RPOA Model-Based Optimal Resource Provisioning

Noha El.Attar

Ras-Elbar High Institute for Specific Studies and Computer,
Ras El- Bar, Damietta, Egypt.

Wael Awad

Department of Mathematics and Computer Science
Faculty of Science, Port Said University
PortSaid, Egypt

Samy Abd El -Hafeez

Department of Mathematics and Computer Science
Faculty of Science, Port Said University
PortSaid, Egypt

Fatma Omara

Departement of Computer Science,
Faculty of Computers and Information, Cairo University,
Cairo, Egypt

Abstract—**Optimal utilization of resources is the core of the provisioning process in the cloud computing. Sometimes the local resources of a data center are not adequate to satisfy the users' requirements. So, the providers need to create several data centers at different geographical area around the world and spread the users' applications on these resources to satisfy both service providers and customers QoS requirements. By considering the expansion of the resources and applications, the transmission cost and time have to be concerned as significant factors in the allocation process.**

According to the work of our previous paper, a Resource Provision Optimal Algorithm (RPOA) based on Particle Swarm Optimization (PSO) has been introduced to find the near optimal resource utilization with considering the customer budget and suitable for deadline time. This paper is considered an enhancement to RPOA algorithm to find the near optimal resource utilization with considering the data transfer time and cost, in addition to the customer budget and deadline time, in the performance measurement.

Keywords—Cloud Computing; Resource Provision; Data Communication; Particle Swarm Optimization

I. INTRODUCTION

One of the main definitions of cloud is that it is a cluster of distributed computers providing on-demand computational resources or services to the remote users over a network [1]. In other words, the cloud computing is a new paradigm for hosting and delivering services on demand over the internet where the users could access services depending on their Quality of services (QoS) requirements regardless where these services are hosted, they only care about how much they will pay, and how much time is expected to provide the required hardware and software resources [2] [3].

Cloud computing has the growing popularity and adoption feature [4]. This means that there is no data center with limited capacity. In other words, if a data center is become overloaded, it may be possible to relocate some workloads to another data center [5]. The workload sharing makes the cloud system has the ability to expand the resource pool and provides more flexible and cheaper resources [1].

Cloud computing delivers three application layers as services that are Infrastructure as a service (IaaS) for delivering IT infrastructure, Platform as a service (PaaS) for software development and deployment, and software as a service (SaaS) for providing software product [6]. depending on the, frameworks, or a final . IaaS rule is to provide Hardware as a Service by offering basic storage, memory and computing capabilities as standardized services over the network. To acquire computing resources, a user launches a server instance on the infrastructure of the cloud provider, thereby specifying the instance's characteristics such as the available processing power, main memory and I/O capacity. Most commonly, the notion of an instance materializes into a virtual machine that is launched on the physical IT infrastructure of the provider. Examples of successful IaaS providers are; Amazon EC2, Joyent, Rackspace. PaaS is a way to rent hardware, operating systems, storage and network capacity over the internet. It delivers a computing platform or software stack as a service to develope applications. This can broadly be defined as application development environment which is offered as a 'service' by the vendors. AppEngine, Bungee Connect, LongJump, Force.com, WaveMaker are examples of PaaS [7]. Instead of executing an application locally, this application could be located on the cloud and can be accessed online via web interfaceby SaaS. The applications are already-created as fully or partially remote services. Sometimes they are in the form of web-based applications and other times they consist of standard non remote applications with Internet-based storage or other network interactions [8]. Yahoo mail, Google docs applications, Salesforce.com CRM apps, Microsoft Exchange Online, Facebook are examples of SaaS [7].

Other important characteristics that distinguish Cloud Computing from other distributed systems are dynamic provisioning, Geo-distribution, ubiquitous network access, and Utility-based pricing [9].

One of main criteria affects data sharing in resource provision process under the cloud environment is the network performance, especially network bandwidth consumption. The network bandwidth can be stated as the amount of data shared over the network in a given time [11]. The relationship between

the characteristics of the offered traffic, the link speed and the resulting Quality of Service (QoS) should be concerned for calculating the overall cost of the resource provisioning. On the other hand, the users ask about the needed resources that achieve their services and deal with the providers with a pay-as-you-go strategy without concerning about the network communication characteristic because the providers do not offer a guaranteed network bandwidth to the users. But sometimes, achieving the requirements of one cloud user may make an influence on another cloud user in the same cloud environment due to the network latency which increases response time and decreases throughput. That may cause time delay in service providing [11]. The possibility of growing the Cloud environments by spreading the resources across wide geographical area using Wide Area Network (WAN), causes a delay for providing the services because of the limitation of the communication infrastructure, which is provided by the Internet [12].

The work in this paper is considered an enhancement of RPOA algorithm which is based on Particle Swarm Optimization to find the near optimal resource utilization with considering the data transfer time and the cost in the performance measurement [3].

The rest of this paper is organized as follows; the related works about the resource provision problem in cloud computing are presented in section 2. The principles of the resource provision in the cloud environment are discussed in Section 3. Section 4 states the influence of data communication on provisioning process. Section 5 represents the enhancement RPOA algorithm and its implementation and evaluation using CloudSim toolkit. The conclusions is presented in section 6.

II. RELATED WORKS

Over the years, the demand of services provisioning through the Internet has been increased . In order to handle these huge numbers of users' applications, Cloud infrastructure providers (i.e., IaaS providers) have established data centers in multiple geographical locations to achieve availability and ensure reliability in case of site failures [13].

Many researchers tried to meet the conditions of Quality of Services (QoS) and agreements of Service Level Agreement (SLA) in the resource provisioning process.

Hu; Y., et. al. [14] have merged two resource allocation policies; Shared Allocation policy (SA) and Dedicated Allocation policy (DA). These allocation policies are evaluated by heuristic algorithm on basis of the smallest number of servers required to meet the negotiated SLA. Chaisiri; S., et. al. [15] have introduced an algorithm for resource provisioning based on the cost optimization by considering the demand price uncertainty in the provisioning stages to manage virtual machines. Byun; E., et. al. [16] have suggested a Partitioned Balanced Time Scheduling algorithm to estimate the minimum number of computing hosts required to execute a workflow within a user-specified finishing time. The main goal of this algorithm is to minimize the resource cost rather than the makespan of the workflow. This algorithm is for the automatic execution of large scale workflow-based applications on dynamically and elastically provisioned computing resources.

