Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.229 – 238

RESEARCH ARTICLE

Improved Service Broker Algorithm Based On Weighted Moving Average Forecast Model for Cloud Computing

Sonali Rathod
Department of CSE, G.H.R.A.E.T
Nagpur (M.S.), India
rathod.sonu01@gmail.com

Prof. Deepak Kapgate
Department of CSE, G.H.R.A.E.T
Nagpur (M.S.), India
deepakkapgate32@gmail.com

Abstract - Proper load balancing is an important aspect while calculating efficiency of cloud computing environment. In Cloud Computing Scenario Load Balancing is composed of selecting proper Data Center for upcoming request (Service Broker) and Virtual machine management at individual Data Center. In this paper, we proposed and implemented Predictive Service Broker (DC selection) dynamic algorithm based on Improved Weighted Moving Average Forecast Model. This algorithm is developed by combining advantages of both proximity service broker algorithm and DC selection algorithm based on weighted moving average prediction model. This study concludes that proposed predictive DC selection algorithm mainly focus on reducing service response time observed at client side. Thus development of effective load balancing algorithm aids in minimizing resource consumption, implementing fail-over, enabling scalability, avoiding bottlenecks etc in cloud computing. The result shows drastic reduction in Response time at client side by using Predictive Improved Weighted Moving Average Forecast DC selection algorithm. Various parameters are also identified such as Data Center request service times, Data Center hourly loading, total Data Transfer and Virtual machine costing and respective values are calculated.

Keywords - Cloud Computing, Response time, Data Center Selection Algorithm, Weighted Moving Average Model, Service Broker Policy

I. INTRODUCTION

The concept behind cloud computing is to offload computation to remote resource providers. The key strengths of cloud computing can be described in terms of the services offered by cloud service providers: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS) [3].

‘‘Cloud computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services’’ [2]. A cluster of computer hardware and software that offer the services to the general public makes up a ‘public cloud’. Computing is therefore offered as a utility much like electricity, water, gas etc. where you only pay per use. Virtualization of resources is a key requirement for a cloud provider for it is needed to create the illusion of infinite resources to the cloud user. Ambrust et al. [2] holds the view that ‘‘different utility computing offerings will be distinguished based on the level of abstraction presented to the programmer and the level of management of the resources’’.

Cloud computing has widely been adopted by the industry, though there are many existing issues like Load Balancing, Virtual Machine Migration, Server Consolidation, Energy Management, security, etc. that are not fully addressed [4]. Central to these issues is the issue of load balancing (service broker policy) that is a mechanism to distribute the workload evenly to all the nodes in the whole cloud to achieve a high user satisfaction and resource utilization ratio. The present problem with cloud computing is that bottlenecks of the system which may occur due to load imbalance, computing resource distribution inefficiently, Minimum resource consumption. So, Proper load balancing techniques not only helps in reducing costs but also making enterprises as per user satisfaction [5] [6]. Hence, improving resource utility and the performance of a distributed system in such a way will reduce the energy consumption require efficient load balancing.

In cloud computing load balancing aspect is divided into two broad categories as Data Center Selection so called **DataAppServiceBroker** and Virtual Machine Management (VMM) at each Data Center so called **DataCenterController**. The paper mainly focuses on implementation of Predictive Service Broker algorithm based on Improved Weighted Moving Average Forecast Model. This research shows that how the proposed predictive service broker algorithm leads to minimization of reduction in response time felt by users and load on data centers.

The rest of the paper is organized as follows: Section II discusses about the existing load balancing (Service Broker) algorithms. Section IV, explains proposed algorithm and flow chart of proposed algorithm. Section V shows the implementation details of algorithm and says about working environment. Section VI shows the results calculated. Finally Section VII concludes the paper with future scope.

II. Literature Survey

Round Robin (RR) and Weighted Round Robin (WRR) are most commonly Static Load Balancing Algorithm used in Cloud Computing. Round Robin Algorithm does not consider server availability, server load, the distance between clients and servers and other factors. In this algorithm server selection for upcoming request is done in sequential fashion. The main problem with this approach is inconsistent server performance which is overcome by WRR. In WRR the weights are added to servers and according to weight amount of traffic directed to servers however for long time connections it causes load tilt.

