Simi Gupta al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 1, Issue 2, Sept 2014, pp. 74-77

Efficient broker scheduling in Cloud Computing

Simi Gupta[1], Dr. Surjeet Dalal[2]

[1]Student, M. Tech, ESEAR, Ambala
[2]Associate Professor, Dept. of CSE, E-Max group of Institutions, Ambala

***Abstract*— Cloud Computing is considered to be a next step in the field of information technology. In this research work most of the researchers are trying to work on scheduling portion in order to gain maximum profit out of it. Therefore in this work we have uses queue based approach to identify and perform scheduling in an efficient manner. Simulation results show the effectiveness of the results.**

***Keywords*— Cloud computing broker Scheduling, resource cost, resource management.**

I. INTRODUCTION

Cloud computing has been described as synonym for distributed computing. It is a cateogry computing which relies on sharing computing resources on pay per usage basis. Therefore, reducing cost of the infrastructure is of greatest importance. But reducing total charge of the cloud is not only the solution to give revenue but multiple features like response time and efficiency also plays a great role. Cloud computing in information communication technology is now coming to a place where a larget horizon of industries stake holder coming to a single point of functions . Not only industry/organizations stakeholder but the users from different fields are using cloud computing at large scale. As we know cloud computing gives a shared pool of resources in order to colloborate with different users requests dynamic users behavior patterns also have an large impact on the efficiency of cloud computing. Therefore in order to survive in the market each cloud vendor must learn their user behavior and match their cycle. Cloud computing is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. The 'cloud' in cloud computing can be defined as the set of hardware, network, storage services and interface that combine to deliver aspects of computing as a service. Cloud service include the delivery of software, infrastructure &storage over the internet based on a user demand.. In computer science Cloud computing is a synonym for distributed computing over a network and means the ability to run a program on many connected computers at the same time. The popularity of the term Cloud computing can be attributed to its use in marketing to sell hosted services in the sense of Application Service Provisioning that run Client server software on a remote location.

II. LITERATURE REVIEW

Zou and Liu[3] use ARMA (Auto-regressive Moving Average) model to predict overloads in a network. This approach helps network managers to prevent communication interruptions, or to take recovery measures beforehand.

Huang and Subhlok[4] define some network transmission patterns, which include stable states, congestion states, and interrupt states. When a network transmission pattern resembles a denoted pattern, this information is utilized for pre-diction. This method has been compared against traditional methods like a simple moving average, exponential weighted moving average, and aggregate measured throughput. This approach performs as well or better than the other methodsin given scenarios Borzemski and Starczewski[5] focus on the regression based algorithms to predict time transfer. Similar to Huangand Subhloks work[4], this study develops a pattern state recognizer to perform the TCP throughput prediction of data transfers originated by clients.

There are some prior approaches aiming at building energy-efficient data centers, such as [6],[7], Bianchini, and Rajamony[8], and [9] ,[6] believe that Cloud Computing with Virtualization is a way to improve the energy efficiency of a datacenter.

[7] dynamically reconfigures a heterogeneous cluster to reduce energy consumption during off-peak hours. Bianchini, and Rajamony[8] identify the techniques for conserving energy in heterogeneous server clusters.

[9] show that using the dynamicvoltage scaling (DVS) on each server node can achieve energy saving of 29% . Moreover, by turning off certain nodes based on workload achieves 40% energy saving.Petrucci et al.[10] propose a control mechanism for turning on/off server nodes according to the client connection number. This control mechanism maintains a pre-definedQoS (Quality of Service) while eliminating unnecessary power consumption.

[11] install several virtual machines (VMs) into a physical machine. In this approach,a network flow forecasting program acts as the manager for controlling the activation/deactivation of each node.