Jung; G., Sim; K. M., [17] have proposed the agent-based adaptive resource allocation algorithm. This algorithm aimed to satisfy customer needs of service in terms of fast allocation time and execution response time where the provider try to allocate each customer request to an appropriate data center among the distributed data centers by offering his resources under the infrastructure as a service model. The authors have proposed an adaptive resource allocation model that took into consideration both the geographical distance between the location of consumer and data centers and the workload of data center.

There is a proposed algorithm for resource provisioning by considering the communication link between the customer and the provider with minimum cost, as illustrated by Gaurav Raj [18]. He has proposed an Efficient Broker Cloud Management as a new provider cloud communication paradigm, explaining communication mechanism between the provider and the cloud using cloud exchange. He obtained optimum route on the basis of cost factor considering hops count, bandwidth and network delay using Optimum Route Cost Finder algorithm.

Gaurav Raj, Ankit Nischal [19] are tried to find an efficient way to utilize the resources within the cloud, and to create virtual machines with consideration to optimum cost based on a performance factor. This performance factor depends upon the overall cost of the resource, communication channel cost, and reliability and popularity factor. They proposed a framework for communication between the resource owner and the cloud using Resource Cloud Communication Paradigm (RCCP) and extended the CloudSim by adding provisioning policies and Efficient Resource Allocation (ERA) algorithm in VMM allocation policy as a decision support for resource provisioning [20].

El-Attar; N., et. al [3] have introduced a Resource Provision Optimal Algorithm (RPOA) based on Particle Swarm Optimization PSO to find a Workload Resource map WM that is commensurate with customer budget and suitable for deadline time. According to this algorithm, it is found that maximizing the performance of computing resource could be achieved by allocating its capacity for the maximum number of workloads using (PSO) algorithm with the utilization function. *let list*)

III. RESOURCE PROVISIONING PROBLEM

Identifying the proper resource to the required service is considered a fundamental problem in Cloud computing. Cloud providers have to achieve the availability by allocating the appropriate resources to the required services without any confliction in the resource demands and with determining the right amount of resources required for the execution of services in an optimal way. This optimization aims to achieve the defined Quality of Service (QoS) that is affected by various parameters related to the application (i.e. security, accessibility, availability and reliability), the resources (i.e. availability, reliability, throughput and utilization) ,as well as, the user-defined requirements (i.e. cost and time) and the provider profit (i.e. resource utilization and revenue) [21].

In other words, the main goal of the provisioning policy is how to spread the applications load on convenient Cloud

resources to achieve the optimization objective of satisfying both service providers and customers QoS requirements (i.e. minimizing both response time and cost of resource utilization and, in the same time, maximizing the provider profit with the available customer's budget) [3]. The Cloud computing service provider's profit is achieved by providing high-quality services to the users through the efficient allocating of the resources on demand [22].

In [3], a Resource Provision Optimal Algorithm (RPOA) has been proposed based on particle swarm optimization to find the near optimal scheduling map of available resources with minimization of user response time and resource usage cost. According to the work in this paper, the RPOA has been enhancement and modified by considering the influence of communication parameters.

IV. DATA COMMUNICATION IN RESOURCE PROVISIONING

One of the main features of Clouds is to spread workloads over distributed virtualization infrastructure and to cover larger geographical areas. The resource providers in the cloud environment rely on the concept of virtualization to supply their computational resources. The aim of virtualization technologies is to hide the underlying infrastructure by introducing a virtual layer between the physical infrastructure and the computational resources. In fact, the large clouds are based on exploitation of distributed virtualization infrastructures of other clouds to provide new types of services "Distributed IaaS, PaaS, and SaaS". On the other hands, the distributed Cloud service is composed of a set of VMs spread over a wide geographical area, coordinated in order to achieve a specific service which is provided on demand to the user in order to meet his requirements [12].

Another scenario of resource provision is to use the federation perspective that allows the Cloud providers to use virtualization infrastructures of other federated Clouds, trying to meet more and more kinds of required services [23, 24]. The federation makes the small and the medium Clouds able to increase their virtualization capabilities by using the virtualization infrastructures of other federated Clouds to achieve the required service.

By using the federation perspective, the resource provision is become more elastic by increasing the virtualization capabilities of Clouds and, on the other hand, to enable the Clouds providers to rent their computational and storage resources when their virtualization infrastructures are unused [12].

Any Cloud user cares about two main criteria; the service response time in the data center, i.e., turnaround time [25], and the cost of the available resources that can be allocated to the service. Generally, the allocation process of any cloud service can be divided into three consecutive stages; scheduling, computation and transmission [25].

On the other hand, the potential growth of Cloud environments due to the distribution of Cloud infrastructures over large scale geographical area makes the provisioning of services face some delay problems due to the latency that happened in transmission through the Internet [12].

Therefore, the transmission time has to be considered as a significant criterion in the allocation process specially when there is a need for transferring a huge dataset.

V. THE ENHANCEMENT OF RPOA

According to the RPOA algorithm, the data centers are distributed over different resource pools. Each pool includes a specific resource type (e.g., computing power or memory capacity), and each computing resource is limited with the available number of hosts and both certain memory capacity and computing power consumption. On the other hand, each task workload is associated with a number of subtasks which require a certain amount of computing power or memory capacity depending on the workload type. In the same time, the RPOA algorithm is considered that the work loads are independent and have the same priority. By these conceptions, the measurement of the implementation's cost and time are become more accurate [3].

The work in this paper is considered a new version of RPOA which has been proposed in [3] with considering the influence of communication bandwidth and time latency beside the cost and time of the resources. This is because the network bandwidth and latency have a critical effect on the network performance, and accordingly on the provisioning process.

The network bandwidth can be defined as the rate at which the number of messages can be transferred from one point to another point in a given amount of time and it can be measured as baud rate, rate of data transfer and bit rate or throughput. The bandwidth that is used in cloud computing is the network bandwidth [11]. On the other hand, the network latency refers to any delay typically incurred in the processing of network data. This delay occurs because of the resources distribution over large number of computing nodes on the network as in the web based cloud applications [26].