Least Connection (LC) and Weighted Least Connection (WLC) is commonly used dynamic load balancing algorithm. In LC the total no of connections on server are identified at run time and the incoming request is sent to server with least number of connections. However LC does not consider service capability, the distance between clients and servers and other factors. WLC considers both weight assigned to service node

$W(S_i)$ and current number of connection of service node $C(S_i)$ [15][16]. The problem with WLC is as time progresses static weight cannot be corrected and the node is bound to deviate from the actual load condition, resulting in load imbalance.

Xiaona Ren et. al. [17] proposed prediction based algorithm called as Exponential Smoothing forecast-Based on Weighted Least-Connection (ESBWLC) which can handle long-connectivity applications well. In this algorithm the load on server is calculated from parameters like CPU utilization, memory usage, no of connections, size of disk occupation. Then load per processor (Load/p) is calculated and this algorithm uses (Load/p) as historical training set, establishes prediction model and predicts the value of next moment. The limitation with this algorithm is this algorithm does not consider the distance between client and servers, network delay and other factors. Deepak Kapgate et. Al. [25] proposed Extended**- ESBWLC** which overcomes above limitation. In this algorithm author directly calculate the response time at client side. This got response time is store for further reference. The response time at time instance ‘t+1’ is predicted by using current response time at time instance ‘t’ and previously predicted response time for time instance ‘t’.

In this paper the author is proposing Improved prediction dynamic service broker algorithm based on Weighted Moving Average Model which improves the results of WMAFM service broker proposed by Deepak Kapgate [28] in terms of reduction in response time, reduction of data centre request timing and reduction of costing of VM and data transfer.

III. Proposed Service Broker Algorithm

The proposed predictive service broker algorithm is based on proximity service broker and DC selection algorithm based on Weighted Moving Average Forecast Model (WMAFM). A weighted moving average forecast model is based on an artificially constructed time series in which the value for a given time period is replaced by the weighted mean of that value and the values for some number of preceding time periods. As you may have guessed from the description this model is best suited to time-series data; i.e. data that changes over time.

Since the forecast value for any given period is a weighted average of the previous periods, then the forecast will always appear to "lag" behind either increases or decreases in the observed (dependent) values. For example, if a data series has a noticeable upward trend then a weighted moving average forecast will generally provide an underestimate of the values of the dependent variable. The weighted moving average model, like the moving average model, has an advantage over other forecasting models in that it does smooth out peaks and troughs (or valleys) in a set of observations. However, like the moving average model, it also has several disadvantages. In particular this model does not produce an actual equation. Therefore, it is not all that useful as a medium-long range forecasting tool. It can only reliably be used to forecast a few periods into the future.

Weighted Moving Average Forecast =

$$\frac{\sum(\text{weight for period } n)\cdot(\text{demand in period } n)}{\sum \text{weights}}$$

In this algorithm we are using Service Proximity Service Broker to select the **earliest region** based on the minimum communication delay and maximum available bandwidth from user base (client) to data center residing region. The region selection is based on the earliest/ highest region in the proximity list and any data center of the selected region is then selected randomly for the user requests to be processed [4] [6]. By using Weighted Moving Average forecast model based service broker we directly calculate predicted values for response time we get at client side. This got response time is store for further reference. The response time at time instance 't+1' are predicted by using current response time at time instance 't' for each Data Centre. The DC with minimum predicted response time is selected to satisfy the upcoming request. This would result in selecting each time DC which would leads to minimum response time observed at client side at each instance.

The flowchart for Proposed Prediction Algorithm:

