A Review on Various Load Balancing Techniques in Cloud Computing

Davinder Singh[1], Gurpreet Singh[2]

[1]M.Tech Student, Department of CSE, [2]Assistant Professor, Department of CSE

Abstract - Cloud Computing is "on-demand" Service. Today there are many challenges in cloud computing environment such as Availability, Security, Resource Allocation etc. Load Balancing is a major issue related to cloud computing systems. It is very difficult to make services more ideal so as to fulfil the different demands of each client individually. In context of mobile computing, various software components can be offloaded from mobile devices to cloud. So it is very important to optimize the deployment by minimizing the network usage. Various Graph partitioning algorithms are designed that allocate the software components to the cloud. Today, ACO (Ant Colony Optimization) has attracted the attention of many researchers and a relatively large amount of successful work has been done using this algorithm. ACO strictly focus on the issue to determine the resources to be used very effectively which are allocated to the various jobs so as to balance the load on the overall cloud system. As, ants search their food themselves by some shortest and easy way and collect their food to their place. In the same way balancing of loads will be done within the network so as to increase the working efficiency of the cloud. In this paper, review of various load balancing techniques is done and proposes a method to get better optimization of graph partitioning using ACO for mobile computing.

Keywords - Cloud computing, Load Balancing, Mobile computing, Graph partitioning, Ant Colony Optimization.

I. INTRODUCTION

Cloud Computing is an internet based model of computer system where different services such as servers, storage and applications are delivered to an organization's computers and devices through the Internet. It is a technique which makes uses of combination of internet and other central remote servers [6]. With this technique, one can maintain data and applications, use these applications without installation and access them at anytime, anywhere. Main advantages of cloud computing are cost efficient, unlimited storage, easy backup and recovery [7]. Inspite of these advantages Security, Performance, Efficient load balancing, Portability, Qos management are the various issues related to cloud computing. Computation in cloud is done with the aim to achieve maximum resource utilization with higher availability at minimized cost. The Mobile Cloud Computing (MCC) term was presented after the idea of Cloud Computing. Fundamentally, MCC alludes to a foundation where both the information stockpiling and the information transforming happen outside of the cell phone. With respect to definition, versatile applications move the reckoning force and capacity from the cellular telephones to the cloud. The mobile devices need other resource providers to perform the execution of its mobile applications. Mobile cloud represents an infrastructure which could allow data storage and processing to occur outside the mobile device. The usage of mobile cloud allows execution of computer intensive applications on low resource mobile devices [18].

The usage of cloud is not beneficial for web-based applications only but for other applications also composed of many service components following the service-oriented programming [16]. Cloud over spilling is also an important scenario in which any company offloads its work from its private infrastructure to a public infrastructure which results to dimension company's infrastructure for the average workload instead of peak workload. The main aim of cloud computing is to provides services transparently among users at massive level. Various software and other data resources are provided to the systems according to their demands.

The architecture of cloud computing is categorized into Front End and Back End. Front End is User or any application i.e. Web browser etc. whereas back end is network of servers with any computer program and data storage system [13]. Fig. 1 shows the structure of cloud computing [7].

Fig.1: Cloud Computing

Cloud computing has three service models:

- Software as a Service (SaaS):- SaaS provides the capability to the user to use the provider's application which is running on cloud infrastructure.
- Infrastructure as a Service (IaaS):- IaaS provides the capability to access processing, storage , network where the user is able to deploy and run any arbitrary software.
- Platform as a Service (PaaS):- PaaS provides capability to user to deploy onto cloud infrastructure and to use applications by using programming languages. The organization of this paper is as follows. Load balancing is presented in section II. Graph partitioning is presented in section III. Ant Colony Optimizations Algorithm is discussed in IV. Performance Analysis is presented in section V and conclusion is discussed is section VI.

II. LOAD BALANCING

Load balancing is a technique to transfer the incoming load or requests among available execution nodes. Today load balancing is the major issue in cloud computing. The various load balancing services are provided by dedicated hardware or software like domain name system server. By dividing the upcoming traffic effectively between servers, it is easy to send or receive the information or data without any time delay [6]. There are different kinds of algorithms are available which are used for load balancing. Load balancing algorithms mainly divided into two categories according to state of system:

Static: Decision making does not depends on state of system i.e. free from current state of system. No more scheduling is done will take place until the work is done. The cloud provider installs homogenous resources. These resources in the cloud are not flexible when environment is made static. Therefore cloud requires prior knowledge of nodes capacity and processing power. Round Robin algorithm provides load balancing in static environment by providing the task on first-cum-first-serve basis [17]. The task which will enter first will be first allocated the resources.

Dynamic: Decision making depends on the present system condition and quickly adapts with workload fluctuations. It depends on current state of system. The cloud provider provides the heterogeneous resources. The resources are flexible in dynamic nature. Though dynamic environment is difficult to simulated yet it is highly adaptable with cloud computing environment.

