Cutting-Edge Load Balancing Algorithms in Cloud Computing

Monika Kushwaha
M.Tech Scholar
Pranveer Singh Institute of Technology
Kanpur,U.P. (208020)
U.P.T.U., Lucknow

Saurabh Gupta
Assistant Professor
Pranveer Singh Institute of Technology
Kanpur,U.P. (208020)
U.P.T.U., Lucknow

ABSTRACT
Cloud computing also known as on demand computing is cutting edge technology today and has gained much popularity in last few years due to rapid growth in technology and internet. Cloud computing provides software, platform and infrastructure as services over the internet which can be accessed using thin client like web browser on pay-per-use basis. With exponential growth of cloud users and demand of better services and performance from cloud provider has brought load balancing in limelight as a key issue. Load balancing ensures proper utilization of resources, scalability and user satisfaction. Several load balancing algorithms are proposed and keeps on evolving due to changing needs of cloud environment. In this paper some latest load balancing algorithms have been discussed along with brief explanation of cloud computing and load balancing.

General Terms
Cloud computing, Load balancing techniques.

Keywords
Cloud computing, Load balancing, Dynamic load balancing, Virtualization.

1. INTRODUCTION
With rapid development in technology and extensive use of internet "Cloud Computing" has become buzzword in academic, research, corporate and IT sector. In a nutshell, cloud computing can be explained by the definition given by National Institute of Standards and Technology (NIST): "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1]." The word Cloud is a metaphor for "the internet" hence the name "cloud computing" explains itself as providing on-demand computing resources over internet on pay-per-use basis which includes providing storage, platform, application, infrastructure as service. Using cloud computing is just like using electricity because you pay for only that much you used. A key technology which is backbone of cloud computing now is "Virtualization" which enabled cloud computing to optimize its resources in scalable way hence making cloud computing cost effective. Using virtualization framework a single physical server is partitioned into logical units called virtual machines (VM). Hypervisor is used to implement virtualization in physical host; it creates multiple VMs in single host and load different operating systems in it. Hypervisor provides hardware abstraction to the running Guest OSs and efficiently multiplexes underlying hardware resources. Virtualization provides various benefits to cloud computing like fast scalability, live migration of VMs, load balancing and consolidation in data center, etc.

Cloud computing has three service models:

1.1. Software as a Service (SaaS)
In this model software application is provided to customers as a service by the vendors over the internet accessible through thin client interface like web browsers. For example Salesforce, Google apps, etc.

1.2. Platform as a Service (PaaS)
In this model, cloud based platform is provided as service by vendors over the cloud to the developers where they can develop, test and deploy there software applications making it quick, simple and cost effective, for example Apprenda.

1.3. Infrastructure as a service (IaaS)
In this model, computing infrastructure such as servers, networking, data center space and storage is provided as a service by vendors over the cloud to users on pay-per-use basis, for example Amazon Web Services, Microsoft Azure, etc.

Cloud computing has four deployment models:

1.4. Public Cloud
It can be accessed by general public all over the world through internet. Some of the public cloud providers are Amazon AWS, Microsoft and Google.

1.5. Private Cloud
It can be accessed only by particular organization. It is owned and managed by that organization or third party vendor or both.

1.6. Community Cloud
It is exclusively used by a specific community of consumers from organizations that have shared concerns [1].

1.7. Hybrid Cloud
It consists of two or more clouds (private, community or public) which exist together sharing their benefits.

2. LOAD BALANCING IN CLOUD COMPUTING
Today is cloud era, everyone is using cloud in one on or other way hence users of cloud have increased rapidly which resulted in the heavy workload on cloud servers which made the load balancing as the key challenge of cloud computing. The main objective of load balancing is to distribute the workload evenly to the entire cloud to ensure at any instant of time no overloaded or under loaded nodes is present in entire network [4]. Load balancing in cloud computing consist of

two tasks: one is to schedule the incoming request to the VM in such a way that no unbalancing of load occurs and second is to monitor the nodes for under loading or overloading and take appropriate action to normalize the load among VMs. Hence load balancing provides user satisfaction, resource optimization, lower downtime, maximized throughput and minimized response time in servers. Load balancing algorithm must be designed in such a way that it runs fast and take care of all its responsibilities.

Load balancing algorithms are mainly of two types:

2.1. Static load balancing

In static load balancing the prior knowledge node's capacity and properties are kept and no current status of node is required. It is non preemptive in nature and don't considers the dynamic changes in the system.