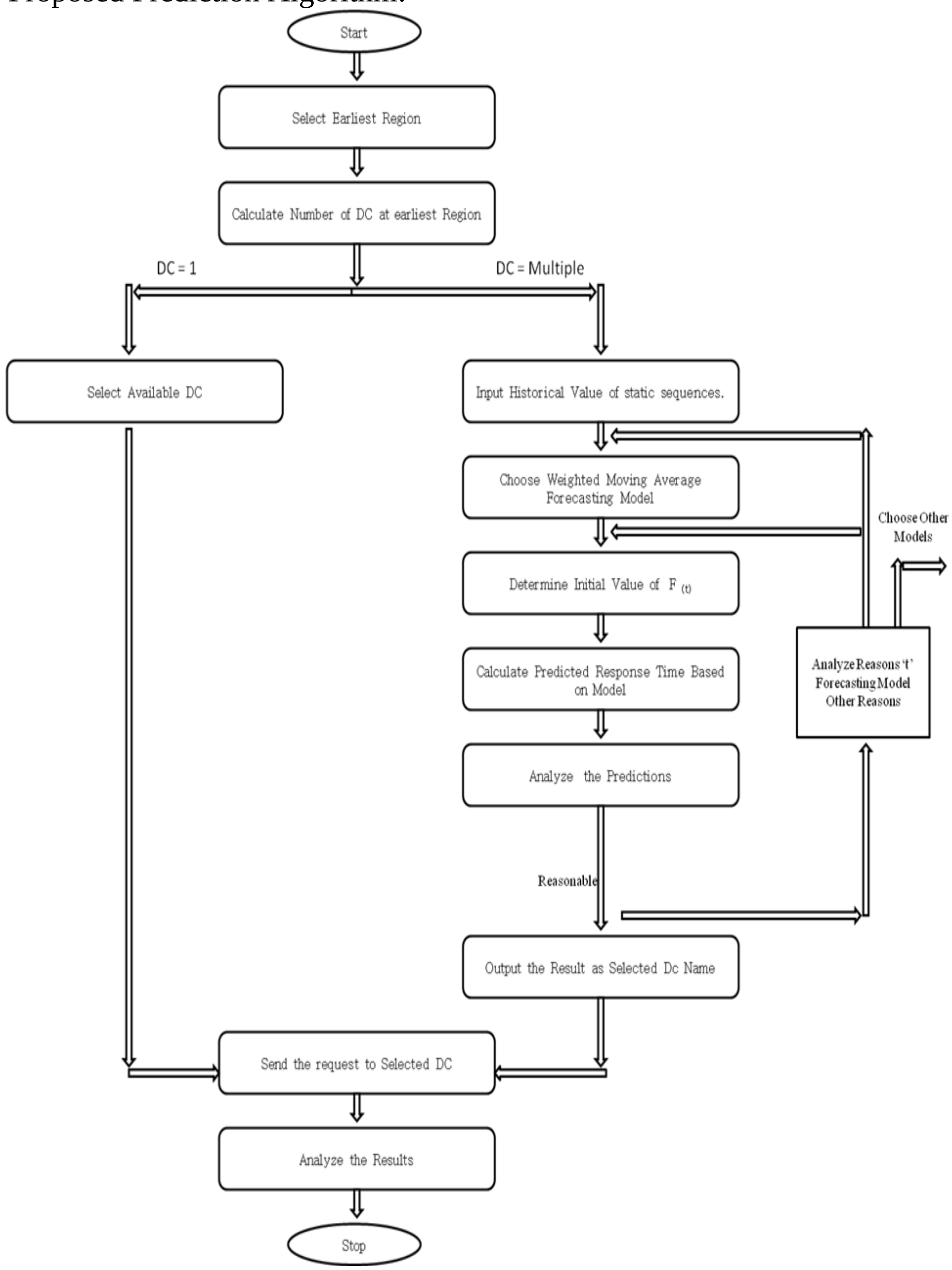

Figure 2. Flowchart of Proposed Algorithm

Proposed Algorithm:

1) Calculate the DC region with minimum communication delay and maximum usable bandwidth between user base (client) and data center as **Earliest Region** by using Proximity Service Broker Algorithm.
2) Calculate number of DC available at Earliest Region.
3) If number of DC = 1 the select the available DC to satisfy upcoming request directly.
4) If number of DC = multiple then follow following steps to select best possible DC.
5) Input historical value of statistical Response Time as a training set.
6) Use Weighted Moving Average Model as a forecasting model.
7) Determine the initial value of predictive model.
8) Calculate Predicted Response Time based on the prediction model.
9) Analyze the predictions.
10) If predictions are reasonable output the result as DC name.
11) Send the upcoming request to selected DC.