A. The Enhancement of RPOA Environment

According to RPOA algorithm, there are 'm' number of available resources, and 'n' workloads that contain 'j' of subtasks, with consideration of some principle assumptions (fixed quantity of resources $'Q'$, with a definite price $'p_j'$ and a default execution time $'t_j'$). On the other hand, every workload $'i'$ has a set of tasks $'j'$ that need a specific resource quantity denoted by 'q', and every customer has the availability of deciding the price of each task that can be paid 'bp_j^i'. All of these consideration with a very important constraint which is ensuring that the total number of all workloads have to be not more than the available resources [3].

On the other hand, the cost and time of data transfer have to be defined, i.e. the communication link type and the available data packet size on the communication channel. The Transfer cost can be calculated according to equations (1, 2, 3) [27]:

$$Cost\ of\ Transfer\ (CoT) = TC + RT + LC + Thr\_c \qquad (1)$$

Where,

$$Transmission\ Cost\ (TC) = TT * Cost\ per\ sec \qquad (2)$$

Throughput Cost (Thr_C)=

Total Bandwidth Price / Packet Size   (3)

Routing Protocol (*RT*) depends on the network distance as shown in Table.1, and both Loading Cost (*LC*) and total bandwidth price are applied from the resource provider.

TABLE I.   DYNAMIC ROUTING PROTOCOL DEFAULT ADMINISTRATIVE DISTANCES [28]

Route Source	Default Distance
Connected interface	0
Static route	1
Enhanced IGRP (EIGRP) summary route	5
Exterior Border Gateway Protocol (BGP)	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
EIGRP external route	170
Interior BGP	200
Unknown	255

Similarly, the transfer latency time will be calculated by equation (4) [29].

Latency (Packet Delivery Time(PDT)) = TT +PD   (4)

Where Transmission Time (TT), and Propagation Delay (PD) are calculated by equations (5, 6):

TT=Packet Size /Bit Rate,   (5)

which are stated in table 2, and

PD=Distance/ Propagation Speed (PS)   (6)

where Propagation Speeds are known as follows [30]:

a) for copper wires= 2*108 meter/sec.

b) for wireless = 3*108 meter/sec.

TABLE II.   List of Network Technology bit rate [31]

Technology	Rate bit/sec
Modem 56k (8000/8000 baud) (V.92)	56.0/48.0 kbit/s
ISDN Basic Rate Interface (single/dual channel)	64/128 kbit/s
IDSL (dual ISDN + 16 kbit/s data channels)	144 kbit/s
HDSL ITU G.991.1 aka DS1	1,544 kbit/s
MSDSL	2,000 kbit/s
SDSL	2,320 kbit/s
SHDSL ITU G.991.2	5,690 kbit/s
ADSL (G.Lite)	1,536/512 kbit/s
ADSL (G.DMT)	8,192/1,024 kbit/s
ADSL2	12,288/1,440 kbit/s
ADSL2+	24,576/3,584 kbit/s
DOCSIS v1.0 (Cable modem)	38,000/9,000 kbit/s
DOCSIS v2.0 (Cable modem)	38,000/27,000 kbit/s
DOCSIS v3.0 (Cable modem)	160,000/120,000 kbit/s
Uni-DSL	200,000 kbit/s
VDSL ITU G.993.1	52,000 kbit/s
VDSL2 ITU G.993.2	100,000 kbit/s
BPON (G.983) fiber optic service	622,000/155,000 kbit/s
GPON (G.984) fiber optic service	2,488,000/1,244,000 kbit/s
Classic WaveLAN (Wireless)	16,384 kbit/s
IEEE 802.11 (Wireless)	16,384 kbit/s
RONJA (Wireless)	81,920 kbit/s
IEEE 802.11a (Wireless)	442,368 kbit/s
IEEE 802.11b (Wireless)	90,112 kbit/s
IEEE 802.11g (Wireless)	442,368 kbit/s
IEEE 802.16 (WiMAX) (Wireless)	573,440 kbit/s
IEEE 802.11g with Super G by Atheros (Wireless)	884,736 kbit/s
IEEE 802.11g with Nitro by Conexant (Wireless)	1,146,880 kbit/s
IEEE 802.11n (Wireless)	4,915,200 kbit/s
IEEE 802.11ac (maximum theoretical speed) (Wireless)	58,1333,053kbit/s

So, depending on RPOA implementation [3], there are *N* users ask for services that are defined

with a matrix q_{nm} subtasks that have to be achieved by the available resources that have a vector Q_m of resources fixed capacity $Q=[Q_1,Q_2,Q_3,.......Q_m]$. The Total processing time t_j^i and cost c_j^i of every workload *i* on the suitable resource *j* is calculated according to equations 7, 8 [3] based on equation 1,4.

$$t_j^i = \sum \left(\frac{q_i}{Q_m} * t_j \right) + PDT_j^i \qquad (7)$$

$$c_j^i = \sum (p_j^i * t_j^i) + CoT_j^i \qquad (8)$$

RPOA uses the elasticity definition to find the existing relation between time and cost of both workloads and resources [32].

B. The Enhancement of RPOA Implementation and Evaluation

The difficulty of obtaining workloads for real applications makes the evaluation of resource provisioning algorithm to be not an easy process. So, the CloudSim simulator is used to overcome this problem.

CloudSim with CloudAnalyst interface is used as Simulation framework to evaluate the Enhancement RPOA algorithm performance. CloudSim is a framework developed by the GRIDS laboratory of University of Melbourne which enables seamless modelling, simulation and experimenting on designing Cloud computing infrastructures [33]. CloudSim is a self-contained platform which can be used to model data centers, service brokers, scheduling and allocation policies of a large scaled Cloud platform. It provides a virtualization engine with extensive features for modelling the creation and life cycle management of virtual engines in a data center. CloudSim framework is built on the top of GridSim framework also is developed by the GRIDS laboratory, and it is written in java 6.1, because Java is a programming language designed for the distributed environment of the internet [34] [35]. CloudAnalyst has an easy graphical user interface that provides a high degree of configurability and flexibility by giving the ability of quickly and easily changing of the parameters that needed to be assumed in the simulation. Also CloudAnalyst gives a graphical output in the form of tables and charts to summarize the potentially large amount of statistics that is collected during the simulation with the ability of repeating the experiments under similar environment. Also, it gives more accurate evaluation by allowing the comparison between different scheduling algorithms [34].

