

Resource Scheduling under Diversified Service Quality Factors (RSDSQ) for IAAS in Cloud Computing

B. Ravindra Babu, M. Veera Sekhar Rao

Abstract: *The Resource Scheduling under Diversified Service Quality Factors for IAAS in Cloud Computing is addressed in this manuscript. This is since the resources under cloud platform are loosely coupled according to the SLA between cloud platform and the resource partakers. This enables the possibility of multiple resources from diversified partakers, those intended to accomplish similar service. The resource scheduling intends to select one resource among available resources to accomplish the scheduled task(s). The contemporary contributions related to resource scheduling are specific to one or more traditional QoS factors, which includes cost, deadline constrain, and power consumption. However, the quality of service often influenced by the contextual factors of the IAAS. Hence, this manuscript portrayed a novel resource scheduling strategy that orders the resources under degree of optimality, which is proposed in this manuscript. Unlike traditional methods of resource scheduling, this manuscript portrayed set of context related factors that are further used to define the heuristic measure called "Degree of Optimality". The experimental study on simulated environment elevating the proposal performance advantage in averse to other existing methods.*

Keywords: *resource management (RM), resource scheduling (RS), resource provisioning (RP), QoS.*

1. INTRODUCTION

The resource management (RM) is the protection activity containing of diverse phases of workloads & resources from the submission of workload to execution of workload. The RM in the cloud contains 2 phases: a) resource scheduling (RS) and b) resource provisioning (RP). The RP is determined to detect the sufficient resources for the specified workload on the basis of QoS pre-requisites described by the consumers of cloud while RS is mapping and performing the workloads of cloud consumer on the basis of chosen resources by RP.

On the basis of QoS pre-requisites, the resource scheduling for the sufficient workloads could be a challenging task. For an effective resource scheduling, it is required for deliberating the requirements of QoS [1]. Hence, there is requirement for uncovering the research tasks in RS to perform the workloads deprived of impacting other QoS pre-requisites.

The RS is an evolving research domain in the cloud because of huge resource cost & execution time. Diverse RS factors and criteria are directed towards divergent classifications of RSAs (Resource Scheduling Algorithms). The effective RS lessens cost of execution, consumption of

energy, performance time and deliberating other QoS essentials such as availability, reliability, scalability & security.

2. RELATED WORKS

The researchers have contributed "Multi-objective optimization scheduling" on the basis of considerations such as economic costs, system execution, confines and consumption of energy. Deliberating computational resources, the scheduling model is suggested, which segregates the budget costs and resources for lessening the tasks length, and hence lessen the completion time of task and enhance the resource utilization of system [2]. The work [3] presents the fast completion time replication algorithms suggestion for the "task-based replication". Initially, the algorithm adapts fuzzy clustering aimed at preliminary resources pre-processing, and later implements the scheduling of task duplication and acyclic graph. Deliberating the execution times of task, utilization rate and resource costs are considered in the environment of cloud utilizing multi-output, multi-input feedback "control dynamic resource scheduling algorithm" for assuring application under time confines for the optimal implementation execution [4]. For indefinite parameters in hybrid environment, the 2 "dynamic resource allocation" algorithms utilizing Pareto optimization model is suggested on the basis of deadline & cost restraints [5]. Nevertheless, the time intricacies of 2 algorithms are maximum and both are higher than or equivalent to $O(n^2)$. The "adaptive workflow scheduling heuristic" model, which considers the confines of time & cost, is proposed, even though the method schedules only data work flow analysis in the hybrid environment of cloud [6]. The work [7] presents that "Multi-objective scheduling" model is proposed on the basis of cost & time optimization objective with the confines as storage & bandwidth.

The model concentrated on enhancing the usage of "private cloud resources" for attaining the balance among costs & performance. The work [8], [9], [10] presents the scheduling issue in the way same to concentrate of current research. For addressing the optimization & IaaS provider benefits, here, "adaptive particle swarm optimization hybrid cloud scheduling algorithm" is proposed. Nevertheless, this model only deliberated the cloud provider's benefits without any deliberation of cost from the users' perception. The work [9] presents further, the researchers have suggested

Revised Manuscript Received on June 10, 2019.

