GROUP BASED RESOURCE MANAGEMENT AND PRICING MODEL IN CLOUD COMPUTING

Shelia Rahman[1], Afroza Sultana[2], Afsana Islam[3], and Md Whaiduzzaman[4]

[1]Institute Of Information Technology,JahangirnagarUniversity,Dhaka,Bangladesh.
[2]Institute Of Information Technology,JahangirnagarUniversity,Dhaka,Bangladesh.
[3]Institute Of Information Technology,JahangirnagarUniversity,Dhaka,Bangladesh.
[4]Institute Of Information Technology,JahangirnagarUniversity,Dhaka,Bangladesh.

ABSTRACT

Cloud computing utilizes large scale computing infrastructure that has been radically changing the IT landscape enabling remote access to computing resources with low service cost, high scalability , availability and accessibility. Serving tasks from multiple users where the tasks are of different characteristics with variation in the requirement of computing power may cause under or over utilization of resources.Therefore maintaining such mega-scale datacenter requires efficient resource management procedure to increase resource utilization. However, while maintaining efficiency in service provisioning it is necessary to ensure the maximization of profit for the cloud providers. Most of the current research works aims at how providers can offer efficient service provisioning to the user and improving system performance. There are comparatively fewer specific works regarding resource management which also deals with the economic section that considers profit maximization for the provider. In this paper we represent a model that deals with both efficient resource utilization and pricing of the resources. The joint resource management model combines the work of user assignment, task scheduling and load balancing on the fact of CPU power endorsement. We propose four algorithms respectively for user assignment, task scheduling, load balancing and pricing that works on group based resources offering reduction in task execution time(56.3%),activated physical machines(41.44%),provisioning cost(23%) . The cost is calculated over a time interval involving the number of served customer at this time and the amount of resources used within this time.

KEYWORDS

Resource Management, Resource Pricing, Task Execution, Load Balancing, Task Scheduling.

1. INTRODUCTION

Cloud computing is a pool of virtual machines with underlying datacenters or physical machines providing various kinds of agile and effective services to the user in a form of virtualization ofevery kind of computing services from infrastructure to software[1] [2].With the development inInternet uses of Internet enabled devices are increasing day by day in fact the number of IoTenabled devices have already outnumbered total human population[3]. This results in more generation of large scale data which requires faster processing and faster task response.Thereforemore devices are now getting connected to the cloud as these devices are limited to storage andprocessing power.However with the increase in users cloud service providers are now more complied to use larger and powerful datacenters as investments were done by many telecommunication companies in order to satisfy their growing customer requirements and avoiding any SLAviolation[4].These datacenters are equipped with powerful hardware and connected with highbandwidth networksand managed with software resources. . As a result the requirement of efficient management procedures for handling such large datacenters are becoming a topic of interest to the researchers.Current research works mostly aims at better service provisioning to the userand improving system performance with only some specific works

DOI: 10.5121/ijcsit.2018.10403

regarding economic and costing model.These models deal with the resource management in cloud computing which assuresefficient resource management that ensures maximization of profit for the provider.As Both userand providers play important role in shaping of cloud computing with clear and distinct promisingfeatures , resource management procedure that holds the end of the bargain for both users andproviders are important.

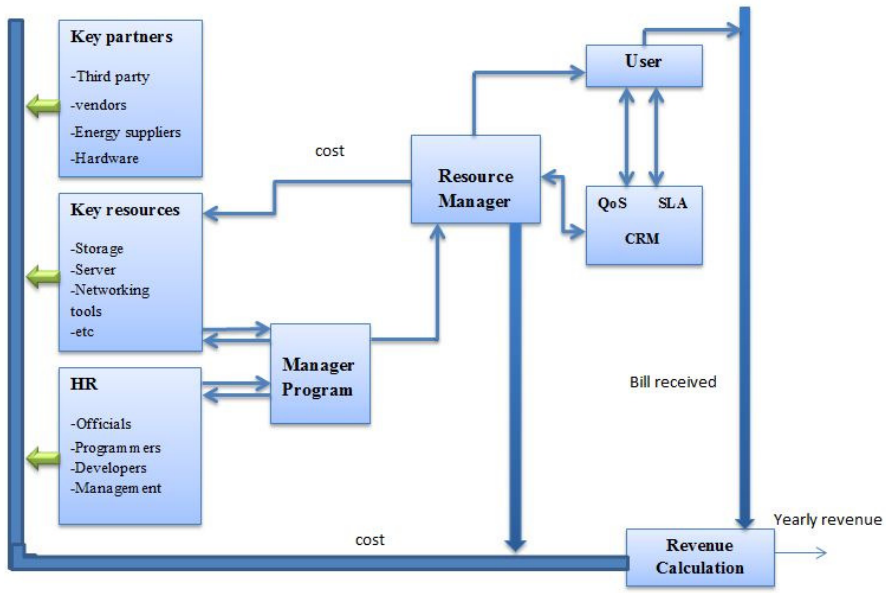


Figure 1: Cost based price management in cloud computing

Resource management in cloud covers a number of issues such as efficient task scheduling, resource provisioning, resource allocation,resource adaption,tasks assignment,load balancing and others.Amongthem three important features of resource management can be considered resource scheduling,load balancing and task assignment to shape the resource utilization in cloud environment. Resource scheduling refers to a timetable of events maintaining events and resources to determinethe time an activity is assigned to execution.Load balancing aims to evenly distributing tasks overphysical machines to reduce traffic in any particular ones and to maximize the number of machines in stand by mode to reduce power usage. Tasks assignment aims to mapping tasks to aVM and generally chooses the VM where it can execute faster.The provisioning and placement ofVMs must be done efficiently taking into account the available resources in cloud .Again,the reconfigurations must be performed fast to resize or release the existing virtual resources because ofthe variability and elasticity of resource demand[5]. Inefficient resource management negativelyaffects performance and cost as well as impairing system functionality.