RPOA is applied for memory provisioning; the user requests the memory on the available data centers. To evaluate the enhancement RPOA algorithm, some resource allocation requests are applied at different intervals of peak hours with different average of peak users. The datacenters that are used in this simulation have (x86) architecture, working on Linux operating system, and with Xen virtual machines that are based on time shared allocation policy.

The simulation is done over a various number of user bases requests on different available data centers. Every user base has its own budget, and asks for a specific amount of memory also it can be divided to independent subtasks; everyone has its own request and budget. On the other hand, every data center has a certain amount of memory, and has a predefined execution time. Also, the available bandwidth and its transmission cost are defined with the type of network communication and the available networks bit rate.

As shown in Figure 1, CloudAnalyst divides the worlds into regions to make the cost and time calculation more accurate.

Fig. 1. Region Boundaries Details in CloudAnalyst

Tables 3 and 4 show the variant parameters of a ten sample from hundred simulated experiments; respectively. Table 3 shows the user bases services request and contains all of the request details as, user region, request size, number of subtasks; if it contains subtasks; and user budget. Other important parameters are that the start and end peak hours that influence the time of server responding.

TABLE III. USER BASES DETAILS

Table 4, also shows a sample of the data centers details that contains data center region, Virtual machine's cost for every

	Region	No. of Subtask per Request	Request Size	User Budget	Peak Hours (start)	Peak Hours (End)
UB1	3	20	8.678532	0.7	3	9
UB2	1	10	8.583069	0.4	4	10
UB3	5	5	7.271767	0.5	2	7
UB4	0	40	3.290176	0.9	1	8
UB5	2	8	23.97537	0.2	3	9
UB6	2	8	8.678532	0.3	4	10
UB7	5	25	8.583069	0.9	2	7
UB8	3	30	7.271767	0.4	5	10
UB9	1	20	3.290176	0.3	3	9
UB10	0	15	23.97537	0.8	1	5

data center, data transfer cost, and Resource size per G. Byte.

TABLE IV. USER BASES DETAILS

	Region	Cost per Proc $/Hr	Memory Cost	Trans. Cost Byte per Sec	Resource Size G. Byte
DC1	0	0.2	0.05	0.01	10.742188
DC2	1	0.08	0.06	0.09	29.296875
DC3	2	0.15	0.07	0.08	9.765625
DC4	4	0.19	0.04	0.12	39.0625
DC5	5	0.3	0.03	0.13	48.828125

Figure 2 is the simulation result interface from CloudAnalyst. It displays the location of every user base on the selected suitable datacenter.

Fig. 2.   Simulation Results

Figures 3 and 4 show the user base and data center configuration, in order. Figure 3 displays the details of user bases configuration (with their subtasks), and virtual machines that used to every user base.
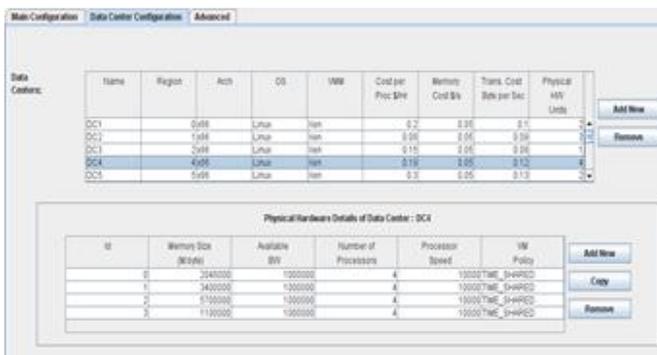

Fig. 3.   UserBase Configuration

Also, Figure 4 displays the data center configurations with showing the physical hardware details of every data center.
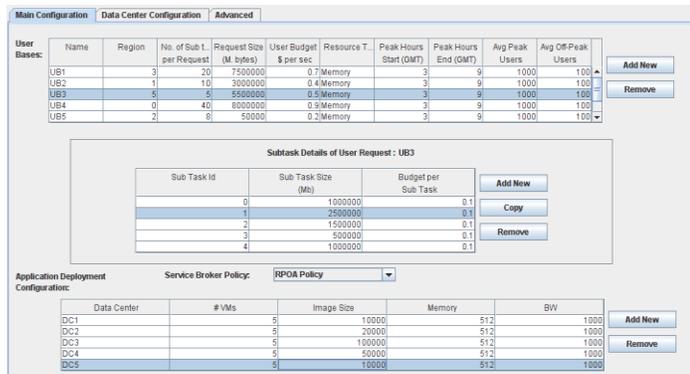

Fig. 4.   Data Center Configuration

Figure 5 shows the chart of statistical results of the executed simulation. Only 50 requests are displayed from the hundred simulated user bases. This is to make the comparison results to be visible and clear. It displays the required amount of every user base's task and its relation with the whole capacity of the available resources.

This relation is to illustrate the importance of using the whole capacity of a data center to one or more of user bases that is because leaving a small amount of capacity in every datacenter will not be useful in another allocation process. But still there is another main condition in the allocation process. This condition is to search for the closest data center to the user base from both the latency and region, for trying to reach the goal of cost and time optimization by using RPOA.

Table 5 displays the overall cost that is paid for every datacenter containing the processing and the transfer cost.

TABLE V.   GRAND TOTAL COST FOR SOME DATACENTERS

	VM cost	Data Transfer Cost	Total Cost
DC1	1.00	10784.16	10785.2
DC2	0.40	706.05	706.447
DC3	0.75	1401.43	1402.18
DC4	0.95	936.79	937.744
DC5	1.50	14991.03	14992.5

The main objective of the Enhancement RPOA algorithm is that find the near optimal allocation of the user requests on available data centers with the user budget restriction and deadline time with putting the influence of communication factors in consideration.

As shown in table 5, the data transfer cost has a significant effect on the total cost, this leads to take the distance between request region and resource region in consideration, and trying to allocate the request to the nearest data center which achieves the user demands.

The Enhancement RPOA algorithm provides a result by taking budget and deadline time into consideration as it can do; the error rate with the proposed sample of a hundred user base is nearly 18% for cost and 24% for time.

VI.   COMPARATIVE STUDY

The first algorithm which we can use to compare with RPOA is Closest Data Center Algorithm (CDCA) [34]. This algorithm is considered the simplest service broker policy. It is based on creating an index table of all data centers indexed by their region. When the user base sends a message with his requirements, the algorithm looks for the data centers which are in the same region of the user base and picks from the data center located at the earliest/highest region in the proximity list and puts the request in the requests' queue of this data center [34].

