A Comparative Analysis of Scheduling Policies in Cloud Computing Environment

Silpa.C.S
Dept of Computer Science, Pondicherry University
Puducherry

M.S.Saleem Basha, PhD.
Dept of Computer Science, Pondicherry University
Puducherry

ABSTRACT

Cloud computing has rapidly gained the popularity of researchers, government organizations and industries in recent years. It uses virtualization technology to provide computing resources to the customers as accessible public utility services. Since it uses virtualized resources, scheduling and resource allocation are the important research topics in the cloud computing. To make use of the capabilities of cloud, efficient scheduling policies are required to minimize the execution cost and to increase the resource utilization. This paper gives an overview of the existing scheduling policies in cloud computing systems. Also gives an analysis and comparison among different scheduling algorithms.

General Terms

Cloud Computing

Key Words

Cloud computing, Job Scheduling, Resource Allocation, Virtual machine.

1. INTRODUCTION

Many companies and institutions are currently running in cloud computing environment by deploying their services, applications and workloads in different cloud providers like Amazon, IBM, Microsoft, etc according to their Quality of Service requirements. The customers make use of the services and resources of the cloud providers in a pay – per – use manner from anywhere and at anytime [1]. Many of the researchers try to define cloud computing from different application aspects. Among the many definitions, the definition provided by U.S. National Institute of Standards and Technology is a relatively more specific definition "The Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resource(e.g, networks, servers, storage, applications and services)that can be rapidly provisioned and released with minimal management effort or service provider interaction[2]. This definition not only defines cloud concept in general, but also specifies essential characteristics of cloud computing delivery and deployment models.

Cloud computing offers everything as a service. There are different methods to deliver the cloud services. Mainly there are three service layers of cloud computing namely, Software as a service (SaaS), Platform as a service (PaaS) and Infrastructure as a service (IaaS) [3]. Efficient scheduling strategies should be used in order to make use of the cloud capabilities effectively. Many researches have been carried out in the field of resource allocation and scheduling in the cloud. The main motivation of these scheduling algorithms is to minimize the execution cost and time and to increase resource utilization.

The objective of this paper is to focus on various scheduling algorithms in cloud computing system. The rest of the sections are organized as follows. Section 2 presents different scheduling policies in cloud computing environment. In Section 3 presents comparison among different scheduling policies. Section 4 and Section 5 provide the analysis and discussion of the scheduling methods respectively. Section 6 concludes the paper with a summary of our future works.

2. EXISTING SCHEDULING ALGORITHMS

The following scheduling algorithms are presently established in the cloud computing environment.

2.1. Fuzzy-Genetic Algorithm based Task scheduling Optimizations

In this paper, an optimized algorithm based on the Fuzzy-Genetic Algorithm optimization[4] which makes a scheduling policy by evaluating the entire group of task in the job queueis proposed. Fuzzy sets were used to represent imprecise scheduling parameters and also to represent satisfaction grades of each objective. Genetic algorithms with different components were developed on the based technique for task level scheduling in Hadoop Map Reduce. To gain a better balanced load across all the nodes in the cloud environment, the scheduler is revised by predicting the execution time of tasks assigned to certain processors and making an optimal decision over the entire group of tasks. Although this method meets user's requirement and gets good resource utilization, the predicted execution time is a disadvantage of this scheduling method since it is not possible to predict the execution time of tasks effectively before executing the tasks.

2.2. The Analytic Hierarchy Process for Task scheduling and resource allocation

Daji Ergu et al. presented a model for task-oriented resource allocation in a cloud computing environment [5]. In this model computing tasks are collected in the Task Pool. These tasks are ranked using the pair wise comparison matrix technique and the Analytic Hierarchy Process giving the available resources and user preferences and are submitted to computing resources distributed in Cloud Computing Nodes. The computing resources can be allocated in terms of the rank of tasks. Besides this, an induced bias matrix is used to identify the inconsistent elements and improve the consistency ratio when incompatible weights in various tasks are assigned. When all tasks are ranked according to available resources this model improves the resource utilization and also meets user requirements. But here it is not possible to allocate resources dynamically.

2.3. A Priority based Job Scheduling Algorithm

A new priority based job scheduling algorithm (PJSC) is proposed [6] in cloud computing environment based on multiple criteria decision making model ,using analytical hierarchy process. Provided a discussion about some issues related to the proposed algorithm such as complexity, consistency and finish time. The proposed algorithm has reasonable complexity. But the main disadvantage is that the finish time cannot be calculated and response time is more .Also for more number of jobs allocations it is not suitable since finding priority of each job is tedious one.