III. PROPOSED WORK

3.1 MATHEMATICAL MODELLING

We model the network as a bipartite graph G = (L|V,E), where L denotes the set of data centers, V denotes the location of customers. For instance, V can be the set of access networks to which customers are connected. Denote by E| L×V the communication paths between customers and data centers. We also assign constant weights dlv to denote the network latency between a data center l ∈ L and a client location v ∈ V . In our framework, we consider a discrete-time system model where time is divided into multiple time periods called reconfiguration periods corresponding to the

timescale at which server placement and routing decisions are made. We assume that there is an interval of interest K = {0, 1, 2, ...,K} that consists of K+1 periods. Let N = {1, 2, ..., n} denote the set of SPs. We assume that at time k ∈ K, each customer location v ∈ V has demand Dvk in terms of average arrival rate of requests from location v at time k. For simplicity, we assume that all the servers leased by each SP have identical size and functionality. For instance, a server can be a virtual machine (VM) that runs a specific application image. We define the state variable xl k| R+ as the number of servers owned by the SP at location l ∈ L at time k. To simplify the model, we assume that xl
k can take continuous values rather than discrete values. This assumption is reasonable for large-scale services that require tens or hundreds of servers, where the weight of each individual server in the overall solution is small. In this case, we can always obtain a feasible solution by rounding up the continuous values to the nearest integer values. Based on this assumption, we can further decouple xl k by defining xlv k ∈ R+ as the number of servers at location l serving demand from v | V .

3.2 System Model

In this fig we can see that user is interesting to CSP by using broker.

It is the accountability of the broker to look at the users location and decide to forward the requests.Forwarding requests is wholly depends upon request size and location .Each data center is composed of several VMS as shown in the fig. which allotted to the
Users.

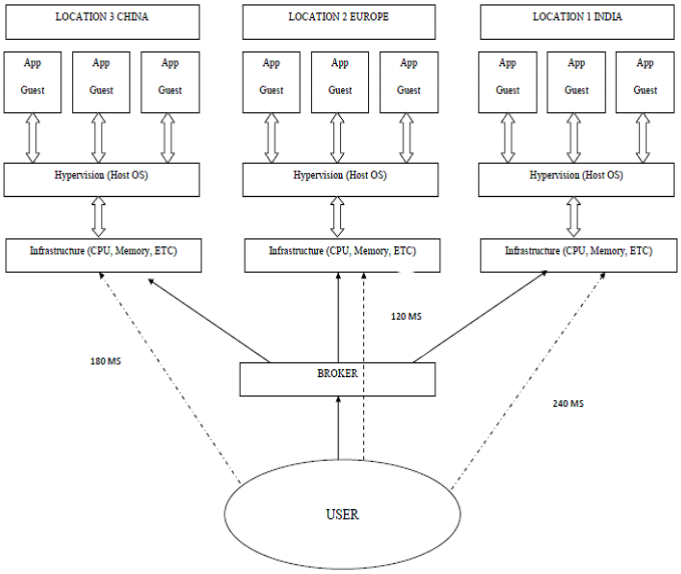


FIGURE 4.3-Block Diagram of Schematic Model Of Proposed Framework

ALGORITHM

- Resources of virtual resource pool are pretreated.
- The reached N tasks are put into buffer, they form a set.
- N tasks are divided into M classes, the special kind of tasks are put together and make up a set, the tasks of large amount of calculation are put together and make up a set, The tasks of small amount of calculation k are put together and makeup a set, it has M sets.
- Choose a task in each queue head; there are M tasks in total.
- M tasks are scheduled to virtual machines at the same time, the tasks of large amount of calculation are scheduled to resources queue whose calculation ability are strong., the tasks of small amount of calculation are scheduled to resources queue whose calculation ability are weak.
- On the basis of cost information available for each virtual machine instance resource is allocated in the corresponding queue.

IV SIMULATIONS AND ASSUMPTIONS

- We consider only one datacenter per cloud service provider, although it can be easily extended to include multiple datacenters.
- For simplicity, we assume that the physical servers in datacenters are of similar pattern and have the capability of running the same number of VM instances.
- The resource requests are in terms of VM instances, although cloud users consume services ranging from Infrastructure-as-a-Service (in terms of physical server instances) to Software-as-a-Service (higher order applications and services). We assume that the "Broker" will translate such requests into VM instances required to provide those services