The comparison between RPOA and CDCA in terms of cost reduction is illustrated in Figure 8. The CDCA's working mechanism is to provide resource with minimizing the cost of the provided service that is because the CDCA searches for the resource in the nearest region of user base that reduces the transfer cost, and so reduces the total cost.
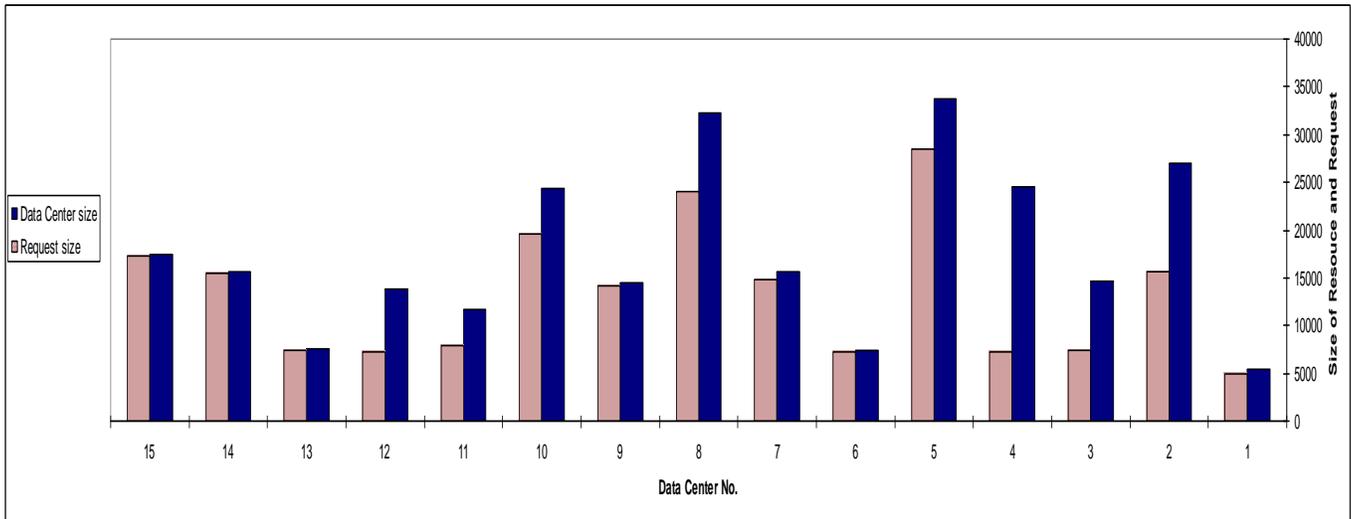
Fig. 5.   Relation between the size of Resource and Request.
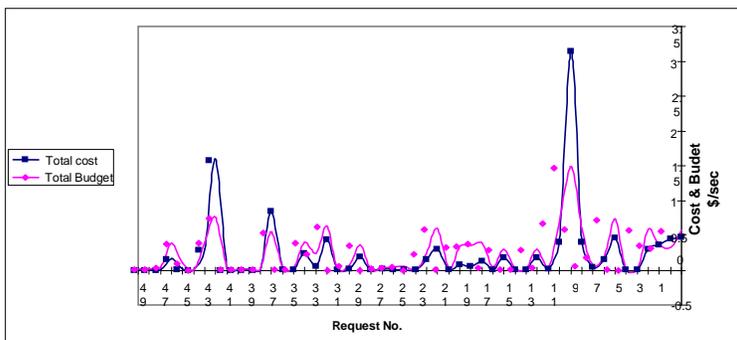

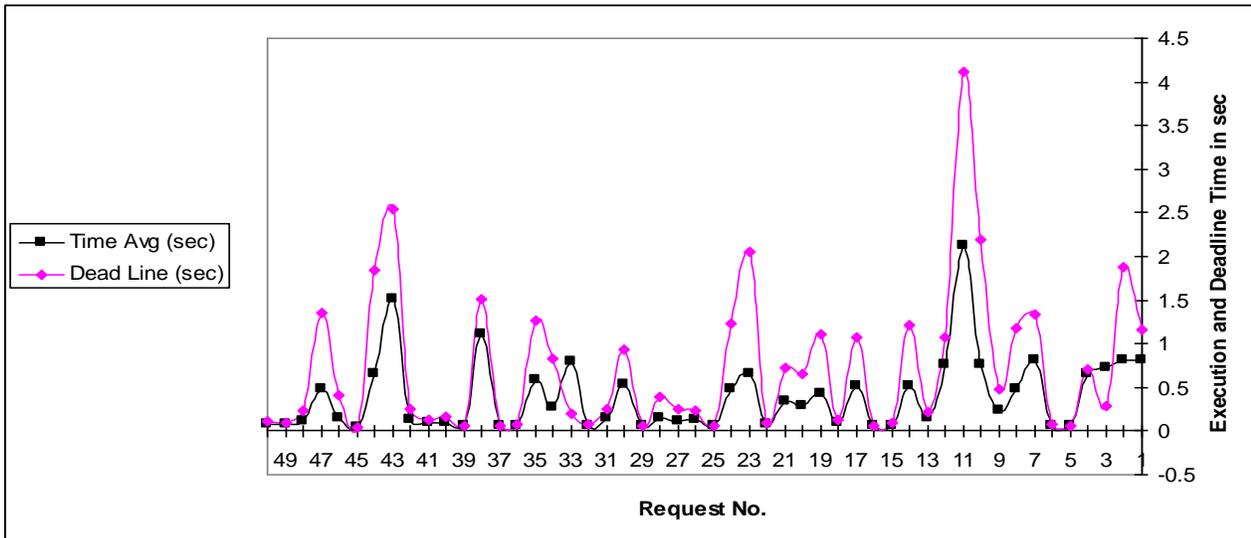Fig. 6.   Comparison between Execution Cost and User Budget In RPOA