III. GRAPH PARTITIONING

Graph partitioning or Graph dividing is the principle issue that has far reaching applications in numerous territories, notwithstanding exploratory figuring, VLSI outline [19] and load balancing [20]. The situation is to segment the vertices of a Graph in p generally identical parts, such that the quantity of edges connecting vertices in divergent parts is minimized. When p=2 this refers to as the min-cut bi-partitioning problem. The graph dividing issue will be NP-hard [21]. Nonetheless, several algorithms are already formulated which find a realistically good partition.

A noteworthy circumstance in this admiration is cloud over spilling. In this circumstance, an organization offloads work from its own particular private framework to an open cloud foundation on crest minutes, as demonstrated in below Fig. 2. This over spilling issue additionally emerges in the milieu of portable registering, where the cloud can be utilized to increase the capacities of a cell phone. Cell phones are getting to be fundamental piece of collaborations in today's universe of correspondence. Portable clients are profoundly experienced of the administrations of versatile applications (e.g., Google applications, and so forth), which run on gadgets or remote servers with the capable methodology of MC as an issue incline in the improvement of IT engineering and additionally business and industry fields.

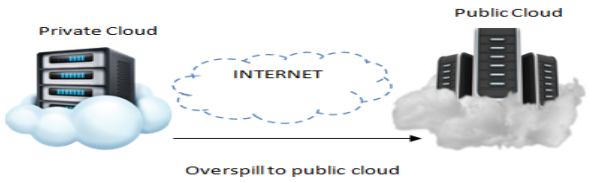

Fig. 2: Work is offloaded from private infrastructure to a public cloud on peak moments, reducing underutilization (and the cost) of the private infrastructure.

Recent work in graph partitioning explores methods based on diffusion [20] or maximum flow. The various methods partition the graph in a predefined number of parts of equal sizes.

IV. ANT COLONY OPTIMIZATION ALGORITHM

Ant colony optimization (ACO) takes motivation from the behavior of some ant species. Ants are very capable to find food. They have the shortest way to find their food. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. Ant colony optimization exploits a similar mechanism for solving optimization problems. This technique now approaches towards cloud computing. It is very effective technique for load balancing.

In early nineties the original ant colony optimization algorithm was proposed which was known as

Ant system. Marco Dorigo, Mauro Birattari and Thomas stutzle [3] focused on swarm intelligence inspired from social behaviors of insects or other animals. The foraging behavior of ants attracts the researchers time to time and at present many successful applications are available. The ethnologists were shocked how even a blind ant was able to follow the same path that was followed by its fellow ants and reaches exactly to the food source location. They found that ants leave a pheromone trail moving from one place to another. Rest ants follow this pheromone and reaches to its destination [2]. This pheromone value depends on various factors like distance of food destination, quality of food source. In various algorithms these factors affects the intensity of pheromone value. The traversal of ants is off two types [14]

- Forward movements: - In this movement, the ants move for searching for food. They continuously move in forward direction to encounter overload or under loaded node.
- Backward movements: - In this technique, after picking up food, ants traverse back to their nest to store food. If ants encounters an overload node when it has previously encountered an under loaded node then it will go backward to under load node to check if the node is still under loaded or not. If it still under loaded then it will redistribute the load to the under loaded node.

V. PERFORMANCE ANALYSIS

In [7] author has applied the Modified approach of ant colony optimization with the main aim of load balancing of nodes. This modified algorithm has an edge over the original approach in which the ants continuously update a single result set rather than updating their own result set.

Two types of pheromone are used by ants for its movements. These are as follows:

1) **Foraging Pheromone**- In this, ants explores new food sources. Ants search for overloaded nodes. The formula for foraging pheromone [1, 2] would:

$$\boldsymbol{FP(t+1)=(1-\beta eva)FP(t)+\sum_{k=1}^{n}\Delta FP}$$

Where,

Beva is pheromone evaporation rate.

FP is foraging pheromone of edge before move.

FP(*t* + 1) is Foraging pheromone of the edge after the move.

ΔFP is change in *FP*.

2) **Trailing Pheromone**- In this, ants discovers its path for going back to their nest. The formula for trailing pheromone [1,2] would:

$$\boldsymbol{TP(t+1)=(1-\beta_{eva})TP(t)+\sum_{k=1}^{n}\Delta TP}$$

Where,

Beva is pheromone evaporation rate.

TP is tracing pheromone of the edge before the move.

TP(*t*+1) is tracing pheromone of the edge after the move.

ΔTP is Change in the *TP*.

The main benefit of this approach lies in its detections of overloaded and under loaded nodes and thereby performing operations based on the identified nodes. The other advantage of the approach lies in the fact that the task of each ant is specialized rather than being general and the task depends on the type of first node that was encountered whether it was overloaded or under loaded.

VI. CONCLUSION

Ant colony optimization technique is a very beneficial technique for load balancing. ACO is inspired by the strategy of real ants of searching their food by short way, stores the food efficiently. FP and TP are two pheromones helps the ants to discover their food. As work is continuously being done on load balancing of cloud computing by many researchers, these two pheromones will be mainly focused and try to design ACO based graph partitioning algorithm for mobile users.

