A Survey on Load Balancing Techniques Using ACO Algorithm

Preeti Kushwah

Department of Computer Science & Engineering,

Acropolis Institute of Technology and Research

Indore bypass road Mangliya square

Abstract— **Cloud computing is a new generation solution for storage and computation, which provide the efficient data transfer, scientific calculations and end user services. For providing the efficient computing experience a huge amount of computational units are processes. In this condition the key issue is to provide the resources to the executing jobs. Here, we are reviewing on cloud based load balancing which helps to resolve heavily loaded cloud host for process execution. This survey paper focuses on load balancing techniques using ACO algorithm, in order to achieve efficient resource scheduling for the cloud environment.**

Keywords — Cloud Computing, Load Balancing, Ant Colony Optimization

I. INTRODUCTION

Cloud computing is rapidly improving the latest technology. This technological trend has enabled the realization of a new computing model called cloud computing, in which shared resources, information, software and other devices are provided according to the client's requirement at specific time. Cloud computing involves virtualization, distributed computing, platform, infrastructure, software and web based services. That's why, Cloud Computing is considered as one of the emerging arenas of computer science in recent times.

According to the NIST- "Cloud computing [1] is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, server, storage, application, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".
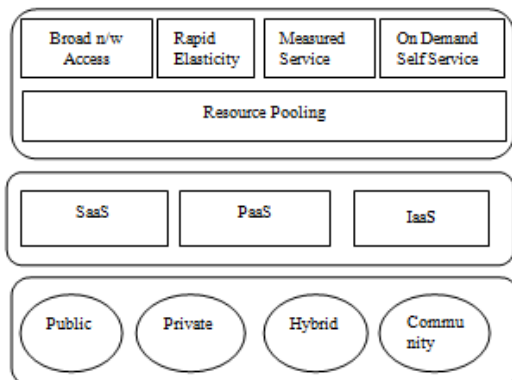

Fig.1 Cloud Computing Architecture

Fig 1 illustrating the three basic service layers that constitute the cloud computing. It provides essential characteristics, deployment models and three basic services.

Essential characteristics:
- Broad network access
- Rapid Elasticity
- Measured service
- On demand self service

Service Models:
- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

Deployment Models:
- Public Cloud
- Private Cloud
- Hybrid Cloud
- Community Cloud

In cloud computing, load balancing is required to distribute the dynamic local workload evenly across all the nodes. It helps to achieve a high user satisfaction and resource utilization ratio by ensuring an efficient and fair allocation of every computing resource.

Load Balancing [2] is the process of distribution of the load among various nodes of distributed system so that the better utilization of resources while also avoiding the situation where some nodes are overloaded and some nodes are idle or under loaded.

Load balancing in cloud computing provides an efficient solution to various issues residing in cloud computing environment set-up and usage.

Load balancing is challenging issues in cloud environment. Various methods are to be used to make a better system by allocating the loads to the nodes in a balancing manner but due to network congestion, bandwidth usage etc, there were problems are occurred. These problems were solved by some of the existing techniques.

A. Goals of Load Balancing :
1. To improve the performance substantially.
2. To have a backup plan in case the system fails even partially.
3. To maintain the system stability
4. To accommodate future modification in the system

B. Need of Load Balancing :

Load balancing is an important aspect of cloud computing environment which is basically used to achieve efficient resource scheduling, maximum utilization of resources & a higher user satisfaction, making sure that no single node is overwhelmed, hence we can say that for improving the

overall performance of the system proper load balancing is needed. It also helps in implementing fail-over, avoiding bottlenecks, enabling scalability, over-provisioning and reducing response time. We use different techniques and algorithms of load balancing for achieving all the requirements.
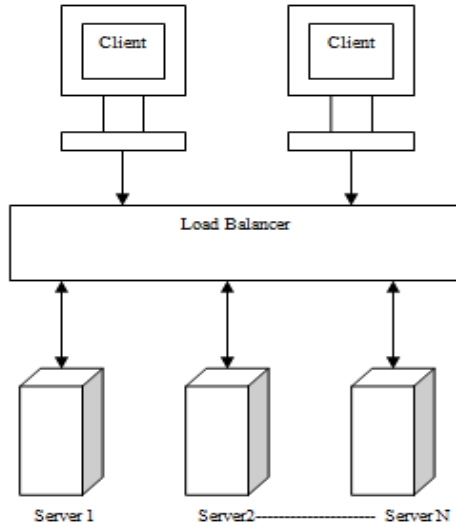

Fig 2 The Concept of Load Balancing

C. Types of Load Balancing Algorithm :

1) Static Load Balancing-
   In static load balancing algorithm the traffic is divided evenly among the servers. This algorithm requires a prior knowledge of system resources, so that the decision of shifting of the load does not depend on the current state of system. This algorithm is proper in the system which has low variation in load.