Ensuring Profit for every cloud provider is a fundamental goal.Traditional procedures such as system optimization tends to aim at different system performance metrics based on system parameters and constraints rather than economic factors(the profit,cost,and revenue)[5].A cost based modelcalculates the price of a service using allover cost for providing the service as well as adding apercentage of the cost as desired profit[5][6].This model uses 1.fixed cost and 2.variable cost tocalculate the total cost[5].Fixed cost includes cost required for setting up hardware, servers andnetwork devices, required human resources. On the other hand variable cost is not constant , itvaries according to the number of sales produced by the service such as energy, bandwidth,costof transferring data between different datacenters and so on.Providers use cost based pricing forlarge scale datacenters to calculate the cost for a service for the incoming users tasks requests[7].

Resources in a cloud server includes both hardwares and software. Cloud computing infrastructure is generally hosted on datacenters and these datacenters can be geographically distributed and connected using high speed network.A hypervisor software emulates server's CPU,memory ,networkand other resources to enable isolation of multiple users when running on the same physical machine.Tasks generated from user devices are assigned to virtual machines with cloud broker. Taskshandled by a server are of hybrid in nature where some tasks require more computation power thanothers and again they can differ in priorities.IaaS-Clouds offer VMs at a specific monetary cost.Excess consumption of virtual resources need to be avoided as underutilized worker-nodes reducethe monetary value the cloud client achieves.The number of serving VMs must never surpass thethreshold over which excess VMs become an overhead for power consumption.Therefore assigning tasks to VMs to ensure both faster execution and better utilization of resources at the sametime is a tricky issue as during service provisioning the provider have to maintain both Quality ofService(QoS) and gaining the maximum revenue.
In this paper we have identified efficient resource management with profit maximization raises three important issues-1.What types of tasks are needed to assign 2.How datacenters should placethe tasks for execution on VMs, 3.How to reduce the number of active servers at a datacenteralong with serving maximum numbers of users,4.How the users are priced so that providers canhave maximized profit. In this research we answer the questions by modeling a resource management system and use cost based pricing model to calculate the price of resources used.In ourproposed model we logically divide the computing resources in a datacenter into groups called ResourceGroup based on what type of tasks will be assigned to them .A number of VMs are assignedto each group where VMs within same group are assigned to execute tasks of requiring similaramount of computing power. Tasks assignment and scheduling works in a sequential manner.Fourdifferent algorithms are presented which works to achieve a overall profit for the provider andfaster task execution for the user.

In the rest part of this article we have described some scenarios of resource management andpricing in cloud in Section 2. After that we have represented our proposed system architecturein Section 3 and algorithms for task assignment, task scheduling and load balancing in Section 4.Later we have evaluated our proposed method with some other existing ones presented in section5.

2. RELATED WORKS

There are several works regarding resource management in cloud computing. Different costing models have been proposed for pricing the services. However there are very few research which collaborates cost management with resource management in cloud computing. The management service observes mechanisms in the virtual servers within the cloud computing sessions which allows it tomonitor,analyse as well as provide rep orts along with alerting pertaining to performance metricsof the various virtual servers[8].In [9] authors proposed an energy efficient adaptive resource management for vehicular cloud to maximize the overall communication and and computing energyefficiency.It meets the application-induced hard QoS requirements with least transmission rates butwith maximum delays and delay-jitters.In [10] author gave their concern on network visualizationof resource allocation dynamically in cloud computing and the importance of meeting QoS. Whenthe best effort falls then the Service Level Agreement (SLA) takes the equivalent actions.Besidesdifferent scheduling and load balancing algorithm has been proposed [11][12][13][14].In [15] researchers proposed to schedule the tasks based on requirement of computation power and [14]represented a load balancing mechanism also based on power requirement.Authors in [16] proposed a resource utilization method based on greedy method.According to [17][18][19] when using mobiledevices augmented with cloudlets for capacity increment the offered service to load ratio is neededto be considered. . These research works considers only limited constraints rather than a complete scenario of resource management.

Besides how providing resource efficiency affects providersrevenue is not discussed which affects in long term provisioning.