Number of cloud service providers	50	60
Number of physical servers per datacenter	100~300	150~350
Maximum Virtual Machines per server	5	10
Resource request quantum	10~50 vms per request	15~60 vms per request
Resource request frequency:	2~5 per minute	5~8 per minute
Duration of resource usage	30~60 minutes	35~65 minutes
Flash-crowd scenario frequency	once every 3 hours	once every 5 hours
Flash-crowd scenario duration	10 minutes	20 minutes
Flash-crowd resource request frequency	15~20 per minute	20~35 per minute

Table 2 The following parameters were considered during simulation

Experimental Setup

In the experimental set up we have evaluated the response time of each of the Data-Center. The datacenter are compared for base algorithm and our proposed algorithm. We have compared the response time of each of the datacenter for each algorithm as shown below. We have compared them on the basis of response time and cost it took

for comparison purpose. The Snap shot explains the User Base response time for the user bases in Figure 5.7.

The following is the configuration we took for the datacenters and the number of users bases from where the request is coming.

Number of Data-Centers - 4 (DC 1, DC 2, DC 3, DC 4)
Number of User Bases - 5 (UB 1, UB 2, UB 3, UB4, UB 5)

In the comparison we have taken round robin algorithm and throttled (queue based) and then compare in the simulation scenarios. It is evident from the simulation results that our framework has outperform on both the parameters (Response time and Cost).

1. The Snap shot explains the User Base response time for the user bases in Figure 6.2.1

Results from Base Algorithm

Overall Response Time Summary

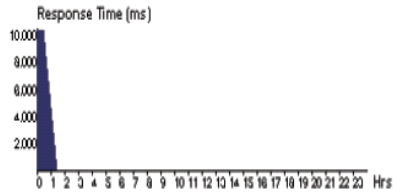
	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	10310.57	7489.51	13285.03
Data Center processing time:	10008.83	7200.01	13000.01

Response Time by Region

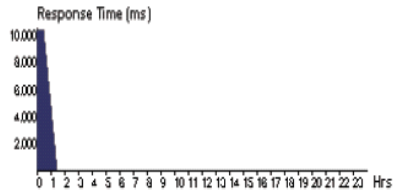
Userbase	Avg (ms)	Min (ms)	Max (ms)
UB1	10380.66	8101.53	13285.03
UB2	10444.29	7489.51	12491.02
UB3	10208.52	7865.51	13100.03
UB4	10145.77	8179.02	12509.02
UB5	10367.04	8176.01	12513.52

User Base Hourly Response Times

UB1



UB2



GRAPH 6.3.1-User Base response time

Result from throttled shown in Graph-6.3.2

Overall Response Time Summary

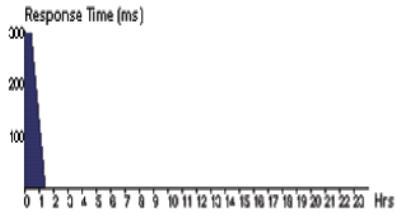
	Avg (ms)	Min (ms)	Max (ms)
Overall response time:	300.13	229.62	373.67
Data Center processing time:	0.36	0.02	0.68

Response Time by Region

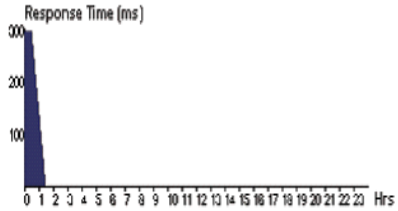
Userbase	Avg (ms)	Min (ms)	Max (ms)
UB1	299.77	246.12	363.14
UB2	301.24	241.64	370.61
UB3	299.31	240.19	369.14
UB4	300.16	234.14	370.61
UB5	300.13	229.62	373.67

User Base Hourly Response Times

UB1

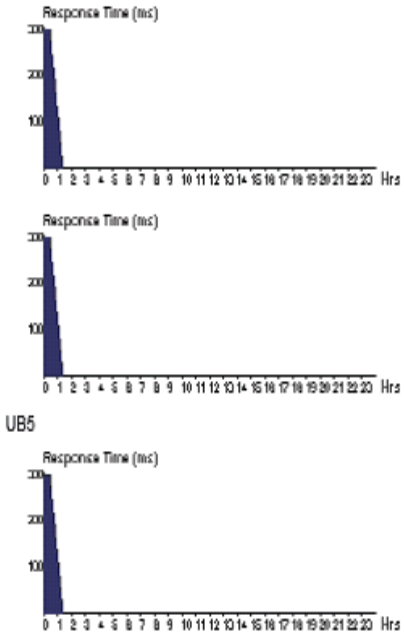


UB2



Total Virtual Machine Cost ($):	1.51
Total Data Transfer Cost ($):	0.32
Grand Total: ($)	1.83