2.4. Market Oriented Scheduling Policies

By considering the time and cost of resource provisioning, two market oriented scheduling policies(MOSP) were proposed[7] that aim at satisfying the application deadline by extending the computational capacity of local resources via hiring resource from Cloud providers. The policies are not having any earlier knowledge about the application execution time. This paper deals with how scheduling policies inside the broker can benefit from resources supplied by the IaaS cloud providers in addition to the local schedulers to get the use of application finished by the deadline and budget. The proposed the Cost Optimization and the Time Optimization scheduling policies increase the computational capacity of the local resources by hiring resources from IaaS providers.

2.5. Online Optimization for Scheduling Preemptable Tasks

JiayinLi et al. proposed a resource optimization mechanism in federated IaaS cloud system which enables preemptable task scheduling [8]. In this model, every data centre has a manager server that knows the current statuses of VMs in it own cloud. And manager servers communicate with each other. When a cloud receives requests from users, its manager server communicate with manager servers of other clouds and distribute its tasks across the whole cloud system by assigning them to other clouds or executing them by itself. The proposed algorithms, dynamic cloud list scheduling (DCLS) and dynamic cloud min-min scheduling (DCMMS) adjust the resource allocation dynamically according to the updated information of actual task execution. Also they have proposed energy aware local mapping mechanism which can reduce the energy consumption in federated cloud system.

2.6. Gang scheduling based Algorithm

Ioannis.A et al. proposed an efficient job scheduling algorithm[9] for time sharing, using gang scheduling. In this paper they have evaluated the performance of distributed cloud computing model based on the Amazon EC2 and estimated both the performance and overall cost of two leading gang scheduling algorithms, Adaptive First Come Fist Served (AFCFS) and Largest Job First Served (LJFS). It utilizes the concept of virtual machines. Depending on the load of the system at a specific time, the virtual machines are added and removed in the system. This model can be used in a cloud environment where the number of virtual machines is varying dynamically.

2.7. Resource Scheduling Strategy based on Genetic Algorithm

Jianhua Gu et al. presented [10] a scheduling strategy on load balancing of virtual machine resources using Genetic Algorithm(RSGA). It uses historical data and current states of VMs. In the proposed method starting from the initialization in cloud itself they look for the best scheduling solution by genetic algorithm in every scheduling and when there are no VM resources in the whole system use the algorithm to choose scheduling solution according to the computed probability. Even though this method can better realize load balancing and proper resource utilization, it does not deals with the dynamic behaviour of resource allocation.

2.8. Job scheduling algorithm based on Berger model

Berger model theory on distributive justice in the field of social distribution is introduced into the job scheduling algorithm in cloud computing [11]. Job scheduling algorithm based on Berger model (JSBM) concentrates on the fairness of the resource allocation. In this paper they classified user tasks by Qos preferences and defined resource fairness justice function to judge the fairness of the resource allocation. The proposed model agrees with the Qos parameters like completion time and bandwidth.

2.9. *Cloud Brokering Mechanisms for Optimized Placement of Virtual machines*

Johan Tordsson et al. proposed a new architecture[12] for cloud brokering and designed algorithms(CBVM) for optimizing the placement of virtual machines across multi cloud environment. The proposed algorithm can be used for cross site deployment of applications and services. The algorithms are based on integer programming formulations. User can guide the VM allocations by specifying maximum budget and minimum performance ,and also constraints with respect to hardware configurations of individual VMs, load balancing, etc. A static approach is used to address the cloud scheduling problem ,where the number of virtual resources are constant. This approach is not suitable for variable services where number of VMs vary dynamically.

2.10. Particle Swarm Optimization based Heuristic for Scheduling Workflow Applications

Particle swarm optimization based Meta heuristic method(PSOHS) was proposed[13]. The jobs are scheduled to cloud resources by taking into account, both computation cost and data transmission cost by focusing to minimize the total execution cost of applications on resources. If the resource cost increases PSOHS minimizes the maximum total cost of assigning all tasks to resources. It is found that PSO based algorithm gives lower cost of execution. It can be used for any number of tasks and resources by simply increasing the dimension of particles and number of resources respectively.

