

Research Article

Resource Optimization Based on Demand in Cloud Computing

¹Ramakrishnan Ramanathan and ²B. Latha

¹Department of Information Technology, Dhanalakshmi College of Engineering, Chennai, India

²Department of Computer Science and Engineering, Sri Sairam Engineering College, Chennai, India

Abstract: A Cloud Computing gives the opportunity to dynamically scale the computing resources for application. Cloud Computing consist of large number of resources, it is called resource pool. These resources are shared among the cloud consumer using virtualization technology. Virtualization technologies engaged in cloud environment is resource consolidation and management. Cloud consists of physical and virtual resources. Cloud performance is important for Cloud Provider perspective predicts the dynamic nature of users, user demands and application demand. The cloud consumer perspective, the job should be completed on time with minimum cost and limited resources. Finding optimum resource allocation is difficult in huge system like Cluster, Data Centre and Grid. In this study we present two types of resource allocation schemes such as Commitment Allocation (CA) and Over Commitment Allocation (OCA) in the physical and virtual level resource. These resource allocation schemes helps to identify the virtual resource utilization and physical resource availability.

Keywords: Cloud computing, resource allocation, resource performance management, virtualization

INTRODUCTION

Cloud computing describes a new supplement, consumption and delivery model for IT services. It involves over-the-internet provision of dynamically scalable and often virtualized resources. Cloud computing environment everything becomes a service. The ultimate goal of the cloud computing is to reduce the cost and dynamic resource allocation and cloud consumer need not worry about the maintenance of the resource. Also scalability, elasticity, multi-tenancy, pay-per-use make cloud computing. Of course some concerns in cloud like security and reliability.

The author proposed the advantages and disadvantages of could computing only for the purpose of user satisfactions. A new model has been introduced in between the cloud provider and cloud user in three ways. At the first way the cloud user must be satisfied on the experience of cloud provider previously. Secondly the user must have an idea about cloud computing SLA. Cloud advantages and disadvantages as well as the different level of security. At the third way user/owner of organization can believe on reliable cloud provider. Different researchers explore cloud computing which is the best way for organization management and control. The main problem in believing the cloud computing may be security and other issues (Shakeel *et al.*, 2012).

As dictated by the future internet vision, an enhanced communication and computational services

with Quality of Service (QoS) requirement have been developed. In the objective of resource mapping procedure a Mixed Integer Programming (MIP) problem have been formulated in optimal networked cloud mapping problem for network allocation (Papagianni *et al.*, 2013; Zhenping *et al.*, 2012). Social network believes that as “apps” will become easier to share their services resources and data. In the presence of user sharing preferences. The construction of cloud computing infrastructures can be allocated for a social community resources.

To support both internal and public cloud users a cloud federation have been established by the multiple cloud providers. For better computing resource utilization simulation results shows the game theory coalition (Zhenping *et al.*, 2012).

The author proposed that the globally scattered distributed cloud resources of different organizations are used in co-operative manner to provide services. Reputation management methods are not effective because it provides single reputation value for each node. Collaboration Cloud Computing (CCC) has become a popular paradigm over the internet to customer for usage of computing resources, storage and bandwidth (Linna and Lifang, 2012).

MATERIALS AND METHODS

In cloud computing resource allocation is batch of jobs. Jobs are allocated non-preemptively. Cloud

Corresponding Author: Ramakrishnan Ramanathan, Department of Information Technology, Dhanalakshmi College of Engineering, Chennai, India