B. RavindraBabu, Research Scholar, JNTUH, Hyderabad, Telangana, India. (Bellam.ravindrababu@gmail.com)

Dr.M. VeeraSekharRao, Research Scholar, JNTUH, Hyderabad, Telangana, India. (Sekhar.muddana@gmail.com)

that the concentration need to be on enhancing the performance of overall system in spite of whether exploring private or public resources of cloud.

Lastly, the tasks outsourcing towards public cloud method is suggested for lessening the outsourcing cost whereas simultaneously increasing the rate of using the data center of internal cloud [10]. Consequently, the research assumed mathematical programming aimed at optimized scheduling. Nevertheless, this method is unable for solving scheduling issues containing huge amount of data and its “optimization objectives” are costly. The work [11] presents the proposal of cloud RM program on the basis of identical objectives of increasing the utilization of resource and lessening the costs. Nonetheless, the method is mostly utilized for migrating on & off the virtual machine and is not implemented for the real instance of optimizing the task scheduling.

Researchers in [12] analysed different job types along with the availability of resources and developed a scheduling strategy that performs at resource broker. However, the model departed due to its computational cost and scheduling overhead. The contemporary contribution [13] portrayed a resource scheduling strategy for IAAS, which is using multiple Quality factors to schedule the resources. However, the contribution estimating the optimality of resource, which is limited to the quality factors such as make-span, price, and availability. The quality factors that are of linked to the context of the target IAAS are not in scope of this contemporary model, and load is the other crucial factor, which is not in the scope of this contemporary scheduling strategy.

In regard to these common constraints observed in these contemporary models, this manuscript aimed to derive a novel resource scheduling strategy that intended to estimate the degree of optimality of the corresponding resource through contextual quality factors defined.

3. RESOURCE SCHEDULING UNDER DIVERSIFIED SERVICE QUALITY FACTORS IN IAAS

The contribution of this manuscript portrayed a novel method that schedules the resources in IAAS of cloud computing under a heuristic measure called degree of resource optimality (*dro*) that estimated by using diversified quality of service factors related to the context of the service called IAAS. The adopted qualities of service factors related to the resources of the IAAS are (i) Degree of Response Time (ii) Degree of Service Denial, (iii) Degree of Realization, (iv) Degree of Load Adoption and (v) Degree of Cost Feasibility.

The information about the task(s) initiated at the SAAS includes the roundtrip time that indicates the arrival and expiration time of the corresponding task, the required service and the acceptable budget of the resource. Based on the information header of the corresponding task(s), the resource broker performs the resource scheduling under proposed scheduling strategy.

QoS factors of the resources

The proposed method of resource scheduling in IAAS of cloud computing estimates the optimum scope of pairing the task(s) initiated at SAAS and the available resources at

IAAS under diversified quality of service factors. Unlike the traditional scheduling strategies, the proposal derived quality of service factors in the context of Infrastructure-As-A-Service (IAAS), which are used further to estimate the optimality of a resources to be scheduled.

Initially the said model is assessing the values of the projected resource’s transmission quality metrics of all available resources and orders these resources according to one of the projected quality metric that considered as primary requirement of the optimum resource utilization. The strategic approach to assess the scope of each optimum resource utilization metric that projected in regard to the available resources is explored in following section.

Upon receiving the task headers, scheduler schedules the respective tasks to optimal resources that accomplish the task successfully. The objective of this manuscript is an optimum resource scheduling in regard to achieve maximal optimality towards resource utilization and task completion. The set of resources that are available to schedule are $R = \{r_1, r_2, r_3, \dots, r_x\}$.

The scheduling of a resource to a task is needed to be resource utilization, and task completion quality specific. The resource selection by Degree of Resource Optimality that scheduled to corresponding task is proposed in this manuscript. The factors of optimum resource utilization are described in the following:

- A resource often reflects diversified scope and divergent quality of service factors.
- The priority of the metrics related to resource utilization quality is different from one context to other.

Hence, it is obvious to argue that a resource that ranked best under a QoS metric is often not optimal under multi-objective quality factors. In the context of this constraint, the proposal considers the multi-objective quality factors to schedule a resource to corresponding task.