IV. Implementation Details

The working environment for cloud computing where the proposed algorithm is implemented is done using cloud analyst simulator which is built above "CloudSim", "GridSim" and "SimJava". Cloud-Analyst is built on the top of Cloud-sim. Cloud-sim is developed on the top of the Grid-sim.

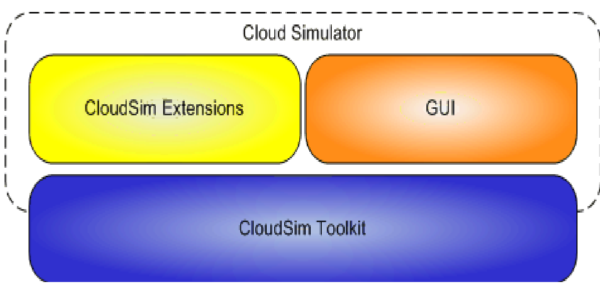

Figure 3. Cloud-Analyst built on top of Cloud-Sim toolkit

- Application users - There is the requirement of autonomous entities to act as traffic generators and behavior needs to be configurable.
- Internet - It is introduced to model the realistically data transmission across Internet with network delays and bandwidth restrictions.
- Simulation defined by time period - In Cloud-sim, the process takes place based on the pre-defined events. Here, in Cloud-Analyst, there is a need to generate events until the set time-period expires.

- Service Brokers - DataCeneterBroker in CloudSim performs VM management in multiple data centers and routing traffic to appropriate data centers. These two main responsibilities were segregated and assigned to DataCenterController and CloudAppServiceBroker in Cloud-Analyst.

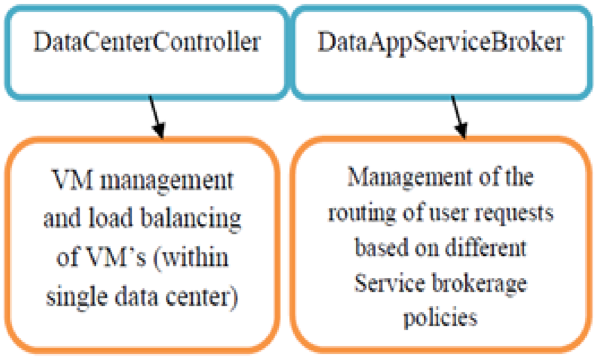

Figure 4. Responsibilities- Segregation

V. Results Calculated

The Proposed algorithm is implemented using simulation Cloud-Analyst. The scenario is taken where the data centers are located at different regions with user bases requesting services from different regions or from same region. The simulation runs approximately 60 min amount of time and the final result screen shown below as -

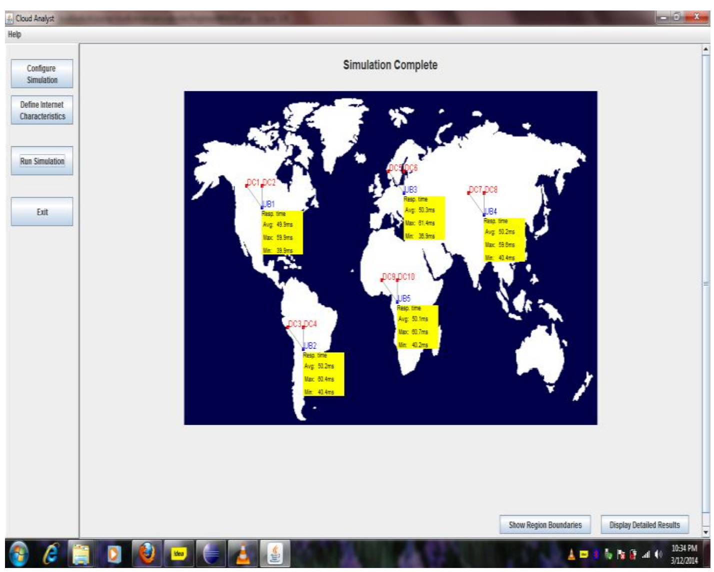

Figure 5. Cloud Analyst Main Result Screen

In above screen fig. 5 the lines show that the user base is requesting service from corresponding data center or server with the values shown at boxes at each user bases represents the response time observed by respected user base. The values are the minimum response time calculated at client side wile requesting service from data center in the duration of simulation was running ,similarly it shows the maximum response time and the average response time from above two calculated values.