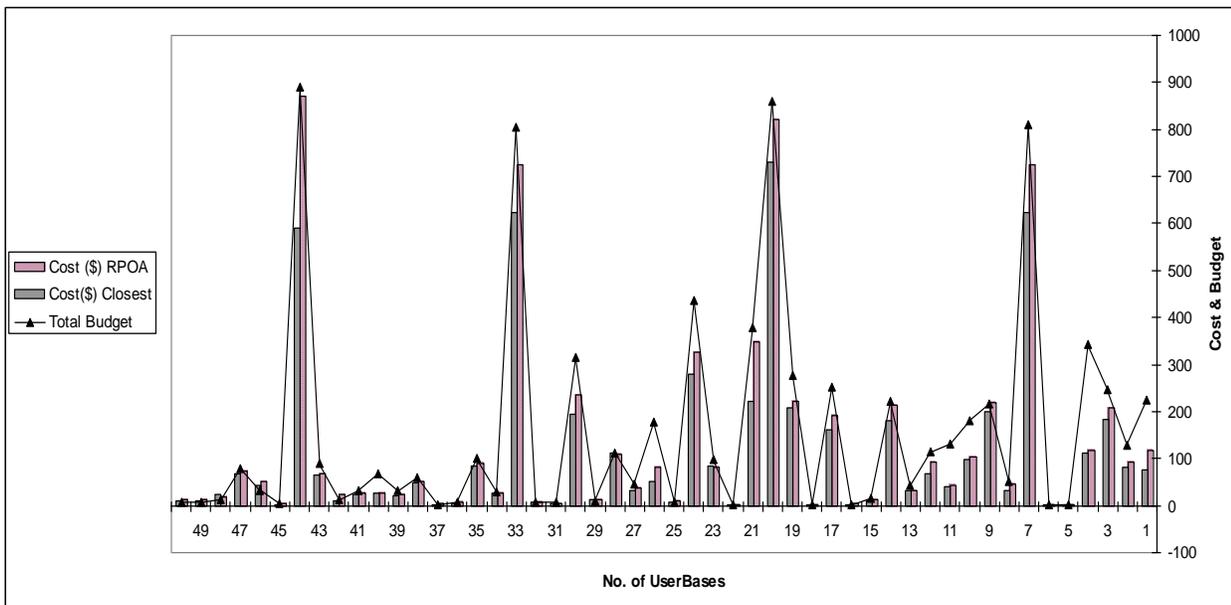Fig. 7.   Comparison between Deadline Time and Execution Time In RPOA

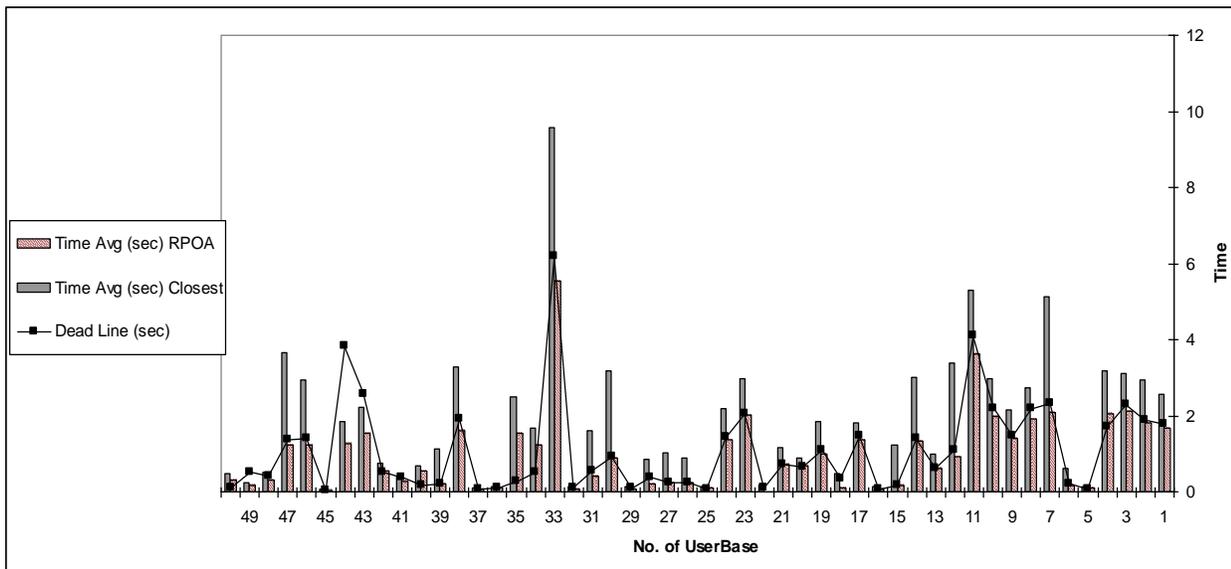Fig. 8. Comparison between Cost and budget in RPOA and CDCA

Fig. 9. Comparison between Execution and Deadline Time in RPOA and CDCA

On the other hand, RPOA gives good results in time optimization comparing with CDCA, as shown in Figure 10. RPOA tries to reduce execution time to fit deadline time, but CDCA does not care about time. It may waste time by letting the task in a long queue in the nearest resource.