REFERENCES

[1] W. Ngenkaew, S. Ono, and S. Nakayama, “Multiple Pheromone Deposition in Ant-Based Clustering as an Ant Foraging Concept”, Proceedings of the 3rd IASTED Int. Conf. on Advances in Computer Science and Technology, Thailand, pp. 432-436, 2007

[2] NW. Ngenkaew, S. Ono and S. Nakayama, “Pheromone-BasedConcept in Ant Clustering”, Proceedings of 2008 3rd International Conference on Intelligent System and Knowledge Engineering, pp. 308-312, 2008.

[3] M. Dorigo, V. Maniezzo and A. Colorni, “Ant System: Optimizationby a Colony of Cooperating Agents”, IEEE Transactionson Systems, Man, and Cybernetics, pp. 29-41, 1996.

[4] M. Dorigo, M. Birattari and T. Stutzle, “Ant ColonyOptimization-Artificial Ants as a Computational IntelligenceTechnique”, IRIDIA – Technical Report Series, pp. 1-12, 2006.

[5] S. Banerjee, I. Mukherjee and P.K. Mahanti, “Cloud Computing Initiative using Modified Ant Colony Framework”, WorldAcademy of Science and Technology, vol. 3, no. 8, pp. 221-224, 2009.

[6] Foster, I., Zhao, Y., Raicu, I. & Lu, S. “Cloud computing and grid computing 360-degreecompared”,IEEE Grid Computing Environment Workshops, 2008.

[7] K.Nishant, P.Sharma, V. Krishna, C. Gupta, K. Pratap Singh, Nitin, R.Rastogi, “ Load Balancing of Nodes in Cloud Using Ant Colony Optimization”, 14th International Conference on Modellingand Simulation, pp.03-08, 2012.

[8] H. Yan, X. Shen, X. Li, M. Wu, "An improved Ant Algorithm for Job scheduling in Grid Computing", IEEE, pp. 2957-2961, 2005.
[9] H. J. A. Nasir, K. R. K. Mahamud, A. M. Din, "Load balancing using Enhanced Ant Algorithmin Grid Computing" IEEE Computer Society, pp. 160-165, 2010.
[10] S. Molaiy, M. Effatparvar, "Scheduling in Grid Systems using Ant Colony Algorithm", I. J. Computer Network and Information Security, pp. 16-22, 2014.
[11] R. S. Chang, J. S. Chang, P. S. Lin, "Balanced Job assignment based on Ant Algorithm for Computing Grids", IEEE Computer Society, pp. 291-295, 2007.
[12] J. Singh, S. Mishra, "Improved Ant Colony Load Balancing Algorithm in Cloud Computing", IJCT, vol. 14, no. 4, pp. 5636-5644, 2015.
[13] Soumya Ray and Ajanta De Sarkar, "Execution analysis of load balancing algorithms in cloud computing environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.5, October 2012 DOI : 10.5121/ijccsa.2012.2501 1
[14] Ekta Gupta, Vidya Deshpande, "A Load Balancing Technique for Cloud Computing", March 2014 Cyber Times International Journal of Technology and Management.
[15] Ekta Gupta Vidya Deshpande, "A Load Balancing Technique using Cloud Partitioning, Scheduling and Optimization of Resource using Ant Colony Optimization",International Journal of Engineering Research & Technology (IJERT) IJERTIJERT ISSN: 2278-0181
[16] T. Verbelen, T. Stevens, P. Simoens, F. De Turck, B. Dhoedt, Dynamic deployment and quality adaptation for mobile augmented reality applications, Journal of Systems and Software 84 (2011) 1871–1882.
[17] Sotomayor, B., Montero, R. S., Llorente, I. M. & Foster, I. (2009). Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing, 13(5), 14-22.
[18] N. Fernando, S. W. Loke, and W. Rahayu, Mobile cloud computing: A survey, Future Generation Computer Systems, 2012.
[19] C. Alpert, Recent directions in netlist partitioning: a survey, Integration, The VLSI Journal 19 (1–2) (1995) 1–81.
[20] H. Meyerhenke, B. Monien, S. Schamberger, Graph partitioning and disturbed diffusion, Parallel Computing 35 (10–11) (2009) 544–569.
[21] T.N. Bui, C. Jones, Finding good approximate vertex and edge partitions is NPhard, Information Processing Letters 42 (3) (1992) 153–159.

AUTHOR'S PROFILE

Davinder Singh is pursuing M.tech from Adesh College of Engineering and Technology, Faridkot, Punjab, India (davinder2607@gmail.com).

Gurpreet Singh is working with Adesh College of Engineering and Technology, Faridkot, Punjab, India. He is with the Department of Computer science and Engineering as the Assistant Professor (aiet.cse.gurpreet@gmail.com).