2) Distributed Dynamic Load Balancing-
   In dynamic load balancing algorithm the lightest server in the whole network or system is searched and preferred for balancing a load. For this real time communication with network is needed which can increase the traffic in the system. Here current state of the system is used to make decisions to manage the load.
   In a distributed system, dynamic load balancing can be done in two different ways: distributed and non-distributed. In the distributed one, the dynamic load balancing algorithm is executed by all nodes present in the system and the task of load balancing is shared among them. The interaction among nodes to achieve load balancing can take two forms: cooperative and non-cooperative. Dynamic load balancing algorithms of distributed nature, usually generate more messages than the non-distributed ones because, each of the nodes in the system needs to interact with every other node. A benefit, of this is that even if one or more nodes in the system fail, it will not cause the total load balancing process to halt. It instead would affect the system performance to some extent. Distributed

dynamic load balancing can introduce immense stress on a system in which each node needs to interchange status information with every other node in the system.

3) Non-distributed dynamic load balancing-
   In non-distributed type, either one node or a group of nodes do the task of load balancing. It can take two forms: centralized and semi-distributed.
   In the first form, the load balancing algorithm is executed only by a single node in the whole system: the central node. This node is solely responsible for load balancing of the whole system. The other nodes interact only with the central node. In semi-distributed form, nodes of the system are partitioned into clusters, where the load balancing in each cluster is of centralized form. A central node is elected in each cluster by appropriate election technique which takes care of load balancing within that cluster.

4) Centralized Load Balancing-
   In centralized load balancing [3] technique all the allocation and scheduling decision are made by a single node. This node is responsible for storing knowledge base of entire cloud network and can apply static or dynamic approach for load balancing. This technique reduces the time required to analyze different cloud resources but creates a great overhead on the centralized node. Also the network is no longer fault tolerant in this scenario as failure intensity of the overloaded centralized node is high and recovery might not be easy in case of node failure.

5) Hierarchical Load Balancing [3]-
   It involves different levels of the cloud in load balancing decision. Such load balancing techniques mostly operate in master slave mode. These can be modelled using tree data structure wherein every node in the tree is balanced under the supervision of its parent node. Master or manager can use light weight agent process to get statistics of slave nodes or child nodes. Based upon the information gathered by the parent node provisioning or scheduling decision is made. Three phases hierarchical scheduling-First phase uses Best Task Order scheduling, second phase uses Enhanced Opportunistic Load Balancing scheduling and third phase uses Enhanced Min-Min scheduling.

D. Ant Colony Optimization :

1) Ant Colony Optimization is a technique for optimization. It is also called ACO in short. We can say that ACO is an algorithm, which is based on the behaviour of the real ants in finding a shortest path from a source to the food. It is inspired from the ant colonies that work together in foraging behaviour. As we know that an ant is a simple creature; collectively a colony of ants performs useful tasks like finding the shortest path to a food source and sharing this information with other ants by depositing pheromone.

2) 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. The ants leave a pheromone trail upon moving from one node to another. With the help of pheromone trails, the ant subsequently came to the food sources.

3) The intensity of the pheromone [4] can vary on various factors like the quality of food sources, distance of the food, etc. The ants use these pheromone trails to select the next node. The ants can even modify their paths upon encountering any obstacles in their path. This phenomenon of the ants was used in many algorithms for optimization where the ants follow each other through a network of pheromone paths. The ants upon traversal from one node to another update the pheromone trail of that path, so a path becomes more feasible if more ants traverse upon it. Paths that have the highest pheromone intensity have the shortest distance between the point and the best food source.