2.2. Dynamic load balancing

In dynamic load balancing the current state of the node is also taken into consideration in combination to prior knowledge of node's attributes. It is preemptive in nature and may dynamically moves the tasks from one node to another.

3. REVIEW OF LOAD BALANCING TECHNIQUES

3.1. Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing [2]

Mohammadreza Mesbahi, Amir Masoud Rahmani and Anthony Theodore Chronopoulos proposed Cloud Light weight (CLW) which is combination of a new architecture and algorithm to balance the workload of cloud computing system. CLW architecture is multi-level event-driven architecture. CLW algorithm uses distributed dynamic approach and uses sender initiated and receiver initiated policies. In Receiver-initiated CLW phase of algorithm, VMs checks its state and if it finds itself under-loaded then it runs the load balancing process and notifies the overloaded VMs for receiving their extra workload. In Sender-initiated CLW phase of algorithm VMs which are overloaded and didn't found out any notification from under-loaded VM for long time initiates the load balancing process. They notify the head VM manager through host manager and VM manager to distribute their extra workload. Average finish time, average response time, average CPU utilization and standard deviation of CPU utilization are taken as performance parameters. It also takes care of quality of service and assures minimum migration time of load but it doesn't consider communication and migration cost which may reduce its efficiency.

3.2. Hybrid approach using Throttled and (ESCE) load balancing algorithms in cloud computing [3]

Vishwas Bagwaiya and Sandeep k. Raghuwanshi proposed Hybrid algorithm for better load balancing in cloud. This algorithm combines the best attributes of Equally Spread Current Execution algorithm (ESCE) and Throttled algorithm. Hybrid algorithm contains two lists one is index list of VM allocation status (available or busy) and other is list to count the allocated request. Comparing size of VM index list of status with allocated request list if former is grater than VM is allocated to request otherwise request is queued until VM is available. The performance parameters used are Response Time, Data Center Processing Time and Data Transfer Cost. It show improvement in all parameters but may face problem of deadlock.

3.3. Genetic Algorithm and Gravitational Emulation Based Hybrid Load Balancing Strategy in Cloud Computing [4]

Santanu Dam, Gopa Mandal et al Hybrid algorithm (GA-GEL) to balance load among VMs. This algorithm is a combination of Genetic Algorithm (GA) and Gravitational Emulation Local Search (GELS) algorithm. It uses dual optimization strategy by combining the global search feature of GA and local search feature of GELS. This algorithm uses two fitness functions, mutation, crossover, force calculation concepts. In this algorithm firstly initial population is generated and then two-point crossover is applied on it following it with mutation of chromosomes. After that gravitational force of current and child chromosome is calculated and added to velocity of that dimension. These steps keep on iterating until primary velocity of each dimension becomes zero or maximum no. of iteration has been reached. This algorithm mainly focuses on reducing number of VMs who can miss their deadlines. It shows improvement in the response time of VMs when simulated but fault tolerance and priority of jobs are not considered here which may be present in real time condition.

3.4. Double Threshold Energy Aware Load Balancing In Cloud Computing [5]

Jayant Adhikari and Sulabha Patil proposed Double Threshold Energy Aware Load Balancing (DT-PALB) algorithm which is extension of original PLAB algorithm. It provides load balancing as well as takes care that minimum power is consumed by nodes. The algorithm consist of three sections first is balancing section which balances the load by instantiating VM on most underutilized node with threshold between 25% and 75% for assigning request. VM of node having utilization below 25% is migrated to other node for power saving. Upscale and Downscale sections used to power on additional computing nodes and shut down idle nodes (nodes below 25% utilization) respectively. Average power consumption is used as performance metrics and shows improvement. It has considered small and medium sized local cloud only, that's a limitation.

3.5. Agent Based Dynamic Load Balancing in Cloud Computing [6]

Jitender Grover and Shivangi Katiyar used Agent Based Dynamic Load Balancing (ABDLB) approach and proposed mobile agent based load balancing algorithm. Mobile agent is an independent software program that runs on behalf of a network administrator. In this algorithm agent takes its first walk from any random server to last server, gathering load information and sorting the servers into under loaded or overloaded list and then in second walk, agent walks back to first server continuing from last server and does load balancing according to information gathered in first walk. CPU time unit consumed, throughput and average waiting time are used as performance measuring parameters. This method shows considerable improvement in all parameters as compared to traditional load balancing algorithms.