Various model for pricing cloud service have been proposed for cloud market scenario e.g costbased pricing [7][20][21] depicts how the cost of the resources used can be calculated,profit maximization [22][23] aiming for increasing profit of the provider, differential pricing [24][15] chargesdifferent price to different client based on their demand. According to [2] charges for the servicerequest are decided by the costing mechanism depending on the time of submission, rate of pricingor the availability of resource and accounting mechanism is used for calculation of the actual usageof resources.A detailed overview of cloud resource management using economic analysis is givenin [5].Techno-economic modelling is used in [20] to assess the cost efficiency of using SDN intothe LTE network.According to [21] Particularly for most demand side management applicationsbecause of the variability of customer behaviours2, the computing requirements fluctuates significantly which discourages providers for setting up new datacenters.The seemingly never-endingresources of a physical cloud, exacting the specific amount of resources required must be based onboth the consumer's intended amount of expenditure and the performance bottlenecks which can bevisible only at runtime[22].They propose to continuously monitoring user application performanceand removing or adding VMs when performance fluctuations is observed in serving the needs ofautonomic systems. To maximize the profit, a cloud provider needs to understand both servicecharges and business costs as well as the way they are by the characteristics of the applicationsand the configuration of a multi-server system[23].Pricing also differs based on the geo-distributionof datacenters.A dynamic pricing along with profit maximization aiming pricing in geographicallydistributed datacenters in cloud is proposed in [25].

3.RESOURCE GROUP MODELING

In this research work we propose that the resources are divided into groups. According to what services the provider is offering the groups are created based on the requirement of both the provider and the user.For example lets think a cloud provider has N geographically distributed datacenters.Based on the job type tasks are divided in three groups where they differs in their requirement of computation power.Based on this the datacenter resources can be divided into three groups. Each groups contains a number of VMs and the groups are further divided into subgroups. Our proposed model partitions all the server resources based on the CPU power required which implies the number of CPU cycles and the amount of internal memory the groups are assigned. The grouping is done logically which implies that they don't necessarily have to be physically separated.

3.1 Resource Group

A ResourceGroup is like a logical container that have certain amount of computing resource assigned to them. Provider can deploy, manage and monitor all the resources for provisioning solution as a group, rather than handling these resources individually. The groups can be divided into subgroups when required. A sequential approach is followed in deploying and balancing theresources.Large group of resources are again divided into subgroups. Groups are different withone another on the basis of amount of resource consumption.Lets the provider's resources in adatacenter is divided into two groups A and B.Group A is for servicing users demanding a largeamount of power and the other one is for the less demanding. The first group offers instances thatis constituted of more numbers of CPU and more Powerful (Speed, MIPS) CPU and RAM powerthen the second group. As the larger groups are split into smaller subgroups control becomes moredistributed. The assignment,scheduling, load balancing algorithms works differently on this subgroups and synchronizes when requires. The subgroups contains a certain number of VMs. EachVM is assigned to user tasks according to their demand. An array is used to contain identificationand status of this VMs at a certain time. Within this array the VMs offers a sequential relation andrelativity among VMs that belongs to the same group. The resources in each subgroup is offeredas instances. For simplifying our description each instance will be addressed as a PM. Resourcesare sharable among the groups. But the sharing should be minimized. A datacenter has ResourceGroup identified as A, B, C. . . .N where N=total number of General Group and threshold ofth1,th2,th3....th1 .Then incoming tasks can be grouped based on the threshold such as tasks requiring power below th1 are assigned to group A ,tasks between th1 and th2 are submitted to group Band goes on.Tasks in the same group need similar amount of computation power but largely differswith other groups but pretty similar within the same group. Subgroups are identified as followsand tasks in the subgroup are similar in requirement of power.A particular number of VMs areassigned to a subgroups , during task moving for load balancing or remapping VMs within samegroups are prioritized. Subgroup within each group= 0, 1, 2. . . . . . n

Service point (VM) in group:

Group A – subgroup 0 VM=0, 1, 2. . . . . . n

Group A – subgroup 1 VM=n+1, n+2. . . . . . n+m

. . . .

. . . .

Group A – subgroup x VM n+j, n+j+1. . . . . . n+j+m

Similarly,

Group B – subgroup 0 VM=0, 1, 2. . . . . . n

Group B – subgroup 1 VM=n+1, n+2. . . . . . n+m

. . . .

. . . .

Group B – subgroup x VM n+j, n+j+1. . . . . . n+j+m

and so on.

Here,n+j =value ID 1 increments than last one in the previous subgroup

A flag indicates if a subgroup is full or not

3.2 Group Based Resource Management

The resource management procedure has different portion that works together for a better cost efficient management of the servers. This system uses Energy efficient management of IaaS[15] cloud which is a integrated approach for VM migration and reconfiguration, and PM power management and Task Scheduling algorithm in Cloud Computing Environment Based on Cloud Pricing models and [28] ,as a base for our algorithms to offer a better load balancing and resource utilization based on resource intensity aware load balancing . The different portion is responsible for different managing work in resource management.

- **Access Log:** The access log is responsible for ensuring the authenticity of the user request.
- **Assignment Manager:** The assignment manager is responsible for assigning user to a freeservice node which in this case will be a VM.It ensures easy identification and servicing ofthe user at that VM as this ID's maintains a sequence accordance with their group.
- **Scheduler:**The scheduler is responsible for assigning a PM which represents an instanceof the resources that is required by the user task. The PM assignment to a VM depends onwhich group they are assigned in. It programs a VM migration to other group when there isnot enough resource in the group.
- **Load Statistics:** This portion check for load in each VM and eventually in each group. This counts the load as percentage of the total resources assigned and the amount occupied. This uses a threshold power to determine if this group can have any new users or new PM could be assigned to avoid any collusion.