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

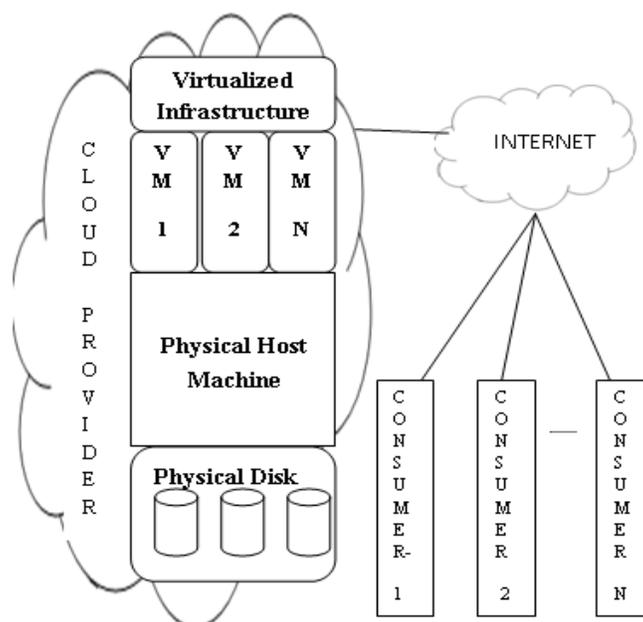


Fig. 1: Cloud architecture

provider need to complete the batch of jobs fully to get gain on the resource. Cloud provider can decide the job request may accept or reject due to other allocating jobs gain (Navendu *et al.*, 2014). Figure 1 shows the cloud environment computing ability, storage capacity and bandwidth are allocated for the resource. If want to allocate the resource simultaneously use the optimal joint multiple resource allocation. The hardware resource installed in different center to support addition resource for large amount of processing. Resource allocation algorithm identified required processing capability or bandwidth (Shin-Ichi, 2011).

The author suggested determination of the resource needs can be classified in two ways one is rely on service benchmark and another is monitoring services as they run to determine their needs (Mark *et al.*, 2010). The cloud user able to scale up and down their usage based on the requirements. The resource to be allocated from the cloud center based on application requirements in dynamically, from that we can optimize the number of server usage in cloud center and cloud center world load can be managed in great manner and improve the cloud center resource utilization.

Predicting future resource need of CPU load on DNS server, measuring the load every minute and predict the load in the next minute. Two types of algorithms for predict the workload on the cloud center such as Fast Up and Slow Down (FUSD) algorithms and Skewness algorithm. FUSD algorithm shows the stability and performance of resource allocation decision. Skewness algorithm combing different types of workloads and improve the overall utilization cloud center resource (Zhen *et al.*, 2013).

Static resource allocation model to allocate fixed resources to an application, the application may be

slowed down sometimes due to insufficient resource, or excess resources are wasted when application is not at its peak load. To overcome this problem the dynamic resource allocation scheme is the best. In the dynamic resource allocation scheme need to know, when to reallocate the resource and how much resource to be adjusted. Wei *et al.* (2011) suggested threshold-based dynamic resource allocation scheme is to monitor and predict the resource needs of the cloud applications and adjust the virtual resources based on applications actual and proposed two functions such as Datacenter and Broker. The broker function runs on cloud user machine with the application. The Datacenter function work as manage of cloud computing resource runs on datacenter. The two functions interact with each other dynamically manage the virtual resources for cloud application. Datacenter function manages the CPU and RAM, waits for request from brokers and provides extra virtual resources or revoke excess virtual resources based on the request from brokers. The broker determine the whether need to more virtual resources or excess virtual resources. Dynamic provisioning techniques to ensure the application achieve QoS and provided two provisioning policies such as time-shared and space-shared (Rodrigo *et al.*, 2011).

Managing virtualized resources and selecting best resource allocation is a difficult task for the cloud providers a service provided with an insufficient number of resources, it can lead two issues such as over provisioned and under provisioned. If these kind of problems to manage need automated resource manager or the system administrator themselves to evaluate

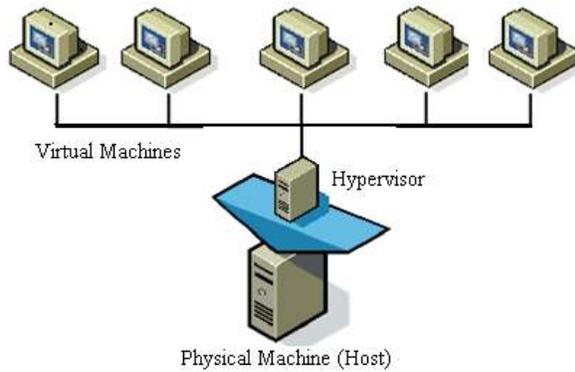


Fig. 2: Test cloud centre

many possible resource allocation quickly and correctly. These techniques require time-consuming, re-adjust and re-validation whenever the workload changes. To overcome this problems the author (Nedeljko *et al.*, 2012) proposing Deja Vu. The DejaVu is to cache and reuse the results of previous resource allocation decision. DejaVu is a framework, which detects the workload conditions, if the work load changes, it enable the cache. DejaVu useful for cloud provider to manage the SLA efficiently when workload changes.