The depicted diversified quality factors of the resources recommended towards resource scheduling in IAAS are,

- Degree of Response Time (*drt*): This metric term the maximum time required to corresponding resource to respond to the resource broker, which is the aggregate of mean value and mean deviation observed from the past anomalies of the response time of the corresponding resource observed by resource broker. This metric found to be critical during the allocation of resource(s) to the deadline constrained task(s).
- Degree of Service Denial (*dsd*): The degree of service denial is another crucial factor of quality of service that indicates the scope of unresponsiveness

$$art(r) = \frac{1}{n} \sum_{i=1}^n rt(t_i, r)$$

$$mdrt(r) = \frac{1}{n} \left(\sum_{i=1}^n \sqrt{(art(r) - rt(t_i, r))^2} \right)$$

$$drt(r) = art(r) + mdrt(r)$$



of the scheduled resource, which is the aggregate of mean and mean deviation of unresponsive schedules against total number of schedules of the corresponding resource.

$$rsr(r_i) = \frac{\sum_{j=1}^{ruc(r_i)} \left\{ \begin{array}{l} 1 \exists \text{reschedule is true} \\ 0 \end{array} \right\}}{ruc(r_i)}$$

- Here in the above equation $rsr(r_i)$ indicates the rescheduling rate of resource r_i , which is the ratio of the count of reschedules observed against the number of times resource r_i scheduled. The notation $ruc(r_i)$ denotes the actual number of times the resource r_i scheduled.
- Degree of Realization (dr): The other quality factor of resource scheduling adopted is Degree of Realization, which is the absolute difference of the mean count of successful task realizations and the corresponding mean deviations.

$$ar(r) = \frac{1}{n} \left(\sum_{i=1}^n 1 \exists \text{task } t_i \text{ realized} \right) // \text{average of realization}$$

$$mdr(r) = \frac{1}{n} \left(\sum_{i=1}^n \left(ar - \left\{ \begin{array}{l} 1 \exists \text{if task } t_i \text{ realized} \\ 0 \end{array} \right\} \right) \right) //$$

mean distance of the realization

$$dr = r - mdr // \text{degree of realization}$$

- Degree of Load Adoption (dla): The expected load on the resource during the stipulated schedule expected by the task(s) is another quality factor of the resources labeled as Degree of Load Adoption. The estimation of load adaption carried as follows:

Find the time interval of the resource in use, which is the average time of the corresponding resource against the total number of times that resource scheduled.

Find the number time intervals of the resource, which is the ratio of total time that resource in service and the time interval

Then find the mean load and mean deviation of the load observed in all of these time intervals. The aggregate of these mean load and mean deviation of the load can denote the Degree of Load Adoption.

$$tin(r) = \frac{1}{n} \sum_{i=1}^n et(t_i, r) // \text{time interval}$$

$$al(r) = \frac{1}{n} \sum_{i=1}^n l(tin_i, r) // \text{mean load}$$

$$mdl(r) = \frac{1}{n} \left(\sum_{i=1}^n \sqrt{(al(r) - l(tin_i, r))^2} \right) // \text{mean load}$$

deviation

$$dla = al(r) + mdl(r) // \text{degree of load adoption}$$

- Degree of Cost Viability (dcv): The resource cost is crucial, resource scheduling carried under Service Level Agreement. The client who initiated

the task accepts the pay per resource, which is certainly lesser than the upper limit concluded in SLA. Hence, the resource with minimal cost would be most viable. However, it is not at the loss of other quality factors. In this context, rather opting a resource with minimal cost, the proposed scheduling strategy adopts a resource which is qualified under other quality factors and having cost of pay per use as lesser than the agreed level of budget. The cost viability of the selected resource can derive as the difference between the max level of the agreed cost and the estimated cost of the resource against pay per use.

$$ac = \frac{\sum_{i=1}^n (ecr(r_i) - mac(t))}{n}$$

$$mdc = \frac{\sum_{i=1}^n \left(\sqrt{(ac - (ecr(r_i) - mac(t)))^2} \right)}{n}$$

$$dcv = ac - mdc$$

The Heuristic Measure (Degree of Resource Optimality (dro))

Let Degree of Response Time (drt), Degree of Service Denial (dsd), Degree of Realization (dr), Degree of Load Adoption (dla), and Degree of Cost Viability (dcv) as a set of QoS metrics denoted for each resource r_i as $M_{(r_i)} = \{drt, dsd, dr, dla, dcv\}$.