- **Predictor:** The predictor gives an estimation on if there is any possibility of conflict and hunger among users can happen.
- **Pricing and revenue:** Calculates the price for the user .Also shows price that the user will find if they use some new resources. With the provider's input in each groups cost the revenue portion calculates the monthly revenue from a particular group which is calculated as per subgroup.

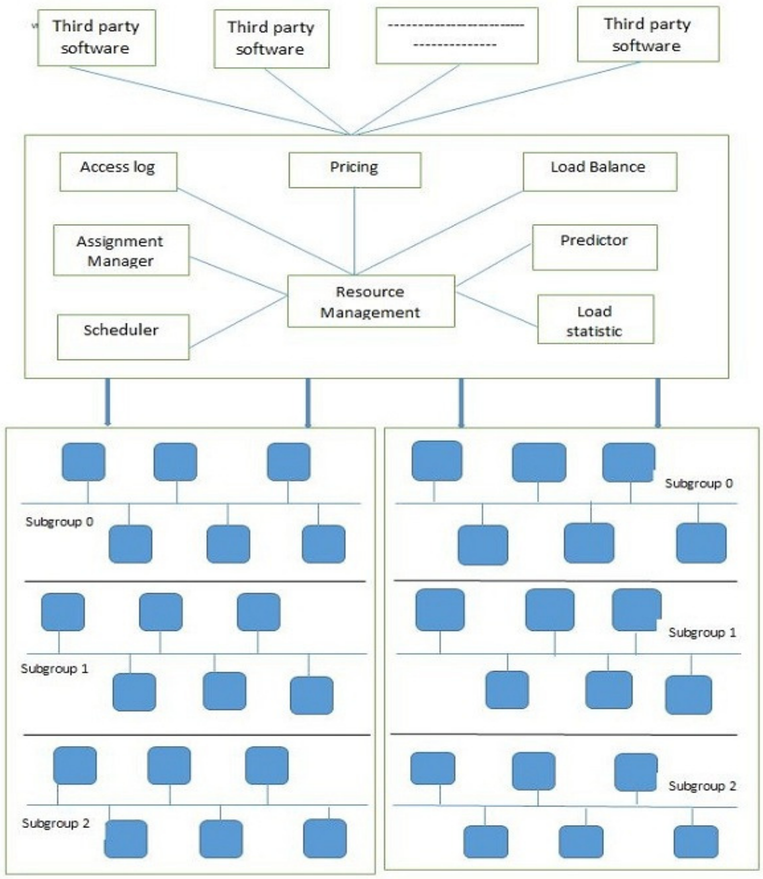


Figure 2: Group based resource management

3.3.Resource Grouping and Profit maximization

The cloud service provider's actual profit refers to the difference between the revenue and thecosts[23].To maximize service providers' profits and ensure load balancing within them, eachcloud provider determines the in-sourcing price as well as the available quota of resources basedon which group the serving resources are resided.The pricing policy proposed in this paper usesthe insourcing price which is set according to the VM costs for a certain type of tasks.Differentgroup has different price calculated within a timing interval.Groups with lower power has lowerprice and higher power requiring groups has price set based on the type of resources assignedto them.According to cost based pricing model services are priced depending on the resourcesconsumed.Therefore each user task in this model is priced based on the group they are assignedto.As a result different tasks can get easily priced based on their power requirement rather thanhaving a common price calculated over time.The proposed constraints ensures that only limitednumber of PMs are activated and at the same time tasks are having low average waiting time inthe queue.Grouping and subgrouping of resources ensures evenly distribution of resources amongthe executing tasks hence reducing under or over consumption of computing power.This reducesthe overall provisioning cost and increasing the gross profit for the service provider.

4. GROUP BASED RESOURCE MANAGEMENT AND PRICING

This section describes how Resource group model is applied for different aspects of resource management which includes task assignment, task scheduling, load balancing along with pricing of provisioning resources.Different algorithms have been designed for each section along with necessary constraints.

4.1. Task Assignment

This section depicts how tasks are assigned to different resource group. The main idea of the enhanced algorithm is allocating the particular VM to the requested tasks according to the following steps:- At first the required computation power of each VM is calculated based on the total number of MIPS allocated to it. Calculate the processing power of each VM_i (i.e.,VM_i's Total MIPS) using equation (1)

$$Total\_MIPS[i] = CPUs * MIPS \qquad (1)$$

CPUs= the number of cores, MIPS=million instruction per second of the single core of the VM. Here i=0, 1, 2. . . . . . , represents the position of the subgroup. Then calculate the requiredpower of each task of all the received tasks from the user using equation (2).If the waiting taskshas m no of 'p' type of tasks then

$$Task_p^m = \frac{\sum_{ts=1}^{m} MIPS_n}{m} \qquad (2)$$

Different threshold values are determined for different groups. Tasks are assigned to the group which matches with its power requirement.Such as for group A if a task of type p requires MIPSsuch that $MIPS_p(i)<Threshold_A$, it's assigned to group A. Therefore all m tasks of type p will beassigned to Group A after inspecting which subgroups to choose.

