



LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING ENVIRONMENT

Poonam

Department of Computer Science
Deenbandhu Chhotu Ram University of Science and
Technology
Murthal, India

Suman

Department of Computer Science
Deenbandhu Chhotu Ram University of Science and
Technology
Murthal, India

Sukhdip Singh

Department of Computer Science
Deenbandhu Chhotu Ram University of science and Technology
Murthal, India

Abstract: Cloud computing is a radical technology which provides computing as-a-service instead of providing computing as a product. It provisions computing resources on demand by paying for fraction of time we are using them. Concept of cloud reduces overall cost and management efforts through efficient utilization of various resources. Cloud computing has developed too much but still it has various research challenges like security, reliability, availability, load balancing, power management and data traffic analysis etc. which need researchers immediate attention. Load balancing is one of the hot research areas nowadays. It is the process of uniformly distributing workload on geographically distributed data centers. Load balancing algorithms provide the mechanism of adjusting load according to the application and node capacity. This paper focuses on load balancing, dynamic and static load balancing algorithms and future work in this direction.

Keywords: Cloud computing, Dynamic algorithms, Load balancing, Static algorithms

I. INTRODUCTION

The concept of cloud has brought a revolution over internet for executing different applications without spending too much on various resources. The concept was firstly introduced 1960's when IBM developed its mainframe computers. Cloud inherited various concepts from mainframe systems like consumption-based pricing model, virtual machines and multi-tenancy etc. and term cloud became popular in 2007. Cloud can be divided in frontend and backend. Frontend is made up of consumers of cloud and backend of cloud providers and cloud developers. Future of cloud is glorious and it is used almost in every area these days like transportation, healthcare, mobile computing, monitoring and education etc. [1] [2].

Cloud is used almost in everything from storage to processing. It provides opportunities to third parties and large or small business organizations to flourish their business without too much initial investment. For efficient working and response from cloud, it's important to balance load on nodes of cloud. Load balancing is the uniform distribution of load on the different nodes of cloud so that none of them gets overloaded or under loaded and we get good results by enhancing various parameters like response time, throughput, energy consumption and processing speed etc. [3] [4] Various types of load balancing techniques can be applied depending on application requirements like static, dynamic, hierarchical, distributed, and centralized etc. Dynamic algorithms are used widely as they consider current state of system, can work in distributed and heterogeneous environment.

This paper is organized as follows section 2 discusses important characteristics of cloud computing, section 3 discuss various load balancing algorithms and section 4 concludes the paper.

II. CLOUD COMPUTING

Cloud inherits features of various existing technologies like distributed computing, grid Computing, utility computing and autonomic computing etc. Main idea which makes cloud different from other technologies is virtualization through which multiple applications can be executed on a single server simultaneously [5].

A. Features of Cloud

Cloud provides various features which makes it different from other technologies. Some features shown in Fig. 1. are listed below.

- **On-demand-services:** cloud resources can be accessed by user at any time and in any amount.
- **Broad access:** services can be obtained from cloud at any time and at any place.
- **Multitenant model:** multi-tenancy means multiprogramming but on pay per use basis.
- **Dignified services:** Cloud can automatically optimize the resources provided by it.
- **Scalable:** More resources can be provisioned on demand as cloud environment is scalable.

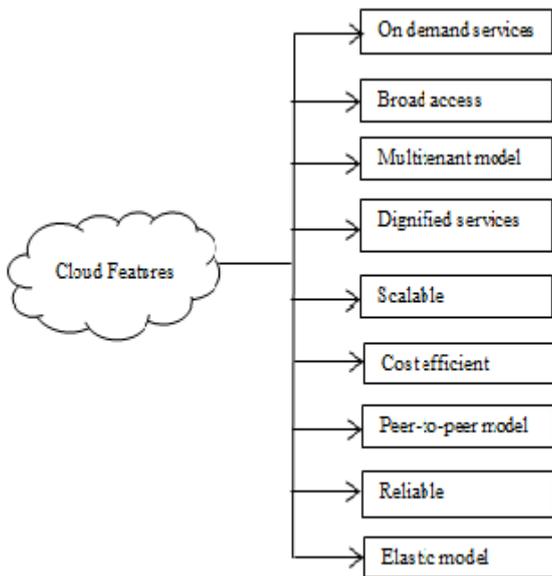


Fig. 1. Features of cloud.