The above comparisons can be summarized in that, RPOA algorithm can enhance the time optimization more than CDCA with 65%, but it has a weakness in reducing execution time with 13% than CDCA.
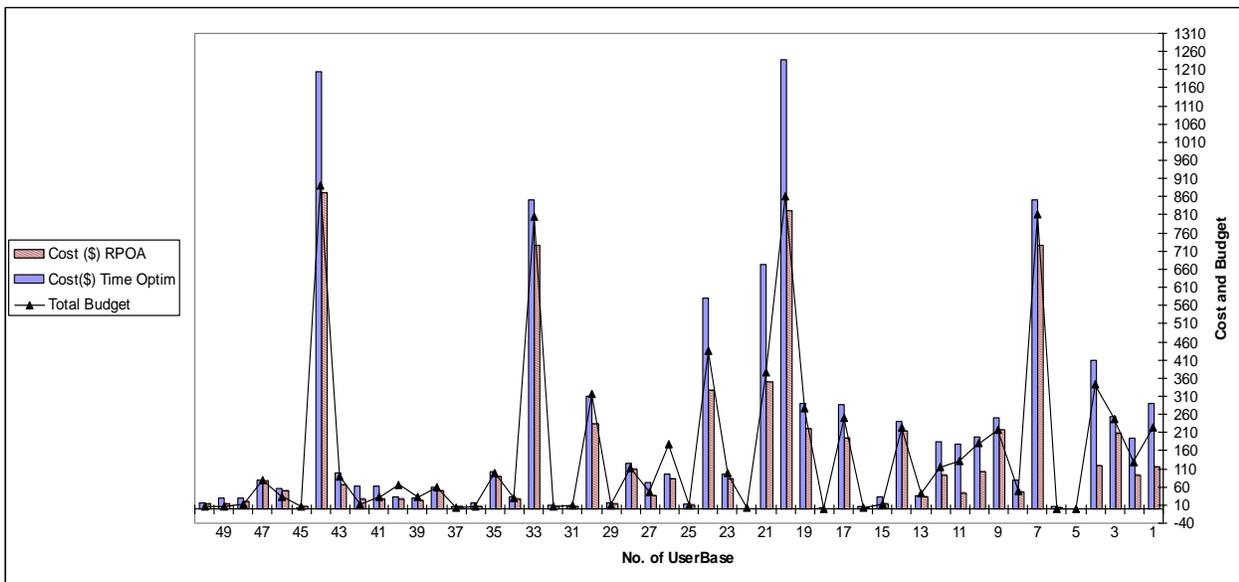
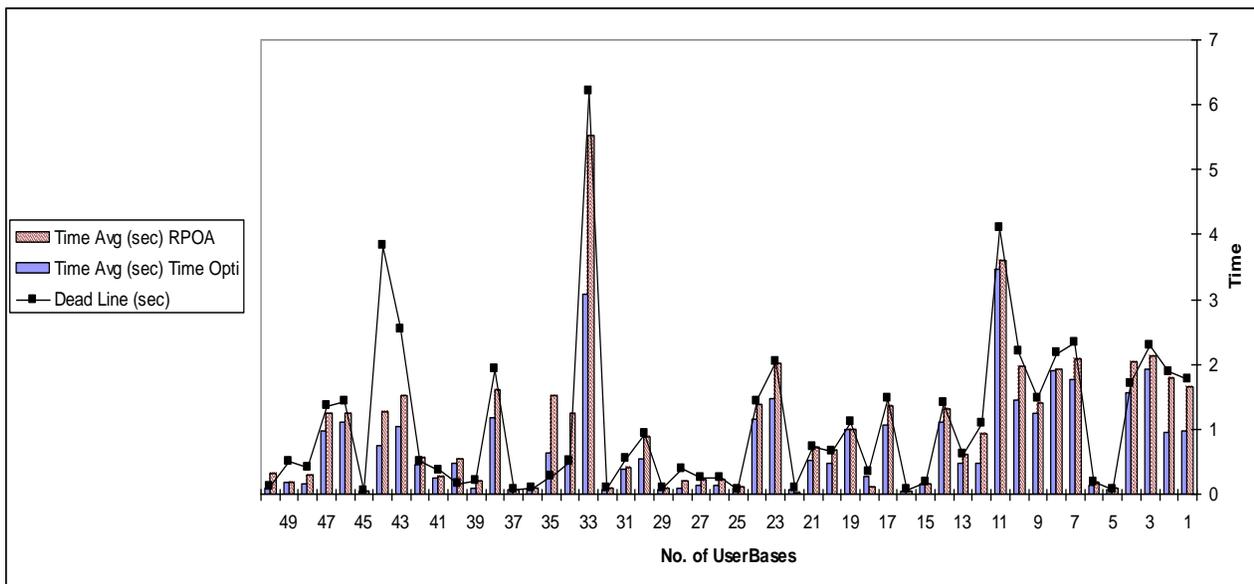Fig. 10. Comparison between Cost and Budget in RPOA and BRTA

Fig. 11. Comparison between Execution and Deadline Time in RPOA and CDCA

The second Comparison will be done between RPOA algorithm and Best Response Time Algorithm (BRTA) [34]. This policy locates all the available data centers and indexes them. After the service broker receives the user's request, it identifies the closest data center (i.e. in terms of latency) and iterates through the list of all data centers and estimates the response time at each data center by querying the last recorded processing time using Internet Characteristics. If this time is recorded before a predefined threshold, the processing time for that data center is reset to 0. This means that the data center has been idle for duration of at least the threshold time. If the least estimated response time is for the closest data center, the broker selects the closest data center. Else, the broker picks either the closest data center or the data center with the least response time with a 50:50 chance (i.e. latency load balanced 50:50) [34].

The comparison between cost and user budget in both RPOA and BRTA algorithms is displayed in Figure 10. BRTA algorithm cares only about how to provide the service with minimizing execution time as much as possible, but without considering the user budget, so it tries only to fit a deadline time without giving a suitable execution cost. Concluding form the BRTA characteristics, RPOA can be considered better than BRTA with 55% of cost reduction, but its performance in the execution time minimization is less than BRTA algorithm with 12% as shown in Figure 11.

VII. CONCLUSIONS

The Enhancement RPOA is an algorithm that is implemented to optimize both response time and cost of

resource provisioning process. On the other hand, the problem is the limitation of the resources capabilities which needs to spread the workloads on different geographical areas around the world. This leads to consider the communication factors, (i.e., communication time and cost) into consideration, which influence the performance of the allocation process. CloudSim is the simulation toolkit which is used to implement and evaluate the Enhanced RPOA algorithm and find the workload –resource map that achieves the user requirements.

REFERENCES

[1] J. Li, et.al., "Online Optimization for Scheduling Preemptable Tasks on Iaas Cloud Systems", J. Parallel Distrib. Comput., Vol. 72, Issue 5, 2012, pp. 666–677.

[2] Q. Zhang, et. al., "Cloud Computing: State-Of-The-Art and Research Challenges", J. of Internet Services and Applications, Vol. 1, No. 1, 2010, pp. 7–18.

[3] N. El.Attar, et. al., "Resource Provision for Services Workloads based on (RPOA)", Int. J. of Computer Science Issue (IJCSI), Vol. 9, Issue 3, 2012, pp. 553- 560.

[4] M. Armbrust, et. al., "Above the Clouds: a Berkeley View of Cloud Computing", Communications of the ACM, vol. 53, Issue 4, 2009, pp. 50-58.

[5] T. Forell, et. al, "Cloud Management Challenges and Opportunities", in: IEEE Inte. Symposium on Parallel and Distributed, 2011, pp. 881–889.