Different organization distributed cloud resources and individual cloud resources are collectively pooled and used in a cooperative manner to provide service is Collaborative Cloud Computing (CCC) (Haiying and Guoxin, 2014). CCC enables resource sharing between clouds and provides huge amount or resource to cloud customers. If the cloud customer need more amount of resource, which is not available in the cloud, it search and uses the resource in other cloud. Monitoring the service and taking runtime decision for dynamic capacity management. Re-allocating the virtual machine from an overloaded host to another not critical host, it improves the resource utilization and better load sharing (Mauro *et al.*, 2010).

The Nephele's architecture, it is a data processing framework using for dynamic resource allocation, in IaaS cloud. Nephele's ability to adapt to resource overload and underutilization during the job execution automatically (Warneke and Odej, 2011). Auction based resource provisioning and the VM allocation in the cloud market (Rajkumar *et al.*, 2011). The auction is effective of dynamic resource provisioning over static resource provisioning (Linqun *et al.*, 2014). Minimizing the both under provisioning and over provisioning problem. The Optimal Cloud Resource Provisioning (OCRP) algorithms minimizing the cost for provisioning resources (Chaisiri *et al.*, 2012).

Problem definition: Cloud resource performance is extremely necessary for cloud provider and cloud consumer, if the cloud resource is not performing then

both will lose their money. Today most enterprise has shifted their attention from cloud security problem to cloud resource performance and availability. The performance refers the application response time, throughput and resource utilizations. The cloud consumers are interested in application response time, cloud network response time.

Evaluating various performances like resource utilization, throughput, response time using different work load types Commitment and Over Commitment schemes. To evaluate the efficiency of resource allocation, simulate various resource allocation policy tested in Cloud Sim (Rodrigo *et al.*, 2011) environment using this simulator cloud provider can measure the performance of the application running on behalf of the customer.

Capacity management: The objective of Capacity Management (CM) is to ensure adequate availability of resources for all service based on their Service-Level Agreements (SLA). CM provides analysis and compares allocated storage to the forecasted storage on a regular basis. CM provides trend analysis of the actual utilization of allocated storage. Figure 2 shows the test performed using configuration with single physical host and five virtual machine.

Performance management: Performance Management (PM) ensures the optimal operational efficiency of all components. Performance analysis is important activity. It provides information on whether a component is meeting the expected performance levels. If the cloud computing resource is not performing well then losing money by the cloud provider as well as cloud consumer. Hence, now enterprises had shifted their focus from cloud security to cloud resource performance. Security was a problem a few years ago but it's now about performance and availability Performance refers the application response time, throughput and resource utilization levels.

We have successfully built an infrastructure cloud computing using VMware ESXi and created Virtual Machine (VM). In the VM being installed with application service which will be tested and measured about its performance using benchmark well known standard benchmark testing tool passmark-cpu test.

Figure 3 shows the overall host machine details from that can extract the information with automated script.

CPU performance: CloudSim has been used to measure the CPU performance for various test cases.

