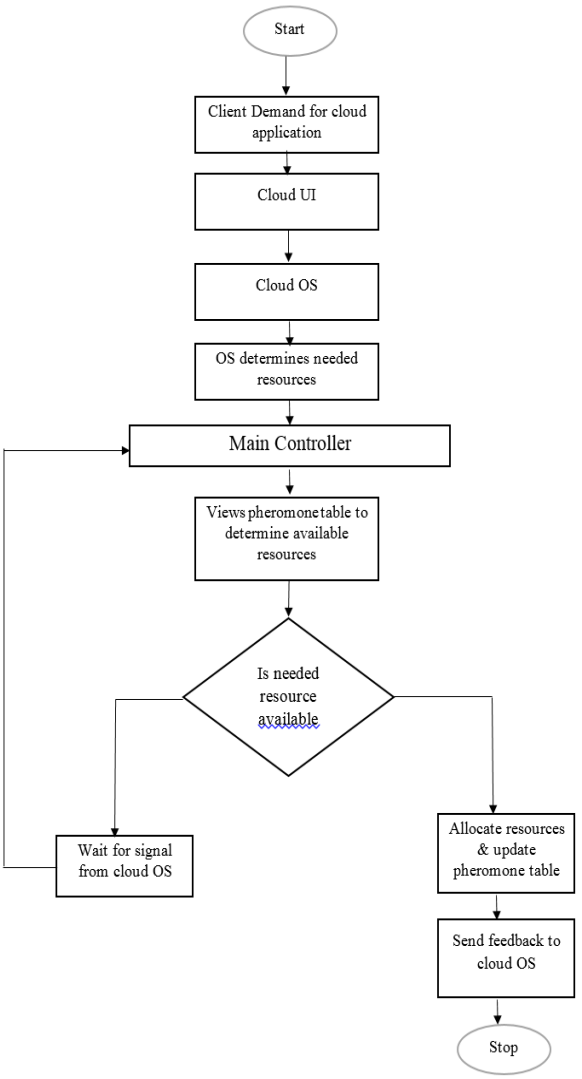

Figure 5: Flowchart of HPAC

5. Conclusion

Load balancing is one of the main challenges in cloud computing. The load balancing techniques distribute the load across the given number of nodes thereby making the computational tasks faster and more efficient. And we may conclude that we can use a particular algorithm according to our requirement/need. This is a modified approach of ant colony optimization that has been applied from the perspective of cloud or grid network systems with the main aim of load balancing of nodes. Therefore, our proposed algorithm for load balancing in cloud computing plays a very important role in future. There is a huge scope of improvement in this area. We have discussed on ACO algorithm that can be applied to clouds for improving the efficiency, resource utilization and different other issues.

To sum up, we can conclude that the High Pace Ant Colony (HPAC) technique will prove to be an efficient algorithm and will improve the performance of the cloud computing environment by balancing the load in an effective way and thereby standing out in the improvement of the various parameters that determine the performance of the cloud computing environment like: resource utilization, throughput, performance increase, fault tolerance, scalability, etc.

References

[1] Michal Armbrust, Armando Fox,and Rean Griffith,et al. “Above the Clouds:A Berkeley View of Cloud Computing, mimeo”,UC Berkeley,RAD Laboratory, 2009, vol.12:67-98.
[2] Shrikant M. Lanjewar, Susmit S. Surwade,Sachin P. Patil,Pratik S. Ghumatkar, Prof Y.B. GURAV, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. VI (Feb. 2014), PP 82-87.
[3] Xin Lu, ZilongGu, “A LOAD-ADAPATIVE CLOUD RESOURCESCHEDULING MODEL BASED ON ANT COLONYALGORITHM”, Proceedings of IEEE CCIS2011.
[4] Shagufta khan Niresh Sharma, “International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)”, Volume 2, Issue 6, November – December 2013 ISSN 2278-6856
[5] Niresh Sharma Professor (CSE) RKDFIST BHOPAL (M.P), India, Khan et al., “International Journal of Advanced Research in Computer Science and Software Engineering” 4(2), February - 2014, pp. 966-973
[6] Y. Fang, F. Wang, and J. Ge, “A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing”, Web Information Systems and Mining, Lecture Notes in Computer Science, Vol. 6318, pages 271-277,2010.
[7] Ruhi Gupta, “Review on Existing Load Balancing Techniques of Cloud Computing”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014.
[8] M. Dorigo, M. Birattari and T. Stutzle, ”Ant Colony Optimization-Artificial Ants as a Computational

Intelligence Technique", IEEE Computational Intelligence Magazine, pp. 1- 12. 2006.
[9] Kumar Nishant, Pratik Sharma, Vishal Krishna, Chhavi Gupta and KuwarPratap Singh, 2012 14th International Conference on Modelling and Simulation, DOI 10.1109

Author Profile

Kushagra Jaiswal is currently perusing his B. Tech (2011-15) from Galgotias College of Engineering & Technology from the Computer Science and Engineering stream.

Navneet Kauris currently pursuing his B. Tech (2011-15) from Galgotias College of Engineering & Technology from the Computer Science and Engineering stream.

Mohit Sharma is currently pursuing his B. Tech (2011-15) from Galgotias College of Engineering & Technology from the Computer Science and Engineering stream.