An array is used to offer theserver nodes in a group. Then using equation (2) it is determined if there is enough resources evenif there is a free node.A Time threshold T_{lowcon}=Low power consumption for time t refers if anode is consuming a very low server power(low-con) for a certain time T then the VM should betransferred to low power consumption group to avoid wastage and vice versa for time T_{high_con}=High Power Consumption for time t

According to the calculations the following algorithm is designed for assignment of the tasks waiting in the queue for execution with considering the power requirement .

Algorithm 1: Task Assignment algorithm

1. **procedure** Assignment(G; ts) //G groups,ts=tasks to be assigned
2. Array[Service_node(ni)] //user assigned identification no
3. Array[a][b]
4. a = user_id
5. Calculate MIPS of all p type of tasks
6. **for** j = 0 to total_numberof_groups**do** //determining the required group
7. Compare Power of $Task_p$ with group threshold
8. **for all** subgroups in group G **do**
9. **if** (**then**load_in_subgroup>T_{safe_load})
10. choose group G
11. **else**Wait in the queue

end if
end for
end for
Check for the next request
end procedure

4.2. Task Scheduling

After selecting a group the tasks are needed to assign to a VM contained within a subgroup.Powerfactor of each VM of a subgroup x is calculated as follows-

$$PF\ of\ VM, VM[i]_{PF} = \frac{Total_{MIPS[i]} - inuse\_MIPSi}{Total\_MIPS[i]} \quad (3)$$

Power factor of each subgroup is calculated using equation (4) which defines the credibility of each subgroup

$$PFofsubgroups\ , subgrp[n]_{PF} = \sum_{q=s}^{e} \frac{Total_{MIPS[q]} - inuse\_MIPSq}{Total\_MIPS[q]} \quad (4)$$

Here,n=no of subgroup,s=first VM in the nth subgroup and e=last VM in nth subgroup. If PFof a subgroup is greater than x the subgroup is not eligible for new task scheduling. Rather thansearching each VM separately for eligibility this mechanism allows a group of VMs to be identifiedat a time which reduces the time cost for searching suitable VM.Tasks are sorted according to theirpriority in the waiting queue.Required power is calculated for the task with equation (2) and it'spower factor is calculated using equation(5).

$$\mathrm{Task}[\mathrm{ts}]_{\mathrm{PF}} = \frac{\mathrm{Task}_{\mathrm{p}}^{\mathrm{m}}}{\mathrm{VM[i]}} \quad (5)$$

Search the requested subgroup sequentially to find a VM that provides processing power equal to or less than the power factor of the task by considering that the difference between the selectedtasks processing power and VM to be minimum.

Algorithm 2: Task Scheduling Algorithm

---

procedure Scheduler(n,ts)
for i = 0 to m **do**
Define MIPS of the VMs
Calculate the power factor of each VM
end for
for i = 0 to n**do**
Define MIPS of the subgroups
Calculate power factor of each subgroup
end for
for i = 0 to m **do**
Calculate the PF of each task
Choose the VM with which task have the highest power factor VM_i
allotment=$Task_{PF}$*requested MIPS
end for
end procedure

4.3. Load Balancing

Overloaded PM transfers VMs running on them which have high consumption on high-power resources and limited consumption on low power resources.Therefore it relieves its load quickly at the same time completely utilizing datacenter resources.Selected PM generally has high competence on the high intensity resources for actively tend to ignore overloading destination PMs in the future.In our work task allocation and VM allocation works based on communication rate. Communication rate refers to the number of contacts between the source VM and destination VM in a unit time period.Communication rate T_{xi}(VM_i of subgroup x) for a VM_{ni} with its local VM_{kh}(VM_h of anysubgroup k) is denoted by,

$$T_{xi} \sum_{i=1}^{n} T_{xikh} \tag{6}$$

k=x if the two communicating VM are in the same group.Task re-allocation takes place within the same subgroup whereas VM re-allocation occurs among different groups.When VMs and tasksare reallocated the performance degrades[27]. We aim to minimize the degradation.To reduce thedegradation the number of switches in the communication path need to be lessened.Reallocationdepends on the results of equation (6) and (7) as follows Load of VM_i comparing with others inthe same subgroup

$$L_{VM_i} = \frac{VM[i]_{PF}}{subgrp[n]_{PF}} \tag{7}$$

If L_{VMi}<$threshold_{load}$(minimum load: below this threshold is considered as low power consuming) then,tasks of VM_i is transmitted to VM_j where

$$L_{VM_j} = \min \left(\frac{VM[j]_{PF}}{VM[i]_{PF}}\right) \tag{8}$$

ifVM_j has enough resources and has low communication rate with the transmitting VM otherwiseit switches to next VM.Therefore the now free VM and its related PMs turned to stand by mode.Similarly if a subgroup is running low consuming VMs they are transferred to the subgroup withminimum communication cost.In our proposed method the number of switches is constrained by thenumber of VMs in a subgroup for tasks reallocation and number of subgroups for VM allocation. Ifone node is consuming more power it’s task is distributed. To distribute the task load free resourcesare at first searched within the group it belongs to. It ensures resource utilization and keeps morePM at standby mode. If a node A is consuming too much CPU power and the consecutivesubgroup is already crossing its threshold then for further assignment the VM is transferred to thenext subgroup that can take it.