Test case:

One host and five virtual machine: In our evaluation five virtual machine running under a physical machine (Host) shows Fig. 2. It has 4 processing unit and

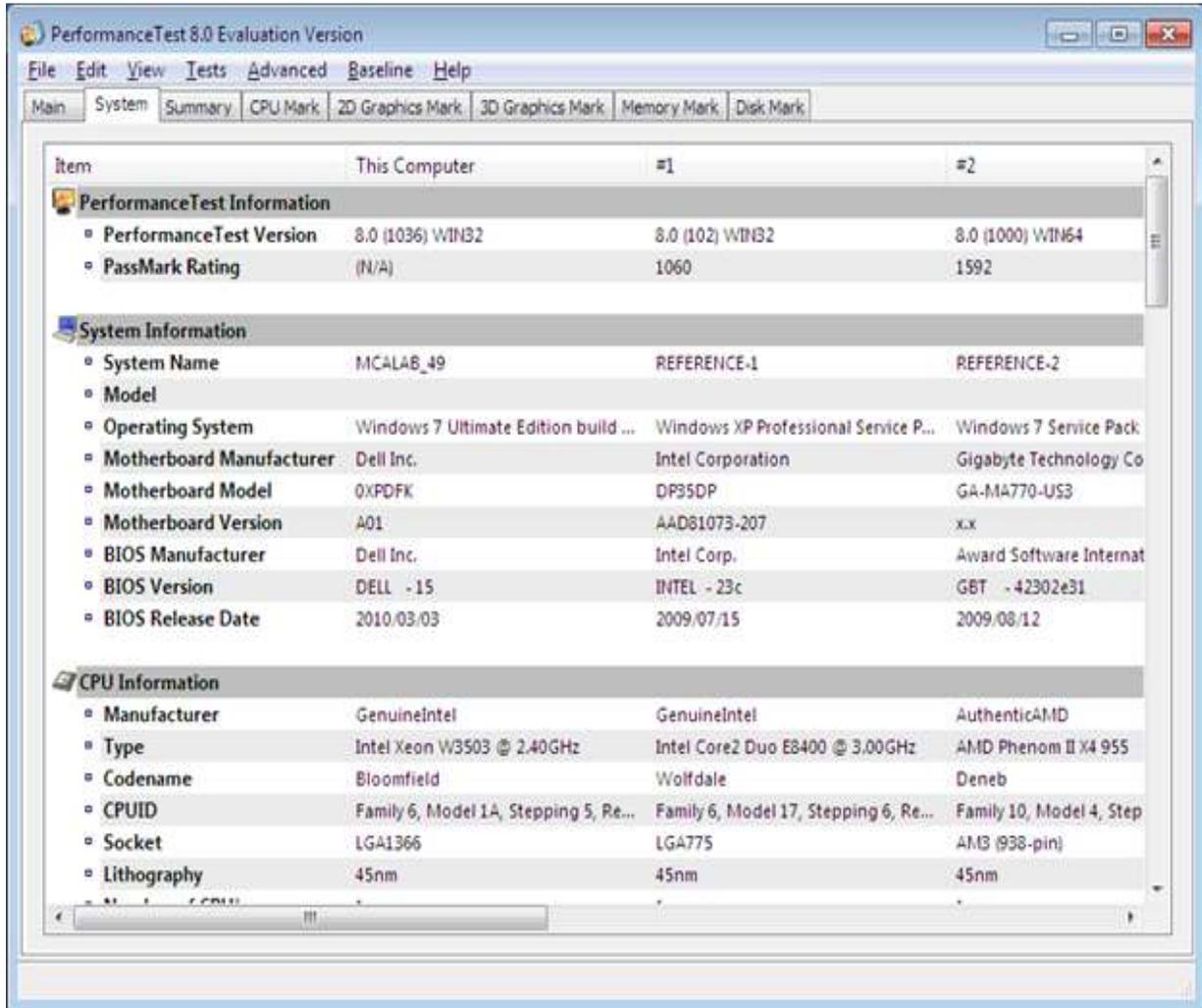
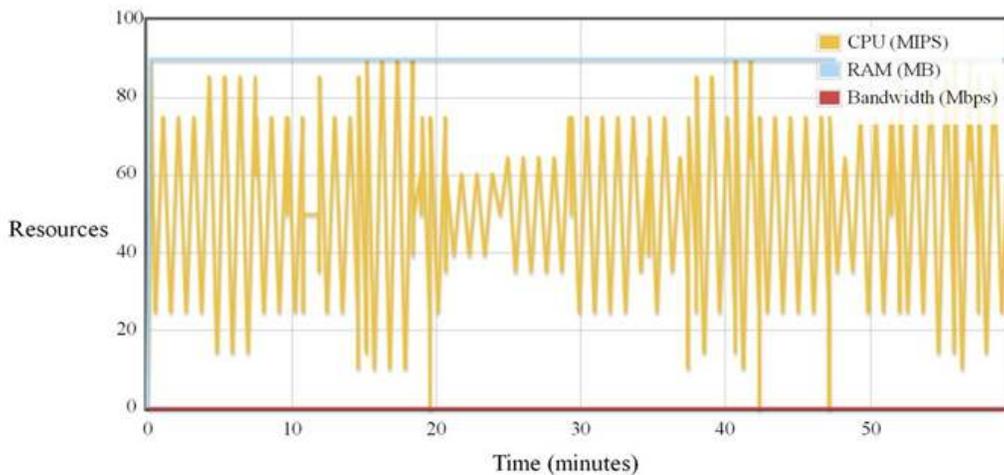
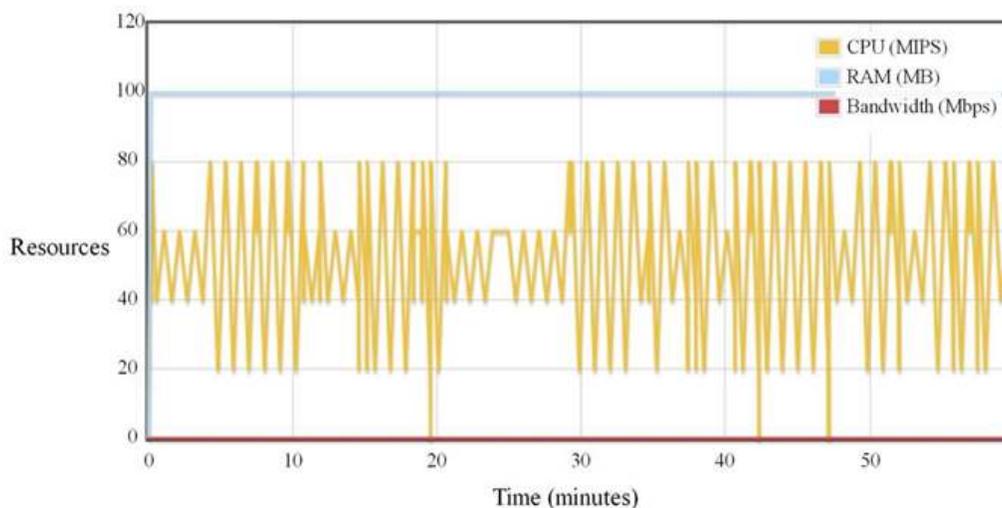