4) In ACO algorithm [4] when the request is initiated the ant start its movement. Movement of ant is of two ways:

   - Forward Movement [4] means the ant in continuously moving from one overloaded node to another node and check it is overloaded or under loaded, if ant find an overloaded node it will continuously moving in the forward direction and check each nodes.

   - Backward Movement- If an ant find an over loaded node the ant will use the backward movement to get to the previous node, in the algorithm if ant finds the target node then ant will commit suicide, this algorithm reduced the unnecessary back ward movement, overcome heterogeneity, is excellent in fault tolerance [4].

5) Application of ACO to combinatorial optimization problems:

   - Travelling salesman problems
   - Quadratic assignment problems
   - Scheduling problems
   - Vehicle routing problems
   - Sequential ordering problems
   - Shortest common super sequence problems
   - Graph coloring problems
   - Frequency assignment problems
   - Bin packing and multi-knapsack problems
   - Constraint satisfaction problems

II. RELATED WORK

Load Balancing is an important aspect of Cloud Computing environment to achieve maximum utilization of resources.

Mayanka Katyal et al [3] discussed various load balancing scheme such as static load balancing, distributed and non-distributed dynamic load balancing, centralized and hierarchical load balancing. On one hand static load balancing scheme provide easiest simulation and monitoring of environment but fail to model heterogeneous nature of cloud. On the other hand, dynamic load balancing algorithm are difficult to simulate but are best suited in heterogeneous environment of cloud computing.

Shagufta Khan et al [5] implemented SALB algorithm. Firstly studied existing ACO's and then develop effective load balancing algorithm using ant colony optimization The main contribution of work is to balance the entire system load while trying to maximize and minimize the different parameter.

Soumya Banerjee et al [6] presented an initial heuristic algorithm to apply modified ant colony optimization approach for the diversified service allocation and scheduling mechanism in cloud paradigm. The pheromone update mechanism of ACO and coefficient τ is modified to $\tau\&$. This modification supports to minimize the make span of the cloud computing based services and probability of servicing the request also has been converged using the modified scheduling.

Cloud task scheduling is an NP-hard optimization problem, and many meta-heuristic algorithms have been proposed to solve it. A good task scheduler should adapt its scheduling strategy to the changing environment and the types of tasks. Kun Li et al [7] proposes a cloud task scheduling policy based on Load Balancing Ant Colony Optimization algorithm. The main contribution of our work is to balance the entire system load while trying to minimizing the make span of a given tasks set.

Al-Dahoud Ali et al [8] proposed ACO algorithm for load balancing in distributed systems. This algorithm is fully distributed in which information is dynamically updated at each ant movement. Multiple colonies paradigm will be adopted such that each node will send a colored colony throughout the network and which are used to prevent ants of the same nest from following the same route and hence enforcing them to be distributed all over the nodes in the system and each ant acts like a mobile agent that carries newly updated load balancing information to the next visited node.

Tushar Desai et al [9] discussed various load balancing techniques for cloud computing. The main purpose of load balancing is to satisfy the customer requirement by distributing load dynamically among the nodes and to make maximum resource utilization by reassigning the total load to individual node. This ensures that every resource is distributed efficiently. So the performance of the system is increased.

Ratan Mishra et al [10] developed an effective load balancing algorithm using Ant colony optimization technique to maximize or minimize different performance parameters like CPU load, Memory capacity, Delay or network load for the clouds of different sizes.