2.11. Optimal Cloud Resource Provisioning Algorithm

In this paper[14], an optimal cloud resource provisioning (OCRP) algorithm is presented to solve the difficulties caused by the uncertainty of consumer's future requirement and provider's resource prices. This OCRP is obtained by formulating and solving a stochastic integer programming problem. The proposed algorithm can provision computing resources for being used in multiple provisioning stages as well as a long term plan. Here the demand and price uncertainty is considered. Various approaches like deterministic equivalent formulation, sample-average approximation and Bender's decomposition are considered to obtain the solution for OCRP algorithm.

2.12. Resource Provisioning Policies to Increase IaaS Provider's Profit

Resource Provisioning Policies to Increase IaaS Provider's Profit(RPIPFC) which help the cloud providers in the federated environment to increase profit, resource utilization, and user satisfaction and benefit from outsourcing requests[15]. Main contribution of this work is proposing policies that help making decisions when providers have different choices regarding incoming requests: rejecting, outsourcing, or terminating spot leases to free resources for more profitable requests etc. These policies only address the possibility of outsourcing on-demand requests. Outsourcing spot requests is not considered in this work, since the proposed policies are not designed to handle highly fluctuating prices of spot VMs.

2.13. Heterogeneity-Aware Resource Allocation and Scheduling

Heterogeneity-Aware Resource Allocation and Scheduling gives (HARS) a metric of share in a heterogeneous cluster to realize a scheduling policy that gives high performance and fairness [16]. The heterogeneity of the environment should be built along with the performance and cost effectiveness. The data analytics system must report for heterogeneity of the situation and workloads. It also needs to provide fairness among jobs when multiple jobs share the cluster. Hence architecture to allocate resources to a data analytics cluster in the cloud proposed.

2.14. Online cost-efficient scheduling of deadline-constrained workloads

Online cost-efficient scheduling algorithms (OCSDCW)for deadline-constrained bag-of-task type applications taking into account data constraints, data locality and inaccuracies in task runtime estimates[17]. Online hybrid scheduling algorithms that operate on larger-scale problems with additional data constraints were presented and evaluated in terms of deadlines met, cost efficiency, computational efficiency, application turnaround time, and robustness with regard to errors in runtime estimates. The proposed algorithms are able to schedule a large number of applications

2.15. Hybrid Energy-Efficient Scheduling

A hybrid energy-efficient scheduling algorithm(HESA)[18] was proposed in private clouds based on their previous work and using dynamic migration of virtual machines. Since private clouds have some unique characteristics and special requirements, it is a challenging problem to schedule virtual machine requests effectively on computing nodes especially with multiple objectives to meet. The proposed scheduling algorithm comprised of a pre-power technique , a min-load first selection algorithm and a min-load first migration algorithm. An expected spectrum set for the left capacity is used. It uses power up command to wake the sleep nodes as well as idle nodes. Here power efficiency is improved.

3. COMPARISON OF DIFFERENT SCHEDULING POLICIES

Table 1 list out the environment, algorithm and scheduling parameters used in different scheduling policies. Also compares various scheduling policies in terms of their advantages and disadvantages.