Fig. 3: Overall host machine details



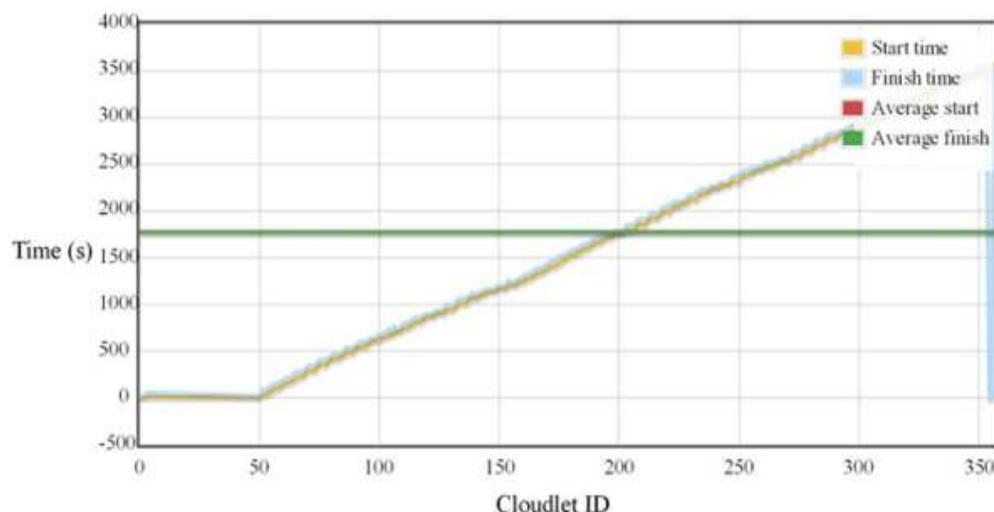
This graph shows the overall resource utilization on this datacenter.