Algorithm 3: Load balancing Algorithm

```
procedureload_balancing
    for i = 0 to ndo
        Check load in each subgroup
        if ( thenLoadinSubgroup[n]>T_accepted)
            whileLoadinSubgroup[n]>T_accepted do
                Select the maximum power consumerist,k
                for  each node in the subgroup do
                    Calculate the    communication rate with k
                    Choose the VM with with highest communication rate and low load
                    Shift task to     the chosen VM
                end for
            end while
        end if
        ifLoadinSubgroup[n]<T_accepted_low then
```

```
            Choose the nearest subgroup s in the array
            if s subgroup can support the load then
             Shift tasks to s
                n =standbymode
            end if
        elseChoose the next nearest subgroup
        endif
    end for
end procedure
```

4.4. Pricing and revenue

Pricing and revenue is calculated as per subgroup basis. Revenue is calculated as how much maintenance cost is implicated and how much gain is received.Pricing is done according to the usage of a particular group resources. VMs from high computational power costs higher price.Each group declares their price . Price for each task is calculated based on which groups the tasks will be assigned to according to their power requirement.Calculated price are then displayed to the customer. This optimization problem is for the cloud provider to choose an appropriate price for each type of jobs at each group ($p_n^p(t)$), the best number of servers to provision each type of VMs in each datacenter ($N_h^d(t)$), the optimal numbers of jobs of each type to schedule and to drop ($u_d^r(t)$) and $G_h^d(t)$,in each t at each datacenter, to maximize its time-averaged profit. The pricing at time t can be calculated as a optimized solution

$$P_n^p(t) = \sum u_d^r(t) * P_G(n) - G_h^d(t) \tag{9}$$

The constraint (udr(t))<i∈n ensures that the activated PMs for the tasks are within the samesubgroup otherwise communication rate will decrease.Revenue for each datacenter is calculated asa factor of serving tasks at tim;8e t and the number of activated server.Using the cost based pricingmodel at time t the cost for serving a p type job is

$$C_p(t) = \sum u_d^r(t) * C_G(n) \tag{10}$$

The revenue for service provisioning at the end of time t(rev) is

$$R(t_{rev}) = \sum_{t=t_0}^{t=t_{rev}} \sum_{job=p}^{job=q} P_p^n(t) - C_p(t) \tag{11}$$

Number of activated PMs are constrained to a certain number which prevents unnecessary scheduling of tasks to new VMs.It enables provisioning most number of users with less number of activated server.

Algorithm 4: Pricing and revenue

```
Procedurepricing_revenue:
for each resource group G do
Determine the unit price for group G
Publish the price
end for
for each task tsdo
Determine the Group
Choose the unit price
end for
   for each resource group G do
  Calculate the number of activated subgroups
```

```
for each activated subgroup in G do
from time t0 to te
Calculate the price of the tasks serving in t0- te using (9)
Calculate the serving cost using unit price CG(n)
Calculate revenue with(11)
end for
end for
end procedure
```

---

5. Evaluation and Result

To evaluate the efficiency of our proposed algorithm we develop a simulator strictly following the system model defined in this paper.The algorithms proposed here are programmed with Javaprogramming language using Eclipse 4.5.0 Mars IDE. Two Different programs are written in Javawhere one is with resources divided in groups and one with resources without dividing in groups.The first group uses an Enumeration that contains 6 groups of variables with different number ofCPU, MIPS, Memoryas signed to it.We have compared our proposed model with different existing algorithms including Enhanced Power based Scheduling (EPBS) and Power intensity LoadBalancing (PBLB) and others.The Amount of CPU power, runtime Cost and re-allocation cost iscalculated. And With these values cost is calculated assigning random cost at values and comparedwith existingalgorithms. The existing algorithm shows runtime complexity ((t − t0) nlog n) where t=starting time and t0=Finishing time and n=no of service node. In our proposed algorithm n reduces to n/kwhere k=number of groups. So the proposed algorithm's runtime complexity reduces to ((t − t0)(n/k) log(n/k) + k).

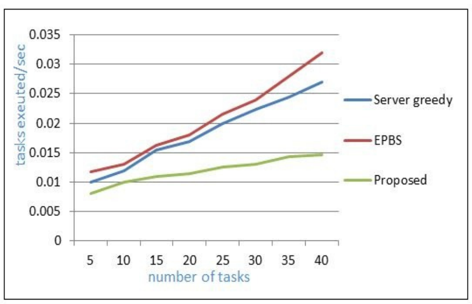


Figure 3. tasks vs task execution rate

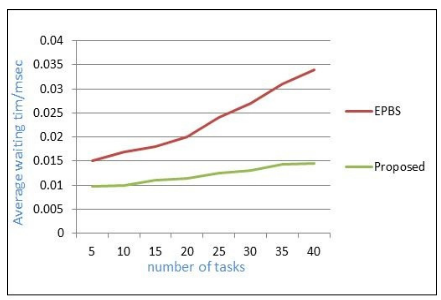
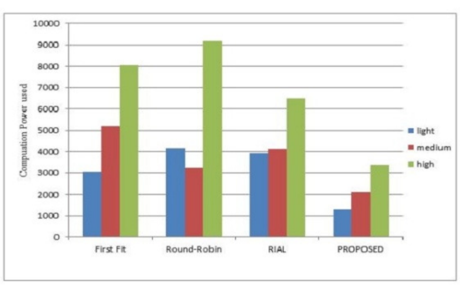