- **Cost efficient:** This model reduces cost by providing shared access to resources in cloud. Cloud environment can be allocated and de-allocated depending on need. When demand decreases, these resources can be de-allocated which leads to significant decrease in cost.
- **Peer-to-peer model:** Cloud is made of distributed nodes without any central authority and all nodes (either service provider or service user) are at same level.
- **Reliable:** Cloud computing model is highly reliable as data is replicated on multiple nodes, if one node goes out of service other can handle its tasks.
- **Elastic model:** User can demand additional resources from cloud according to their requirement as cloud provides flexible environment.

B. Cloud Services

Three types of services are provided by cloud to its user software as a service, platform as a service and infrastructure as a service [6].

- **Software as a service:** This model provides different software which can be directly used by user without any concern about their infrastructure and implementation. Gmail, Facebook, Microsoft online are example Software as a service hosted over cloud [7].
- **Platform as a service:** This model provides platform for development of new applications. Google App Engine is widely used for developments of new apps without installation of any infrastructure. Other examples are Windows Azure and Heroku.
- **Infrastructure as a service:** This model provides resources to users like memory, processing, bandwidth on lease for processing of applications. Amazon Elastic Compute Cloud, Amazon simple storage service are examples of it.

C. Deployment Models

Four type of clouds are provided by cloud service providers depending on user’s demand about different parameters like security, reliability etc. as in Fig.2. [8].

- **Public cloud:** This cloud is owned and managed by large organizations like Google, Microsoft etc. and available for general public. Public cloud is less secure because of its public access.
- **Private cloud:** Private cloud is secured one and exclusive for the use of a single organization. This cloud is costly as compared to public cloud.
- **Community Cloud:** This cloud is owned and managed by multiple organizations that have same working procedures. This is cost effective way of using cloud as cost is shared among organizations.
- **Hybrid cloud:** This cloud is combination of public, private and community cloud. This cloud is secure and more flexible as it combines benefits of all three clouds. Different types of cloud deployment models are listed in Table 1.

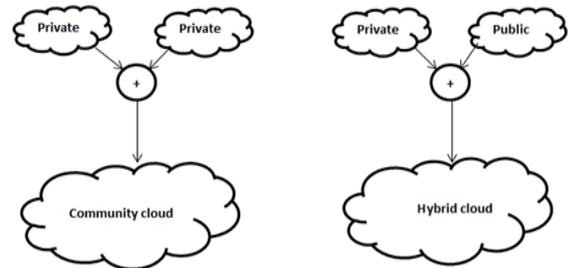


Fig. 2. Different Type of Cloud Models.

Table 1. Deployment models

Deployment model	Advantage	Disadvantage	Example
Public cloud	Cost Effective, High availability and scalability	Security and privacy issues, unreliability	Amazon EC2, Microsoft Office 365, IBM Smart Cloud, Microsoft Windows Azure
Private cloud	High security	Costly	Ubuntu Enterprise Cloud (UEC), Eucalyptus, Amazon VPC
Community cloud	Good for organizations having same policies	Costly as compared to public cloud	Google Apps for Government, Microsoft Government Community Cloud
Hybrid cloud	Cost effective and secure	Privacy and integrity issues	Windows Azure, VMware cloud

III. LOAD BALANCING

Cloud is made of large no of nodes or servers distributed geographically. Whenever a request arrives for processing of an application it needs to be distributed uniformly among servers of cloud without depleting its resources. This process of efficiently distributing load across servers without getting overloaded or under loaded is called load balancing [9][10]. Existing load balancing algorithms can be broadly classified into various categories like static, dynamic, centralized, distributed, hierarchical etc. [11][12]. Comparison between dynamic and static algorithms is shown in Table 2.

A. Static Algorithms

Static algorithms rely on history and don't consider present state of system. These algorithms are easy to implement and works for homogenous environments. These algorithms are obsolete now and not used much. Few of the static algorithms are compared in Table 3.

B. Dynamic Algorithms

These algorithms consider both present and previous state of system. They are fault tolerant and can handle busy

traffic efficiently even in heterogeneous environments. These algorithms are used often for balancing load on cloud. But these algorithms incur high cost and take more execution time. Comparative analysis of important dynamic algorithms is presented in Table 4.

Table 2. Comparison of static and dynamic algorithms.

Parameters	Static algorithms	Dynamic algorithms
State	Depends on previous state	Depends on previous as well as present state
Environment	Homogenous	Heterogeneous
Load	Can't handle heavy load	Can handle heavy load
Nature	Simple	Complex
Fault tolerant	Not fault tolerant	Fault tolerant
Example	FIFO, Round-Robin etc.	ACO, GA etc.