3.6. Optimal Load Balancing in Cloud Computing By Efficient Utilization of Virtual Machines [7]

Shridhar G.Damanal and G. Ram Mahana Reddy presented a novel VM-assign load balancer algorithm. It allocates incoming jobs to available virtual machines in such that overloading of VMs doesn't takes place. In this algorithm

whenever a new request arises to data center, the load balancer checks for least loaded VM from index table, if VM is available and it is not used in the previous assignment then request is assigned to this VM otherwise next least loaded VM is being searched. On comparing it with existing Active-VM load balance algorithm it shows improvement in distributing load efficiently. Limitation of this algorithm is it considered static predefined number of incoming request.

3.7. A Dynamic Load Balancing Strategy For Cloud Computing Platform Based On Exponential Smoothing Forecast [8]

Xiaona Ren, Rongheng Lin and Hua Zou proposed algorithm named Exponential Smoothing forecast-Based on Weighted Least-Connection (ESBWLC). They developed this algorithm taking into consideration the unique features of long-connectivity applications for example spike in online shopping services, internet auction, etc. ESBWLC uses weighted least connection(WLC) algorithm and improve its limitation by using Single Exponential Smoothing Prediction Algorithm hence it optimizes the number of connections and static weights in WLC to actual load and service capability in ESBWLC. The experiment showed that server work better when ESBWLC is used.

3.8. Autonomous Agent Based Load Balancing Algorithm in Cloud Computing [9]

Aarti Singh, Dimple Juneja and Manisha Malhotra proposed an Autonomous Agent Based Load Balancing Algorithm (A2LB) for load balancing in cloud environment. A2LB consist of three agents first is Load agent which maintains all detail of a data centre, second is Channel agent which initializes migration agent on getting request from load agent and keep record of messages received from migration agent and third is Migration agent which is an ant (a special category of mobile agent), it move to other data centers in search of desired VM and give all information to channel agent. Now algorithm works in the way that load agent keep updating fitness value of each VM and whenever fitness of VVM become below or equal to threshold of 25% then load balancing is done by notifying to channel agent which initiates migration agent to find VM of similar configuration. The advantage of using ant as agent is that it doesn't come back to its source and destroys itself in destination reducing unnecessary traffic. Major benefit of this algorithm is proactive load calculation of VM in a data center and whenever load of a VM reaches near threshold value, load agent initiates search for a candidate VM from other data centers.

3.9. A Bee Colony based Multi-Objective Load Balancing Technique for Cloud Computing Environment [10]

Ashish Soni, Gagan Vishwakarma and Yogendra Kumar Jain proposed bee colony based Multi-Objective load balancing algorithm to efficiently balance the workload of cloud. They specifically took into consideration the different resource requirement of each user in the form of task priority, task size and resources needed by them. According the algorithm cloud manager stores the request and maintains a table which stores task size, priority and resource requirements specified by user. Then bee colony algorithm is applied to calculate fitness value. Then resources are scheduled according to output of bee colony algorithm. Load execution ratio, priority requests execution ratio and unhandled task ratio are used as performance parameters and showed that it also fulfills user specific requirements.

4. RESEARCH ISSUES IN DEVELOPMENT OF LOAD BALANCING ALGORITHM

There are some research issues which should be taken care while developing a load balancing algorithm so that optimal solution could be obtained. These are as following [11]:

- While developing a load balancing solution distance between the cloud nodes should be taken into consideration. Solution developed should work efficiently for nodes which are far away from each other like in internet as well as for nodes which are close to each other like intranet.
- Implementation and operation of the algorithm being developed should not be complex because it may degrade its performance.
- Single point of failure should be avoided in proposed solution of load balancing.
- Solution being proposed should consider all possible scenarios in the cloud that is peak hours of workload, long connectivity applications producing spikes, etc. algorithm should be developed to work well in all types of workload.
- While developing solution power consumption reduction should also be taken into consideration which facilitates Green computing.

5. LOAD BALANCING ALGORITHM SUMMARY TABLE

Here summary of the key features of above discussed algorithm with their pros and cons are presented to get a simplified view of discussion.