Figure 4. tasks vs average waiting time in the queue



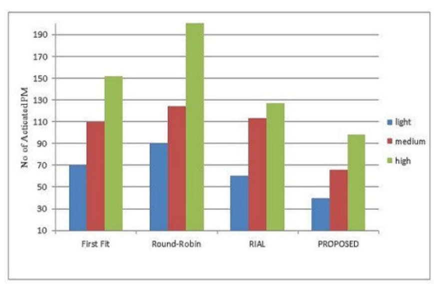


Figure 5. Tasks load vs power used

Figure 6. Load vs activated PM number

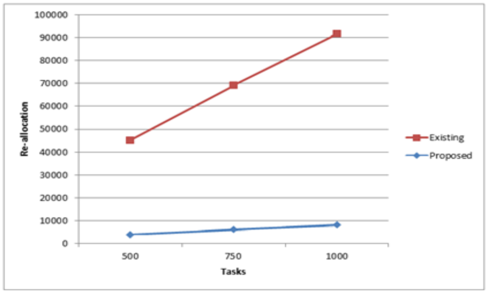


Figure 7. task vs task reallocation cost

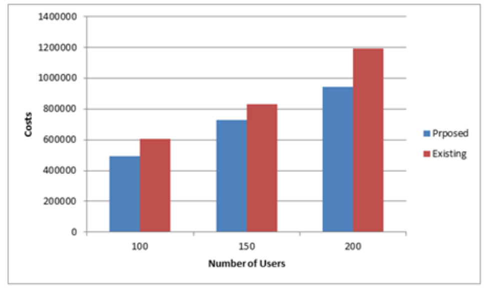


Figure 8. user vs provisioning cost

The first figure compares the task execution rate of our proposed system against the server greedyand EPBS and has a gain of 56.3% over server greedy and of around 67% over EPBS. In figure(4)we see the average waiting time for a task to be served tend to havea steady nature for our proposedmodel comparing to EPBS with a gain of 38% lower average waiting time than EPBS.Figure(5)shows the power consumed by the server with varying load for round robin, first-fit , RIAL andour proposed system. With the increase of load power consumption increases at a high rate.Inthe proposed work the consumption of power is restricted by the structure of the subgroup so ithas lower power consumption rate comparing to the others(nearly 20% against first-fit,40%aginstround-robin,34% against RIAL)with the increase in load.It also shows around 41.44% of less activated number of PMs when the load increases(figure-6). In the last two figures(7)and (8) the cost fortask reallocation and provisioning represents the proposed systems efficiency over existing models.Comparing with intensity based load management we find our proposed work has around 68%less reallocation cost.Again it can serve the same number of users for the same tasks with a reduction of 23% in the provisioning cost.This in turn results in the profit maximization for the serviceprovider.

5.CONCLUSIONS

In this paper we propose a group based resource management and pricing model in cloud computing that logically divides the computing resources into groups, based on computation power requirement of heterogeneous tasks from various user.Dividing large datacenter into smaller groups enables better a synchronization among them by reducing the task execution time, waiting time in the queue, power consumption and overall provisioning cost which in turn increases the revenue of a cloud provider. Assigning tasks into power-consumption based groups makes sure a high power consumerist is never assigned to a low instance which may cause the excess abuse of power and similarly assignment of low power requiring tasks to high power servers never causes under-utilization of resources. In this grouping system tasks of all similar power-consumerist belongs to similar groups resulting in lower SLA violation. Our proposed model put constraints on PM assignment based on subgroups to keep minimum number of PM are activated and at the same time maximum numb e r of user assignment. Simulation results shows that our proposed algorithms outperforms existing system in various aspects such as execution time, average task waiting time in the queue,used power,number of activated PM,task reallocation cost and overall provisioning cost to a notable extent. We have achieved reduction in task executiontime(56.3%),activated physical machines(41.44%),provisioning cost(23%) when comparing our system with existing models.

REFERENCES

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emergingit platforms: Vision, hype, and reality for delivering computing as the 5th utility, FutureGeneration computer systems 25 (2009) 599–616..

[1] M. Whaiduzzaman, M. N. Haque, M. RejaulKarimChowdhury, A. Gani, A study on strategicprovisioning of cloud computing services, The Scientific World Journal 2014 (2014)

[2] M .Alba, IoT Devices to Outnumber Humans in 2017, https://www.engineering.com/IOT/ArticleID/15594/IoT-Devices-to-Outnumber-Humans-in-2017.aspx,2017. [Online; ac-cessed 8-May-2018].

[3] S. Akter, M. Whaiduzzaman, Dynamic service level agreement verification in cloud computing, IJCSIS (2017).