Table 3. Static algorithms

S. no.	Load balancing approach	Objective	Merit	Demerit
1	FIFO [13]	Efficient and simple algorithm	Fair allocation based on arrival time	Large response time
2	Round-Robin [14]	Uniform resource utilization	Free from starvation	Sometimes lead to load imbalance
3	Weighted Round-Robin [14]	To improve RR algorithm	Better performance than Round-Robin	Execution time prediction is required
4	OLB [15]	To utilize each node efficiently	Efficient resource utilization	Large makespan
5	Minimum Execution Time [15]	To execute jobs speedily	Improved makespan	May lead to load imbalance as not considering current state
6	Minimum Completion Time [15]	To execute jobs quickly	Better than minimum execution time algorithm as it considers load also	Large makespan
7	LBMM [15]	Take benefit of both Min-Min and Max-Min	Increased resource utilization	Based on theoretical concepts
8	OLB+LBMM [15]	To decrease execution time and storage cost	Reduced execution time as compared to OLB and LBMM	Network management cost
9	MIN-MIN [16] [17]	To reduce processing time	Better response time	May lead to starvation

S. no.	Load balancing approach	Objective	Merit	Demerit
10	Max-Min [16][17]	To reduce processing time	Reduced makespan	Poor load balancing

Table 4. Dynamic algorithms

S. no.	Load balancing approach	Objective	Merit	Demerit
1	Biased Random Sampling [18]	To achieve load balancing through random sampling	Good performance in large networks	Randomness may lead load imbalance
2	ACO [7], [19]	To improve computing power	Efficiently balance the load	May lead to stagnation
3	GA [20]	To minimize makespan and cost	Improved performance	This may fall in local optima
4	Particle Swarm Optimization (PSO) [21]	Better load distribution	Cost effective and scalable algorithm	Poor convergence
5	Improved Particle Swarm Optimization (IPSO) [22]	To improve execution time	Reduced makespan Elude local optima	Complex
6	ALO [23]	Reduce makespan and improve completion time	High search space exploration and exploitation	Required to test on different scales of cloud
7	Cloud Light Weight [24]	Achieve load balance without violating SLA	Minimize migration time and Scalable	Expensive approach
8	Honey Bee Foraging [25]	To meet dynamically demands	Good in heterogeneous environments	Poor improvement in throughput
9	Multi Cluster [10]	Load balancing with improved security	Uniform load balancing with high security	Complex approach
10	ESCE algorithm [26]	Allocation of VMs on priority basis	Unbiased distribution of load	Leads to central point of failure

IV. CONCLUSION

Use of cloud is ubiquitous and increasing at faster rate. Cloud is used nowadays in almost every area. Cloud computing technology has provided an opportunity for business organizations to flourish without making much investment in infrastructure. Due to ubiquitous use of cloud, data on cloud is increasing at exponential rate and it needs to be managed properly for efficient use of services. Load balancing techniques are used for uniform distribution of data so as to balance load on different servers. Millions of users are using cloud computing services and their number may increase in future. Increase in load on servers will lead to increase in response time resulting in user dissatisfaction and performance degradation. So by applying efficient load balancing technique for distributing user’s requests uniformly over nodes of server, system performance and

user’s satisfaction can be increased significantly. This paper presented an analysis of static and dynamic load balancing algorithms on cloud. Dynamic algorithms are more efficient and used more frequently for handling runtime traffic as their decision is based on both current and previous state of system. Different algorithms consider different parameters like throughput, response time, bandwidth etc. for balancing the load on nodes of cloud. Future work will be oriented towards design of an efficient load balancing algorithm considering energy consumption, VM migration etc and combining best features of static and dynamic algorithms so as to provide cloud services efficiently and with ease in different environments.

V. REFERENCES

[1] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J. and Brandic, I., “Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility,”