Table1. Comparison between different scheduling policies

Sl no	Paper title/ Author	Algorithm/technique used	Scheduling parameters considered	Advantage	Disadvantage
1	Task scheduling optimizations for the cloud computing system, Sandeep Tayal	Genetic algorithm based scheduling	Execution time of tasks	Meet user requirements and improved resource utilization	Execution time is more
2	The analytic hierarchy process: Task scheduling and resource allocation in cloud computing environment, DajiErgu, Gang Kou, YiPeng, YongShi,YuShi	Ranking of tasks is done by using reciprocal pair wise comparison matrix and analytical hierarchy process	Response time, task expense	Improves resource utilization	Cannot allocate tasks dynamically
3	A Priority based job scheduling algorithm in cloud computing, shamsollah Ghanbari,Mohamed Othman	Based on the theory of Analytical hierarchy process	Make span	Since priority is considered important task will not be lagged	Increased make span
4	Adapting market oriented scheduling policies for cloud computing, Mohsen Amini Salehi,Rajkumar Buyya	Deadline budget constraint based Time and cost optimization scheduling policy	Response time, execution time, cost	Increase the computational capacity of the local resources by hiring resources from IaaS providers	Increased completion time
5	Online optimization for scheduling preemptable tasks on IaaS cloud systems, JiayinLi, MeikangQiu, ZhongMing, GangQuan, XiaoQin, Zonghua Gu	Based on cloud list scheduling and cloud min-min greedy algorithm for scheduling	Arrival time and execution time	The dynamic procedure provides significant improvement in the fierce resource contention situation.	Preemption leads to increased response time and overhead to the cloud providers
6	Evaluation of gang scheduling performance and cost in cloud computing system, Ioannis A, Moschakis, Helen.D, Karatza	Gang scheduling approach based shortest queue first, adaptive first come first served and largest job first algorithm	Waiting time, response time, cost	Improved resource utilization	Not considered the priority among the tasks
7	Anew resource scheduling strategy based on genetic algorithm in cloud computing environment, Jianhua Gu, Jinhua Hu, Tianhai Zhao, Guofei Sun	Based on genetic algorithm and spanning tree principle	Number of virtual machines, execution time	This method can better realize load balancing and proper resource utilization	It does not deals with the dynamic behaviour of resource allocation.
8	Job scheduling algorithm based on Berger model in cloud Environment, Baomin Xu, Chunyan Zhao, Enzhao Hu, Bin Hu	Based on Berger model theory on distributive justice in the field of social distribution	Completion time	Better completion time and fairness.	Comparatively lesser overall execution efficiency

9	Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers,Johan Tordsson, Ruben S.Montero, Rafael Moreno-Vozmediano, Ignacio M.Llorente	Based on integer programming formulations	Cost, throughput, execution time	Lower cost	This approach is not suitable for variable services where number of VMs vary dynamically.
10	A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing systems, Suraj Pandey, Linlin Wu, Siddeshwara Mayura Guru, Rajkumar Buyya	Based on particle swarm optimization	Computation cost ,data transmission cost	Reduces the cost, dynamic	Execution time is increased
11	Optimization of Resource Provisioning Cost in Cloud Computing, Sivadon Chaisiri, Bu-Sung Lee,, Dusit Niyato	Stochastic programming model	Cost	Reduces cost	Limitation of stochastic programming.ie, Stochastic programming does not address distributions of scenarios
12	Resource Provisioning Policies to Increase IaaS Provider's Profit in a Federated Cloud Environment, Adel Nadrajan Toosi, Rodrigo N.Calheiros, Ruppa K.Thulasiram, Rajkumar Buyya	Different policies to examine how decisions made by providers influence their profit and reputation with customers	Cost	Meet user requirements efficiently	proposed policies are not designed to handle highly fluctuating prices of spot VMs.
13	Heterogeneity-Aware Resource Allocation and Scheduling in the Cloud, Gunho Lee, Byung-Gon Chun, Rangy H.Katz	Progress share is calculated based on Computation Rate per slot	Response time, cost, fairness	High performance and fairness	Low resource utilization
14	Online cost-efficient scheduling of deadline-constrained workloads on hybrid clouds, Ruben Van den Bossche, Kurt Vanmechelen, Jan Broeckhave	Public cloud scheduler and hybrid cloud scheduler	Cost	Able to schedule a large number of applications	Not considered the dynamic behaviour of resources
15	A Scheduling Algorithm for Private Clouds, Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang	Energy efficient algorithm using dynamic migration	Execution time, energy consumption	Power efficiency is improved	Cannot meet user requirements efficiently

4. ANALYSIS

Analysed the feasibility of existing scheduling algorithms in terms of various parameters like execution time, resource utilization and cost and compared among different scheduling policies. Analysis result is shown in Fig 1, Fig 2 and Fig 3 based on the values in Table2,Table3 and Table4 which give comparison of execution time of scheduling algorithms, comparison between resource utilization rate of different scheduling algorithms and comparison between cost of different scheduling algorithms respectively.