[4] N. C. Luong, P. Wang, D. Niyato, Y. Wen, Z. Han, Resource management in cloud networking using economic analysis and pricing models: A survey, IEEE Communications Surveys & Tutorials 19 (2017) 954–1001.

[5] R. weber, Cost Based Pricing,https://onlinelibrary.wiley.com/doi/abs/10.1002/ 0470867175.ch7, 2003. [Online; accessed 9-May-2018].

[6] K. H. Prasad, T. A. Faruquie, L. V. Subramaniam, M. Mohania, G. Venkatachaliah, Resource allocation and sla determination for large data processing services over cloud, in: Services Computing (SCC), 2010 IEEE International Conference on, IEEE, pp. 522–529.

[7] D. Di Spaltro, A. Polvi, L. Welliver, Methods and systems for cloud computing management, 2016. US Patent 9,501,329.

[8] M. Shojafar, N. Cordeschi, E. Baccarelli, Energy-efficient adaptive resource management for real-time vehicular cloud services, IEEE Transactions on Cloud computing (2016)

[9] E. Oppong, S. Khaddaj, H. E. Elasriss, Cloud computing: resource management and serviceallocation, in: Distributed Computing and Applications to Business, Engineering & Science (DCABES), 2013 12th International Symposium on, IEEE, pp. 142–145.

[10] D. Ajmire, M. Atique, Grouping based load balancing in cloud computing, International Journal of Innovative Research and Development 5 (2016).

[11] S. Abrishami, M. Naghibzadeh, D. H. Epema, Deadline-constrained workflow scheduling algorithms for infrastructure as a service clouds, Future Generation Computer Systems 29 (2013) 158–169.

[12] Z. Tang, L. Qi, Z. Cheng, K. Li, S. U. Khan, K. Li, An energy-efficient task scheduling algorithm in dvfs-enabled cloud environment, Journal of Grid Computing 14 (2016) 55–74.

[13] J. M. Galloway, K. L. Smith, S. S. Vrbsky, Power aware load balancing for cloud computing, in: Proceedings of the World Congress on Engineering and Computer Science, volume 1, pp. 19–21.

[14] E. Ibrahim, N. A. El-Bahnasawy, F. A. Omara, Task scheduling algorithm in cloud com-puting environment based on cloud pricing models, in: Computer Applications & Research (WSCAR), 2016 World Symposium on, IEEE, pp. 65–71.

[15] H. K. Ala'a Al-Shaikh, A. Sharieh, A. Sleit, Resource utilization in cloud computing as an optimization problem, Resource 7 (2016).

[16] M. Whaiduzzaman, A. Naveed, A. Gani, Mobicore: Mobile device based cloudlet resource enhancement for optimal task response, IEEE Transactions on Services Computing (2016).

[17] Whaiduzzaman, Md. "Performance enhancement framework for cloudlet in mobile cloud computing." PhD diss., FakultiSainsKomputerdanTeknologiMaklumat, Universiti Malaya, 2016.

[18] Whaiduzzaman, Md, Abdullah Gani, and AnjumNaveed. "Pefc: performance enhancement framework for cloudlet in mobile cloud computing." In Robotics and Manufacturing Automation (ROMA), 2014 IEEE International Symposium on, pp. 224-229. IEEE, 2014.

[19] N. Zhang, H. Hämmäinen, Cost efficiency of sdn in lte-based mobile networks: Case finland,in: Networked Systems (NetSys), 2015 International Conference and Workshops on, IEEE, pp. 1–5

[20] Z. Cao, J. Lin, C. Wan, Y. Song, Y. Zhang, X. Wang, Optimal cloud computing resource allocation for demand side management in smart grid, IEEE Transactions on Smart Grid 8 (2017) 1943–1955.

[21] K. Tsakalozos, H. Kllapi, E. Sitaridi, M. Roussopoulos, D. Paparas, A. Delis, Flexible use of cloud resources through profit maximization and price discrimination, in: Data Engineering (ICDE), 2011 IEEE 27th International Conference on, IEEE, pp. 75–86.

[22] J. Cao, K. Hwang, K. Li, A. Y. Zomaya, Optimal multiserver configuration for profit maxi-mization in cloud computing, ieee transactions on parallel and distributed systems 24 (2013) 1087–1096.

[23] D. M. Divakaran, M. Gurusamy, M. Sellamuthu, Bandwidth allocation with differential pricing for flexible demands in data center networks, Computer Networks 73 (2014) 84–97.

[24] J. Zhao, H. Li, C. Wu, Z. Li, Z. Zhang, F. C. Lau, Dynamic pricing and profit maximization for the cloud with geo-distributed data centers, in: INFOCOM, 2014 Proceedings IEEE, IEEE, pp. 118–126.

[25] H. Shen, Rial: Resource intensity aware load balancing in clouds, IEEE Transactions on Cloud Computing (2017).

[26] W. Voorsluys, J. Broberg, S. Venugopal, R. Buyya, Cost of virtual machine live migration in clouds: A performance evaluation, in: IEEE International Conference on Cloud Computing, Springer, pp. 254–265.

[27] Whaiduzzaman, Md. "Performance enhancement framework for cloudlet in mobile cloud computing." PhD diss., FakultiSainsKomputerdanTeknologiMaklumat, Universiti Malaya, 2016.