Fig. 4: Resource utilization on data centre



This graph shows the overall resource utilization of this customer.

Fig. 5: Resource utilization of customer



This graph shows the amount of time each of this customer's cloudlets took to be executed.

Fig. 6: Customer virtual machine response time

Processing capacity (MIPS) 9600, storage capacity 11 TB and total amount of RAM 40 GB. Figure 4 shows the overall resource utilization on the datacenter and Fig. 5 shows customer utilization and Fig. 6 shows customer virtual machine response time.

Network performance: Network bandwidth can be effect on the performance, if the bandwidth is too low provide to cloud user, performance will be low too. Throughput is the measurement of bandwidth consumed during the process.

Resource allocation: In cloud computing resource allocation is one of the challenges. Resource allocation scheme decides the amount of resource to be allocated a particular or set of virtual machine. Cloud computing is using virtualization technique, by which easily can optimize the usage of resource, management of infrastructure, software and reduce the hardware requirements. Our research focus on Infrastructure as a Service (IaaS), resource allocated by cloud provider for cloud consumer. The main objective is minimizing the number of servers or the server cost.

In cloud computing, virtualization technology used to create/destroy virtual machine to dynamically allocate/reduce resource for an application.

In this study not implemented in any particular virtualization technology in physical Machine. Most of the virtualization techniques allow different allocation policy of resources to virtual machines. Basically resource allocation policies are implemented by virtualization techniques.

Allocation schemes decide the amount of resource share, which will be allocated to the virtual machine. We have implemented two schemes Committed Allocation (CA) and Overcommitted Allocation (OCA).

The Committed Allocation resource share can be up-to maximum pre-defined limit. The Over Commitment Allocation share can be higher than pre-defined limit.

Resource allocation for application: Cloud computing providers charge their customer for running their application on the physical machine. Application can be for example running web server, database server, batch processing and etc., in cloud computing load generated by the application. This load may be generated by application itself or serving client request.

We are testing, the allocated shortage and awaiting values of the application at the virtual machine level. In

Table 1: Sample test of CA and OCA

Tests	Input		Output					
			Commitment Allocation (CA)			Over Commitment Allocation (OCA)		
	Time	Demand	Current (instance)	Shortage (instance)	Awaiting (instance)	Current (instance)	Shortage (instance)	Awaiting (instance)
1	0.0	225	225	0	0	225	0	0
	1.0	150	150	0	0	150	0	0
	2.0	90	90	0	0	90	0	0
	3.0	45	45	0	0	45	0	0
	4.0	250	250	0	0	250	0	0
2	0.0	210	210	0	0	210	0	0
	1.0	420	250	170	170	420	0	0
	2.0	120	250	40	40	120	0	0
	3.0	210	250	0	0	250	0	0
	4.0	520	250	270	270	500	20	20
3	0.0	320	250	70	70	320	0	0
	1.0	400	250	150	220	400	0	0
	2.0	500	250	250	470	500	0	0
	3.0	700	250	450	920	500	200	200
	4.0	250	250	0	500	450	0	0
	5.0	0	250	0	250	0	0	0