Table 2 Execution time required for different scheduling policies

No. of jobs	MOSP	DCMMS	DCLS	LJFS	JSBM	PSOHS	OCSDCW	HESA
50	12	11	10	33	27	25	25	32
100	15	14	12	42	31	28	37	39
150	29	13	16	43	32	30	38	46
200	34	18	21	51	38	43	45	54
250	46	28	36	65	47	49	53	75

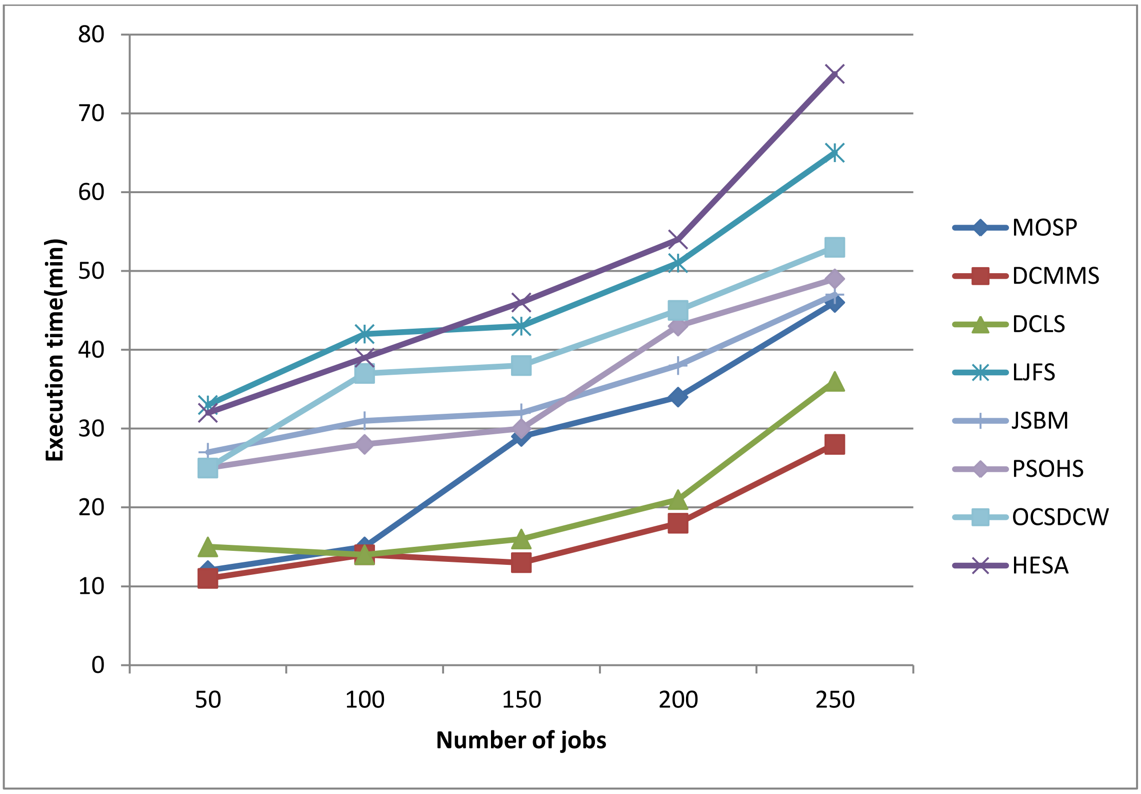

Fig.1. Comparison between execution time of different scheduling policies

Table 3 Resource utilization rate of different scheduling policies

No of Requests	RSGA	RPIPFC	HARAS
20	65	83	42
40	58	79	46
60	68	75	51
80	62	69	55
100	71	70	62