Data Center	VM Cost $	Data Transfer Cost $	Total $
DC3	0.50	0.11	0.61
DC2	0.50	0.10	0.61
DC1	0.50	0.11	0.61



Data Center Request Servicing Times

Data Center	Avg (ms)	Min (ms)	Max (ms)
DC1	0.36	0.02	0.64
DC2	0.37	0.02	0.68
DC3	0.36	0.02	0.66

GRAPH 6.3.2-Result from throttled (User Base response time)

Experimental Results.

CloudSim is the the core of the simulation and in the results we can observe that in base algorithm the average response time is 10000ms while as in our proposed frame work it is 300-400 ms (average). Further we can see that the cost of the prosed algorithm is also very limited hence it is capable to generate more revenue for the users as well as CSPs.

V. CONCLUSION AND FUTURE WORK

Migration of virtual machines is a well-organized system used to implement cost saving and load balancing in virtualized cloud computing data center. In this paper, we study the request allocation of multiple virtual machines from experimental perspective and investigate different resource reservation methods in the cost saving process as well as other complex migration strategies such as parallel migration and workload-aware migration. Experimental results show that:

- Migration of virtual machine brings some performance overheads.
- Queuing and pre-analyzer provides further enhancement in decision making
- Resource reservation in target machine is necessary to avoid the migration failures and performance cost.

In this work we have tried to implement our technique and compared that technique with existing commonly used. In this process we are able to conclude that queuing and pre-analyzing plays an important role in minimizing the response time and sla violations.

Future Work will include the impact of heuristic approaches in the proposed framework. The broker will be loaded with historical results in order to respond to the user's requests and the effect of this change will be observed by comparing existing techniques.

REFERENCES

[1] http://en.wikipedia.org/wiki/Cloud_computing
[2] Pamlin, D. (2008) The Potential Global CO2 Reductions from ICT Use: Identifying and Assessing the Opportunities to Reduce the First Billion Tonnes of CO2 , Vol. May. WWF, Sweden.
[3] Zou and Liu Data Centre Energy Forecast Report. Final Report, Silicon Valley Leadership Group, July.
[4] Huang and Subhlok Metrics to Characterise Data Centre & IT Equipment Energy Use. Proc. Digital Power Forum , Richardson, TX, USA, September.
[5] Borzemski and Starczewski ORGs for scalable, robust, privacy-friendly client cloud computing. IEEE Internet Comput., September, 96–99.
[6] Fan, X., Weber, W.-D. and Barroso, L.A. (2007) Power provisioning for a warehousesized computer, Proc. 34th Annual Int. Symp. Computer Architecture , San Diego, CA, USA, June 9–13, 2007. pp. 13–23. ACM, NewYork.
[7] D. Nurmi, R. Wolski, C. Grzegorczyk, G. Obertelli, S. Soman,L. Youseff, and D. agorodnov, "The eucalyptus open-source cloud-computing system," in Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid- Volume 00, pp. 124–131,2009.
[8] Bianchini, and Rajamony "Virtual infrastructure management in private and hybrid clouds," IEEE Internet Computing, pp. 14–22, 2009.
[9] The green grid consortium, 2011. URL: http://www.thegreengrid.org.
[10] K. Ye, X. Jiang, D. Ye, and D. Huang, "Two Optimization Mechanisms to Improve the Isolation Property of Server Consolidation in Virtualized Multi-core Server," in Proceedings of 12th IEEE International Conference on High Performance Computing and Communications, pp. 281–288, 2010.
[11] Amazon Elastic Computing Cloud, aws.amazon.com/ec2
[12] Amazon Web Services, aws.amazon.com