The Results calculated as values of Response time observed at each client side, Data Center Service Request Times, Total VM and Data Transfer Cost. These are shown as below:

Results of the Simulation Completed at: 12/03/2014 22:25:34

Overall Response Time Summary

	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	50.12	36.88	61.38
Data Center processing time:	0.49	0.02	0.90

Response Time by Region

Userbase	Avg (ms)	Min (ms)	Max (ms)
UB1	49.83	39.86	59.86
UB2	50.21	40.40	60.39
UB3	50.23	36.88	61.38
UB4	50.11	40.38	58.64
UB5	50.21	40.16	60.66

Data Center Request Servicing Times

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC10	0.53	0.04	0.90
DC1	0.46	0.09	0.86
DC2	0.44	0.03	0.88
DC3	0.49	0.03	0.89
DC4	0.49	0.04	0.89
DC5	0.49	0.02	0.88
DC6	0.51	0.05	0.88
DC7	0.52	0.02	0.88
DC8	0.50	0.02	0.88
DC9	0.50	0.03	0.90

Cost

Total Virtual Machine Cost ($):	5.02
Total Data Transfer Cost ($):	0.32
Grand Total: ($)	5.34

Data Center	VM Cost $	Data Transfer Cost $	Total $
DC6	0.50	0.03	0.53
DC5	0.50	0.03	0.53
DC4	0.50	0.03	0.54
DC3	0.50	0.03	0.53
DC2	0.50	0.03	0.53
DC10	0.50	0.03	0.53
DC1	0.50	0.03	0.53
DC9	0.50	0.03	0.54
DC8	0.50	0.03	0.53
DC7	0.50	0.03	0.53

Figure 6. Main Result Values

VI. Comparative Analysis

A. Experiment 1 – Comparison of Response time observed by user.

The graph shows drastic reduction in average response timing observed by user for proposed algorithm as compared to traditional service broker algorithm based on weighted moving average forecast model.

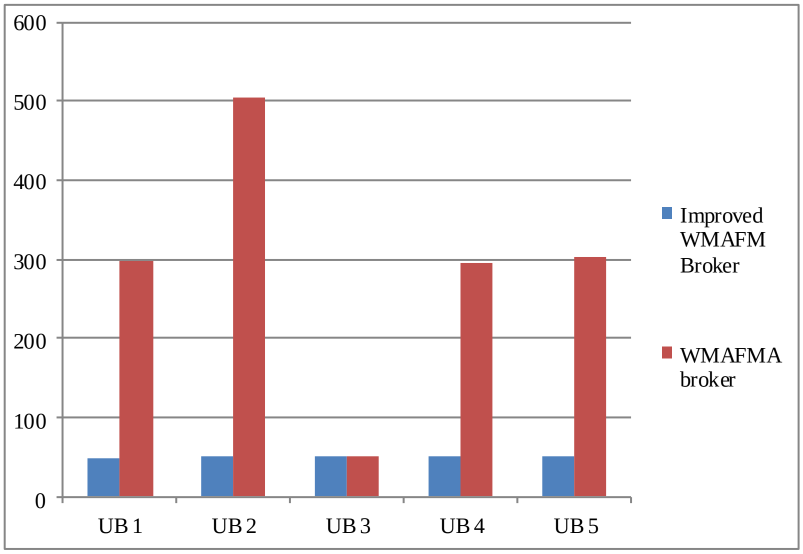

Figure 10. Response Time observed by user

B. Experiment 2 – Comparison of DC Request Service Times.

The graph shows increase in average DC Request Service Times for proposed algorithm as compared to traditional round robin service broker algorithm.