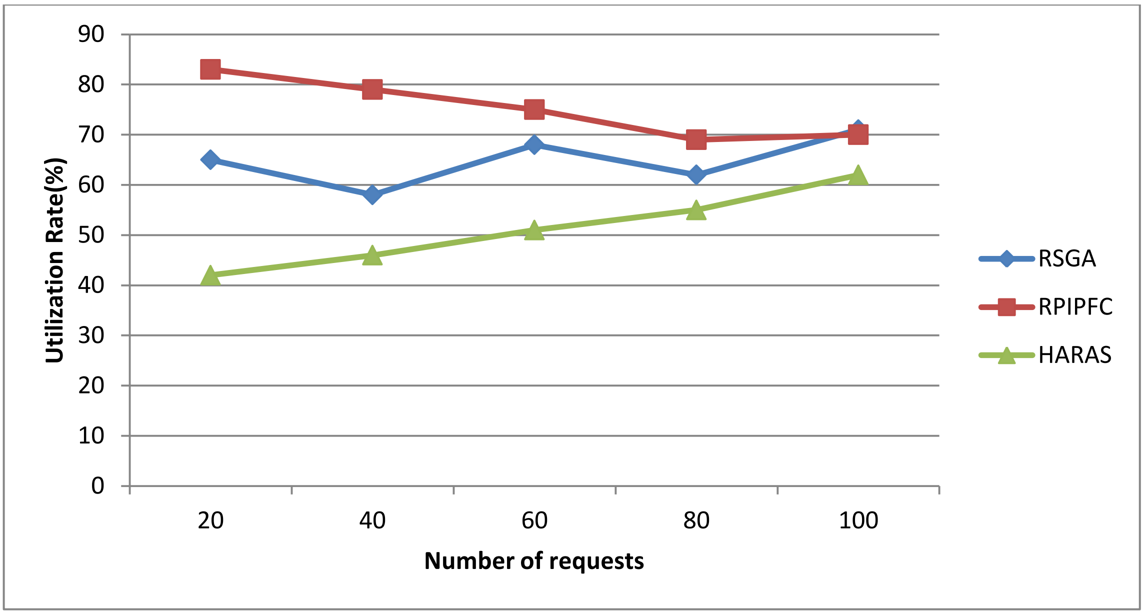

Figure.2. Comparison between utilization rates of different scheduling policies

From fig1 and Table 2, it is found that maximum number of jobs is executed with comparatively less execution time in DCLS,DCMMS and OMOSP. RPIPFC shows higher resource utilization rate compared to other scheduling policies in fig 2 and Table 3.Also by fig 3 and Table 4 it is shown that RPIPFC executes with less cost as per increasing user demands.

Table 4 Cost of different scheduling policies

No of required VMs	ORPC	CBVM	PSOHSW	RPIPFC
10	2100	2550	3050	1030
20	2350	2750	3110	1700
30	2550	3025	3250	2400
40	2750	3075	4050	2500
50	3060	3100	5010	3050

The outcome of this analysis shows that the cloud environment with federated multi cloud providers helps to increase user satisfaction and to improve resource utilization. The cost of deploying applications in multi cloud environments is cheaper compared to the deployment in each cloud provider alone. From the user point of view cloud federation creates an environment where the users can deploy their applications with less expense and less completion time. Also from the provider point of view resource utilization also improved in multi cloud systems. Thus by analysing table 1 ,fig 1,fig 2 and fig 3 it is clear that RPIPFC is a better scheduling strategy to achieve user satisfaction and resource utilization

5. DISCUSSION

Analysis and comparison between various existing scheduling policies in the cloud computing environment has done depending upon various parameters. Resource provisioning policies adapted in the federated cloud environment gives efficient scheduling and helps to meet user satisfaction and improve resource utilization. But it is not designed to handle highly fluctuating prices of spot VMs[15]. In the time optimization policy[7] it does not spend money to request more resources from IaaS providers and decreases completion time, whereas cost optimization policy takes more completion time. DLS and DCMMS [8] leads to better improvements in user satisfaction, where as in gang scheduling approach[9] since it not considered the priority of jobs it is not met with user satisfaction efficiently.

Genetic algorithm based scheduling policy[10] is not dealing with the dynamicity of cloud resources and is not applicable in real time environment where varying user demands coexist. Cloud brokering mechanism[12] reduces the execution cost reasonably but still dynamicity is missing there. Execution efficiency is comparatively less in Berger model based scheduling mechanism[11]. To achieve better efficiency and to improve user satisfaction and resource utilization it is better to adapt a scheduling policy in which multiple IaaS cloud providers are exist with an agreement policy among themselves.