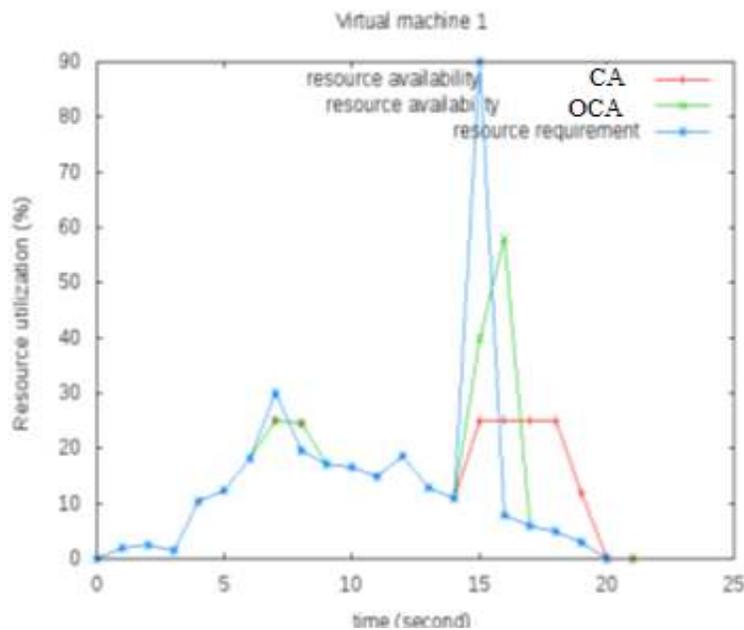


Fig. 7: Utilization of virtual machine 1

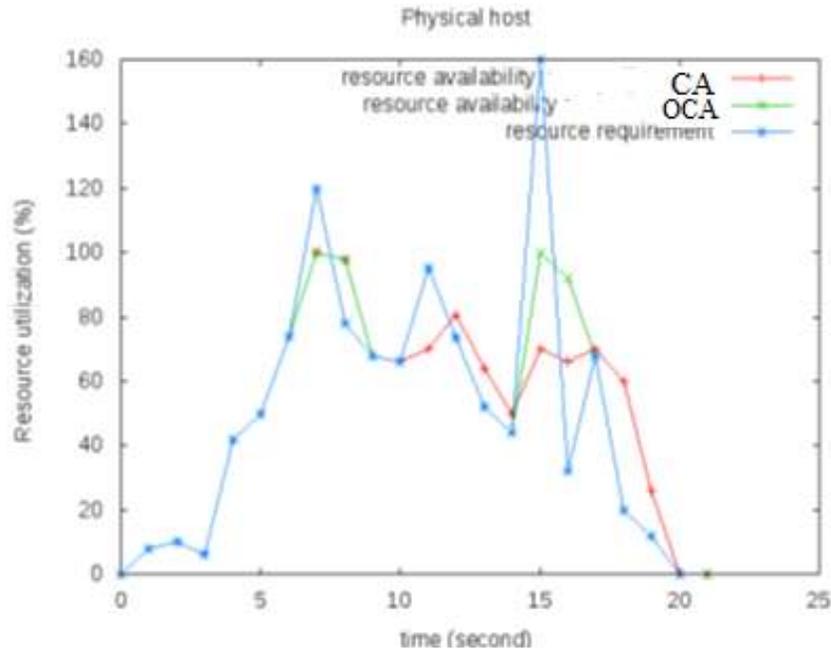


Fig. 8: Utilization of physical machine

committed allocation and allocated resource share can be up-to maximum pre-defined limit. But Over Commitment allocation, allocated resource share can be higher than pre-defined limit. Table 1 shows the sample test of input and output. Need to test the efficient of resource utilization at virtual machine and physical machine level varying the allocation policy.

RESULTS AND DISCUSSION

In efficient of resource utilization at virtual and physical machine level varying the allocation schemes. The common setup is 5 virtual machine in a single physical host. Physical machine configuration 2000 inst/sec (processing), 2048 GB (RAM), 40 MB/sec (disk i/o) 40 MB/sec (Network i/o). The Virtual Machine configuration 500 inst/sec (processing), 512 MB (RAM), 10 MB/sec (disk i/o) and 10 MB/sec (network i/o). Figure 7 and 8 shows the resource demand availability of resource at virtual and physical level with committed allocation and over commitment allocation schemes.