[6] R. Bossche, et. al, "Online Cost-Efficient Scheduling of Deadline-Constrained Workloads on Hybrid Clouds", Future Generation Computer Systems, Vol. 29, Issue 4, 2013, pp. 973–985.

[7] R. Grewal, P Pateriya, "A Rule-based Approach for Effective Resource Provisioning in Hybrid Cloud Environment", Int. J. of Computer Science and Informatics ISSN, Vol. 1, Issue. 4, 2012, pp. 2231 –5292.

[8] Y. Chen, S. Tai, "Optimal Provisioning of Resource in a Cloud Service", IJCSI, Vol. 7, No. 6, 2010, pp. 95-99.

[9] J. lmeida, et.al. , "Joint Admission Control And Resource Allocation In Virtualized Servers", Journal of Parallel and Distributed Computing, Vol. 70, No. 4, 2010, pp. 344-362.

[10] F. Silva, et al., "Application Execution Management on the InteGrade Opportunistic Grid Middleware", J. of Parallel and Distributed Computing, Vol. 70, No. 5, 2010, pp. 573- 583.

[11] K. Chandrakanth, S. Gayathri, "Examining Bandwidth Provisioning on Cloud Computing and Resource Sharing via Web", Int. J. of Engineering Science & Advanced Technology, Vol. 2, Issue-5, 2012, pp. 1372 – 1376.

[12] A. Celesti, et. al, "Virtual Machine Provisioning through Satellite Communications in Federated Cloud Environments", Future Generation Computer Systems, Vol. 28, Issue 1, 2012, pp. 85–93.

[13] R. Buyya, et. al., InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services, in 10th Int.. Conf. on Algorithms and Architectures for Parallel Processing (ICA3PP), pp 13- 31, May 21-23, 2010.

[14] H. Wong, et. al., "Resource Provisioning for Cloud Computing", Conference of Center for Advanced Studies on Collaborative Research, 2009, pp. 101- 111.

[15] S. Chaisiri, et. al., "Optimization of Resource Provisioning Cost in Cloud Computing", IEEE Transactions on Services Computing , Vol. 5, Issue 2, 2012,   pp. 164-177 .

[16] E. Byun, et. al, "Cost Optimized Provisioning of Elastic Resources for Application Workflows", Future Generation Computer System, Vol. 27, No. 8, 2011, pp. 1011-1026.

[17] G. Jung, K. Sim, "Agent-based Adaptive Resource Allocation on the Cloud Computing Environment", International Conference on Parallel Processing Workshops (ICPPW), 2011, pp. 345-351.

[18] G. Raj, et. al., "An Efficient Broker Cloud Management System", Int. Conf. on Advances in Computing and Artificial Intelligence, 2011, pp. 72-76.

[19] G. Raj, A. Nischal, "Efficient Resource Allocation in Resource provisioning policies over Resource Cloud Communication Paradigm", Int. Journal on Cloud Computing: Services and Architecture (IJCCSA),Vol.2, No.3, Issue 3, 2012, pp. 11-18.

[20] R. Calheiros, et. al, "CloudSim: A Novel Framework for Modeling and Sim9ulation of Cloud Computing Infrastructures and Services", http:/arxiv.org/ftp/arxiv/papers/0903/0903.2525.pdf, 2009.

[21] T. Varvarigou, et. al., "A Study on the Effect of Application and Resource Characteristics on the QoS in Service Provisioning Environments", Int. J. of Distributed Systems and Technologies, Vol. 1, Issue 1, 2010, pp. 55-75.

[22] H. Zhou, "Dynamic Resource Provisioning for Interactive Workflow Applications on Cloud Computing Platform", proceeding of: Methods and Tools of Parallel Programming Multicomputers - Second Russia-Taiwan Symposium, MTPP 2010, Vladivostok, Russia, May 16-19, 2010, pp. 116-125.

[23] J. Goiri, J. Torres, "Characterizing Cloud Federation for Enhancing Providers' Profit", in: Cloud Computing, IEEE International Conference on Cloud Computing, 2010, pp. 123–130.

[24] R. Ranjan, R. Buyya, "Decentralized Overlay for Federation of Enterprise Clouds", Handbook of Research on Scalable Computing Technologies, 2010, pp. 191–217.

[25] X. Nan, et. al, "Optimal Resource Allocation for Multimedia Cloud Based on Queuing Model", IEEE 13th International Workshop on Multimedia Signal Processing (MMSP), 2011, pp. 1-6.

[26] Arista Whitepaper, Architecting Low Latency Cloud Networks, 2007, http://www.computerlinks.co.uk/FMS/22489.cloud_network_latency.pdf , visited on Sunday 16-June-2013 at 11 pm.

[27] H. Lehpamer, Microwave Transmission Networks: Planning, Design, and Deployment, chapter 1- Transmission Network Media, 2004, McGrow-HillCompanies, USA.

[28] Cisco, Configuring IP Routing Protocol-Independent Features, http://www.cisco.com/en/US/docs/ios/12_2/ip/configuration/guide/1cfin dep.html, visited on Friday, 14- June- 2013 at 10 am.

[29] B. Forouzan, Data Communication and networking, fourth edition, Chapter 3 Data And Signals, McGraw-Hill Forouzan Networking Series, 2007.

[30] S. Alshaban, et. al., "Measurement of Transmission Time Delay and Efficiency of ATM LANE", Research Journal of Applied Sciences, Engineering and Technology, Vol. 2, Issue 2, pp.176-179, 2010

[31] L. Surhone, et. al., List of Device Bit Rates, Paperback Edition, Chapter-1, pp. 1-18, 2010, VDM Verlag Dr. Mueller AG & Co. Kg.

[32] I. Png, C. Cheng, Managerial economics, http://www.comp.nus.edu.sg /~ipng/mecon/sg/03elas_sg.pdf, 2001, visited on June - 8- 2013, 12:30 am.

[33] R. Buyya, et. al, Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities, in the 7th High Performance Computing and Simulation (HPCS 2009) Conf., June 21 - 24, 2009

[34] B. Wickremasinghe, "CloudAnalyst: A CloudSim-based Tool for Modelling and Analysis of Large Scale Cloud Computing Environments", Distributed Computing Project, Csse Dept., University Of Melbourne, 2009, pp. 433-659.

[35] R. Calheiros, et. al., "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms", J. of Software: Practice and Experience (SPE), Vol. 41, Issue 1, 2011, pp. 23-50.