- Future Generation computer systems, vol. 25(6), pp.599-616, 2009.
- [2] Mondal, B., Dasgupta, K. and Dutta, P., "Load balancing in cloud computing using stochastic hill climbing-a soft computing approach," *Procedia Technology*, vol. 4, pp.783-789, 2012.
- [3] Singh, A., Juneja, D. and Malhotra, M., "Autonomous agent based load balancing algorithm in cloud computing," *Procedia Computer Science*, vol. 45, pp.832-841, 2015.
- [4] Gupta, S. and Sanghwan, S., "Load balancing in cloud computing: A review," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 4(6),2015.
- [5] Sriram, Ilango, and Ali Khajeh-Hosseini, "Research agenda in cloud technologies," arXiv preprint arXiv:1001.3259, 2010.
- [6] Alam, M.I., Pandey, M. and Rautaray, S.S., "A comprehensive survey on cloud computing," *International Journal of Information Technology and Computer Science*, vol. 2, pp.68-79,2015.
- [7] Hayes, B., "Cloud computing," *Communications of the ACM*, vol. 51(7), pp.9-11, 2008.
- [8] Zhang, Q., Cheng, L. and Boutaba, R., "Cloud computing: state-of-the-art and research challenges," *Journal of internet services and applications*, vol. 1(1), pp.7-18, 2010.
- [9] Sangwan, S., Singh, P. and Patel, R.B., "Uivh-algorithm for seamless mobility in heterogeneous wireless network," In *Proceedings of the CUBE international information technology conference* (pp. 210-215). ACM, 2012.
- [10] Sangwan, S. and Sangwan, S., "An Effective Approach on Scheduling Algorithm in Cloud Computing," 2014.
- [11] Roy, S., Bose, R. and Sarddar, D., "Fuzzy based dynamic load balancing scheme for efficient edge server selection in Cloud-oriented content delivery network using Voronoi diagram," In *Advance Computing Conference (IACC), 2015 IEEE International* (pp. 828-833),2015.
- [12] Gandhi, R., Liu, H.H., Hu, Y.C., Lu, G., Padhye, J., Yuan, L. and Zhang, M., "Duet: Cloud scale load balancing with hardware and software," *ACM SIGCOMM Computer Communication Review*, vol. 44(4), pp.27-38,2015.
- [13] Vibha M B, R.RajuGondkar, "Smart Load Balancing Algorithm towards Effective Cloud Computing," *International Journal of Advanced research in Computer and Communication Engineering*, vol. 5(4) : 543-547, 2016.
- [14] Tong, Ruixia, and Xiongfeng Zhu., "A load balancing strategy based on the combination of static and dynamic," In *Database Technology and Applications (DBTA), 2010 2nd International Workshop on* :1-4,2010.
- [15] Wang, S.C., Yan, K.Q., Liao, W.P. and Wang, S.S., "Towards a load balancing in a three-level cloud computing network," In *Computer Science and information technology (ICCSIT), 2010 3rd IEEE International Conference on* (Vol. 1, pp. 108-113). IEEE, 2010.
- [16] Belkhouraf, M., Kartit, A., Ouahmane, H., Idrissi, H.K., Kartit, Z. and El Marraki, M., "A secured load balancing architecture for cloud computing based on multiple clusters," In *Cloud Technologies and Applications (CloudTech), 2015 International Conference on* (pp. 1-6). IEEE,2015
- [17] Sran, N. and Kaur, N., "Comparative analysis of existing load balancing techniques in cloud computing," *International Journal of Engineering Science Invention*, vol. 2(1), pp.60-63,2013.
- [18] Rahmeh, O.A., Johnson, P. and Taleb-Bendiab, A., "A dynamic biased random sampling scheme for scalable and reliable grid networks," *INFOCOMP*, vol. 7(4), pp.1-10,2008.
- [19] Kaur, Davinder, and Sarpreet Singh., "An Efficient Job Scheduling Algorithm using Min-Min and Ant Colony Concept for Grid Computing," *International Journal Of Engineering And Computer Science*, vol. 3(07), 2014.
- [20] Dasgupta, K., Mandal, B., Dutta, P., Mandal, J.K. and Dam, S., "A genetic algorithm (ga) based load balancing strategy for cloud computing," *Procedia Technology*, vol. 10, pp.340-347, 2013.
- [21] Kaur, G. and Sharma, S., "Research Paper on Optimized Utilization of Resources Using PSO and Improved Particle Swarm Optimization (IPSO) Algorithms in Cloud Computing," *International Journal of Advanced Research in Computer Science & Technology (IJARCS 2014)*, vol. 2(3), 2014.
- [22] Pan, K. and Chen, J., "September. Load balancing in cloud computing environment based on an improved particle swarm optimization," In *Software Engineering and Service Science (ICSESS), 2015 6th IEEE International Conference on* (pp. 595-598). IEEE, 2015.
- [23] Mirjalili, S., "The ant lion optimizer," *Advances in Engineering Software*, 83, pp.80-98, 2015.
- [24] Mesbahi, M., Rahmani, A.M. and Chronopoulos, A.T., "Cloud light weight: A new solution for load balancing in cloud computing," In *Data Science & Engineering (ICDSE), 2014 International Conference on* (pp. 44-50). IEEE, 2014.
- [25] Randles, M., Lamb, D. and Taleb-Bendiab, A., "A comparative study into distributed load balancing algorithms for cloud computing," In *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on* (pp. 551-556). IEEE, 2010.
- [26] Suguna, S., and R. Barani., "Simulation of Dynamic Load Balancing Algorithms," *Bonfring International Journal of Software Engineering and Soft Computing* 5(1): 1, 2015.