In regard to explore the proposed model, let the QoS factors $dcv(r_i), dla(r_i)$ are principle metrics, which are using to find the compatibility scope of each resource. These key metrics are used to order the resources as described in following algorithm.

Initial process normalizes the cost feasibility and degree of load adoption (dla) as follows:

foreach{ $r_i \exists r_i \in R \wedge i = 1, 2, 3, \dots, |R|$ } // Begin

$ndcv(r_i) = dcv(r_i)^{-1}$ // degree of cost viability in normal form $ndcv$, which is between 0 and 1.

$ndcv_{abs} \leftarrow abs(ndcv(r_i))$ //The set $ndcv_{abs}$ contains the absolute values of the corresponding degree of cost viability in normal form observed for each resource

End // of step 1

foreach{ $r_i \exists r_i \in R \wedge i = 1, 2, 3, \dots, |R|$ } // Begin

$ndla(r_i) = dla(r_i)^{-1}$ // the degree of load adoption in normal form $ndla$ which is in range of 0 to 1

$diff_{abs} \leftarrow abs(ndla(r_i))$ //The set $diff_{abs}$ contains the absolute values of the entries in $diff$

End // of step 5

foreach{ $r_i \exists r_i \in R \wedge i = 1, 2, 3, \dots, |R|$ } Begin

$km(r_i) = 1 - (ndcv(r_i) \times ndla(r_i)) \exists (ndcv(r_i) < 1 \parallel ndla(r_i) < 1)$ //The

resultant product is subtracted from 1, which is since, to obtain higher value, as the product of two decimal fractions delivers the other decimal fraction that surely less than the decimal fractions involved in multiplication.

End // of step 9

Then the available resources are rated in regard to each metric, such that each resource will have individual rating for each metric. For each metric, resources will be rated in ascending order of corresponding metric values, if higher values are optimal, such that the resource having lowest value for corresponding metric will be rated as 1, and the resource having highest value for corresponding metric will be rated as $\{n \exists n \leq x\}$, here the notation x represents number of resources. If lowest values are optimal, then the resources will be rated in descending order of corresponding metric values, such that the resource is having highest value for corresponding metric will be rated as 1, and the resource having lowest value for corresponding metric will be rated as $\{n \exists n \leq x\}$.

Upon completion of the process, each resource reflects multiple ratings in regard to specified quality metrics. Further these ratings will be used as input to measure the Degree of Resource Optimality $dro(r_i)$ as follows.

For each resource $\{r_i \exists i = 1, 2, \dots, x\}$ Begin

$$\mu(r_i) = \frac{km(r_i) + drt(r_i) + dsd(r_i) + dr(r_i)}{4}$$

// the above equation represents the average of the ratings obtained for different metrics of resource r_i

$$dro(r_i) = \left[\frac{\sqrt{(\mu(r_i) - km(r_i))^2} + \sqrt{(\mu(r_i) - drt(r_i))^2} + \sqrt{(\mu(r_i) - dsd(r_i))^2} + \sqrt{(\mu(r_i) - dr(r_i))^2}}{4} \right]^{-1}$$

Degree of Resource Optimality $dro(r_i)$ is the inverse of root mean square distance of the ratings assigned to resource r_i , which is since the lowest distance is most optimal.

Upon completion of assessing Degree of Resource Optimality for given resources, the resources will be sorted in descending order of their rating obtained for key metric.

Further select the set of resources having optimal rating in regard to key metric, which is in accordance to given threshold.

The resources selected are sorted in descending order of their Degree of Resource Optimality $dro(r_i)$, which helps to project the best resource in first place of the ordered list. The same order is the preferred order to choose resources in regard to schedule the task.