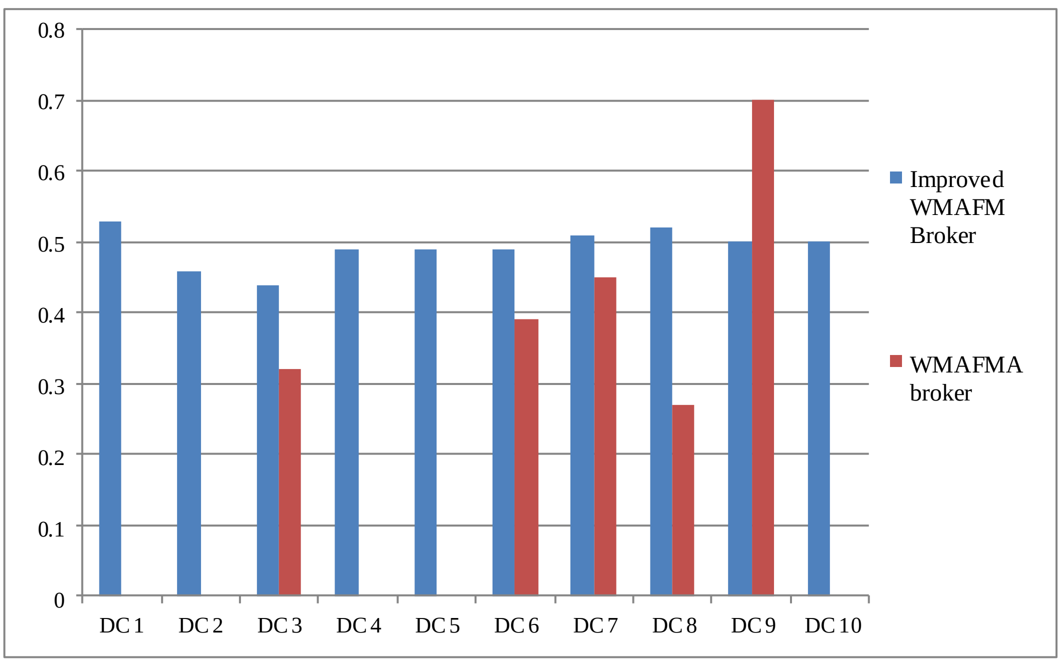

Figure 10. Data Center Request Servicing Times

VII. Conclusion

By results Proposed Improved Predictive service broker algorithm based on Weighted Moving Average Forecasting Model (WMAFM) drastically minimizes the response time observed by user which leads to improvement of service request timing. The proposed DC selection algorithm is made by combining the advantages of Prediction Techniques, WMAFM model and closest DC selection algorithm (Proximity service Broker). Finally, experiments show that proposed algorithm improves the performance of existing Round Robin algorithm. The future work may include design and development of effective service broker algorithm for multimedia and live streaming web applications.

References

[1] Vogels, "A head in the clouds the power of infrastructureas as service", Proceedings of the 1st Workshop on Cloud Computing and Applications, CCA'08.

[2] Armbrust, A.Fox, R.Griffith, A.Joseph, R.Katz, A.Konwinski, G.Lee,DPatterson, A.Rabkin, I.Stoica, , "Above the clouds : A Berkeley view of cloud computing, Technical Report UCB / EECS -2009-28,2009.

[3] J.Carolan, S.Gaede, J.Baty, G.Brunette, A.Licht, J.Remmell, L.Tucker, J.Weise, "Introduction to cloud computing architecture", whitepaper, 2009.

[4] Rimal B.P., Choi E. and Lumb I.**, "**A Taxonomy and Survey of Cloud Computing Systems**"**, IEEE 5th International Joint Conference on INC, IMS and IDC, pp - 44-51, 2009.

[5] Mata-Toledo R. and Gupta P. (2010)**, "**Cloud Load Balancing Techniques: A Step towards Green**"** Journal of Technology Research, Volume No-2 issue - 1, 1-8.

[6] Alakeel A. M., "A Fuzzy Dynamic Load Balancing Algorithm for Homogenous Distributed Systems" International Journal of Computer Science and Network Security, 10(6), 153-160, 2010.

[7] E.Marinelli, "Hyrax: cloud computing on mobile devices using Map Reduce", Master's Thesis, Carnegie Mellon University, 2009.

[8] M.Satyanarayanan, P.Bahl, R.Caceres, N.Davies, "The case for VM-based cloudlets in mobile computing, IEEE Pervasive Computing, volume – 8, 2009, pp - 14–23.

[9] Mei,W., Chan,T. Tse,A , "Tale of clouds : paradigm comparis on sandsome thoughtson research issues", ProceedingsoftheAsia-PacificServices Computing Conference, APSCC' IEEE, 2008, pp - 464–469.