Table 1. Summary of load balancing algorithms

Algorithm	Key features	pros	cons
CLW [2]	• multi-level event-driven architecture • distributed dynamic algorithm • uses sender and receiver initiated approach	• ensures quality of services • minimum migration time	• didn't considered communication cost
Hybrid algorithm [3]	• combines features of ESCE and Throttled algorithm	• improved response time • minimized data transfer cost	• may face problem of deadlock
GA-GEL [4]	• combines features of GA and GELS algorithm	• improved response time of VMs	• didn't considered fault tolerance

			and priority of jobs
DT-PALB [5]	• extended the PLAB algorithm	• minimum power consumption by nodes	• didn't considered large cloud
ABDLB [6]	• mobile agent based load balancing	• improved throughput • less CPU time consumed	
VM-assign [7]	• Active VM load balancing algorithm is taken as base and improved	• Optimal distribution of load	• Dynamic situation of incoming request are not considered
ESBWLC [8]	• Combines WLC algorithm and Single Exponential Smoothing Prediction Algorithm	• Give better performance when Long-connectivity application in web	
A2LB [9]	• Uses ant as mobile agent • Load agent and channel agent are static	• Reduced network traffic • Proactive load calculation of VM	• Communication cost can increase in worst case scenario
Multi-Objective load balancing algorithm [10]	• Uses bee colony algorithm • Take care of user specific requirements	• Unhandled task are reduced • Task executed according to priority given by user	

6. CONCLUSION AND FUTURE SCOPE

This review paper shows cloud computing as game changer technology which provided cost effective solutions to the corporate world as well as individuals. Load balancing appeared as critical issue in cloud computing because of heavy workload the servers may get under or over loaded. Some latest load balancing algorithms have been discussed in this paper with their advantages and limitations and some research issues are also described which should be taken care of while developing a load balancing algorithm for cloud computing. The above review of algorithms in this paper opens doors for various improvements in load balancing algorithm for cloud in future. The reviewed algorithms performs well in their specified scenario bur have some limitations. As a future scope researchers can develop new algorithm or modify the above discussed algorithms to improve the performance and resource utilization of cloud.

7. REFERENCES

[1] Lee Badger ,Tim Grance, Robert Patt-Corner and Jeff Voas, "Cloud Computing Synopsis and Recommendations", National Institute of Standards and Technology, Information Technology Laboratory, NIST Special Publication 800-146.

[2] Mohammadreza Mesbahi, Amir Masoud Rahmani and Anthony Theodore Chronopoulos, "Cloud Light Weight: a New Solution for Load Balancing in Cloud Computing", in proc. International Conference on Data Science & Engineering (ICDSE), IEEE, pp. 44-50, August 2014.

[3] Vishwas Bagwaiya and Sandeep k. Raghuwanshi, "Hybrid approach using Throttled and (ESCE) load balancing algorithms in cloud computing", in proc. International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), IEEE, pp. 1-6, March 2014.

[4] Santanu Dam, Gopa Mandal, Kousik Dasgupta and Paramartha Dutta, "Genetic Algorithm and Gravitational Emulation Based Hybrid Load Balancing Strategy in Cloud Computing", in proc. third International Conference on Computer, Communication, Control and Information Technology (C3IT), IEEE, pp. 1-7, February 2015.

[5] Jayant Adhikari and Sulabha Patil, "Double Threshold Energy Aware Load Balancing In Cloud Computing", in proc. fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE, pp.1-6, July 2013.

[6] Jitender Grover and Shivangi Katiyar, "Agent Based Dynamic Load Balancing in Cloud Computing", in proc. International Conference on Human Computer Interactions (ICHCI), IEEE, pp. 1-6, August 2013.

[7] Shridhar G.Damanal and G. Ram Mahana Reddy, "Optimal Load Balancing in Cloud Computing By Efficient Utilization of Virtual Machines", in proc. sixth International Conference on Communication Systems and Networks (COMSNETS), IEEE, pp. 1-4, January 2014.

[8] Xiaona Ren, Rongheng Lin and Hua Zou, "A Dynamic Load Balancing Strategy For Cloud Computing Platform Based On Exponential Smoothing Forecast", in proc. International Conference on Cloud Computing and Intelligence Systems (CCIS), IEEE, pp. 220-224, September 2011.

[9] Aarti Singh, Dimple Juneja and Manisha Malhotra, "Autonomous Agent Based Load Balancing Algorithm in Cloud Computing", International Conference on Advanced Computing Technologies and Applications (ICACTA), Elsevier B.V., pp. 832 – 841, 2015.

[10] Ashish Soni, Gagan Vishwakarma and Yogendra Kumar," A Bee Colony based Multi-Objective Load Balancing Technique for Cloud Computing Environment", International Journal of Computer Applications (0975 – 8887), Volume 114 – No. 4, March 2015.

[11] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi, "A Survey of Load Balancing in Cloud Computing:Challenges and Algorithms", in proc. Second Symposium on Network Cloud Computing and Applications (NCCA), IEEE, pp. 137-142, December 2012.