4. EXPERIMENTAL SETUP AND EMPIRICAL ANALYSIS RESULTS

The empirical study for comparing the results of RS suggested scheme called RSDSQ and other existing methods “job scheduling with efficient resource monitoring (JS-ERM) [12]” & “multi-objective scheduling method based on ant colony optimization (MOSACO) [13]” that is simulated utilizing Cloudsim [14], which allows for simulating high dimensional CC network. Synthesizing input jobs, so that

there could be no priority sequence applicable to corresponding jobs. The confines are executed for performing simulation from 1 processor towards other, and pre-emption is not enabled. The scheduling the resources utilizing the QoS factors considered by proposed & other existing methods are chosen for analysing the performance. Moreover, we noticed the metrics of performance that are discussed in next segment at distinct intervals of time.

The proposed RSDSQ is evaluated by comparing with other contemporary models JS-ERM [12], and MOSACO [13]. The performance was scaled under QOS metrics rescheduling rate, task completion rate, and resource utilization rate.

The resource utilization rate observed for RSDSQ is high and stable that compared to other benchmarking models JS-ERM, and MOSACO (see Figure 3). The rescheduling rate noticed for RSDSQ could be minimal & linear when compared with other benchmarking models (see Figure 1). Since, the resource rescheduling rate is observed low in RSDSQ; obviously that delivers the best job completion rate (see Figure 2). The process complexity is minimal for RSDSQ, which is low due to the scalable approach adapted for Degree of Resource Optimality assessment.