TABLE I
ANALYSIS OF LOAD BALANCING TECHNIQUES USING ANT COLONY OPTIMIZATION ALGORITHM

S.No	TITLE	PURPOSE	METHOD USED	ADVANTAGE	DISADVANTAGE
(1)	Effective Scheduling Algorithm for Load balancing using Ant Colony Optimization in Cloud Computing	1.To develop an effective load balancing algorithm using ACO 2.To balance the entire system load while trying to maximize and minimize the different parameters	1.Scheduling Algorithm for Load Balancing(SALB)	1.Efficient to finding the overloaded node in minimum time 2. To balance the node with efficiency and maximum utilization of resources.	1.Less performance(can be increased by varying different parameters)
(2)	Cloud task Scheduling based on Load balancing ACO.	1.To minimize the task completion time that are distributed among different virtual machines	1.Load Balancing ACO Algorithm is used. to find the optimal resource allocation for each task in dynamic cloud system.	1. Nodes are balanced dynamically 2. LBMM Balanced the entire system whether the entire task are the same or not that can handle all type of conditions.	1. No precedence constraints between different task means we can say the task are mutually independent. 2. The task are Computationally intensive which is not realistic for cloud system. 3. It is not Heterogeneity of system.
(3)	Cloud computing initiative using modified ACO framework	1. To minimize the make span (throughput of heterogeneous computing system) 2. Uses Modified ACO 3. The main modification in pheromone Formula.	1. based on heuristics algorithm uses modified ACO.	1.Better utilization of resources. 2. It is not dynamic in nature. 3.Probability of servicing the request also has been converged using modified Scheduling.	1. Fault tolerance is not good. 2. overhead Increases.
(4)	Load Balancing of Distributed Systems Based on Multiple Ant Colonies Optimization	1. Multiple colonies paradigm will be adopted such that each node will send a colour colony throughout the network	1. ACO algorithm for load balancing in distributed systems	1.The use of multiple nests, or ant colonies in the search process, helped in raising the rate of information exchange all over the nodes in the system	1. Less Efficiency
(5)	Ant colony Optimization: A Solution of Load Balancing in Cloud.	1. To develop an effective load balancing algorithm using ACO technique to maximize or minimize different performance parameters	1. A heuristic algorithm based on ant colony optimization has been proposed.	1. The pheromone update mechanism has been proved as a efficient and effective tool to balance the load. This modification supports to minimize the make span of the cloud computing based services.	1. This technique does not consider the fault tolerance issues. 2. Less efficient due to the absence of cluster in our cloud.

III. CONCLUSIONS

As we know that Load Balancing is an important aspect of cloud computing environment which provides an efficient solution using different techniques and algorithms. In this paper, we have surveyed various load balancing techniques using Ant Colony Optimization (ACO) algorithm, in order to achieve efficient resource scheduling and maximum utilization of resources for the cloud environment. We also included the short description of load balancing types such as distributed and non-distributed dynamic load balancing. Here, the main focus on Ant Colony Optimization Algorithm (ACO). We conclude load balancing techniques using Ant Colony Optimization (ACO) algorithm in cloud environment by analysing the existing work in table 1.

REFERENCES

[1] The NIST definition of cloud computing - http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf
[2] "Ant Colony Optimization for Effective Load Balancing in Cloud Computing" , Shagufta khan, Niresh Sharma, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 6, November – December 2013.
[3] "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment" Mayanka Katyal, Atul Mishra.
[4] "Improving Efficiency by Balancing the Load Using Enhanced Ant Colony Optimization Algorithm in Cloud Environment", Ashwini L, Nivedha G, Mrs A.Chitra, IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 2, Apr-May, 2014.
[5] "Effective Scheduling Algorithm for Load balancing using Ant Colony Optimization in Cloud Computing", Shagufta Khan, Niresh Sharma, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014.
[6] "Cloud computing initiative using modified ACO framework", Soumya Banerjee, Indrajit Mukherjee, and P.K. Mahanti, World

Academy of Science, Engineering and Technology Vol:3 2009-08-27.

[7] "Cloud Task scheduling based on Load Balancing Ant Colony Optimization", Kun Li, Gaochao Xu, Guangyu Zhao, Yushuang Dong, Dan Wang, 2011 Sixth Annual ChinaGrid Conference, 2011 IEEE.

[8] "Load Balancing of Distributed Systems Based on Multiple Ant Colonies Optimization", Al-Dahoud Ali, Mohamed A. Belal and Moh'd Belal Al-Zoubi, American Journal of Applied Sciences 7 (3): 428-433, 2010 ISSN 1546-9239..

[9] "A Survey Of Various Load Balancing Techniques And Challenges In Cloud Computing", Tushar Desai, Jignesh Prajapati, International Journal of Scientific & Technology Research volume 2, issue 11, november 2013.

[10] "Ant colony Optimization: A Solution of Load balancing in Cloud" Ratan Mishra, Anant Jaiswal, International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.2, April 2012.

[11] "Reviews of Load Balancing Based on Partitioning in Cloud Computing" Ms. Shilpa D.More1, Mrs. Smita Chaudhari2, International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3965-3967.