[10] Xiaona Ren, Rongheng Lin,Hua Zou, "a dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast", IEEE ccis2011, pp – 220 – 225, 2011.

[11] Qingfeng Liu, Xie Jian, Jicheng Hu, "An optimized solution for mobile environment using mobile cloud computing", IEEE International Conference on Wireless Communications, Networking and Mobile Computing , 2009

[12] Qi Cao, Zhi-Bo Wei, Wen-Mao Gong, "An Optimized Algorithm for Task Scheduling Based On Activity Based Costing in Cloud Computing", International Conference on Bioinformatics and Biomedical Engineering , 2009, Page(s): 1 – 3.

[13] Luo Yongjun, LiXiaole, and Sun Ruxiang, "Load Balancing Algorithms Overview," Information Development and Economy, vol. 18, 2008, pp. 134–136.

[14] M. Randles, D. Lamb, and A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing," 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, 2010, pp. 551–556.

[15] Tian Shaoliang, Zuo Ming, and Wu Shaowei, "An improved load balancing algorithm based on dynamic feedback," Computer Engineering and Design, vol. 28, 2007, pp. 572–573.

[16] Zheng Qi, "Load balancing algorithm based on dynamic feedback," Computer Age, 2006, pp. 49–51.

[17] Xiaona Ren ; Rongheng Lin ; Hua Zou, " A dynamic load balancing strategy for cloud computing platform based on exponential smoothing forecast", 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Page(s): 220 – 224

[18] Elghoneimy, E.; Bouhali, O.; Alnuweiri, H., "Resource allocation and scheduling in cloud computing", International Conference on Computing, Networking and Communications (ICNC), 2012 , Page(s): 309 – 314.

[19] Khazaei H., Misic J. "Performance Analysis of Cloud Computing Centers Using M/G/m/m+r Queuing Systems", IEEE Transactions on parallel and distributed systems, Volume: 23, Issue : 5,2012, PP: 936 – 943.

[20] Murata Y., Egawa R., Higashida M., Kobayashi H.,"A History-Based Job Scheduling Mechanism for the Vector Computing Cloud" , IEEE/IPSJ International Symposium on Applications and the Internet(SAIT), 2010, Page(s): 125 – 128.

[21] Chih-Yung Chen; Hsiang-Yi Tseng, "An Exploration of the Optimization of Excutive Scheduling in the Cloud Computing" , IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA), Page(s): 1316 – 1319,2012.

[22] Hu Wu; Zhuo Tang; Renfa Li , "A priority constrained scheduling strategy of multiple workflows for cloud computing", IEEE international conference on Advanced Communication Technology (ICACT), Page(s): 1086 – 1089, 2012.

[23] Chih-Yung Chen; Hsiang-Yi Tseng, "An Exploration of the Optimization of Excutive Scheduling in the Cloud Computing" , IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA), Page(s): 1316 – 1319,2012.

[24] Gotoda, S.; Ito, M.; Shibata, N., "Task Scheduling Algorithm for Multicore Processor System for Minimizing Recovery Time in Case of Single Node Fault" , IEEE/ACM International Symposium Cluster, Cloud and Grid Computing (CCGrid), Page(s): 260 – 267, 2012

[25] Deepak Kapgate et. al., "predictive load balancing strategy for reduction of latency in mobile cloud computing", IJCCER, Vol No - ,Issue No - ,2013

[26] Bhathiya Wickremasinghe CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments MEDC Project. 433-659 distributed computing project, csse dept., university of melbourne June 2009.

[27] Bhathiya Wickremasinghe, Rodrigo N. Calheiros, and Rajkumar Buyya. CloudAnalyst: A CloudSim-based Visual Modeller for Analysing Cloud Computing Environments and Application. The Cloud Computing and Distributed Systems (CLOUDS) Laboratory, The University of Melbourne, Australia, 2009.

[28] D. Kapgate, "Efficient Service Broker Algorithm for Data Center Selection in Cloud Computing", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.1, January- 2014.

[28] D. Kapgate, "Weighted Moving Average Forecast Model based Prediction Service Broker Algorithm for Cloud Computing", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.2, 2014.