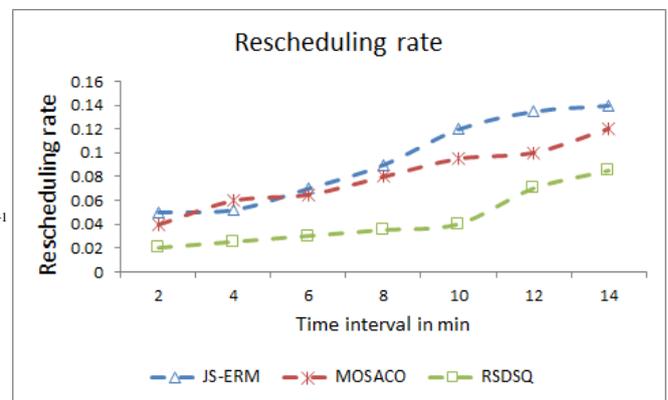


Figure 1: Resource rescheduling rate observed

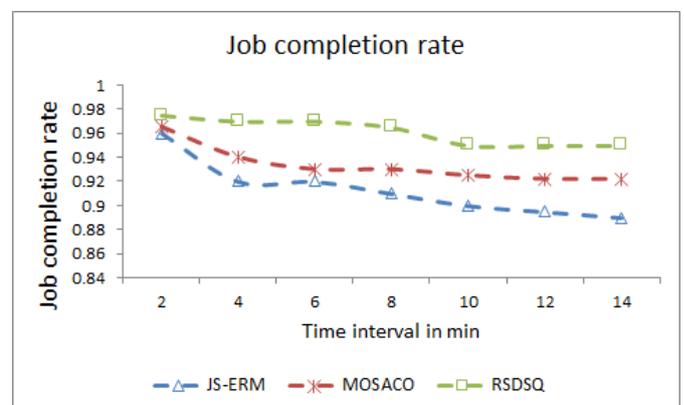


Figure 2: Job Completion Rate observed



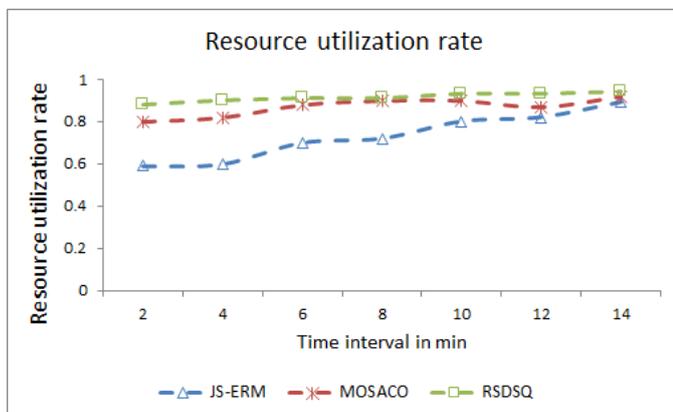


Figure 3: Resource utilization rate observed

Figure 1 is the representation of rescheduling rate noticed at diverse intervals of time of the simulation. The observed rescheduling rate is against the task load. Figure 1 is indicating that the proposed RSDSQ is significantly best to minimize the rescheduling rate that compared to other benchmark models. Figure 2 is indicating that the job completion rate observed RSDSQ is also considerable and significant that compared to JS-ERM and MOSACO, which is since, the minimal rescheduling rate maximizes the task completion rate, hence the considerable improvement in task completion rate observed for RSDSQ. Figure 3 depicts the RSDSQ advantage over other 2 existing methods in respect to resource utilization rate, which is prime objective of the resource scheduling strategies.

5. CONCLUSION

In this manuscript, a quality aware scheduling algorithm, which is for scheduling the resources and consequently that endeavoured to optimize the task completion in cloud computing environment. The manuscript addressed a novel scale called Degree of Resource Optimality, which indicates the fitness of the resources under divergent service quality metrics proposed. The outcomes attained from the method suggested are compared with other 2 existing methods JS-ERM [12] & MOSACO [13]. The performance analysis exhibiting that the suggested method surpassed when compared with other 2 existing methods, in respect to divergent metrics of quality. Here, empirical analysis of suggested method might impact further research for developing a scheduling and load balancing strategy to achieve optimal scheduling of virtual machines as resources in cloud computing.

REFERENCES

1. Chana, Inderveer, and Sukhpal Singh. "Quality of service and service level agreements for cloud environments: Issues and challenges." *Cloud Computing*. Springer, Cham, 2014. 51-72.
2. Liu, Zhaobin, et al. "Resource preprocessing and optimal task scheduling in cloud computing environments." *Concurrency and Computation: Practice and Experience* 27.13 (2015): 3461-3482.
3. Zhu, Qian, and GaganAgrawal. "Resource provisioning with budget constraints for adaptive applications in cloud environments." *IEEE Transactions on Services Computing* 5.4 (2012): 497-511.
4. Abrishami, Saeid, Mahmoud Naghibzadeh, and Dick HJ Epema. "Cost-driven scheduling of grid workflows using

- partial critical paths." *IEEE Transactions on Parallel and Distributed Systems* 23.8 (2011): 1400-1414.
5. Shifrin, Mark, Rami Atar, and Israel Cidon. "Optimal scheduling in the hybrid-cloud." *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*. IEEE, 2013.
6. Duan, Rubing, RaduProdan, and Xiaorong Li. "Multi-objective game theoretic scheduling of bag-of-tasks workflows on hybrid clouds." *IEEE Transactions on Cloud Computing* 2.1 (2014): 29-42.
7. Rahman, Mustafizur, Xiaorong Li, and Henry Palit. "Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment." *2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*. IEEE, 2011.
8. Chopra, Nitish, and Sarbjeet Singh. "Survey on scheduling in hybrid clouds." *Fifth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2014.
9. Javadi, Bahman, JemalAbawajy, and RajkumarBuyya. "Failure-aware resource provisioning for hybrid Cloud infrastructure." *Journal of parallel and distributed computing* 72.10 (2012): 1318-1331.
10. Van den Bossche, Ruben, Kurt Vanmechelen, and Jan Broeckhove. "Cost-optimal scheduling in hybrid iaas clouds for deadline constrained workloads." *2010 IEEE 3rd international conference on cloud computing*. IEEE, 2010.
11. He, Sijin, Li Guo, and YikeGuo. "Real time elastic cloud management for limited resources." *2011 IEEE 4th International Conference on Cloud Computing*. IEEE, 2011.
12. Loganathan, S., and S. Mukherjee. "Job Scheduling with Efficient Resource Monitoring in Cloud Datacenter." *TheScientificWorldJournal* 2015 (2015): 983018-983018.
13. Zuo, Liyun, et al. "A multi-objective hybrid cloud resource scheduling method based on deadline and cost constraints." *IEEE Access* 5 (2017): 22067-22080.
14. Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." *Softw. Pract. Exper* 41 (2011): 23-50.