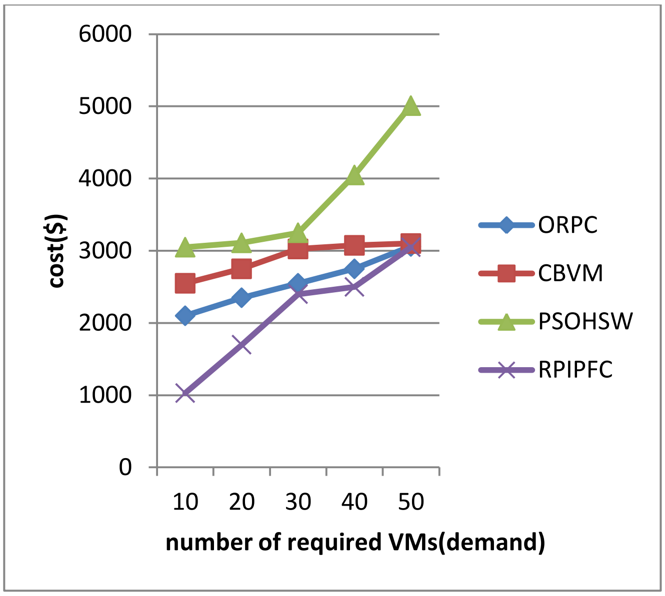

Fig.3. Comparison between Costs of Algorithms

6. CONCLUSION

Resource allocation and scheduling is a key issue in the cloud computing environment. In this paper various existing scheduling policies in cloud computing systems have been surveyed and compared them. Scheduling framework should meet user requirements as well as it should improve resource utilization so that it can achieve both user and resource provider satisfaction. In the above study it has found that no paper has specified about the effect of pre-empting VMs while different priority requests arrive. Our future work will be based on the above finding to develop a more efficient policy for scheduling the requests which will reduce the response time of requests as well as increases the profit.

7. ACKNOWLEDGEMENT

This work is a part of the Research Project sponsored under the Major Project Scheme, UGC, India, Reference No: F. No. 41-619/2012(SR), dated 17 July 2012. The authors would like to express their thanks for the financial support offered by the Sponsored Agency.

8. REFERENCES

[1]. Introduction to cloud computing architecture whitepaper, Sun Microsystems, 1st edition, June 2009.

[2]. The NIST definition of cloud computing, NIST special publication 800-145.

[3]. Introduction to cloud computing whitepaper, Dialogic making innovation thrive, July 2010.

[4]. Sandeep Tayal, "Task scheduling optimization for the cloud computing system" International Journal of Advanced Engineering sciences and Technologies,Vol No.5,Issue No .2,111-115, 2011.

[5]. Daji Ergu, Gang Kou, Yi Peng, Yong Shi, Yu Shi, "The analytic hierarchy process: Task scheduling and resource allocation in cloud computing environment" Springer Publications, May 2011.

[6]. Shamsollah Ghanbari, Mohamed Othman, "A priority based job scheduling algorithm in cloud computing"

International Conference on Advances Science and Contemporary Engineering, 2012.

[7]. Mohsen Amini Salehi, Rajkumar Buyya, "Adapting Market-Oriented Scheduling Policies for Cloud Computing".

[8]. Jiayin Li, Meikang Qiu, Zhong Ming, Gang Quan, Xiao Qin, Zonghua Gu, "Online optimization for scheduling preemptable tasks on Iaas cloud Systems" Elsevier Publications, February 2012.

[9]. Ioannis A, Moschakis, Helen D, Karatza, "Evaluation of gang scheduling performance and cost in a cloud computing system" Springer Publications, 2010.

[10]. Jianhua Gu, Jinhua Hu, Tianhai Zhao, Guofei Sun, "A new resource scheduling strategy based on genetic algorithm in cloud computing environment" Journal of Computers,Vol.7,No.1, January 2012.

[11]. Baomin Xu, Chunyan Zhao, Enzhao Hu, Bin Hu, "Job scheduling algorithm based on Berger model in cloud environment" Elsevier Publications, April 2011.

[12]. Johan Tordsson, Ruben S.Montero, Rafael Moreno-Vozmediano, Ignacio M.Llorente, "Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers" Elsevier Publications, July 2012.

[13]. Suraj Pandey, Linlin Wu, Siddeshwara Mayura Guru, Rajkumar Buyya, "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments".

[14]. Sivadon Chaisiri, Bu-Sung Lee,, Dusit Niyato, "Optimization of resource provisioning cost in cloud computing" IEEE transactions on services Computing,Vol.5, No.2, April- June 2012.

[15]. Adel Nadrajan Toosi, Rodrigo N.Calheiros, Ruppa K.Thulasiram, Rajkumar Buyya, "Resource provisioning policies to increase IaaS provider's profit in a federated cloud environment ".

[16]. Gunho Lee, Byung-Gon Chun, Rangy H.Katz, "Heterogeneity-aware resource allocation and scheduling in the clouds".

[17]. Ruben Van den Bossche, Kurt Vanmechelen, Jan Broeckhave, "Online cost efficient scheduling of deadline-constraint workloads on hybrid clouds" Elsevier Publications, December 2012.

[18]. Jiandun Li, Junjie Peng, Zhou Lei, Wu Zhang, "A Scheduling Algorithm for Private Clouds" Journal of Convergence Information Technology, 2011.