CONCLUSION

In this study describes the cloud computing resource allocation schemes and performance measurements. In Simulation, the tests have been conducted and validated using the resource allocation schemes. The results can be used as reference model for cloud provider and cloud consumer. The series of tests has been conducted for this paper and proved in CloudSim. It has to be evaluated the proposed schemes

in real cloud computing environment such as VMware ESXi Server hypervisor, Xen Hypervisor and Ubuntu Eucalyptus Cloud (UEC). In the future this study to be implemented and tested in the private cloud center.

REFERENCES

- Chaisiri, S., L. Bu-Sung and D. Niyato, 2012. Optimization of resource provisioning cost in cloud computing. *IEEE T. Serv. Comput.*, 5(2): 164-177.
- Haiying, S. and L. Guoxin, 2014. An efficient and trustworthy resource sharing platform for collaborative cloud computing. *IEEE T. Paralle. Distr.*, 25(4): 862-875.
- Linna, H. and T. Lifang, 2012. Model design on sharing information resource of Cangzhou city area based on cloud computing. *Proceeding of the 2nd International Conference on Computer Science and Network Technology (ICCSNT, 2012)*, pp: 1869-1872.
- Linquan, Z., L. Zongpeng and W. Chuan, 2014. Dynamic resource provisioning in cloud computing: A randomized auction approach. *Proceeding of the IEEE INFOCOM*, pp: 433-441.
- Mark, S., S. David, V. Frederic and C. Henri, 2010. Resource allocation algorithms for virtualized service hosting platforms. *J. Parallel Distr. Com.*, 70(9): 962-974.
- Mauro, A., C. Michele and M. Michele, 2010. Dynamic load management of virtual machines in cloud architectures. In: Avresky, D.R. *et al.* (Eds.), *CloudComp 2009. LNICST 34*, Springer-Verlag, Berlin, Heidelberg, 34: 201-214.

- Navendu, J., M. Ishai, N. Joseph (Seffi) and Y. Jonathan, 2014. A truthful mechanism for value-based scheduling in cloud computing. *Theor. Comput. Syst.*, 54(3): 388-406.
- Nedeljko, V., N. Dejan, M. Svetozar, K. Dejan and B. Ricardo, 2012. DeJaVu: Accelerating resource allocation in virtualized environments. *Proceeding of the 17th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS' 12)*, pp: 423-436.
- Papagianni, C., A. Leivadreas, S. Papavassiliou, V. Maglaris, C. Cervello-Pastor and A. Monje, 2013. On the optimal allocation of virtual resources in cloud computing networks. *IEEE T. Comput.*, 62(6): 1060-1071.
- Rajkumar, B., K.G. Saurabh and N.C. Rodrigo, 2011. SLA-oriented resource provisioning for cloud computing: Challenges, architecture and solutions. *Proceeding of the International Conference on Cloud and Service Computing*, pp: 1-10.
- Rodrigo, N.C., R. Rajiv, B. Anton, A.F.D.R. Cesar and B. Rajkumar, 2011. CloudSim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software Pract. Exper.*, 41(1): 23-50.
- Shakeel, A., A. Bashir, M.S. Sheikh and M.K. Rashid, 2012. Trust model: Cloud's provider and cloud's user. *Int. J. Adv. Sci. Technol.*, 44: 69-80.
- Shin-Ichi, K., 2011. Optimal joint multiple resource allocation method for cloud computing environments. *Int. J. Res. Rev. Comput. Sci.*, 2(1).
- Warneke, D. and K. Odej, 2011. Exploiting dynamic resource allocation for efficient parallel data processing in the cloud. *IEEE T. Parall. Distr.*, 22(6): 985-997.
- Wei, W.L., Z.W. James, L. Chen and Q. Deyu, 2011. A threshold-based dynamic resource allocation scheme for cloud computing. *Proc. Eng.*, 23: 695-703.
- Zhen, X., S. Weijia and C. Qi, 2013. Dynamic resource allocation using virtual machines for cloud computing environment. *IEEE T. Parall. Distr.*, 24(6): 1107-1117.
- Zhenping, L., W. Xiangming and S. Yong, 2012. A game theory based resource sharing scheme in cloud computing environment. *Proceeding of the World Congress on Information and Communication Technologies (WICT, 2012). Trivandrum*, pp: 1097-1102.